Identifying Network Correlates of Memory Consolidation

by

Quinton Skilling

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biophysics)
In the University of Michigan
2020

Doctoral Committee:

Professor Michal R. Zochowski, Chair
Associate Professor Sara J. Aton
Assistant Professor Kevin Wood
Assistant Professor Qiong Yang

Quinton M Skilling
qmskill@umich.edu
ORCID iD: 0000-0002-5659-6558

To my family, for always showing me unwavering support as I pursue my dreams.

ACKNOWLEDGEMENTS

Growing up, I never really thought about the future or what I would do for a living. In college, I decided that graduate school sounded like fun and that it would put me on track for a career in science and teaching, but I didn't really think about what being a graduate student would be like. While I had some immense challenges in graduate school, it has been the most rewarding experience I have ever had and the people and institutions that have supported me over the last six years deserve much of the credit.

First, I would like to thank my advisor, Michal Zochowski. You have pushed me to be a better scientist and educator. Under your guidance I have learned nearly everything I know about computational neuroscience, from devising a plan and testing it, to finalizing manuscripts and dealing with the nuances of scientific politics. I will always be grateful to you for showing praise in my successes and patience in my mistakes. My dissertation committee, Sara Aton, Kevin Wood, and Qiong Yang, also deserve much thanks. They have helped me become a better scientific communicator by listening to updates on my projects and providing much-needed feedback.

I would also like to thank Sara Aton and her (now graduated) students Nicolette Ogjanovski, Jaclyn Durkin, and Brittany Clawson for providing experimental data so rich that it defined a large part of my PhD work. I would especially like to thank Sara for her patience in my struggle through her data but also for helping me refine my understanding of information encoding and for helping me to better link abstract ideas to biological reality.

Next, I would like to thank the Biophysics program and the Rackham Graduate School at the University of Michigan, for their financial support throughout my graduate studies. Because of them, I was able to travel to many national and international conferences and I was even lucky enough to secure a Predoctoral Fellowship from Rackham during my fifth year of graduate school. I would especially like to thank the

Biophysics graduate student coordinator, Sara Grosky, for enthusiastically helping me navigate the administrative requirements of obtaining a PhD.

Moving to Ann Arbor for graduate school was frightening at first but I owe a lot to members of my cohort for helping me establish a new way of life in a new place and to members of my lab for making work life enjoyable. Kelsey, Katie, and Hayden, getting to know all of you was an absolute blast and I hope that our paths cross again soon. Jeff, I count late night sessions of Arkham Horror with you as some of the best moments of graduate school; hopefully I can stop being lazy soon and develop an app so that our gameplay can continue when we take separate paths. To my lab members Bola, Jack, James, Jiaxing, Maral, and Yihao, you all made day-to-day life in office environment easy and I hope that whatever comes next will have others people as easy-going, friendly, and scientifically-minded as you all have been.

When things at the university or outside it became stressful, I knew that I always had my family for support. To my mother, Connie, sister, Ashley, and near-brother, Shaun, you have all been so encouraging and have always pushed me to try new things, learn from my mistakes, and to be a better person. Because of each of you, I am hopeful for everything the future will bring and I am incredibly happy to know that you will be part of it.

I've come to realize that many people can't say that they've been friends with someone since childhood and so I count myself lucky that I have eight such friends. Aaron, Bryce, Daniel, Justin, Keith, Luke, Matt, and Tyee have been the best friends anyone could ask for. To Cody and Josh, though I have known you for less time, I count you both among my closest friends. I hope you all know that I consider you family and always will.

Finally, a huge shout-out goes to my amazing wife Jordan. You have been the most fun, loving, and supportive person I have ever known. You've helped me through good and bad times and have taught me to slow down and live in the moment instead of constantly worrying about the future. Thanks to you, Jordan, we've had amazing animals Henri, Bode, and Jupiter, and they have made it so much better. Our next adventure is on the horizon and I am the luckiest person to be taking it with you.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF APPENDICES

APPENDIX

LIST OF ABBREVIATIONS

Ach	Acetylcholine
AMD	Average Minimal Distance
AMPA	α-3hydroxy-5methyl-4-isoxazoleproprionic acid
AMPAR	AMPA receptor
CFC	Contextual Fear Conditioning
CFM	Contextual Fear Memory
E/I	Ratio of excitation and inhibition
FCM	Functional Connectivity Matrix
FSM	Functional Stability Matrix
FuNS	Functional Network Stability
GABA	γ-aminobutyric acid
GABAR	GABA receptor
NREM	Non-Rapid-Eye Movement
NMDA	N-methyl D-aspartate
NMDAR	NMDA receptor
REM	Rapid-Eye Movement
SD	Sleep Deprivation
STDP	Spike-Timing Dependent Plasticity
SWS	Slow-wave sleep

ABSTRACT

Neuronal spiking activity carries information about our experiences in the waking world but exactly how the brain can quickly and efficiently encode sensory information into a useful neural code and then subsequently consolidate that information into memory remains a mystery. While neuronal networks are known to play a vital role in these processes, detangling the properties of network activity from the complex spiking dynamics observed is a formidable challenge, requiring collaborations across scientific disciplines. In this work, I outline my contributions in computational modeling and data analysis toward understanding how network dynamics facilitate memory consolidation. For experimental perspective, I investigate hippocampal recordings of mice that are subjected to contextual fear conditioning and subsequently undergo sleep-dependent fear memory consolidation. First, I outline the development of a functional connectivity algorithm which rapidly and robustly assesses network structure based on neuronal spike timing. I show that the relative stability of these functional networks can be used to identify global network dynamics, revealing that an increase in functional network stability correlates with successful fear memory consolidation in vivo. Using an attractor-based model to simulate memory encoding and consolidation, I go on to show that dynamics associated with a second-order phase transition, at a critical point in phase-space, are necessary for recruiting additional neurons into network dynamics associated with memory consolidation. I show that successful consolidation subsequently shifts dynamics away from a critical point and towards sub-critical dynamics. Investigations of in vivo spiking dynamics likewise revealed that hippocampal dynamics during non-rapid-eye-movement (NREM) sleep show features of being near a critical point and that fear memory consolidation leads to a shift in dynamics. Finally, I investigate the role of NREM sleep in facilitating memory consolidation using a conductance-based model of neuronal activity that can easily switch between modes of activity loosely representing waking and NREM sleep. Analysis of model simulations revealed that oscillations associated with NREM

sleep promote a phase-based coding of information; neurons with high firing rates during periods of wake lead spiking activity during NREM oscillations. I show that when phase-coding is active in both simulations and in vivo, synaptic plasticity selectively strengthens the input to neurons firing late in the oscillation while simultaneously reducing input to neurons firing early in the oscillation. The effect is a net homogenization of firing rates observed in multiple other studies, and subsequently leads to recruitment of new neurons into a memory engram and information transfer from fast firing neurons to slow firing neurons. Taken together, my work outlines important, newly-discovered features of neuronal network dynamics related to memory encoding and consolidation: networks near criticality promote recruitment of additional neurons into stable firing patterns through NREM-associated oscillations and subsequently consolidates information into memories through phase-based coding.

Chapter 1

Introduction

As we transverse through our waking lives, we observe our local surroundings, process the information we gather through our senses, and make decisions based on input we receive. Often, our decisions are guided by our memories, prior experiences that shape the way we process and respond to incoming information. How are these memories formed? What gives our memories their longevity and our ability to recall them seemingly at will? These are long-standing questions in the field of neuroscience and considerably shaped the work I present here.

It is widely held that memories are constructs of neural activity in the brain: neurons are excitable cells that connect to form networks and subsequently processes information by generating patterns of electrochemical activity. Howgvbfever, the exact nature of neural activity patterns as they pertain to memory formation and consolidation is still being investigated. Paramount to these studies are the modeling and theoretical tools used to classify neural activity. Luckily, decades of research in fields like graph theory and statistical physics [1] provide a fundamental starting point of analytics, whether it be a pseudo-reconstruction of the underlying networks [2] or classifying dynamical state of neural activity [3]. These tools and others are becoming ubiquitous in studies of systems-level neuroscience while the ability to apply these frameworks to a vast array of data raises another question: which data should be analyzed to understand memory consolidation? The answer is largely dependent on which type of memories are being studied. In the work presented here, I focus on sleep-dependent memory consolidation. As I will show, dynamics of neural activity during sleep can be greatly different than waking dynamics, providing active and necessary mechanisms for memory consolidation. To begin, it is important to understand how neural activity patterns emerge, i.e. through the activations of neurons connected in networks.

1.1 Neurons and Networks

Neurons, the primary information processing units in the brain, have semi-permeable membranes that maintain electrochemical gradients of ions between the intracellular space and the extracellular space, resulting in a resting membrane potential $\boldsymbol{V_{membrane} = V_{rest} \sim -60\ mV}$. Semi-permeable in this context refers to neurons' utilization of transmembrane proteins to help facilitate the blockage and selective passage of polar ions and large macromolecules across the membrane. Ion channels, for example, are transmembrane proteins that can change their conformation from on open state, where a select type of ion can flow into or out of the cell, to a closed state, where ion movement across the membrane is not possible. The dynamics of these ion channels and other transmembrane proteins govern neural activity [4].

As positive ions flow into the cell (or negative ions out of the cell), the voltage across the membrane increases and activates (i.e. opens) voltage-gated ion channels permeable to sodium (Na+) ions. The intracellular concentration of Na+ is quite low compared to the extracellular concentration (5mM inside compared to ~100 mM outside), causing a rapid influx of sodium into the cell and resulting in a rapid increase in the membrane potential to $\boldsymbol{V_{membrane} \sim 40\ mV}$. As voltage continues to increase, sodium channels start to inactivate (due to their reversal potentials, the voltage where ionic flux is not possible), effectively slowing sodium intake, as voltage-gated potassium (K+) channels activate, allowing potassium to flow from the intracellular space outward, along the K+ concentration gradient. The result is a rapid decrease in the membrane potential, past the resting membrane potential ($\boldsymbol{V_{membrane} < V_{rest}}$) called a hyper-polarization. This entire process, called an action potential, lasting from initial depolarization to eventual hyper-polarization, typically takes on the order of a millisecond. Given the large, abrupt fluctuations in membrane potential over such a short period of time, action potentials are generally referred to as "spikes" of activity. Each time a neuron spikes, its signal is propagated to other neurons along anatomical connections called synapses, providing the primary input neurons receive and forming expansive neuronal networks [4].

1.1.1 Synaptic Connections Form Neuronal Networks

While neuron morphology varies greatly depending on function and brain region, a vast majority of neurons in the brain are so-called "multi-polar": dendrites receive

incoming information and are spatially separated, via the soma of the neuron, from the axonal terminals, which send outgoing information to other neurons. The soma is commonly implicated as a site of integration (though it is known that already dendrites can act as nonlinear integrators of the incoming signals), where incoming signals are summed via subcellular mechanisms, producing spikes in axon hillock (the beginning of the axon, just past the soma) which subsequently propagate to the axonal terminals and out to dendrites of post-synaptic neurons. Connections, called synapses, between neurons thus predominantly form between the pre-synaptic axonal terminals and the post-synaptic dendrites. As a neuron spikes, it releases neurotransmitters into the synaptic cleft which can then bind to and activate voltage- and ligand-gated ion channels in the post-synaptic neuron, initiating the passage of ions into the post-synaptic neuron. The nature of these connections can either lead to subsequent spiking or suppression in post-synaptic neurons and relies heavily on whether the pre-synaptic neuron is excitatory or inhibitory.

Excitatory neurons predominantly release the neurotransmitter glutamate and its analogs, which bind to N-methyl D-aspartate receptors (NMDARs) and α-3hydroxy-5methyl-4-isoxazoleproprionic acid receptors (AMPARs) on the post-synaptic cell, each of which permits cationic flux into the neuron (Na^+ through AMPARs and Ca^{2+} through NMDARs) and out of it (K^+ in both NMDARs and AMPARs). Though both these ion channels help facilitate excitatory synaptic connections, the timescales on which they act and the mechanisms they facilitate vary. AMPARs are thought to facilitate rapid signaling between neurons (on the order of tens of milliseconds at most): glutamate (or its competitive agonist analog AMPA) binds to the receptor, causing a conformational change and subsequent ionic flow. NMDARs, on the other hand, are blocked by magnesium (Mg+) ions unless the postsynaptic neuron is depolarized to $\boldsymbol{V_{membrane} \sim 0\ mV}$, in which case Mg+ is expelled from the channel and ionic flow through NMDARs can proceed. Given that the primary depolarizing agent of a neuron is the activity of other neurons, NMDARs are thus activated when both pre- and post-synaptic neurons are active, indicating the importance of NMDARs in the relative strengthening and weakening of synaptic connections, which will be described shortly [4].

Converse to excitatory neurons, inhibitory neurons release the neurotransmitter γ-aminobutyric acid (GABA), which binds to the respective receptors (GABARs) on the

post-synaptic neuron, leading to a signaling cascade which eventually causes a conformational change and a rapid influx of the negatively charged Chloride (Cl-) into the post-synaptic neuron, effectively lowering the membrane potential. The characteristic timescale of GABARs varies drastically depending on whether the receptor is ionotropic or metabotropic. Ionotropic receptors activate (possibly in a voltage-dependent manner) on the direct binding of a ligand to a receptor and lead to the direct flow of ions into or out of the cell. Metabotropic synapses, on the other hand, involve a much more complicated signaling mechanisms and act on the order of 100 milliseconds to seconds or longer, but aid in the diversity of synaptic signaling and contribute to important cognitive processes.

In metabotropic synaptic signaling, ligands released pre-synaptically bind to post-synaptic receptors and initiate a signaling cascade involving g-proteins and second messengers. Eventually, these second messengers bind to post-synaptic ligands intracellularly, causing those receptors to undergo a conformational change and subsequent flow of permeable ions along their concentration gradients. Metabotropic synapses, like ionotropic ones, can be excitatory or inhibitory and include relevant ligands such as serotonin (responsible for happiness), dopamine (implicated in reward feedback), acetylcholine (important for attention), and others. Taken together, ionotropic and metabotropic synapses support a multitude of signaling mechanisms between neurons and give the neurons great control over how and under which circumstances electrochemical activity is communicated. Going further, these connections can strengthen and weaken in time, referred to as synaptic plasticity, a process thought to be crucial for memory encoding.

1.1.2 Plasticity: The Strengthening and Weakening of Synaptic Connections

Decades of research support that neuronal networks are not static. Not only can new connections form and old connections be pruned, but the efficacy of existing connections can also be modified through plasticity mechanisms. From a biophysical perspective, there are multiple mechanisms that modify synaptic efficacy, effecting both the pre-synaptic terminals and post-synaptic dendrites. In post-synaptic neurons, plasticity can increase (potentiation) or decrease (depression) the number of available receptors, leading to a correlative size change in synaptic spine density [5]. NMDARs play a vital role in this process: as both pre- and post-synaptic neurons spike within a small time window of each

other, Ca2+ influx through unblock NMDARs in the post-synaptic cell aid in the trafficking of additional AMPARs into the post-synaptic membrane. In pre-synaptic neurons, feedback from post-synaptic neurons induces plasticity by increasing the number of neurotransmitters produced and released [6].

Plasticity can thus affect both the pre-synaptic axonal terminals and post-synaptic dendrites, but what dynamical mechanisms facilitate plasticity in the first place? Much evidence indicates that the relative spike timings between neurons play a major role in plasticity as it relates to information encoding [7, 8]. This process, called spike-timing dependent plasticity (STDP), dictates that the spike-timing between pre- and post-synaptic neurons, along with the order the spikes occur, determines whether a synaptic connection is strengthened or weakened. Evidence has shown that STDP has an active time window around 60 ms, where a pre-synaptic neuron firing before a post-synaptic neuron leads to a net increase in the synaptic weight; by comparison, the opposite ordering of spikes (post-before pre-) leads to a net decrease in the synaptic weight [7]. Other experiments suggest that plasticity rules change based on neuron firing rate [8] and differ between excitatory and inhibitory neurons [9]. Regardless, activity-based plasticity is thought to be evidence of Hebbian plasticity, that neurons who fire together, wire together [10], forming a neural correlate of memory.

1.1.3 Synaptic Plasticity and Memory Formation

Hebbian Plasticity plays a vital role in information encoding, preferentially strengthening connections among neurons that are co-active in time, called memory engrams, and are thought to represent a locale of memory. Early studies on the engram found that memories are widely distributed, so much so that removing targeted brain regions are not necessarily sufficient to block memory recall [11]. Recent studies using sophisticated biophysical techniques indicate that memory can be identified by a small group of neurons that reliably activate over extended periods of time during memory encoding and recall [12]. In this context, these identified neurons represent only a subsection of the distributed engram: the impermanence of synapses through plasticity [13], along with the confirmed recruitment of new neurons into an engram [14-16] show the robustness of an engram. Activity patterns of vast networks are thus responsible for encoding information and partial reconstruction of the activity pattern may be sufficient to

successfully reconstruct the full pattern, leading to a recall of the encoded memory, and thus forming a type of dynamical attractor [17]. This type of associative recall is familiar to most people, for example when smelling a certain type of food and remembering the restaurant you ate that food in, and who you were dining with.

While a definitive explanation of how engrams are formed and consolidated is still lacking, it is obvious that neural activity is at the heart of the problem. What is it about the activity that permits plasticity-induced engram formation? How are various forms of activity classified? Do different dynamical classes of activity neuronal patterns provide differential mechanisms for memory consolidation and/or mediate different aspects of memory formation? These outstanding questions outline the importance of developing theoretical frameworks for understanding neural activity, including classifying the dynamical state of activity, how effective networks based on that activity arrange and subsequently help coordinate activity to form new memories.

1.2 Dynamical States and Processes of Neural Activity Mediating Memory Formation

The brain represents a complex system of neuronal units that interact so that we may understand the world around us. From the point of view of activity dynamics, it is not fully understood how this occurs. Many theories exist which help explain how we are able to convert perceptions into neural activity, forming long-lasting memories, and doing so in an efficient manner. In the following, I outline two of these phenomena: self-organized critical dynamics in neural activity and the role of synchronization and oscillations in information encoding.

1.2.1 Self-Organized Criticality: A Dynamical Classification of Neuronal Activity

Classification of system dynamics was first introduced through statistical mechanics to describe the relative behavior of atoms in various stages of matter (i.e. solid, liquid, water, etc.) and so borrows many of its formalisms and concepts from that field. Chief among classification schemes, criticality has emerged as a ubiquitous description of the dynamical state of many systems, ranging from granular matter [18] to earthquakes [19] and neuroscience [3]. Criticality refers to system proximity to an unstable fixed point (or phase) in the system, where even slight deviation in constraints results in phase space can result in significantly different behavior, depending on the direction of deviation. In

water, for example, a critical point exists at the triple point [1], where deviation in volume, pressure, or temperature dictates if the water attains liquid, solid, or gaseous phase. At this critical point, water exists in each phase and in fact the size of each pocket of each phase arrange as power laws, indicating independence from a characteristic length scale. These changes occur simultaneously, but randomly, throughout the water and are thus said to have long-range correlations.

In the 1980's, criticality emerged as an explanation for pink noise in dynamical granular systems, showing that bursts of activity, termed avalanches, arrange as power laws [18]. This description of critical dynamics eventually was used to show that bursts of neural activity in cortical slices likewise arrange as power laws [3], indicating that neural activity may be distinctly positioned at a phase transition between other, non-critical states. This raises two important questions: (1) What are the characteristic difference between critical and non-critical dynamics class and (2) what benefit would critical dynamics lend to the brain?

To answer the first question, critical dynamics are typically distinguished between subcritical and supercritical dynamics. In terms of neuronal avalanches, subcritical dynamics produce small, temporally separated bursts of activity, indicating potentially random spiking among neurons and subsequently diminished signal propagation through a network. Supercritical dynamics, on the other hand, represent an "over-activation" of the system, where neuronal firing occurs nearly simultaneously between neurons. Hence, critical dynamics are thought to separate random firing states from completely synchronized firing states, and are often characterized as asynchronous and irregular but not random. But what benefit would such dynamics provide to neuronal networks? The answer to this question has driven exploration of the critical brain for nearly two decades.

The seminal work by *JM Beggs and D Plenz* [3] showed that information transfer between neurons is maximized near criticality. Subsequent studies showed that critical dynamics provide a host of other benefits pertaining to neural computations, including increased metastable dynamics [20], increased dynamic range [21], maximized information storage capacity [22], increased complexity of spiking relationships among neurons [23], and that critical dynamics may help mediate temporal filtering of information [24]. At the same time, opponents of the critical theory of the brain argue that the traditional

means of detecting criticality, namely power-law distributions, can be explained by other phenomena or are inherent features of how the data is collected [25, 26]. Still, the field continues to progress and attempt to take these criticisms into account. Recent studies suggest that the brain may be not critical, but slightly sub-critical [23, 27] and that more clear distinctions between dynamical states can be achieved through regression analysis [26, 28], in part a response to the aforementioned criticisms.

Critical dynamics is thus considered by many to be an important feature of information encoding [29, 30] and may lead to new insights into how memories are formed and consolidated. At the same time, however, the critical theory of the brain doesn't intrinsically account for network-level interactions, which are known to mediate information processing. It is with this in mind that we will now consider network-level mechanisms, namely synchronization and oscillations, and their role in information processing. Namely, we will show in Chapter 3 that universally diverging susceptibility and correlation length mediate storage of new memory engrams during presentation of the stimulus.

1.2.2 Oscillations and Synchrony as Dynamical Mechanisms of Binding Neuronal Representations

Up to this point, I have claimed that external representations are internalized by the activity patterns of neurons in the brain, but which features of these activity patterns mediate information processing? Can individual neurons represent features, propagating information through the network one neuron at a time, or are simultaneous interactions responsible for binding physical constructs to digital constructs in the brain? This question has historically been mired by contention [31-33], with decades of research from the inception of neurophysiology supporting a rate-coding phenomenon of individual neurons now being upended by evidence of temporal-coding of networks.

Rate-coding is where neurons or groups of neurons increase their firing rate in response to a stimuli, external or internal, and thus act as a representative of information. A pioneer study in the 1920s showed for the first time that neurons stimulated by muscles increased firing rate as the muscle was extend [34]. Since then, external sensations such as sight [35], hearing [36], and touch [37] have been shown to trigger neural activations following this scheme (though see the discussion on temporal coding, below). Proponents

of rate-coding have argued that it preferentially supports activity-based plasticity, like STDP, as evidence has shown that plasticity onset may require bursts of activity [38]. On the other hand, studies have pointed out that rate-coding may be insufficient to bind external features to neural activity within the cortex due to limits in the number of neurons [31, 32]: if one neuron increases its firing rate in the presence of a green apple and another for a red cherry, it follows that presentation of a red apple will require the activation of yet another neuron and ad infinitum for each object we've ever encountered. An alternative mechanism, colloquially referred to as temporal coding, proposes that the relative coactivation of neurons in time is responsible for binding features of the physical world to neural activity.

Temporal coding schemes assume that the population activity of many neurons within a given instant of time facilitates feature binding, i.e. that synchronization among the activity of neurons is what encodes information [31, 32]. The time scale of temporal coding is generally assumed to be on the order of 10ms or shorter and neurons that are consistently mutually active within this time window are said to be synchronized. However, the firing of individual neurons can be highly variable over short time scales (~seconds), indicating that (1) neurons can participate in many different activity patterns and (2) that specific patterns of synchronized activity may collectively representing feature binding. In the past few decades, the role of synchronized activity has come to the forefront of neuroscience in the form of neuronal oscillations.

Oscillations in the human brain were first identified in the 1920s [39]. Since then, the analysis of neurological recordings across animal species has led to the identification of multiple forms of oscillatory activity of neuronal firing. In mammals, biologically conserved bands of oscillatory activity have been identified including slow (delta [0.1-4Hz], theta [4-12Hz], beta [20-30Hz]) and fast (gamma [30-80Hz], sharp-wave ripples [120Hz]) rhythms. Generally, slower oscillations are characterized by larger amplitudes and therefore indicate a greater number of neurons synchronized over larger spatiotemporal extents. Among all oscillatory modes, gamma and theta oscillations are commonly associated with feature binding across cortical sites [40, 41], as well as in learning and memory [42, 43]. Ongoing evidence suggests the synchronization of various brain rhythms (give the examples), called cross-frequency coupling [44], indicating that, while neuronal

synchrony produces oscillatory behavior, small groups of individual neurons may "ride the wave" of an oscillation, increasing their firing rate and providing precisely timed input at particular phases of the main oscillation (possibly forming a link between rate- and temporal-coding). This form of temporal coding is often referred to as phase-coding and has been most noticeably attributed to the identification of place and grid cells [45], but also has been identified in information transfer between brain regions and is thought to be important for memory consolidation [46-48].

Taken together, while rate-coding has been historically attributed to feature binding, the role of temporal-coding through synchronization of spike times and the production of neural oscillations has provided much dynamical insight into how information is processed, and memories are formed. By switching between oscillatory modes, the brain is able to accomplish many different cognitive processes. This dynamical switching emerges as a natural process of the sleep-wake cycle, and so elucidating the interplay between wake and sleep in memory encoding and consolidation is essential.

1.3 Sleep – A Behavior Mediating Memory Consolidation

During periods of wakefulness, neurons throughout the brain increase their spiking firing rate in response to different stimuli, both in the peripheral nervous system and in subsequent regions (e.g. the visual and auditory cortices) in the brain. It has long been known that the corresponding neural activity emerges as high frequency oscillations in the gamma band of frequencies (20Hz-80Hz) with low amplitude, indicating a relatively low level of synchronization of spike timings among neurons. Still, STDP during wakefulness plays an important role in modifying neuronal networks and acts to form an initial memory engram, selectively strengthening connections among neurons firing in response to a given stimulus. These engrams are internal, network-level representations of synthesized extern input, including sensations such as sight, hearing, texture, and smell. However, our waking lives are separated by periods of sleep, when processing of external input is diminished, and neural activity is governed solely by internal, network-level dynamics. If new information is not being processed, then what role is sleep playing in information encoding in general? Is it just a passive process, allowing our brains to "turn off" for a while? Ongoing work suggests a much more active role of sleep in information processing and memory consolidation.

1.3.1 Sleep: Not Just a Passive Process

While neural activity during waking is important for initial memory engram encoding, sleep has long been implicated in memory consolidation. Early studies showed that declarative memory tasks markedly improved after subsequent sleep [49], and many studies since have shown sleep to be vitally important for many cognitive functions, supporting memory consolidation and information processing in a variety of processes [50].How does this occur? What processes present during sleep help facilitate memory encoding and consolidation? There are two leading hypotheses that help to explain the role of sleep in memory.

One hypothesis claims that sleep plays primarily a homeostatic role [51] by universally downscaling synaptic connections that were strengthened through activity-driven plasticity during waking. In this sleep homeostasis hypothesis, important (i.e. highly strengthened) connections are maintained relative to "unimportant" connections, potentially conserving engrams for future use. This may serve not only to strengthen memories but also to prime the brain for subsequent engram formation in subsequent periods of wakefulness [50]. However, recent studies in sleep-dependent memories [52] have shown that sleep provides a necessary potentiation mechanism, paving the way for a potentially more active role of sleep in the formation and consolidation of memories.

This more active role of sleep is often attributed to another hypothesis, active systems consolidation, which posits that features of sleeping dynamics help facilitate information transfer between regions of the brain [46, 47], e.g. between the hippocampus and neocortex, while simultaneously providing a system state conducive to feed-forward memory replay [53]. Accumulating evidence supports active systems consolidation during sleep and raises some important questions, namely, what is it about the dynamics of sleeping neural activity that promotes memory consolidation?

1.3.2 Dynamical States of Sleep

Sleep is generally broken down into two phases, rapid-eye movement (REM) sleep and non-REM (NREM) sleep. The most notable macroscopically observable distinction in dynamics between the two phases is the change in oscillatory tone: REM sleep, like wake, is characterized by gamma oscillations, as well as slower theta oscillations (4-12Hz), while NREM sleep is characterized by slow-wave, large-amplitude delta band oscillations (0.1-

4Hz), indicating a slow, synchronized rhythmicity of neural activity. At the same time during NREM sleep, bursts of neural activity called sleep spindles (10-15 Hz range) and sharp-wave ripples (>100Hz) are thought to coordinate activations across brain regions [46, 48, 54]. How does the brain know to switch between the oscillating regimes in and across sleep? One hypothesis is that changes in the biochemical environment may be an explanation. In fact, growing evidence supports the role of acetylcholine in governing individual neuron and network-level dynamics.

The relative concentration and production of acetylcholine (Ach) changes between waking and the different phases of sleep. During wake, Ach concentration levels are high and have been linked to attentive wakefulness and learning and memory [55]. During NREM sleep, the concentration of Ach drops significantly but rebounds during REM sleep; perhaps not by coincidence, REM sleep is associated with dreaming, where the brain simulates waking life. Growing evidence from our lab [53, 56] and others [57] support that Ach can have diverse effects on cell excitability, leading to various network-level activity features [50]. Ach reduces the slow potassium current through muscarinic acetylcholine receptors, effectively increasing the excitability of the neuron and switching input response from being T2 in NREM sleep, when Ach is low, to T1 in wake and REM, when Ach is high[57]. While the concentration of other neuromodulators also changes during sleep (e.g. changes in serotonin), Ach may provide a unique insight into how properties of individual neurons produce emergent oscillations via network-level interactions.

At the same time, the fact that internal networks control brain dynamics during sleep has led some to posit that sleep dynamics are viable for being self-organized critical states. Indeed, studies show that neural activity during sleep across multiple organisms show near-critical dynamics [27, 58], while prolonged periods of wakefulness have been shown to disrupt internal dynamics [59]. If true, dynamic criticality during sleep would indicate that neural networks benefit from dynamical properties of systems undergoing a phase transition in activity: long-range correlations [1], increased dynamical range [21], and increased memory storage capacity [22].

The rich dynamical state of sleep, together with the functional importance it plays in cognition, makes sleep an ideal candidate to study how memories are encoded in neural activity. One sleep-dependent memory consolidation paradigm, contextual fear

conditioning, is an encouraging model system to study, as learning is rapid and relies on the dynamics associated with sleep.

1.4 An Example of Rapid Memory Formation: Contextual Fear Conditioning

Contextual fear conditioning is a relatively simple, yet powerful training procedure that requires only one session. In this training session, mice are taken from their home cage and placed in a novel environment. Shortly thereafter, they are given a mild foot shock, the fear stimulus to be consolidated to memory. Mice are returned to their home cage for 24 hours and engage in normal activity: they sleep, eat, explore, etc. Following the 24 hours, the mice are reintroduced to the novel environment and monitored for a change in behavior. Normally, mice are inquisitive by nature and routinely explore new environments. Upon successful consolidation of the foot-shock fear memory, however, mice remember they were shocked in the novel environment and no longer explore but instead cower in place. This "freezing" behavior is thus a macroscopic observable of successful contextual fear memory (CFM) consolidation.

Numerous studies have shown that CFM is a sleep-dependent memory, consolidated within the hippocampus. Disruption of fear memory consolidation is simply achieved by sleep depriving the mice for six hours following fear exposure [60]. While evidence of a fear memory engram has been demonstrated [12], experts in CFMs seem to agree that the memory is distributed from engram to non-engram neurons via synaptic connections [61].

In order to investigate the network level dynamics associated with CFM consolidation, collaborators in the Aton lab recorded neural activity from by implanting stereotrodes into hippocampal area CA1. These recordings track the activity of a few dozen neurons in the 24 hours prior to and following fear training, representing a sparse representation of the system. Probabilistically, these recordings don't include engram cells directly, but their effect may be distinguished through a detailed understanding of network-level activity.

My PhD work focused on bridging the gap between this network-level activity and macroscopic observations of contextual fear conditioning. To this end, I developed multiple analytical frameworks in attempt to understand the dynamical context on neural activity related to memory consolidation. I likewise employed computational modeling to

help provide understanding of these processes not directly available from experimental data. I outline my findings in the following chapters.

1.5 Outline of the Dissertation

Taken together, my dissertation focuses on elucidating the network-level dynamics and processes that account for memory encoding and sleep-dependent consolidation. In the following chapters, I combine computational simulations with characterization of neural activity dynamics of both in vivo and in silico spiking data to outline processes that are important for memory consolidation.

In the second chapter, I discuss a functional connectivity algorithm, called Average Minimal Distance, and how the relative stability of functional connectivity can be used to inform on memory consolidation. I show that functional network stability (FuNS) can be utilized to investigate various forms of dynamical correlates of memory, from distinguishing real, anatomical connections from non-anatomical connections, to characterizing how initial correlates of memory shape networks dynamics in a spatiotemporal manner. This work also features in latter chapters, where I start to identify mechanisms related to information encoding and consolidation. This work was originally published in the *Journal of Neuroscience Methods* [62].

In the third chapter, I investigate the role of the dynamical state of neuronal networks and successful memory consolidation. I show that so-called critical dynamics are an important feature of neuronal activity, as this dynamical regime supports a maximized response to incoming information to be encoded and are subsequently required for successful consolidation of this information. I go on to show that successful consolidation causes a shift in the dynamical state away from criticality and toward a highly stable, sub-critical regime. Importantly, these features are also identified in experimental recordings of hippocampal CA1, indicating that the brain likely utilizes near-critical dynamics in a similar way that I show in theory. This work was originally published in the journal *Entropy* [63].

Finally, in the fourth chapter, I elucidate another potential mechanism important for sleep-dependent memory consolidation, phase-coding. In this work, I use a biophysically realistic model that can account for easy modification of the dynamical state of individual neurons in the system. I show that the initial correlates of a memory produce

oscillations in the network activity and serve to substantially increase FuNS. Subsequent plasticity intervals reveal that neuronal firing rates during periods of high Ach arrange neurons by phase in periods of low Ach: high firing rate neurons during waking lead slower neurons in terms of their phase-of-firing during an oscillation. By assessing the asymmetry of spiking data during bursts of activity recorded in vivo, I show that there is convincing evidence of a phase-based coding mechanism present during sleep which helps facilitate memory consolidation.

Chapter 2

Functional Network Stability and Average Minimal Distance – A framework to rapidly assess dynamics of functional network representations

In this work that I coauthored with Jiaxing Wu, we developed the analytical AMD and FuNS frameworks and applied them to surrogate data, simulation data of spiking neural networks, and in vivo data recorded from mouse hippocampus. We show how FuNS can track global changes in network dynamics related to memory encoding. This work was published in the Journal of Neuroscience Methods in 2018 [62].

New multisite recording techniques have generated a wealth of data on neuronal activity patterns in various brain modalities [64-67]. An unresolved question is how, using such data sets, one can correctly identify large-scale network dynamics from populations of neurons which either may, or may not, include neurons involved in a particular cognitive process of interest. This is due in part to the fact that even high-density recordings sample only a sparse subset of the neural system responsible for the modality in question. It is also complicated by the inherent separation of temporal scales over which neural vs. behavioral measurements occur.

In response to this question, multiple linear and non-linear techniques have been developed over the years to assess functional connectivity between neurons, and to possibly infer from it structural connectivity (see for example: [68-75]). More recent approaches utilize network theory to establish links between recorded data and the underlying connectivity (see for example: [76-99]). The idea is that, by estimating networks based on functional interactions, one can potentially gain insight into global dynamics, which reflect the general property of the whole network, instead of a specific subset of neurons. While all these approaches can provide insightful information, they share some the same problems. These methods are often limited by under-sampling (and potentially

unrepresentative sampling) of neuronal recordings, and are not optimized for monitoring changes in network structure across extended time periods (i.e., those associated with behaviors of interest, such as memory formation).

Here we propose a novel technique that rapidly estimates functional connectivity between recorded neurons. Then, rather than characterizing details of the recovered network, the metric measures changes in the network dynamical stability over time. The technique is based on an estimation of Average Minimal Distance (AMD) between spike trains of recorded neurons, a metric which has previously been compared to other clustering algorithms [100]. Here, we expand on this work and show that the analytic estimation of AMD for the null case, when the two cells are independent, allows for rapid estimation of the significance of pairwise connections between the spike trains, without need for time-expensive bootstrapping.

Further, Functional Network Stability (FuNS) is introduced and is monitored over timescales relevant for behavior. We show that FuNS measures global change in network dynamics in response to localized changes within the network. This, in part, alleviates the problem of sparse sampling so prevalent in neuroscience.

Below, the statistical underpinnings of AMD and FuNS are detailed. We compare AMD and cross-correlation (CC) on both surrogate data and model simulation data. Model results show the applicability of AMD and FuNS on excitatory-only networks, as well as on mixed networks of excitatory and inhibitory neurons poised near a balance between excitation and inhibition, a regime thought to be a universal dynamical state achieved by brain networks, resulting in enhanced information processing properties [101-105]. We end by analyzing experimental data recorded from the mouse hippocampus during contextual fear memory formation. Our results indicate that AMD yields results comparable to that of the gold-standard CC, but, importantly, it is orders of magnitude faster and reports statistically significant increases in FuNS due to behavioral-based network topological changes compared to CC FuNS.

2.1 Assessing Functional Connectivity via Average Minimal Distance

Pairwise functional connectivity is estimated using average minimal distance (AMD) [100] (Figure 2.1) separating the relative spike times between neurons. AMD is calculated as follows: given the full spike trains $\{S_1, S_2, \ldots, S_n\}$ for n neurons within a

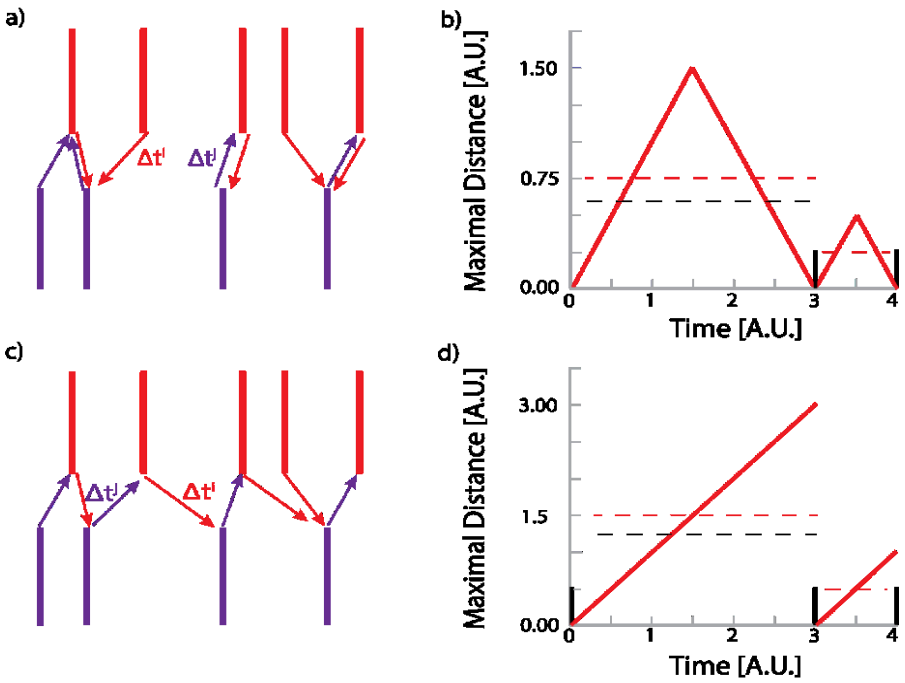

Figure 2.1: Calculation of AMD and analytical significance. The average minimal distance algorithm calculates shortest temporal length between spikes emitted by a neuron to the closest spikes in a reference neuron, looking in either both temporal directions (a), or in a single temporal direction (b), e.g. forward in time. The maximal possible distance between spikes is either half the interspike interval (c) or the full interspike interval (d), when looking in either both temporal directions or a single temporal direction, respectively. The measurements require a collective average timing sequence to be below one quarter (bidirectional) or one half the interspike interval (unidirectional) in order to be considered significant.

network, the pairwise functional relationship, FC_{ij}, of the i^{th} and j^{th} neurons is evaluated by comparing the average temporal closeness of spike trains S_i and S_j to the expected sampling distance of train S_j (Figure 2.1a). That is,

$$\boldsymbol{AMD_{ij}} = \frac{\mathbf{1}}{\boldsymbol{N_i}} \sum_{\boldsymbol{k}} \boldsymbol{\Delta t_k^i},$$

where N_i is the number of events in S_i and $\boldsymbol{\Delta t_k^i}$ is the temporal distance between an event k in S_i to the nearest event in S_j. With AMD measured, the functional connectivity (FC) is calculated as $\boldsymbol{FC_{ij}} = \sqrt{\boldsymbol{N_i}} * \left(\boldsymbol{AMD_{ij}} - \boldsymbol{\mu_j}\right)/\boldsymbol{\sigma_j}$, which is expressed in terms of probabilistic significance of connectivity between pair ij. The mean and standard deviation, μ_j and σ_j, of the expected sampling distance, assuming that the spike trains are independent, can be calculated from either: 1) boot-strapping (i.e. randomizing the spike trains multiple times and reassessing the AMD for the null hypothesis being statistically independent of the two spike trains), or 2) numerical estimation of expected values given the distribution of inter-spike intervals (ISIs) on S_j. Hereafter, the analytical method is referred to as "fast AMD" and the bootstrapping method as "bootstrapped AMD". For a system with n

neurons, the functional connectivity value between each pair of spike trains is calculated, generating an n-by-n Functional Connectivity Matrix (FCM).

In the fast AMD approach, the maximal distance between an input spike and any spike in the spike train to be analyzed is $\frac{ISI_i}{2}$. Then, the expected mean distance between spikes in the independent spike trains is $\boldsymbol{\mu_i} = \frac{ISI_i}{4}$, where ISI_i is the corresponding interspike interval of spike train *i* (Figure 2.1b). Calculating the first and second raw moments from the maximal distance then yields $\boldsymbol{\mu_1^L} = \frac{1}{4}\boldsymbol{L}$ and $\boldsymbol{\mu_2^L} = \frac{1}{12}\boldsymbol{L^2}$ for a specific ISI with length L. Taking into account the probability of observing an ISI with length L over the recording interval T, $\boldsymbol{p(L)} = \frac{L}{T}$, the first and second moment for sampling the whole spike train randomly are then $\boldsymbol{\mu_1} = \sum_L \frac{L}{T}\boldsymbol{\mu_1^L} = \frac{1}{4T}\sum_L \boldsymbol{L^2}$ and $\boldsymbol{\mu_2} = \sum_L \frac{L}{T}\boldsymbol{\mu_2^L} = \frac{1}{12T}\sum_L \boldsymbol{L^3}$, respectively. The expected mean and standard deviation of a random spike train are then calculated as $\boldsymbol{\mu} = \boldsymbol{\mu_1}$ and $\boldsymbol{\sigma} = \sqrt{\boldsymbol{\mu_2} - \boldsymbol{\mu_1^2}}$.

2.1.1 Unidirectional AMD for Causality Detection

The bidirectional AMD described above (i.e. the temporal distance between spikes of two different neurons, measured in either direction) can be extended to be unidirectional to identify causality between the two spike trains. In this scenario, the temporal distance is measured only forward in time and the mean delay time expected within the null hypothesis (i.e. independence of both spike trains) is only set to $\boldsymbol{\mu_i} = \frac{ISI_i}{2}$, assuming a maximal temporal distance equal to the ISI (Figure 2.1c and 2.1d). The calculation of first and second moment change accordingly to $\boldsymbol{\mu_1} = \frac{1}{2T}\sum_L \boldsymbol{L^2}$ and $\boldsymbol{\mu_2} = \frac{1}{3T}\sum_L \boldsymbol{L^3}$; the mean and standard are then calculated in the same manner as above.

2.1.2 Functional Stability Matrices (FSMs) and Functional Network Stability (FuNS)

The fast AMD metric offers critical advantage over the bootstrapped AMD method, as well as over the standard CC method, for quantifying functional connectivity measured over behaviorally-relevant timescales (i.e., hours to days). It allows rapid analysis of functional connectivity that can then be used to link neuronal activity with behavior.

The speed of the fast AMD metric is utilized to introduce Functional network stability (FuNS) as a way of measuring the dynamics of functional connectivity over time. Namely, we want to assess the stability of functional connectivity between the neurons within the network rather than to characterize the detailed network connectivity, which, again, is usually based on extremely sub-sampled systems. The remainder of this section is focused on characterizing the stability metric. Later, we show that changes in stability provide information about gross structural changes in the network.

Calculating the stability of network-wide functional connectivity patterns across time requires a division of the data sets into at least two time-windows; the remaining theoretical discussion assumes two time-windows for simplicity. The functional connectivity matrices are denoted as F^A and F^B where A and B represent the first and second time windows, respectively. The functional stability between these data sets is then calculated using cosine similarity, $\boldsymbol{C_{A,B} = cos\,\theta_{AB} = \frac{<F^A,F^B>}{\sqrt{<F^A,F^A>*<F^B,F^B>}}}$, with an absolute value of 1 denoting no change (maximum similarity) and 0 indicating great change (no similarity; orthogonality) between the time intervals (Figure 2.2a). Functional stability can thus be calculated in a pairwise manner across all time bins for a given recording in order to generate what we call a functional stability matrix (FSM; Figure 2.2b, see also Figure 2.8), or only on directly-adjacent time windows (Figure 2.2a), to generate a single measure of stability: $\boldsymbol{FuNS = \frac{1}{T}\sum_{t=0}^{T-1} C_{t,t+1}}$.

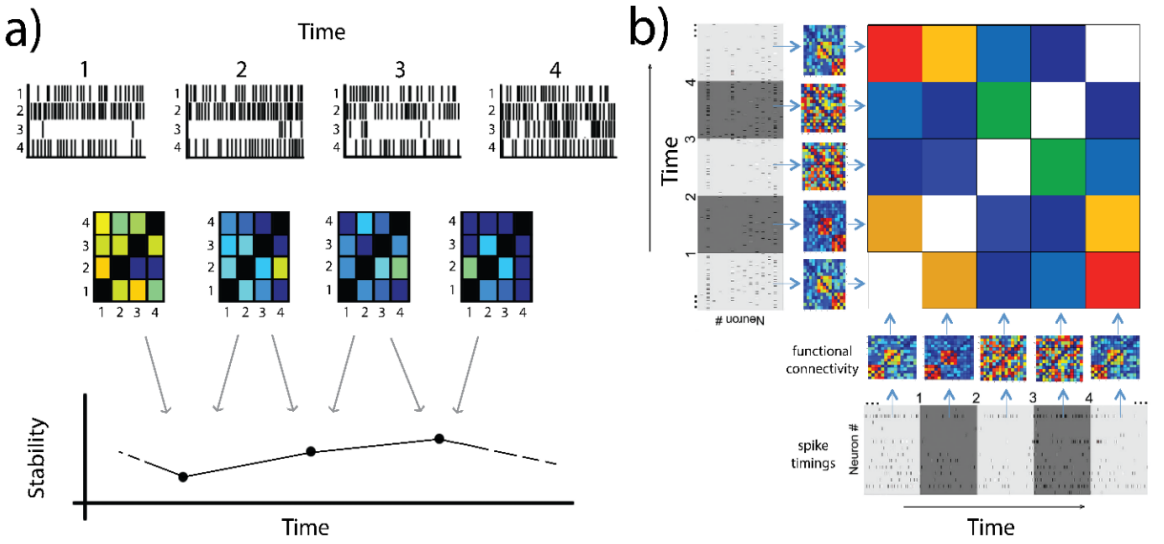

Figure 2.2: Calculation of Functional Network Stability and Construction of Functional Stability Matrices. a) Given the spike time series of neurons (top), the functional connectivity matrices (FCMs) are calculated over each interval (center), whereupon FuNS is calculated by measuring the mean cosine similarity between each consecutive time interval (bottom). b) Alternatively, similarity can be calculated in a pairwise manner across all time intervals to yield the functional stability matrix (FSM).

FuNS can also be used to determine the effect behavior has on neural network dynamics. In this scenario, stability is calculated before and after the presence of a synaptic heterogeneity (see Methods 2.2), $\boldsymbol{FuNS_{A,B}}$ and $\boldsymbol{FuNS_{C,D}}$, respectively. The significance of stability increase over many simulations is then given as a Z-score: $\boldsymbol{Z_s} = (\boldsymbol{\mu_{C,D}} - \boldsymbol{\mu_{A,B}}) \left(\frac{\boldsymbol{\sigma^2_{A,B}}}{\boldsymbol{N}} + \frac{\boldsymbol{\sigma^2_{C,D}}}{\boldsymbol{N}}\right)^{-\frac{1}{2}}$ with values greater than 2 indicating a significant increase in stability due to behavioral effects and values less than -2 indicating a significant decrease in stability. Here, μ and σ represent the mean and standard deviation of functional network stability, respectively, taken over many simulations or recordings.

2.2 Computer Simulations

2.2.1 Simulations of Integrate and Fire Networks

Neural activity is simulated using leaky integrate-and-fire model neurons with dynamics given by $\dot{\boldsymbol{V}} = -\boldsymbol{\alpha V} + \sum_j \boldsymbol{\omega_{ij} X_j} + \boldsymbol{I_\xi}$.

The summation represents the total input from recently fired (within ~20ms) pre-synaptic neurons with connectivity strength $\boldsymbol{\omega_{ij}}$ and input dynamics given by the double exponential

$\boldsymbol{X_j = exp(-(t - t_j^{spk})/3.0) - exp(-(t - t_j^{spk})/0.3)}$, where $\boldsymbol{t_j^{spk}}$ represents the timing of the last pre-synaptic spike.

In addition to synaptic input, each neuron receives noisy input $\boldsymbol{I_\xi = 0.15 + 10H(\xi - p)}$, where H is the Heaviside step function, $\boldsymbol{\xi = 10^{-5}}$, and $\boldsymbol{p} \in \{[\mathbf{0}, \mathbf{1}]\}$ is real-valued, random variable generated at every integration step from a uniform distribution.

Networks are formed using 1000 excitatory neurons arranged on a ring network. Connection densities range from 1.5% to 4.5% of the network population and connection weights range from ω = 0.02 to ω = 0.045 unless stated otherwise. The networks are initially connected locally and subsequently rewired with probability p_r. This parameter is varied from zero to unity, changing network topology from completely local connections to completely random. Each simulation is completed using the Euler integration method.

Additional network are simulated using a mixed population of excitatory/inhibitory cells. In this scenario, connections are completely local ($\boldsymbol{p_r = 0}$), have a connection density of 2%, and synaptic weights are pulled from a uniform distribution $\boldsymbol{\omega_{ji} \in \{[0, 0.2]\}}$. These networks follow the same dynamics as the excitatory only networks, except that 225 inhibitory neurons are added to the existing networks, evenly spaced among the excitatory cells, with inhibitory output connectivity strength $\boldsymbol{\omega_{ji}^* = -\beta\omega_{ji}}$. The variable $\boldsymbol{\beta}$ is used to investigate network dynamics when excitation or inhibition dominate. We calculate the ratio of excitation to inhibition, E/I, as the ratio between total excitatory to inhibitory synaptic input, averaged over all neurons not in the heterogeneity. Balance between excitation and inhibition (E/I ~ 1) occurs at $\boldsymbol{\beta = 3.0}$.

2.2.2 Introduction of Synaptic Heterogeneities and Their Long-range Dynamical Effects

Sensory input causes topological changes in anatomical network structure through both the strengthening and weakening of synapses [106, 107] as well as through the introduction of new synapses [108] and depletion of unused synapses [109]. Here, we focus solely on the strengthening of synaptic coupling for simplicity. The effect of synaptic strengthening is mimicked by introducing a discrete heterogeneity in network connectivity, i.e. a small, localized region spanning 10% of the network, with increased synaptic

connectivity between nodes. To simplify comparing networks with and without these synaptic heterogeneities, the underlying pairwise connectivity and synaptic strengths are conserved.

To analyze the potential long-range effects of such a heterogeneity, we calculate the mean synaptic distance to the heterogeneity for each neuron not in the heterogeneity. The mean synaptic distance here is the average number of steps that need to be taken from neurons in the heterogeneity to any other neuron in the network, along synaptic connections. The calculation of the distance is adopted from Newman 2010 [78]. In the simplest way, the synaptic distance between every neuron can be measured by calculating A^N, where A is the adjacency matrix and the power N is the number of synaptic steps necessary to reach every other neuron. With each successive multiplication of A, new non-zeros entries appear, representing new long-range (i.e. not directly connected), multi-unit synaptic connections. The synaptic distance d is the number of multiplications of A with itself, necessary to give rise to the new long-range connection. With the full synaptic distance matrix populated, the mean synaptic distance to the heterogeneity is calculated by averaging over all heterogeneity distances calculated for a given neuron. The mean synaptic distance to heterogeneity, and indeed between any two neurons, changes based on the size and connectivity density of the network. We thus normalize the distance to heterogeneity with a value of 1 representing neurons farthest from the heterogeneity, incorporating the entire network, and a value of 0 representing the minimum degree of separation from the heterogeneity (i.e. within the heterogeneity). It should be noted that by definition of d, the shortest distance to heterogeneity would be for a neuron not in the heterogeneity but connected to every other neuron within the heterogeneity, attaining a normalized value of $= \frac{\mathbf{1}}{\boldsymbol{N}}$.

2.3 Experimental Design

2.3.1 Recordings from Mouse CA1 Before and After Contextual Fear Conditioning (CFC)

To test the effects of memory formation on network dynamics in vivo, C57BL/6J mice (age 1-4 months) were implanted with custom-built drivable headstages (see [110,

111]) with bundles of stereotrodes targeting hippocampal area CA1. For a full description of the experimental procedure, please refer to Appendix A.1.

2.4 Benchmark Testing of the Analytical AMD Algorithm

We first compare the bootstrapped and fast AMD metrics for different distributions of ISIs (Figure 2.3): Gaussian, Poisson, uniform, and exponential. To measure the performance of the metrics, a single spike train following any one of these distributions is generated and cloned, with clones receiving a bidirectional jitter of their spike times equal

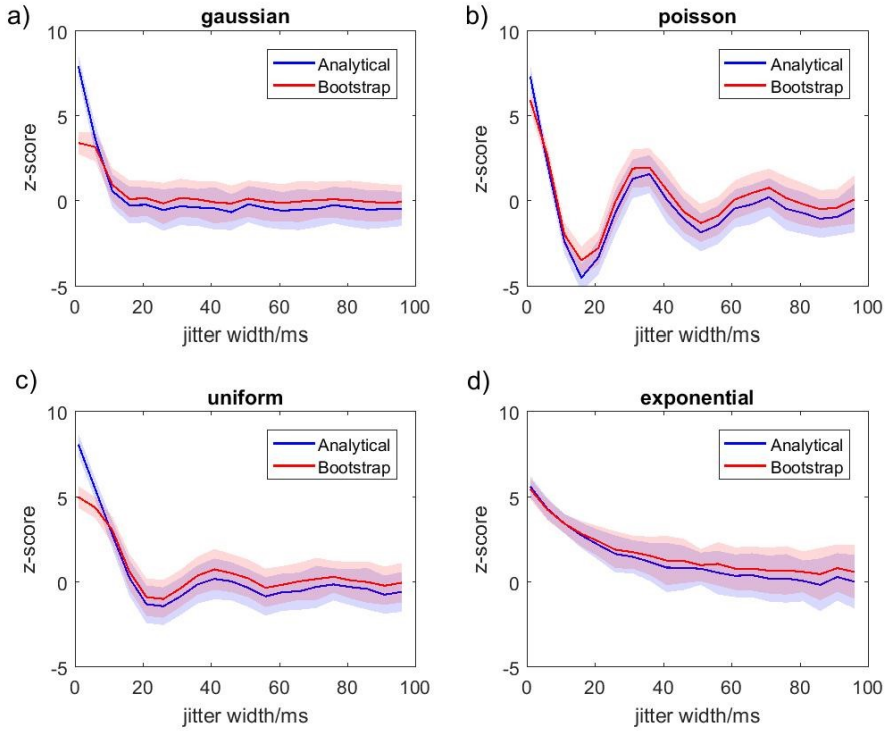

Figure 2.3: Comparison of bootstrapped and fast AMD metrics for rapid estimation of functional connectivity (FC). Two identical spike trains were artificially generated using various distributions of inter-spike intervals: a) Gaussian, b) Poisson, c) uniform, and d) exponential; the second spike train was jittered using the same type of statistical distribution, with various jitter widths (x-axis) to progressively de-correlate the spike trains. Each set of spiking data represents a 1s long recording (the time length is arbitrary, however all values are scaled to length) and contains 30 spikes. The analytical value of pairwise functional connectivity (FC_{21}) is calculated using the method described in the text (Methods 1.1). For all the distributions, AMD detects the significant functional connectivity when jitter width is small. The average value at which FC loses significance is a quarter of mean ISI, ~ 8ms. For a Poisson distribution (b), due to the fact that the mean value and standard deviation are controlled by the same parameter, when the jitter width equals around 17ms, the mean value of jitter is also around 17ms, and the maximal value of the AMD and therefore FC has the most negative value. The same reasoning applies when the jitter width is around 33ms. The significance from bootstrapping was obtained by shuffling the ISIs of the second train 100 times. As before, the Z-score of the AMD values represents the FC. The results agree with the analytical values.

to the jitter width depicted on the x axis (Figure 2.3). The jitter from every spike is drawn

from the same distribution as the original spike train, of which the standard deviation serves as the jitter width. For all cases, the mean ISI is arbitrarily chosen to be ~33ms (this ISI gives a 30Hz signal, representative of awake brain oscillations). Figure 2.3 depicts the mean z-score and its standard deviation, calculated as a function of the jitter width for the two approaches. In all four instances, the two AMD methods perform nearly identically.

Next, we compare the performance of fast AMD to CC, using the same distributions as above, i.e. Gaussian, Poisson, etc., with jittering (Figure 2.4). To calculate CC between two spike trains, the two spike trains are convolved with a Gaussian having one of three different widths, σ = 1ms, 5ms, or 33ms. Both metrics are calculated for 0 temporal shift between the spike trains. Importantly, we note that AMD does not have any free parameters and, at the same time, better captures finer characteristics for Poisson spike distributions compared to CC with any Gaussian convolution width.

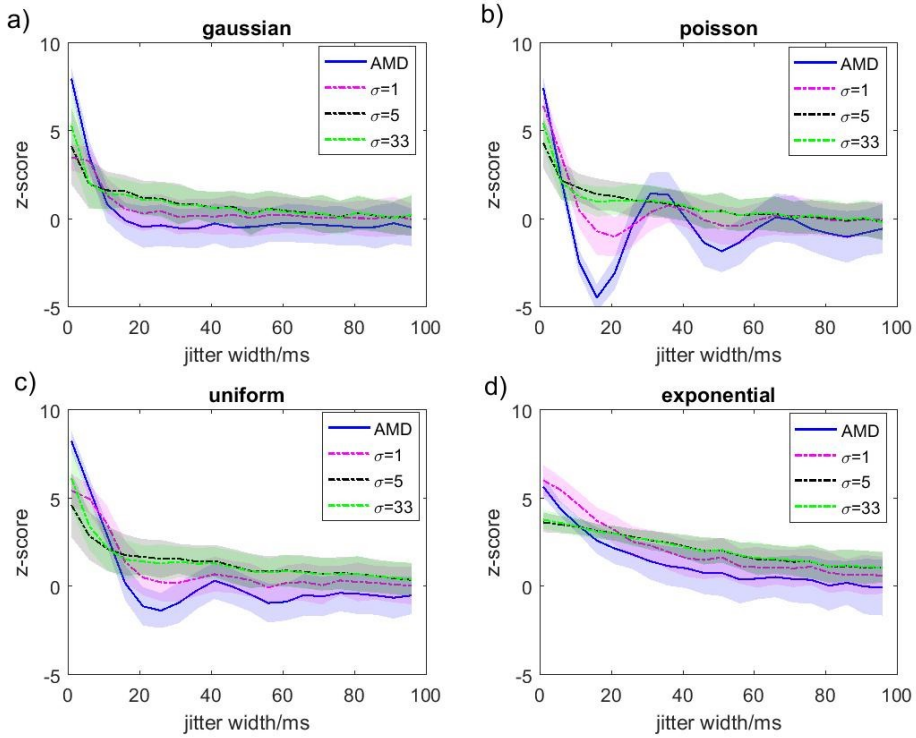

Figure 2.4: Comparison between fast AMD and CC. We compared the traditional cross-correlation (CC) method to fast AMD using a) Gaussian, b) Poisson, c) uniform, and d) exponential distributions, as in Figure 2.3. For the CC calculation, spike trains are convolved with a Gaussian waveform having a standard deviation σ as a free parameter. We used sigma σ = 1ms, 5ms and 33ms respectively. As before, the Z-score of CC was based on bootstrapping. AMD and CC results are equivalent for σ = 1ms. For larger σ, CC cannot capture the specific features of ISIs distributions, but behaves generally in a similar manner as AMD for increasing jitter width.

Critically, the fast AMD approach provides a rapid estimation of the significance of pairwise functional connectivity. Figure 2.5 shows the computing times of fast AMD, bootstrapped AMD, and CC with zero time-shift and bootstrapping for spike trains having various numbers of spikes and network sizes. The reduction of the computing time for fast AMD is very significant (up to 10000 times faster) which may be crucial for multiscale data analysis.

2.5 *Comparison of Bidirectional and Unidirectional AMD Performance*.

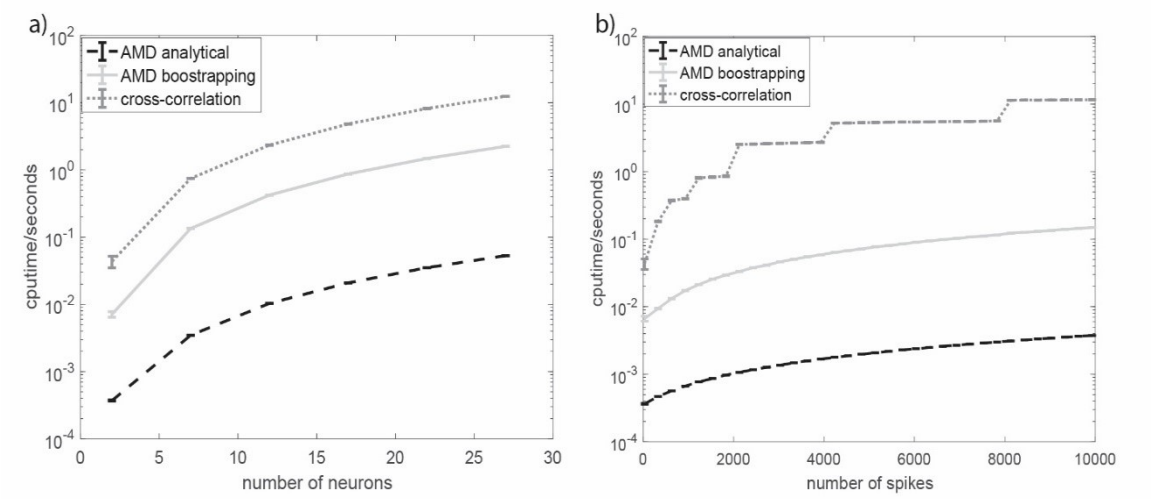

Figure 2.5: Comparison of computation speeds obtained for fast AMD, bootstrapped AMD, and bootstrapped CC. We measured the calculation time (recorded by CPU time from MATLAB) for three methods: CC, bootstrapped AMD and fast AMD. a) Calculation time for increasing the number of cells in the system. B) Calculation time for increasing the number of spikes in a two-cell system. Fast AMD is more than 20 times faster than bootstrapped AMD, and 200 times faster than CC calculation. For two-cell systems with different number of spikes (b), the advantage of fast AMD is more significant for larger spike trains, up to four orders of magnitude less than CC when the number of spikes is 10000. The sharp increases in CC computation time is most likely related to the memory allocation of the computer. The results for fast and bootstrapped AMD were averaged over 200 realizations, whereas 10 realizations were used for CC. The reported results are based on shuffling the ISIs 100 times for CC and bootstrapped AMD calculations.

Next, the performance of unidirectional fast AMD and bidirectional fast AMD on surrogate data sets is compared (Figure 2.6). A set of 5 spike trains are generated such that they are: 1) coincident (but not causal) with respect to each other, or 2) are causal, with FCMs calculated in each scenario. First, one spike train is generated from a Gaussian distribution. In the case of coincidence, the "master" spike train is copied and each spike is subsequently jittered following the same distribution. This process is repeated, with subsequent spike trains copying the previously jittered spike train. In the case of causality, copied spike trains retain the same interspike intervals as the original master copy, but are delayed slightly in time. Figure 2.6 depicts the result of bidirectional AMD (Figure 2.6a)

and unidirectional AMD (Figure 2.6b) estimation for coincident spike trains. As expected, bidirectional AMD reports highly significant temporal relations between the two trains whereas the unidirectional AMD estimation reports lack of causality (i.e., the significance is lower that one standard deviation). Figure 2.6c and 2.6d depict similar calculations for causally related spike trains. Here both measures report high temporal coincidence, however unidirectional AMD provides additional information about causal relationships.

2.5.1 Functional Stability between Functional Connectivity Matrices (FCMs).

We sought to determine how functional stability between FCMs can capture the similarity between different functional connectivity patterns in the network. Changing functional connectivity patterns are constructed by jittering five copies of a master spike train. For increasing jitter amplitude, all spike trains become increasingly de-correlated, resulting in different functional connectivity patterns. The FCM is first calculated using the

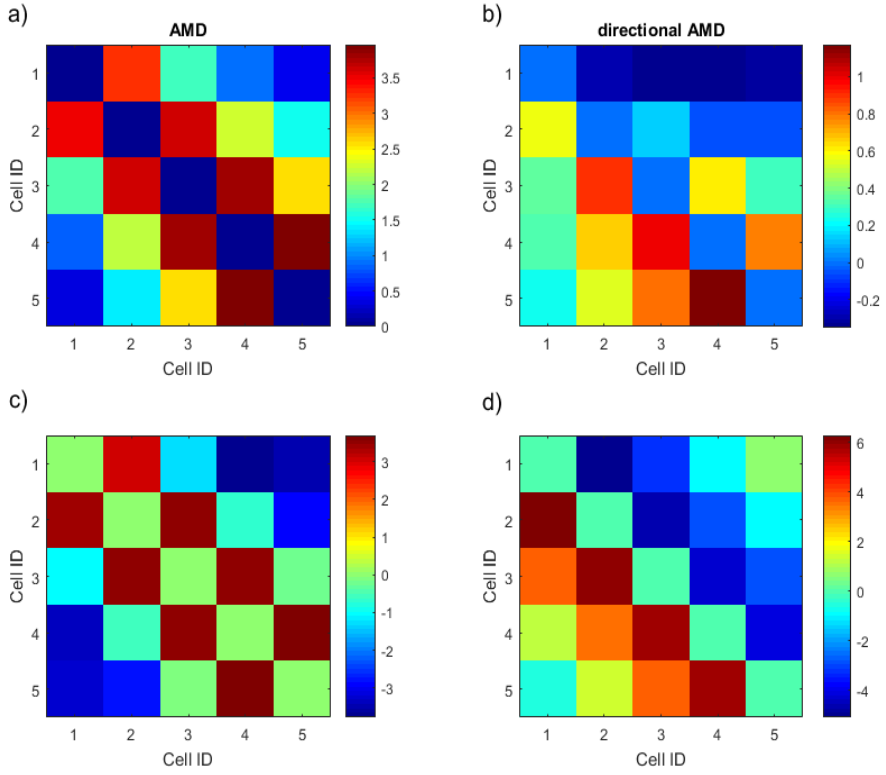

Figure 2.6: Bidirectional AMD and unidirectional AMD FCMs. An example of functional connectivity matrices (FCMs) calculated using two AMD methods for coincidence (a, b; bidirectional time lags taken into account) and causality (c, d; unidirectional time-lags taken into account) of functional connectivity (FC) patterns. Color represents the significance of fast AMD. In the case of coincidence, the FCM calculated by bidirectional AMD is almost symmetric and captures the functionally connected neurons (a), but unidirectional AMD does not (b); conversely in the causality case, the anti-symmetric FC matrix given by unidirectional AMD indicates the causal relationship (d), while bidirectional AMD does not differentiate from the coincidence case (c). The results were averaged over 100 realizations.

fast AMD metric for the five spike trains. Then, the stability between FCMs for different realizations of the spike trains having various jitter width is determined (Methods 1.3). Figure 2.7 shows the functional stability as a function of jitter of the compared spike trains. For small jitter, the FCMs yield stability values close to one, indicating high similarity

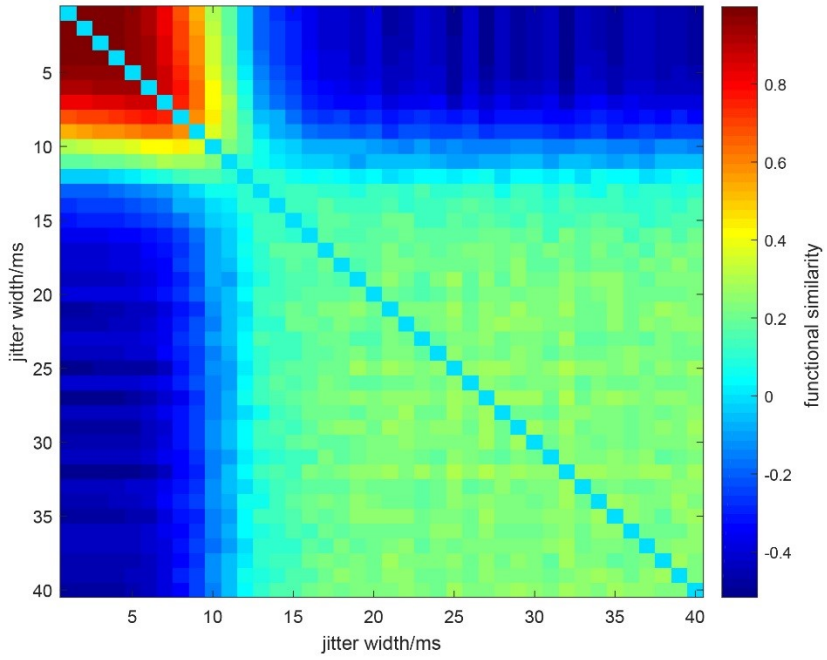

Figure 2.7: Similarity between FC patterns. A five-cell system is simulated, where the other four spike trains were jittered from the master train with same jitter width. Each train contains 30 spikes and time recording is set arbitrarily to 1 second. After calculating the functional connectivity matrix (FCM) for each jitter width, the similarity between each pair of FCMs is measured. The result is averaged over 100 realizations. Similarity is high when both jitter widths are small as the AMD values are small for both cases. There is a transition to negative values as one of the jitter widths gets significantly larger. For the pair of FCMs, both with high-valued jitter width, FC patterns are relatively random and similarity is low (~0.2).

between the FCMs. On the other hand, when a small jitter FCM is compared to a high jitter FCM, similarity rapidly declines to negative values. This is due to switching from a well-defined network structure to a random one. Finally, when two largely random states are compared (i.e. both FCMs have high jitter and are de-correlated) the stability value hovers around 0.2. Taken together, these results indicate that functional stability reasonably quantifies the similarity between functional connectivity in the network.

2.5.2 Functional Stability Matrix (FSM) and FuNS as a Monitor of Changes in Functional Connectivity Patterns

Following the data generating procedure used in Figure 2.8, a five cell system is simulated to demonstrate the applicability of the Functional Stability Matrix (FSM) (Figure 2.8) in monitoring changes in dynamical network states over time. A bidirectional jitter with a width of 8ms is applied during the first and last 7 seconds of the spike train, while a bidirectional jitter of width 15ms (Figure 2.8a and 2.8c) or unidirectional jitter of width 8ms (Figure 2.8b and 2.8d) is applied during the middle 7 seconds. After segmenting the time series into 21 bins of equal size and calculating the 5-by-5 FCMs using the fast AMD algorithm, the FSM is obtained by calculating the functional stability between each pair of FCMs (Figure 2.8a and 2.8b). For both cases, significantly positive stability values in region I and III and low values in region IV indicate the temporal relationship between different functional connectivity patterns in the network. Importantly, region V in both cases demonstrates that the functional connectivity returns to the same pattern observed in the first 7s, subsequent to the changes occurring during the 8-14s time window. In the bidirectional case, the network loses stability during the middle 7s in region II (Figure 2.8a), while in unidirectional case region II (Figure 2.8b), due to the corresponding unidirectional shifts, the stability between FCMs attains a high value. Hence, FSM gives effective information to keep track of the similarity in functional connectivity patterns in the network at any time point. Figure 2.8c and 2.8d illustrate the functional stability trace over time, with the red line indicating FuNS, i.e. the mean of the stability values (0.4362 for bidirectional and 0.7720 for unidirectional). As expected, the minimum similarity in both cases happens at the point when FC changes, at the end of 7s and 14s, respectively.

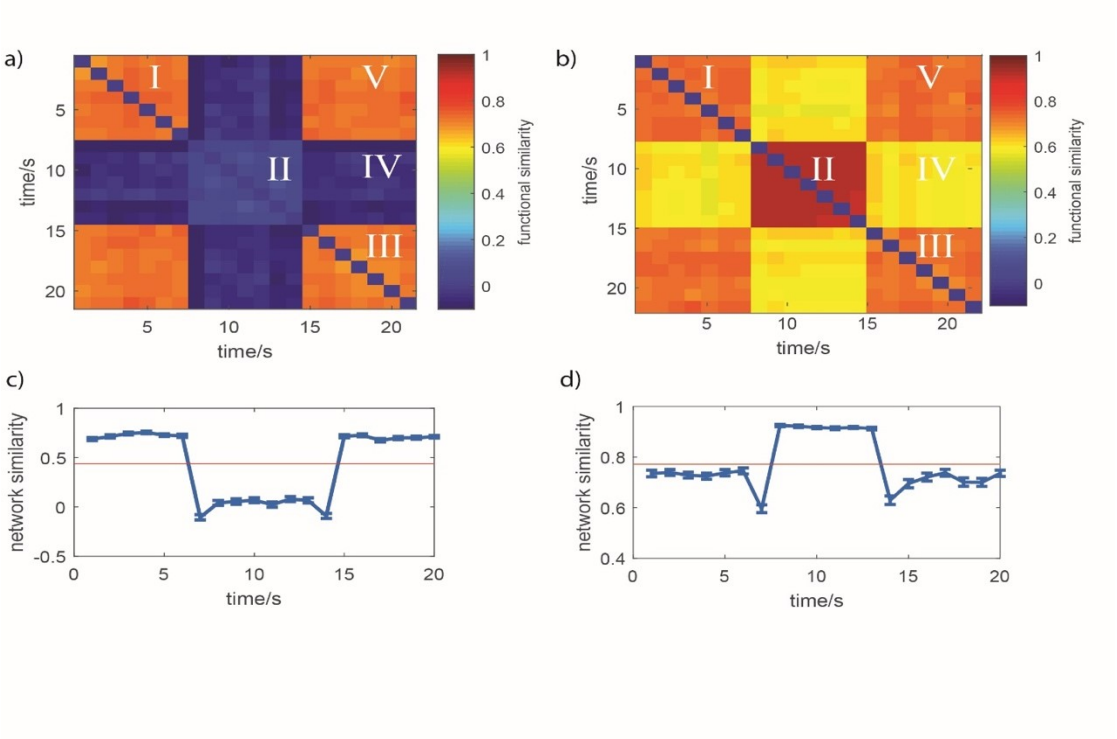

Figure 2.8: Functional Similarity Matrix (FSM) and similarity trace over time. Simulation of temporal changes in spike relationships between five neurons. The spike trains are jittered bidirectionally with a jitter width of 8ms from the master train during the first and last 7 seconds of the spike train. During the middle 7 seconds, jitter was bidirectional with width 15ms (a,c) and unidirectional with width 8ms (b,d), respectively. The spike trains were binned into 21 time windows and a five by five functional connectivity matrix (FCM) was calculated by bidirectional AMD for each window. (a,b) Similarity value between each pair of FCMs. FCMs originating from spike trains having common properties show high similarity (c,d). Similarity trace over time, within which only FCMs in adjacent time windows are compared. Red line indicates the functional network stability (FuNS), the average of the similarity trace values.

These results thus give a reliable way to track functional network changes in time, which may be due to cognitive processing, for example.

2.5.3 *Estimation of Fast AMD for Functional Connectivity for Mutually Delayed Spike Trains*.

We tested the performance of fast AMD on spike trains with applied time delay (Figure 2.9). Two random spike trains with Gaussian ISIs are generated with a jitter width of 5ms. Time delay is added to the second train by shifting each spike time by a constant value. In Figure 2.9a, the FC and standard deviation between two trains are estimated by fast AMD for different time delays. Around a delay of 7.5ms, FC is around 0 due to the fact that the second train is shifted to one quarter of the average ISI (33ms). FC values become negative with the increase of the delay time, indicating an anti-correlation between two trains.

Next, fast AMD is utilized to detect the delay and to recover the original, non-delayed z-score. The estimated delay time (DT) from S_i to S_j, as given by fast AMD, is defined as $\boldsymbol{DT_{ij}} = \frac{\mathbf{1}}{\boldsymbol{N_j}}\sum_{\boldsymbol{k}}(\boldsymbol{t_k^j} - \boldsymbol{t'^i_k})$, where $\boldsymbol{t_k^j}$ is the temporal value of the k^{th} spike in S_j and $\boldsymbol{t'^i_k}$ refers to the temporal value of the nearest event to $\boldsymbol{t_k^j}$ in S_i. Then, S_j is shifted by $-DT_{12}$ and FC is re-calculated using fast AMD. The red line on Figure 2.9a depicts the FC values after the shift, and as a function of original delay time. Fast AMD reliably detects the delay and restores FCs back to the level of no delay (indicated by the black dash line). As a comparison, we also calculated the FC with and without subtracting the estimated DT for non-delayed spike trains (Figure 2.9b). There is no significant difference after subtracting DT, indicating that no spurious correlations were introduced for non-delayed spike trains.

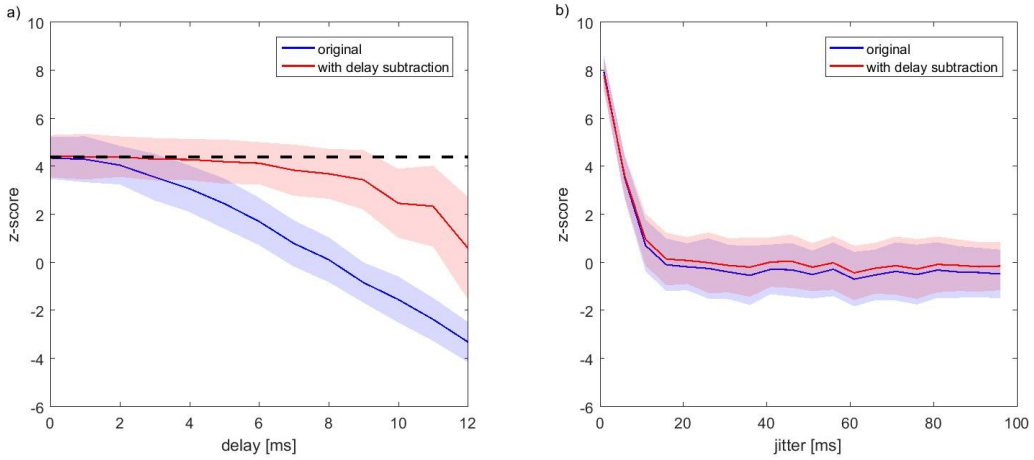

Figure 2.9: Fast AMD can be adjusted to account for time delays. A copy of the original random spike train having Gaussian ISIs is jittered with a jitter width of 5ms and then is shifted by variable time-delay. a) The FC between two trains is estimated by fast AMD (blue trace). Next the second train is shifted back by the amount of delay that is estimated by fast AMD algorithm, and FC is re-calculated (red trace). The black dashed line shows the FC for non-delayed spike times. b) The same analysis is applied to spike trains with no delay, and the FCs show no significant differences. The results were averaged over 100 realizations.

To further test performance of the tools on the delayed spike trains, we calculated FuNS for a 5-cell system with jittering and applied variable time-delay (Figure 2.10); with mean delay time in the system denoted on the y-axis. The total recording time duration was 10s. FuNS was calculated from 10 equal length time bins. The top row indicates the system without delay. When the system is strongly connected (i.e., a small jitter width), FuNS is highly robust to delays, reporting nearly identical values as the case without delay. For bigger jitter width, as expected, FuNS is low when the delay time is around one quarter of

the average ISI, i.e. when the FC between spike trains loses significance. Thus, even though FC values can be affected by delays, FuNS can still quantify the stability level of the system effectively.

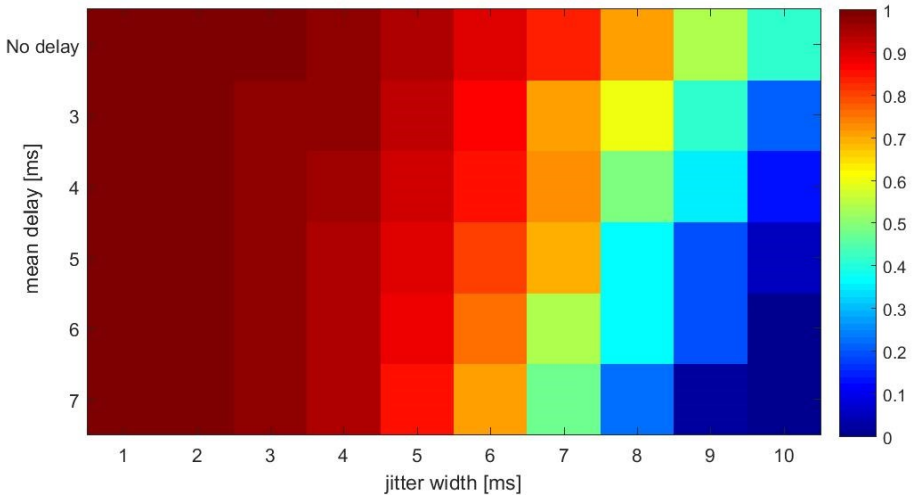

Figure 2.10: Functional Network Stability (FuNS) of the delayed dataset. A 5-cell system is simulated by adding jitter and time delay to spike trains to randomly generate spike train using Gaussian distribution of ISIs. The color scale represents FuNS calculated after binning the data into 10 one-second time windows. A control realization, where no delay is added to the spike trains, is indicated by the top row. The results were averaged over 10 simulations. FuNS reports robust stability despite the delay.

2.6 Effects of Localized Network Heterogeneity in Model Networks

Using the statistical tools introduced above, we investigate networks of leaky integrate-and-fire neurons for dynamic stability. The focus here is to establish how the new metrics help to elucidate network connectivity structure, as well as potential changes in network dynamics, due to the formation of localized network heterogeneities. As noted previously, these heterogeneities represent the formation of localized cognitive representations (e.g. memories) within the network.

2.6.1 *Identification of Direct Structural Connections within the Network.*

We first constructed sparsely connected, excitatory only networks to investigate whether, and for what ranges of connectivity parameters, is it possible to statistically separate sets of neurons with direct structural connectivity from those who lack direct connections. This corresponds to adjacency matrix entries of 1 and 0, respectively. We use the bidirectional, fast AMD metric to measure the functional connectivity between pairs of neurons that share direct structural connections and those that do not. The distribution of

FC values are then characterized (i.e. their mean and variance are calculated) for the two populations and we subsequently calculate the statistical separation between groups in terms of a Z-score: $\boldsymbol{Z_s = (\mu_{wc} - \mu_{nc})\left(\frac{\sigma_{wc}^2}{N} + \frac{\sigma_{nc}^2}{N}\right)^{-\frac{1}{2}}}$, where μ_{wc}, σ_{wc} and μ_{nc}, σ_{nc} represent mean and standard deviation of the distributions of functional connectivity values for directly coupled pairs and non-coupled pairs, respectively. Figure 2.11 shows the Z-Score comparison between these two populations (Figure 2.11a). Each colored panel represents the statistical separation of the two populations as a function of network topology for increasing synaptic connectivity. The obtained results indicate that there is a well-defined parameter region where the two populations can be separated with a large degree of accuracy. As expected, weak network connectivity prohibits this separation (Figure 2.11b). Also, the statistical significance is lower in networks deviating from local to random connectivity (Figure 2.11b-f). Importantly, significance between the groups is seen even under very strong connectivity, though eventually the response is saturated and no new network parameter values result in an increase in significance (Figure 2.11e and 2.11f).

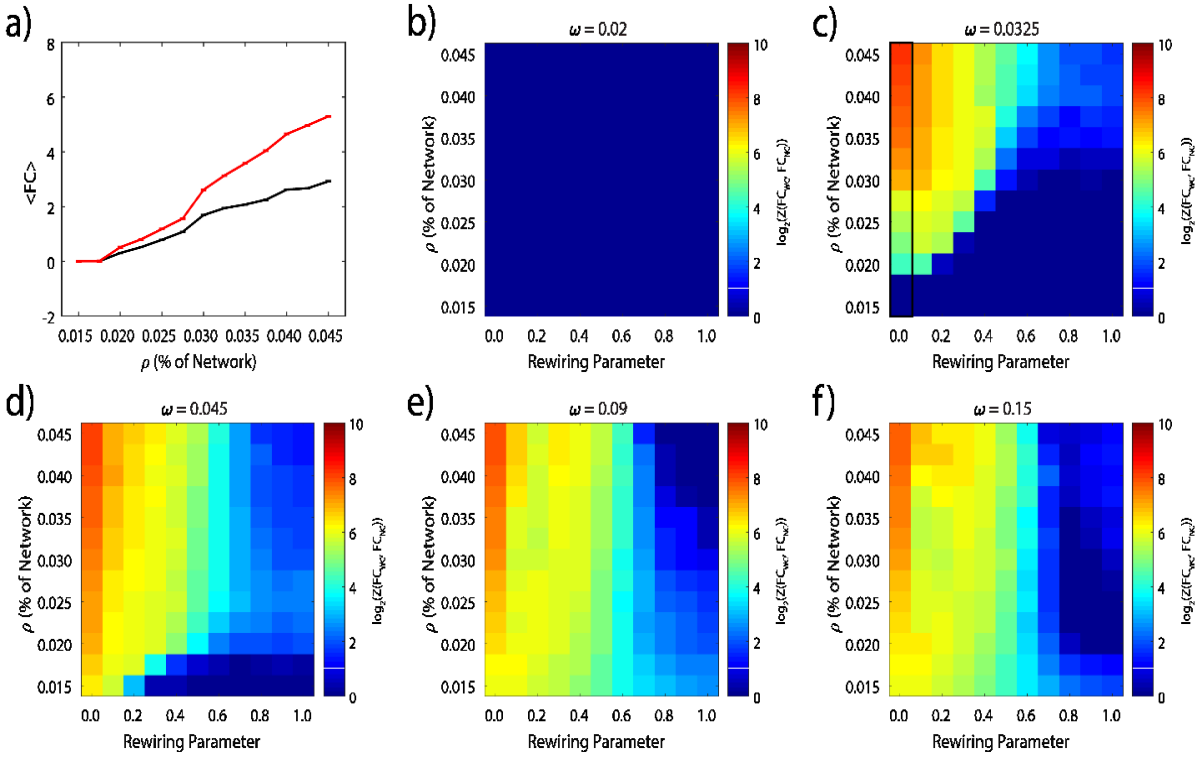

Figure 2.11: Z-Score significance between functional connectivity matrices as a function of network topology. Functional connectivity matrices (FCMs) were parsed based on the existence or non-existence of synaptic connections between neurons and fast AMD results for these two groups were generated. a) Mean grouped-averaged functional connectivity as a function of connectivity density for a connection strength of ω = 0.0325 and rewiring parameter equal to 0 (red traces: directly connected neurons; black traces: unconnected neurons). Error bars represent standard error of the mean. Lack of variation in network structure (i.e. there is no rewiring of local connections) results in uniformly small standard error; the network for each simulation is exactly the same and so responds to random input in nearly the same manner. b-f) Color images indicate the logarithmically scaled significance, with warmer colors indicating a greater significance, with the white bands indicating the level above which the Z-score is significant (consistent with two standard deviations from the mean). As synaptic connectivity strength ω increases from very low values (b) through moderate values (c) to higher values (d), significance increases between the parsed groups over an increasingly large topological parameter region. As ω further increases, more than half of the parameter region has a significant separation between groups (e) but saturates, admitting no additional significant parameters (f). The black box in panel (c) indicates the range of data used to generate panel (a). All results were averaged over five trials.

2.6.2 *Changes in Functional Connectivity and Stability of the Network with Introduction of Network Heterogeneity.*

It is still not clear how localized changes in network structure (i.e. inclusion of a network heterogeneity) affect network-wide dynamics. To address this, functional connectivity and the subsequent stability of these matrices is measured between the neurons that are not included in the heterogeneity, using the fast AMD method. Simulations are cut into two parts and we subsequently measure both the change in FC as well as FuNS, both given as a function of network topology and connectivity density (Figure 2.12). Figure 2.12a depicts FuNS in the same network before the heterogeneity is introduced (black line)

and after its introduction (red line) as a function of the connection rewiring parameter. Significant changes in network stability are observed for localized network topologies with significance decreasing as the topologies become more random. Figure 2.12b depicts changes of network stability upon the introduction of a heterogeneity, as a function of both connectivity density and network topology, and compares it to changes in mean value of FC, averaged over all pairwise indices of the corresponding FCM, for the network (Figure 2.12c). We note that while FuNS changes are quite significant for a wide parameter range (up to Z-score of 64, noting the logarithmic scale), the changes in mean functional connectivity are quite insignificant and provide a less clear picture of how the FCM itself changes. This leads us to conclude that measuring the changes of FuNS is a more tenable indicator of global change in network dynamics in response to introduction of network heterogeneity compared to FC.

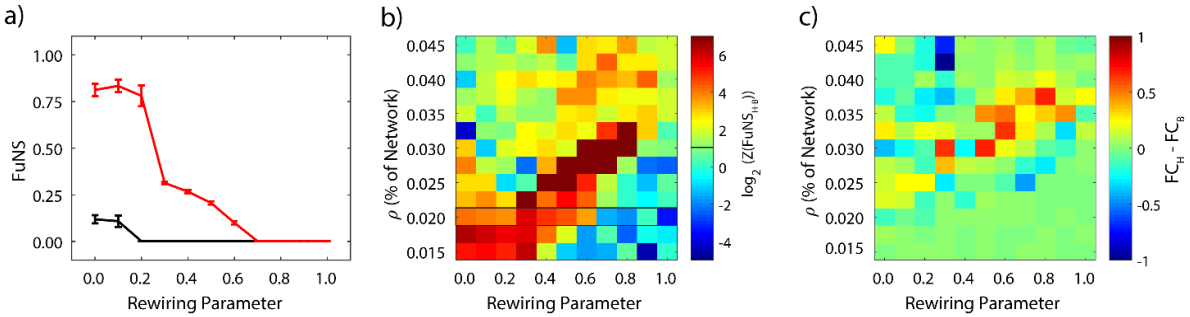

Figure 2.12: Functional Network Stability detects dynamic changes due to synaptic heterogeneities over a large topological parameter region. a) FuNS as a function of connection rewiring parameter for networks before (black trace) and after (red trace) introduction of a synaptic heterogeneity. Synaptic heterogeneities are defined as spatial regions within the network, where connections between neurons only in the region were appointed a greater synaptic connectivity compared to the rest of the network. Error bars indicate the standard error of the mean. b) Z-score of FuNS as a function of connection density ρ and rewiring parameter, scaled using a logarithm of base two. Warmer (cooler) colors denote an increase (loss or no change) in stability due to the introduction of a synaptic heterogeneity. The black bar on the color scale indicates the minimum value needed to be considered significant. The black box in the main panel shows the parameter region used to generate FuNS curves in the left panel. c) Difference in average FC over the entire FCM as a function of ρ and rewiring parameter is less robust than analyzing FuNS. All results shown are for a synaptic coupling strength of A = 0.03, averaged over five trials.

2.6.3 *FuNS as a Global Measure of Structural Network Changes*.

We have shown above that FuNS is sensitive to the introduction of a discrete network heterogeneity. Thus, it allows the identification of the existence of structural network changes without the requirement of measuring specific cells that participate in that

change. This is of paramount importance in the situation when the experimental

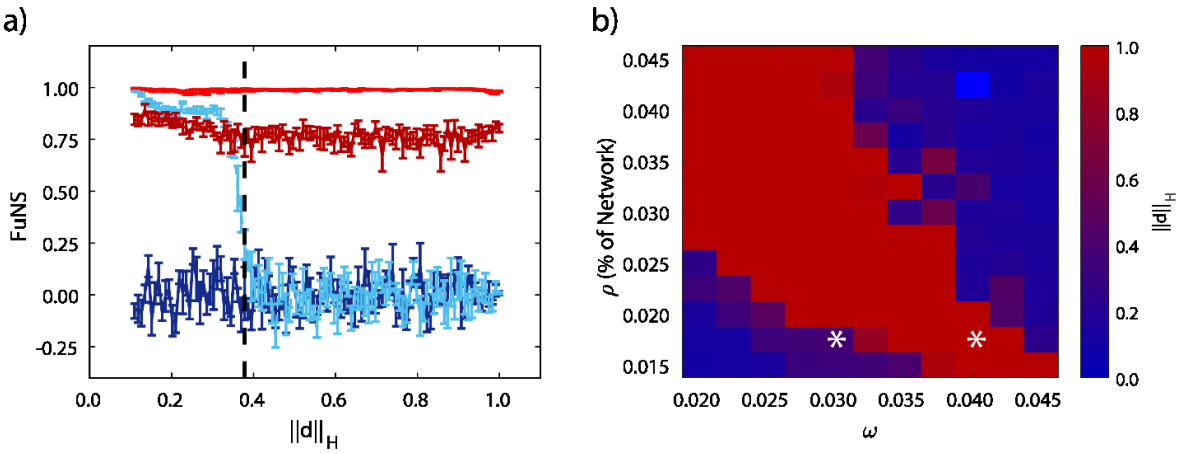

Figure 2.13: Local synaptic heterogeneities globally increase FuNS. a) Example FuNS traces as a function of normalized synaptic distance from the heterogeneity for networks before (darker colors) and after (lighter colors) introduction of a synaptic heterogeneity. Some values of the simulation parameters result in a distance dependent decrease or no change in FuNS Z-scores (blue traces), while others result in consistent, network-wide significance (red traces). The black, dashed line indicates the normalized distance where FuNS loses significance in the example case shown. Error bars indicate standard error of the mean. b) Normalized distance from the heterogeneity where FuNS significance is lost. Values of one indicate that the global network observes an increase in FuNS due to a localized synaptic heterogeneity. All results averaged over five trials.

measurement is critically under-sampled and there is no way to identify either the specific neurons participating in the structural network reorganization or the anatomical network structure. To quantify the long-range effects of synaptic heterogeneities, we set out to measure the synaptic distances from network heterogeneity where significant changes in network stability are observed.

Neurons are grouped depending on their mean synaptic distance from heterogeneity (Methods 2.2). Functional connectivity matrices for each group of cells is calculated separately, whereupon we determine the mean change in FuNS within each group due to introduction of heterogeneity (Figure 2.13). Figure 2.13a shows an example of change in FuNS as a function of mean distance from heterogeneity, normalized by the maximum possible distance to the heterogeneity. Some network parameters results in a persistently significant separation of FuNS at long distances from the heterogeneity, while other parameters result in a rapid decline of FuNS away from the heterogeneity. Thus, the Z-score of FuNS is calculated between networks with and without synaptic heterogeneity at each synaptic distance in order to determine the normalized distance where significance is lost. Figure 2.13b depicts the normalized mean distance from the heterogeneity at which the results become insignificant, as a function of connection density and strength. Here, a

value of one corresponds to the situation where we can detect changes in FuNS throughout entire network. We observe that localized heterogeneity has global dynamical effects on the system under a large array of network topologies, giving credence to the notion of dynamical attractors in neural networks.

2.6.4 *FuNS Sensitivity to Structural Heterogeneity in Mixed Excitatory and Inhibitory Networks*.

Finally, we measure changes in FuNS in response to introduction of network heterogeneity in mixed inhibitory and excitatory networks. Specifically, FuNS is measured as a function of the ratio of total excitation and inhibition generated by neurons in the network (i.e. E/I ratio; Methods 2.1). Generally, we observe that for low values of E/I ratio the reported FuNS is low regardless of the presence of a heterogeneity and, at the same time, a high E/I ratio saturates FuNS in both cases (Figure 2.14). The greatest response of the networks, in terms of stabilizing dynamics in presence of heterogeneity, is near a balance between excitation and inhibition, i.e. E/I ~ 1 (Figure 2.14b). Thus, only near such an E/I balance can the dynamics of the network respond in a distributed manner to the introduction of heterogeneity. This provides another piece of evidence that mixed networks near E/I balance increase their dynamic range in response to even localized structural

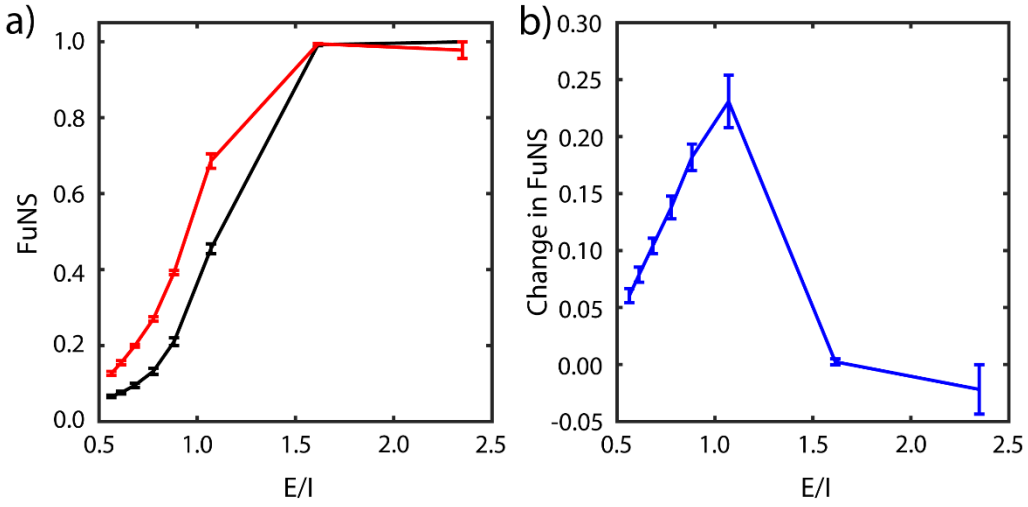

Figure 2.14: Introduction of synaptic heterogeneities maximize increased Functional Network Stability near a balance between excitation and inhibition. a) FuNS as a function of the ratio between excitation and inhibition. Introduction of synaptic heterogeneities (red traces) increases stability over networks missing a synaptic heterogeneity (black traces). b) Difference in FuNS between networks containing and not-containing synaptic heterogeneities. All error bars indicate the standard deviation of the mean, taken over five trials.

network changes, in agreement with previous studies [21, 103].

We further compare FCM and FuNS measurements between the fast AMD approach and the CC approach (Figure 2.15) at the point where FuNS observes a maximum increase in Figure 2.14b, i.e. E/I ~ 1. As expected, the FCM analysis for both methods is very similar and, indeed, does not show a significant difference between networks with and without a synaptic heterogeneity (Figures 2.15a and 2.15b). However, we observe a significant increase in FuNS for the fast AMD method (Figure 2.15d) but not for the CC method (Figure 2.15c), assuming a non-normalized Gaussian distribution for both. Thus, though the resulting FCMs are similar, FuNS more accurately picks up on discrete changes

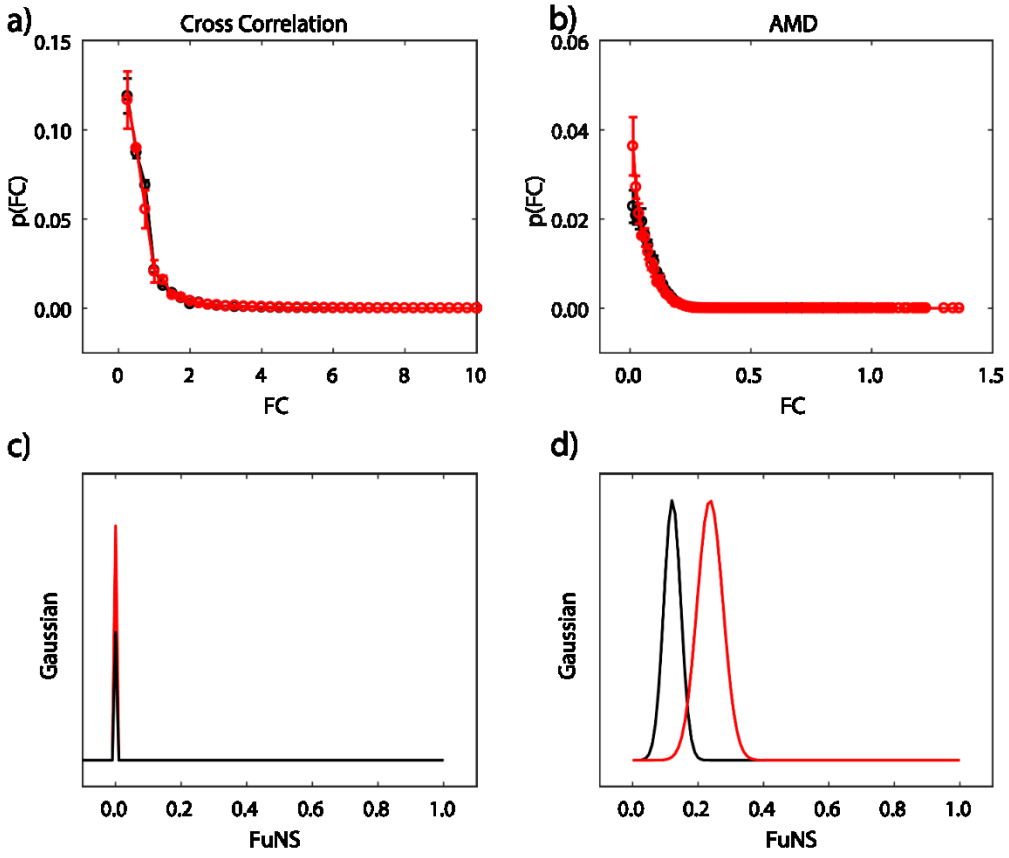

Figure 2.15: Comparing FC and FuNS between AMD and CC near the E/I Balance. The probability of observing a mean FC value was measured for functional structures from both CC (a) and AMD (b) derived methods, only over the excitatory neurons in the mixed networks, before (black) and after (red) adding a network heterogeneity. The distributions were not significantly different, within a 5% confidence interval (K-S test; CC: $p = 0.83$, AMD: $p=0.54$). Similarly, non-normalized Gaussian distributions of FuNS were constructed for CC (c) and AMD (d) before (black) and after (red) introduction of a synaptic heterogeneity. Calculating FuNS for AMD yielded significantly different distributions whereas FuNS for CC did not, within a 5% confidence interval (K-S test; CC: $p = 0.09$, AMD: $p = 7 \times 10^{-9}$). Error bars represent standard error of the mean, whereas Gaussian widths stem from the standard deviation. Averaged over 5 trials.

in functional network topologies generated using AMD compared to cross-correlation.

2.7 FuNS Applied to in vivo Data

Finally, we wanted to know whether functional connectivity and stability changes could be detected following network reorganization in vivo (Figure 2.16). We hypothesize that synaptic plasticity in hippocampal area CA1 following single-trial contextual fear conditioning (CFC) [112] is a plausible biological model to investigate how rapid structural network changes underlying memory formation affects network dynamics. CA1 network activity is necessary for fear memory consolidation in the hours following CFC [113]. For this reason, we recorded the same population of CA1 neurons from C57BL/6J mice over a 24-h baseline and for 24 h following CFC (placement into a novel environmental context, followed 2.5 min later by a 0.75 mA foot shock) to determine how functional network dynamics are affected by de novo memory formation. CFC affects many aspects of CA1 network dynamics; for a detailed description of the obtained results, please refer to Appendix A.1.

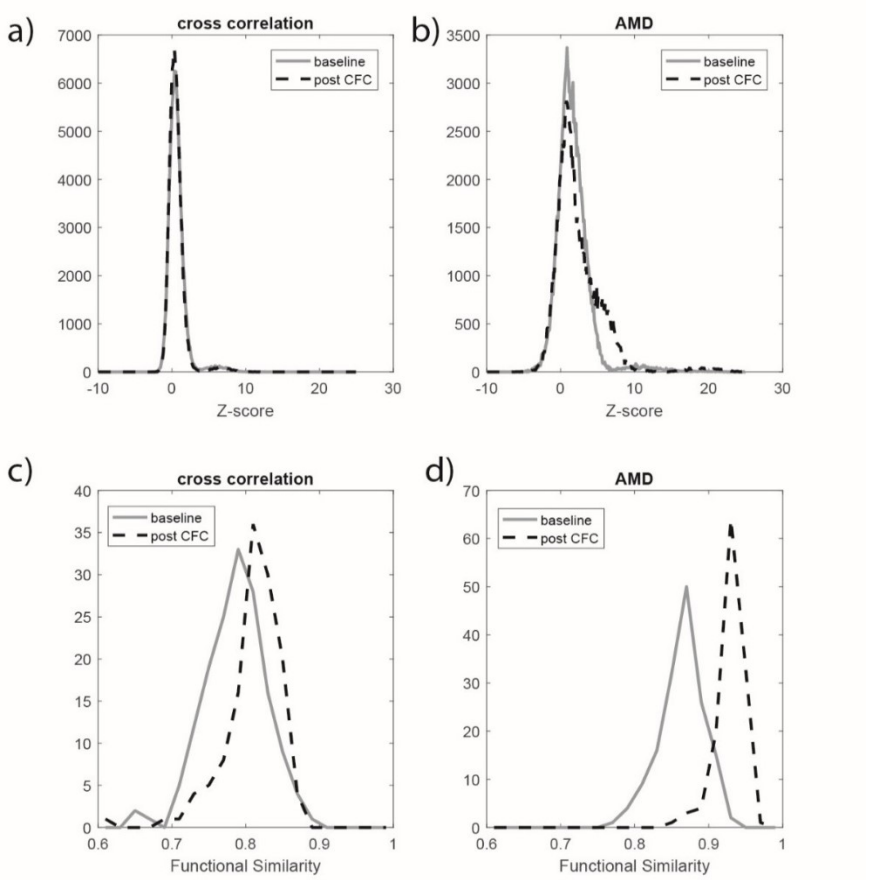

Figure 2.16: Application of AMD and FSM to in vivo mouse data. We extracted spike data from intervals of slow wave sleep across 6-hour recordings for both before (baseline) and after (post stimulation) the contextual fear conditioning. The spike trains were first divided into multiple one-minute bins, then the functional connectivity (FC) pattern for each bin is calculated by bidirectional AMD and CC. We compared the distribution of both FC values (a,b) and stability values (c,d). For FC values, the elements are extracted collectively from all the FCMs. The histogram shows us that AMD is able to capture the functional connectivity changes from baseline to post contextual fear conditioning more sensitively than CC. For stability values, the elements were extracted from the FSMs for baseline and post-stimulation respectively, and calculated as described in Results 1.4. We observe significant shift in stability for both CC and AMD calculation, however, AMD gives a more statistically significant separation between the two distributions.

The results presented here focus on comparing performance of the metrics (fast AMD and CC, both together with FuNS assessment) for the case when mouse was subjected to successful memory consolidation (success was determined by observing behavioral changes 24 hours after training). First, the 6-hour baseline and 6-hour post stimulation are divided into 1-minute time windows, and FCMs are calculated in each bin, which are further used to calculate FSM. Figure 2.16 shows comparisons of the distribution of functional connectivity values (Figures 2.16a and 2.16b) and stability values (Figures 2.16c and 2.16d). Comparing with CC, AMD is shown to be more sensitive to capture the change

of functional connectivity and stability in the network during memory consolidation. Furthermore, the more significant shift of similarity distribution indicates that stability is a better measurement of the change in global network properties.

2.8 Discussion

The advent of new recording techniques allowing for prolonged recordings from an increasing number of neurons in the brain drives the necessity to develop new analysis tools to meaningfully process data. Two underlying issues however need to be overcome. First, there is a severe under-sampling problem: how is it possible to identify universal properties of neuronal dynamics during information processing if the number of recorded cells remain tremendously small in comparison with number of cells participating in the computation? Second, and related to the first question, is how to characterize the data so that the (small) recorded population provides a representative picture of the dynamics of whole modality? Solutions to the latter attempt to bridge the timescales between neuronal activity and behavioral states which they encode, while prolonged recordings on freely behaving mice are now possible, they generate enormous data sets which need to be meaningfully processed in a finite amount of time.

In this paper we have addressed both of these problems - we have introduced a framework, based on the AMD between spikes in individual neurons' recorded spike trains, which allows for rapid assessment of network functional connectivity structure throughout extended time periods. We showed that we can extend the developed metrics so that we can rapidly estimate significance of functional connectivity between neuronal pairs, based on analysis of distribution of ISI intervals of the neurons in question, not only without loss of resolution, but often with improved sensitivity as compared to cross-correlation based methods. At the same time, rapid assessment of significance allows us to speed up functional connectivity reconstruction by a couple of orders of magnitude, primarily due to the fact that we can bypass typical bootstrapping methods without loss of accuracy (Results 1.1-1.5).

Further, we used this fast, AMD-based method to reconstruct instantaneous functional connectivity within the network and subsequently introduced Functional Network Stability (FuNS), a measure that assesses the temporal stability of functional connectivity networks. We showed that FuNS is especially useful in detecting changes in

network-wide dynamics due to discrete changes in network structural connectivity, referred to here as synaptic heterogeneities. Namely, we show that localized and relatively small heterogeneities can induce dynamical changes throughout the entire network, as is evidenced by high FuNS in neuronal groups distantly connected to the heterogeneity region (Figure 2.14). This in turn allows for robust detection of such changes experimentally, even in the conditions of severe under-sampling [111]. These results indicate that while reconstruction of functional connectivity between the recorded neurons may yield ambiguous results as the functional relation of the recorded cells to the computational task is unknown, the changes in the global dynamics of the representations is a more robust measure of local network changes in response to computational tasks. (Results 2.1-2.4) To better exemplify this point, we used both model simulations and in vivo experimental recordings to show that discrete changes to network structure may yield ambiguous results in terms of reconstruction of detailed changes in functional network connectivity, but at the same time show robust stabilization of dynamical network representations (Results 3.1).

Finally, we investigate whether observed stabilization of dynamical network representations can inform us about universal network properties that are underlying the computation. Here, we show that in mixed excitatory-inhibitory networks, the highest sensitivity (in terms of changes in global network representations) to introduction of localized heterogeneity is achieved near a balance between excitation and inhibition (E/I balance; Results 2.5). This result is in line with other existing results which have shown that E/I balance emerges naturally in neural networks [114] and that neurons operating in networks near E/I balance exhibit faster linear responses to stimulation, and greater dynamic range [114]. Recent findings have also shown that E/I balance is required for heightened neuronal selectivity [105].

Altogether, we believe that the introduced framework for rapid calculation of functional network connectivity allows for robust analysis of multiunit recordings. Numerous linear and nonlinear, methods have been developed over the last decade to reconstruct and characterize functional network connectivity. We have earlier compared the performance of functional grouping based on AMD assessment to some of these methods [100]. Many of the developed tools require assessment of functional adjacency

matrix. We believe that the algorithm proposed here provides a robust alternative for the commonly used cross-correlation method. Further we believe that fast AMD together with evaluation of FuNS helps to overcome two major constraints in neuroscience: under sampling and the difficulty of bridging diverse timescales of neuronal dynamics and cognition. We believe that this framework will be widely applicable to numerous problems in systems neuroscience.

Chapter 3

Critical dynamics mediate learning of new distributed memory representations in neuronal networks

In this chapter, I explore the role of critical dynamics of neuronal networks as they pertain to the encoding of new information. I show these dynamics are necessary for new information to be encoded and subsequently consolidated but that consolidation shifts dynamics toward sub-criticality. I go on to show that these features, critical dynamics during learning and sub-critical dynamics post-learning, are distinguishable in mouse hippocampal related to fear memory consolidation. This work was published in the journal Entropy in 2019 [63].

Phase transitions and critical phenomena are of central importance to statistical physics and there is growing evidence that also it can play a crucial role in living systems [29, 115, 116]. Here we investigate how near critical network dynamics may recruit neurons and facilitate formation of a new distributed memory in a situation where the incoming input must compete with the already stored (native) memories for neuronal resources.

It is widely hypothesized that new information is encoded in brain circuits through activity dependent, long-term synaptic structural changes [106] which are a putative substrate for memory formation [117, 118]. While features of memory traces can be localized to specific cell populations (e.g., location information encoded in place cell activity), in general, tracing so-called "engrams" to neural circuits has been an elusive task [117]. Attempts at disrupting well-established memories through brain lesions [11] or, more recently, through optogenetic silencing [119] have shown that they are robust to alterations in communication between individual neurons or brain areas. A parsimonious and longstanding explanation of these phenomena is that a process termed "systems consolidation" leads to diffuse, widespread memory encoding and storage. However,

despite more than a century of study, it is not well understood how engrams are initially formed and subsequently stored across vast distances (in terms of numbers of synaptic connections between neurons) in the brain.

A major problem to understanding the mechanisms for systems consolidation is that very little is known about how the formation of new memories (i.e., learning) impacts neural network dynamics. The general, long-accepted assumption is that either strengthening of existing synaptic connections, or the de novo creation of additional synapses (i.e., formation of a discrete structural heterogeneity), leads to the formation of a dynamical attractor [17]. If this is the case, then the dynamical state of the network must support long-range correlations across the network, as the number of neurons actively involved in encoding a specific memory trace is thought to constitute only a small fraction of the total neuronal population [12]. Moreover, individual synapses in regions like the hippocampus have a surprisingly brief lifetime (~1-2 weeks on average [13]), necessitating rapid dissemination and consolidation of information. These requirements raise two questions: 1) how do permanent and widely-distributed neural engrams form from initial, transient changes to a discrete subset of the network's synapses during learning, and 2) what mediates transformation of local representations of disparate features to global memory representation? New experimental [111] and computational work shows that theta band oscillatory patterning and/or dynamics associated with sharp wave ripples can mechanistically coordinate neuronal activity recruiting them into the representation [61].

In this work, we show computationally that in addition to large scale temporal pattering of neuronal activity, near-critical dynamics in the brain could be an important factor in facilitating memory consolidation. Specifically, we show that storage of new information that is weakly and/or sparsely impinged on the network is mediated through plastic, state-dependent changes in network connectivity and can be successfully consolidated (which is associated with attractor formation) near criticality - a point associated with second order phase transitions [1]. This storage is followed by a subsequent shift from critical to sub-critical dynamics.

The idea that the brain operates at or near dynamical criticality is not new (see Ref. [29, 120, 121] for comprehensive reviews) and it was experimentally observed in in vivo and in vitro preparations [3, 27, 58, 103, 122-127]. A large body of work also investigated

the potential functional benefit of operating in a near-critical regime [21-23, 30, 128]. Here, we specifically identify very basic, underlying importance for the brain to reside near criticality and demonstrate that near-critical dynamics may be essential for a system-wide consolidation of new memories in a situation when the sensory input is weak and/or sparse in comparison with signals generated by memories native to the network (i.e. those previously stored).

To substantiate these hypotheses, we analyze in vivo recordings associated with contextual fear memory (CFM) consolidation. Contextual fear conditioning (CFC) is an optimal experimental paradigm in this regard as it allows for rapid formation and consolidation of memory (i.e. after single-trial learning) [110, 111]. In this particular case, the CFM consolidation is associated with normal sleep, which has been shown to play a vital role in various types of memory consolidation [110, 111, 129, 130]. Here, we first characterize hippocampal dynamics in mice subjected to CFC and show that: 1) the hippocampus operates in a near critical regime pre- and post-CFC training, and 2) successful, behaviorally-verified consolidation of fear memory leads to an underlying shift in hippocampal dynamics towards a subcritical state, similar to what we predict in our model simulations.

Together, these results indicate that novel learning may occur preferentially near a critical regime and leads to universal widespread stabilization of network activity patterns, which in turn drives the formation of widely-distributed engrams (i.e., systems memory consolidation).

3.1 Attractor Neural Networks Elucidate Mechanisms for Storage and Consolidation of New Memories

We modeled a neuronal network with easily controllable dynamics using a mean-field, Hopfield-like formalism [17]. In this context, instantaneous neuronal states are modeled as binary variables, $\boldsymbol{S_i = \pm 1}$, corresponding to a firing (+1) and a quiescent (-1) neuronal state, respectively. Instantaneous states are updated based on a neuron's input

$$\boldsymbol{h_i = \frac{1}{k}\sum_{j=1}^{k} J_{ij} S_j}\ , \tag{1}$$

which serves to align that neuron's state with that input, so that $\boldsymbol{sgn(S_i) = sgn(h_i)}$ with probability

$$\boldsymbol{P(h_i, \beta)} = \frac{\mathbf{1}}{\mathbf{1}+\boldsymbol{exp}(-\mathbf{2}\boldsymbol{\beta}|\boldsymbol{h_i}|)}, \quad (2)$$

where k is the incoming degree of each neuron, $\boldsymbol{sgn(x)}$ is the sign function, and J is the connectivity matrix, discussed in detail below. The term $\boldsymbol{\beta} = \boldsymbol{T^{-1}}$ is a control parameter which directly controls the dynamical state of the system: when $\boldsymbol{\beta} \ll \mathbf{1}, \boldsymbol{P(h_i, \beta)} \rightarrow \frac{\mathbf{1}}{\mathbf{2}}$ and, conversely, $\boldsymbol{\beta} \gg \mathbf{1}, \boldsymbol{P(h_i, \beta)} \rightarrow \mathbf{0}$, with the critical point typically located near $\boldsymbol{T} = \mathbf{1}$. These dynamics describe properties of the standard Hopfield model in the absence of an external field [17, 131, 132].

The network we use here consists of $\boldsymbol{N} = \mathbf{10000}$ neurons, arranged in a directional, small-world network (10% chance of rewiring a local connection) [133] with ~2% incoming connectivity, but with no self-connections allowed. Initially the network is seeded with $\boldsymbol{p}$ native memories (hereafter collectively referred to as the native state and designated by the superscript n) defined by a random configuration of states $\{\boldsymbol{\xi_i^n} = \pm\mathbf{1} \mid \boldsymbol{i} \in [\mathbf{1}, \boldsymbol{N}]\}$ for each memory, and with the weighted connectivity matrix indices defined as

$$\boldsymbol{J_{ij}} = \frac{\mathbf{1}}{\boldsymbol{p}}\sum_{\boldsymbol{\mu}=\mathbf{1}}^{\boldsymbol{p}} \boldsymbol{\xi_{i,\mu}^n \xi_{j,\mu}^n}, \quad (3)$$

for all incoming connections (hence, $\boldsymbol{J_{ii}} = \mathbf{0}$).

With $\boldsymbol{p}$ memories already embedded in the network through Eq. 3, we want to investigate how the network responds to, and possibly consolidates, a new representation with randomly configured states similar to the native memories, $\{\boldsymbol{\xi_i^e} = \pm\mathbf{1} \mid \boldsymbol{i} \in [\mathbf{1}, \boldsymbol{N}]\}$ (the superscript e hereafter representing a configuration of states associated with the new memory). However, throughout evolution of the network the new representation influences only a small subset, $\boldsymbol{N'}$, of network neurons; here, the neurons belonging to $\boldsymbol{N'}$ are randomly selected from the full network. The instantaneous states of these neurons do not change throughout the simulation and are set to $\boldsymbol{S_i(t)} = \boldsymbol{\xi_i^e}, \forall\, \boldsymbol{t}, \boldsymbol{i} \in \boldsymbol{N'}$. In addition, a connection emanating from these input neurons is modified to:

$$\boldsymbol{J_{ij}} = \frac{\boldsymbol{w^e}}{\boldsymbol{p}}\boldsymbol{\xi_i^e \xi_j^e}\ \forall\, \boldsymbol{j} \in \boldsymbol{N'}, \quad (4)$$

where the $\boldsymbol{w^e}$ term is the additional weight of the connections corresponding to the new state relative to the native connectivity (Eq. 3). We formulated the input in this way to mimic real biological processes of learning and memory. The subset of the input neurons

and their corresponding connectivity is to roughly represent the memory backbone formed rapidly during the presentation of the new input (associated with the new representation). At the same time the freezing of the dynamics of these neurons is to correspond to input constancy during the experience, or, reactivation of these neurons during sleep that was observed experimentally [134, 135].

Although the majority of neurons in the network encode for the native memories, those receiving input from neurons representing the new state will align with it if the new state is fractionally stronger than the native state at any time. The competition between the native and new states are encapsulated in the total input a neuron receives,

$$\boldsymbol{h_i(t) = \sum_{j \notin N'} J_{ij} S_j(t) + w^{ext} \sum_{l \in N'} J_{il} \xi_l^e \equiv h_i^n(t) + h_i^{ext}}, \qquad (5)$$

where $\boldsymbol{h_i^n(t)}$ is given by Eq. 1 and $\boldsymbol{h_i^{ext}}$ represents input from the constant external field, here facilitated by fixed neuron states.

We assessed the presence of attractors in the network by measuring the overlap of the final state of the network with one of the native configurations ($\boldsymbol{m_\mu^n = \left|\frac{1}{|N-N'|}\sum_{i \in N-N'} \xi_{i,\mu}^n S_i\right|}$) and/or the new configuration ($\boldsymbol{m^e = \left|\frac{1}{|N-N'|}\sum_{i \in N-N'} \xi_i^e S_i\right|}$), where the averages are over all non-fixed neurons in the network (i.e. the relative compliment of $\boldsymbol{N}$ and $\boldsymbol{N'}$).

3.1.1 Critical Dynamics Mediate New Memory Encoding and Consolidation through Plasticity

First, we examined the overlap of the network with the new configuration (red solid line in Figure 3.1a; with $\boldsymbol{|N'| = 700}$) and the native ones (black solid line in Figure 3.1a) when only one native memory is stored in the system ($\boldsymbol{p = 1}$) as a function of temperature T. The universal dynamical properties of the system at criticality maximizes the susceptibility to the external input at the critical point. In this work, we define $\boldsymbol{T_c}$ as the point where the order parameter (i.e. $\boldsymbol{m_\mu^n}$) reaches a half-maximal value (when quantified, it is found via linear interpolation between point pairs). This point coincides with the half maximal value achieved by stability, another order parameter of the system (black line, Figure 3.3b). Here, the critical point has been well-characterized as a second order phase

transition that separates the phase of high-stability dynamics ($\boldsymbol{T < T_c}$; characterized by convergence to stored attractors) from disordered dynamics ($\boldsymbol{T > T_c}$) [132]. The result is a well-defined regime, where the network overlap with the new configuration, impinged on the system through external input, is higher than that of native memories. When the system is sub-critical ($\boldsymbol{T < T_c}$), the overlap with one of the native memories dominates the system. In contrast, the super-critical network ($\boldsymbol{T > T_c}$) is in a disordered state where neither the native configuration nor the new configuration dominates dynamics. At ($\boldsymbol{T \sim T_c}$) the attractor associated with native memory becomes unstable, and at the same time magnetic susceptibility peaks making the overlap of the network with the new representation significantly higher. However, if the states of the input neurons are set to the values of the natively stored configuration, i.e. $\boldsymbol{\xi_i^e = \xi_i^n, i \in N'}$ (during a memory recall event, for example), the stability of the native memory is extended over the critical range (Fig. 1a dashed black line), shifting the phase transition towards higher temperatures. Thus, depending on the input configuration, at ($\boldsymbol{T \sim T_c}$) both, native memory can be stabilized or new memory representation can be fractionally successfully impinged on the system. This theoretically provides the network with agility to store a new memory or to retrieve a known one. Such shift away from criticality in presence of structured input was also observed in self organizing recurrent networks (SORNs [136]), and may explain slightly subcritical brain state observed in vivo [27].

We next wanted to investigate how proximity to the external input (through numbers of connections) effects the corresponding overlap, $\boldsymbol{m^e}$ for different temperature ranges. We measure fractional overlap of the final network state with the new configuration as a function of the number of connections that neurons receive from the external input neurons (Figure 3.1b); those neurons receiving higher native input should align with the native configuration, whereas neurons with higher non-native input should be driven to adopt the new configuration (under the right dynamical state, given by the control parameter $\boldsymbol{\beta}$). We observe, as predicted, that the mean overlap of neuronal states with the new configurations is significant and highest (Figure 3.1b, red curve) for neurons receiving the external input at criticality, as compared to sub-critical (black points), and super-critical (blue points) regimes. Thus, at criticality, as opposed to sub-critical and super-critical

regimes, even sparse and/or weak input can lead to global changes in the network, providing a plausible mechanistic explanation for the distributed nature of memory traces.

We investigated whether application of a type of activity-dependent synaptic plasticity rule observed experimentally [7] can lead to consolidation of the new configuration. Here by consolidation we mean whether a) the overlap between the new configuration and stability of the network in the presence of input can be increased, and b) whether the stable (in absence of the external input) attractor representing the new configuration can be successfully formed.

We implement these synaptic changes in the model by introducing state-based changes in connectivity strengths,

$$\Delta \boldsymbol{J_{ij}(t)} = \boldsymbol{\varepsilon S_i(t) S_j(t)}. \quad (6)$$

During the learning phase, both the neural states and the connections were updated (with $\boldsymbol{\varepsilon = 0.1}$), with the exception of neurons pertaining to the external input (i.e. those neurons remain fixed and so receive no relevant input).

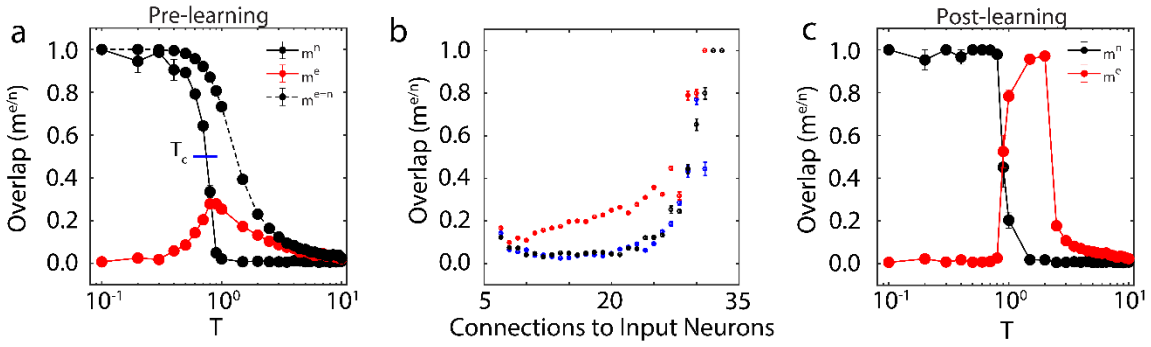

Figure 3.1 New memory consolidation occurs only near criticality. a) Overlap of the system with the native configuration without external input (solid black line) and with external input (dashed black line), as well as overlap with the new configuration (red) represented by external input, as a function of temperature before learning. Note that maximal susceptibility of the new configuration only occurs near the initial critical temperature of the system, where overlap with the native configuration declines. Here, we define the critical temperature to be the temperature where the order parameter (Overlap) reaches its half-maximal value, as indicated by the blue line. b) Overlap of the new configuration after learning for neurons grouped based on their number of connections to the input. Colors represent pre-learning sub- (blue) super- (black) and critical (red) temperatures. c) Overlap of the system with the native (black) and new (red) configurations as a function of system temperature after learning. Few changes in overlap occur before the initial critical temperature, after which (near criticality) the system aligns to the new configuration. Note also that the new configuration overlap occurs for larger values of temperature, indicating consolidation and a shift in critical temperature due to learning. All error bars in (a-c) represent the standard error of the mean.

We investigated the range of the control parameter, T, for which the network is able to successfully store the new configuration (i.e. the emergence of a new attractor with a large value of $\boldsymbol{m^e}$). We found that the system successfully consolidated the new configuration starting near $\boldsymbol{T_C}$, indicated by an increase in $\boldsymbol{m^e}$ post-learning (Figure 3.1c).

This shows that new memory consolidation occurs only near criticality, when susceptibility to external input drives the increased overlap with new configuration (Figure 3.1c). In addition, we observed that consolidation shifts $\boldsymbol{T_C}$ to higher values of T, causing an initially critical regime to become sub-critical. These changes (due to the unbounded learning rule) lead to an increase of the overall magnitude of synaptic coupling, resulting in a stronger external field and ultimately leading to a peak in $\boldsymbol{m^e}$ at $\boldsymbol{T > T_C}$ after learning.

3.2 Memory Consolidation at Criticality is Robust to Input Type and Strength

We next investigated how the consolidation depends on the number of input neurons, i.e. the size of $\boldsymbol{N'}$, and the magnitude of the weight of the connections stemming from the input (Eq. 4). We varied both parameters and monitored maximal change in magnitude of new-state overlap from pre-learning (as exemplified on Figure 3.1a) to post-learning (Figure 3.1c). These results are presented in Figure 3.2a – one can observe that the number of input neurons can be as small as 4% of the total network size to observe meaning full change in the overlap over the rest of the network. Conversely the $\boldsymbol{w^{ext}}$ can be as low 2.3 to observe increase of the overlap. Hence, even weak and sparse input can have noticeable impact on network dynamics, but only near criticality. The asterisk represents the parameter configuration used to generate results presented on Figure 3.1.

Up to this point, we have examined network response to external input represented by fixed neuronal states in the network. Alternatively, instead of the new memory being represented by specific neurons, we can represent the new memory as persistent input to all neurons in the network by defining an extra term for the observed input,

$$\begin{aligned} \boldsymbol{h_i(t)} &= \boldsymbol{h_i^n(t) + h_i^{ext}} \\ \boldsymbol{h_i^{ext}} &= \boldsymbol{\frac{w^{ext}}{p}\xi_i^e} \end{aligned} \quad , \tag{7}$$

with $\boldsymbol{h_i^n(t)}$ again being represented by Eq. 1. We ran the simulation in the presence of fixed external field applied to all neurons with learning (a pre-learning phase followed by a learning phase), followed by an additional phase with $\boldsymbol{h_i^{ext} = 0}$ and then subsequently calculated the difference in the final overlap between the new and the native memory. We found that the system only consolidates the new configuration given sufficiently high external field strength, and only near the critical temperature (Figure 3.2b). Hence, the system is able to adapt to the new configuration, regardless of its source, only near the

critical regime, in support of previous studies [22]. Importantly, higher field magnitude increases the range of temperatures for which the new memory is consolidated (top of Figure 3.2b).

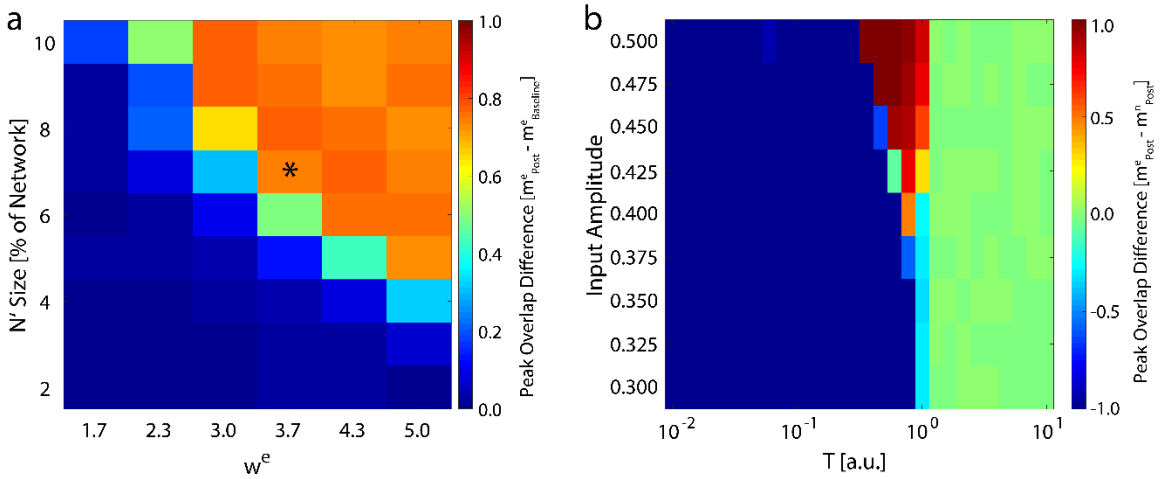

Figure 3.2 Robustness of new memory consolidation as a function of input strength. a) Peak change in overlap of the new state between pre- and post-learning as a function of input size (percentage of fixed nodes in the network) and strength (w^e). The asterisk represents the parameters used to generate the data showed in Figure 3.1. b) Change in overlap (color) between the new and native configurations post-learning as a function of temperature for increasing values of external field strength applied during learning. Blue colors represent cases where the native configuration is still stable after learning, red colors are where the new configuration is stable, and green is where neither configuration is stable. Note that for sufficiently high field strength, we see a slight increase in the maximal critical temperature.

3.3 Consolidation of New Memory Subsequently Shifts Dynamical State towards Sub-criticality

These results show evidence of the possible importance of near-critical dynamics in storing new memories. To be more robust with our findings, we further examined the properties of new memory consolidation. We first calculated the amount of time (i.e. the number of iterations) needed for the network to align with the new configuration, so that $\boldsymbol{m^e \geq \frac{1}{2}}$ (Figure 3.3a). Near the initial (i.e. pre-learning) value of $\boldsymbol{T_C}$, only a fraction of the learning time was required to consolidate the new configuration, and with increasing T, the consolidation time increased exponentially before abruptly increasing to an interval greater than the simulation time. In contrast, sub-critical and super-critical states were marked by prohibitively long consolidation periods (left- and right-hand sides of Figure 3.3a, respectively).

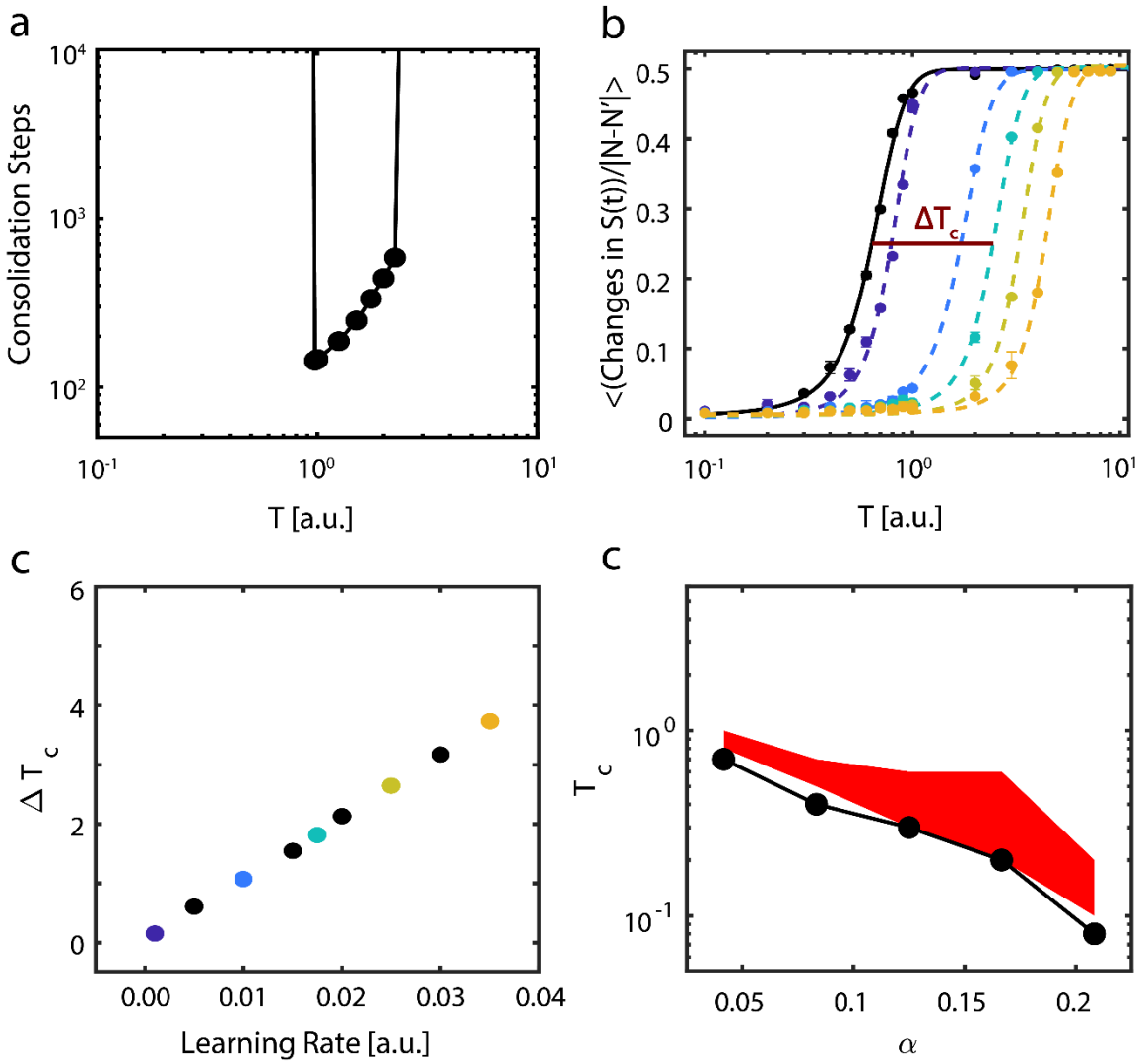

Figure 3.3 Dynamical properties of consolidating new information. a) Time (steps) required for the system to consolidate the new configuration, as a function of temperature. Values not shown (on the left and right sides) indicate timescales longer than the simulation runtime, *i.e.* that it takes a prohibitively long time to consolidate a new memory. b) Data and fit sigmoidal functions for mean number of changes in the neurons' state S_i per iteration as a function of temperature, pre- (solid black line) and post-learning (dashed lines); the learning rate ε increases left-to-right and from darker to lighter colors of the dashed lines. Error bars represent the standard error of the mean. The horizontal line labeled Tc represents the half-maximal point of the where we calculate the critical temperature via linear interpolation. c) Change in observed critical temperature $\boldsymbol{T_c}$, calculated using the sigmoidal half-maximum values (b) as a function of the learning rate ε. Colors correspond to the curves shown in (b). d) Critical temperature $\boldsymbol{T_c}$ as a function of the memories per degree distribution $\boldsymbol{\alpha}$ before (black points) and after (red shaded region) learning for a new-state connectivity strength of $\boldsymbol{w^e = 3.0}$. Note that the minimal value of the critical temperature for the new configuration post-learning closely matches the critical temperature pre-learning, but that the effect of learning is a broadening of the stable regime.

Next, we examined how changes to the learning rate ($\boldsymbol{\varepsilon}$ in Eq. 6) affects both the consolidation of the external input representing the new configuration and the dynamical properties of the system. To assess the transition point (i.e. $\boldsymbol{T_C}$), we measured the network's

configuration stability, $\boldsymbol{f(T)}$, as a function of temperature for different values of $\boldsymbol{\varepsilon}$. Stability here is defined as the mean number of changes from active (+1) to quiescent (-1) states occurring in the network for fixed simulation length; the expected number of these activity changes is 0 in the sub-critical regime and ~N/2 in the super-critical regime, due to Eq. 2. We subsequently fit sigmoidal functions to the transition numbers as a function of temperature, taking the form $\boldsymbol{f(T) = \frac{0.5}{1+exp(-\frac{T-T_{c,i}}{\mu})}}$, where the slope μ represents the change in stability due to changing regime and we designate $\boldsymbol{T_{c,i}}$, the temperature where the transitions reach their half-maximum value, as a proxy for critical temperature (Figure 3.3b); as previously mentioned, $\boldsymbol{f(T)}$ is thus an order parameter of the system. We next calculated the change in the critical temperature due to learning, $\boldsymbol{\Delta T_{c,i} = T_{c,i}(t_0) - T_c(t_{final})}$, and found that consolidation of new information shifts the stability, and therefore the critical regime, of the system approximately linearly with the learning rate (Figure 3c).

Finally, we investigated the behavior of the system when it is loaded with multiple native configurations, and when the location of the critical point is a function of both memory loading α and temperature T [131]. We thus pre-loaded additional native configurations into the network. It is known that memory recall fails for T=0 at $\boldsymbol{\alpha_{max} = \frac{p_{max}}{N} \sim 0.14}$ (with $\boldsymbol{p_{max}}$ being the maximal number of configurations stored and N number of neurons in the network) for a fully connected network [132], but this value changes for a sparsely connected system and is proportional to nodal degree k, $\boldsymbol{\alpha_{max} = \frac{p_{max}}{<k>}}$.

We found that regardless of the number of memories pre-loaded into the system (below the loading limit), successful consolidation of new configuration always occurs near $\boldsymbol{T_C}$ (Figure 3.3d). Here, the black curve represents the location of the pre-learning critical point, estimated as the first point where rapid decline of stability for the native configuration occurs (i.e. $\boldsymbol{m^n < 0.45}$), whereas the red area is the parametric space where the new memory is consolidated ($\boldsymbol{m^e \geq 0.45}$).

Taken together, the model simulations outline how the process of learning is affected by dynamics near criticality. Here, the system is highly susceptible to network input and subsequently consolidates new configurations through state-based plastic changes in network connectivity strengths. If, on the other hand, the input corresponds to

one of the native memories their stability is extended over the critical temperature range. Thus, the critical state on one hand provides metastability to native configurations allowing their retrieval in presence of correct external input, but also provides dynamical substrate for storage and consolidation of the new configurations.

Further, during learning, the synaptic plasticity shifts the critical point, extending the sub-critical regime post-learning. To test whether these are general principals of learning in neuronal networks in vivo, we next analyzed spike data recorded from neurons in mouse hippocampal area CA1 during consolidation of a fear memory.

3.4 Evidence of Criticality and Consolidation-induced Dynamical Shift in Recordings of Mouse Hippocampus

We analyzed spiking data recorded from hippocampal area CA1 of mice subjected to contextual fear conditioning (CFC) in order to investigate the effect of learning on network dynamics. Specifically, mice are placed in the novel environment that they are allowed to explore briefly. They are subsequently exposed to electric shock while in the novel environment (induction of CFC) or not (sham) through the wire mesh placed in the floor. The mice exposed to the shock exhibit a freezing behavior (i.e. they stop moving) in the novel environment on subsequent presentation while the sham mice do not. Using CFC, long-lasting fear memories (CFMs) can be successfully consolidated in mice in the hours following a single training trial, consisting of placement in a novel environmental context paired with a foot shock. This single-trial learning, unlike more elaborate training procedures (e.g. object recall or track learning), provides clear boundaries between baseline and post-conditioning and allows for direct comparisons of network dynamics. Further, memory consolidation in general [130] and fear memory consolidation in particular [110, 111, 129] is known to rely on sleep, a vigilance state characterized by internally driven dynamics and thus allowing the possibility for truly self-organized neural behavior [137].

Successfully consolidated CFMs manifest as visual changes in behavior, where mice cower in place (i.e. freezing behavior) instead of adopting their normally inquisitive/explorative nature [110, 111]. The level of success of memory consolidation is quantified by a percent change in this behavior as compared to baseline, which we hereafter refer to as the learning score. In this study, we thus compare the behavior and analysis of hippocampal recordings across two groups of mice: 1) contextual fear conditioned mice

(CFC) that are given a fear stimulus in a novel environment and have ad lib sleeping patterns in the 24 hours following the stimulus; and 2) sham mice that are introduced to the novel environment but do not receive a foot shock and are not sleep deprived. More information about the experimental procedure can be found in Appendix A.1. We indeed found that CFC mice had higher learning scores post-stimulus compared to their Sham counterparts (Figure 3.4a).

In order to substantiate our model hypothesis that (a) near-critical dynamics may be important for memory consolidation and (b) that consolidation actually stabilizes the system, CA1 neurons' spiking data was analyzed for proximity to a critical state by calculating the branching parameter [3]. While other metrics have been used to determine dynamical states - namely, power-law-distributed avalanches [3, 137]- the benefit of the method described here is that it better controls for spurious correlations between the data and can account for slowly varying dynamical changes [28]. A previous study by the Priesemann group addressed this issue by showing a more accurate branching parameter can be determined by taking into account the relationship between the variance and covariance of the branching parameter and by eliminating data sets that showcase non-stationarities [28]. In this study, we used the python package associated with their study to calculate the optimal branching parameter σ (Python Package Index - mrestimator v 0.1.4; https://pypi.org/project/mrestimator). For each Slow Wave sleep (SWS) interval, we binned hippocampal spiking data into sub-intervals of 16 ms, calculated the avalanche size (i.e. the number of spikes) in each interval, and then used the provided software to calculate the branching parameter. Data sets that failed tests of non-stationarity (e.g. due to fast fluctuation between up and down states or from external drive; see [28]) were removed, and the average branching parameter was calculated.

We calculated σ from CA1 spike data recorded in the two groups (CFC and Sham) from every bout of SWS sleep during 24h time interval post CFC. We analyzed only SWS as during wake the mice are constantly swamped with new input making assessment of intrinsic hippocampal dynamical state impossible. At the same time REM sleep bouts in mice are few and short in duration making branching parameter estimate unstable (i.e. it failed many criteria set forth in [28]). We found that mice in both groups had branching parameters near $\sigma \sim 1$ (Figure 3.4b), indicating that the mouse brain naturally has near-

critical dynamics. After the learning interval, we observed a noticeable decrease in σ in most CFC compared to Sham mice (Figure 3.4c). Indeed, we found that an increase in learning score generally exhibited a decrease in their branching parameter (Figure 3.4c) away from a critical state and that CFC mice exhibited a more significant reduction due to learning ($p < 0.02$) compared to Sham animals ($p < 0.10$). These data indicate that (a) typical in vivo dynamics lie near criticality and that (b) consolidation of memory in vivo causes a deviation from critical to sub-critical behavior, as predicted by modeling.

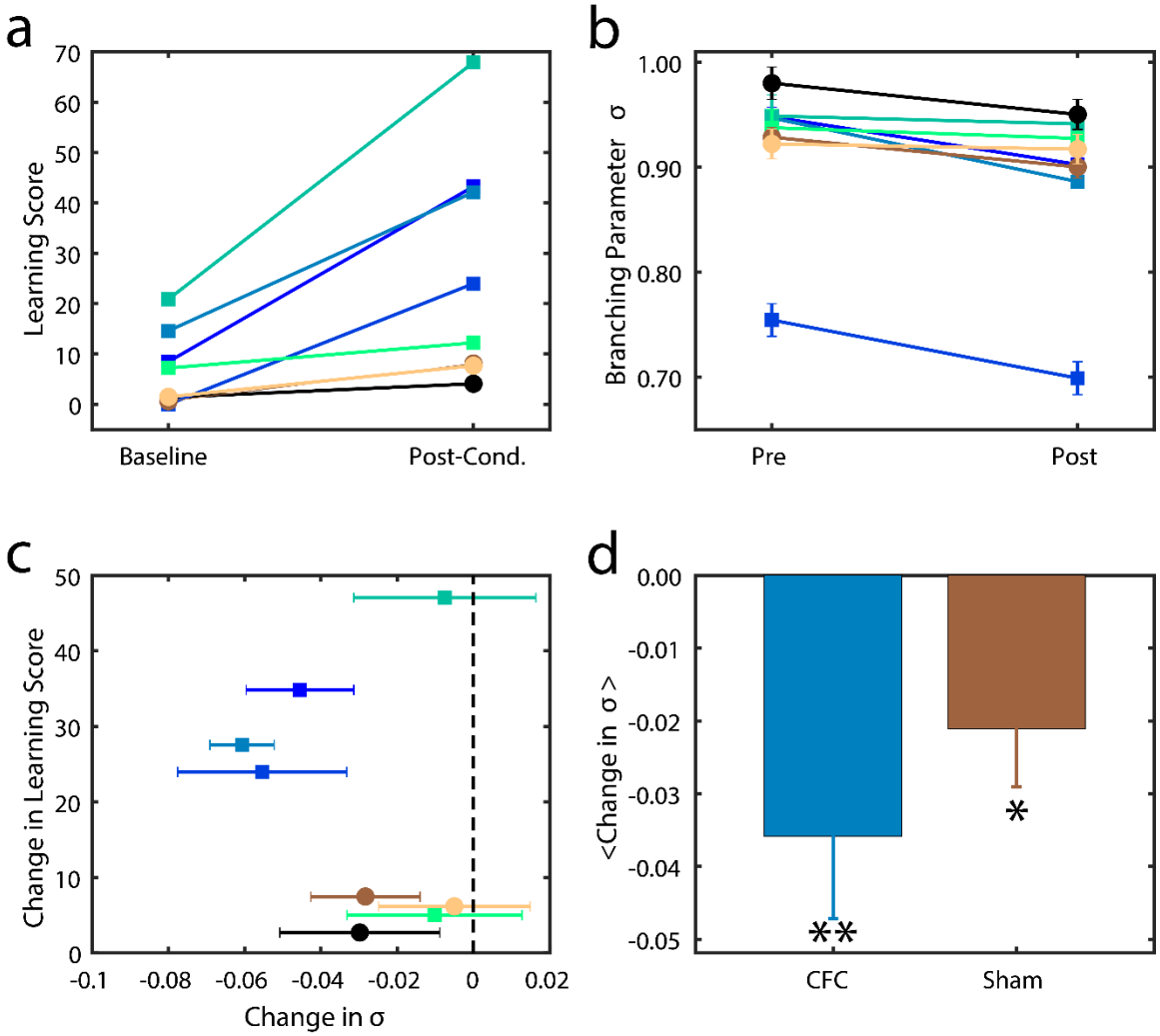

Figure 3.4 Branching parameter and its changes as a function of quality of memory consolidation during SWS. a) Percentage of freezing behavior observed in mice before (baseline) and after learning (post-cond.) for sham (circles) and CFC (squares) groups. Different colors represent different mice. b) Branching parameters σ during SWS before (baseline) and after learning (post-cond.). Colors and shapes are conserved as in (a). Error bars represent the standard error of the mean, calculated for each mouse over all intervals. c) Change in freezing behavior vs change in branching parameter across the learning interval. Error bars represent the propagation of standard errors between Pre and Post in (b). d) Mean change in branching parameter within each group. Error bars represent the standard error of the mean. * $p < 0.10$ confidence interval that the reduction was significant; ** $p < 0.02$ confidence interval that the reduction was significant, using the one-way T test.

The smaller drop in branching parameter in sham group may also be associated with (smaller) degree of consolidation of the new environment even without the electric shock.

3.5 Discussion

The question we address in this paper is how relatively sparse input can dynamically compete with already stored representations, to be stored and later consolidated into a distributed memory (engram). Through computational modeling work and analysis of in vivo hippocampal recordings, we show that criticality may play a pivotal role in mediating stabilization and subsequent storage of the new memory as a distributed representation. Namely, we show in a reduced attractor network, that only when the system is near a critical point can the new representation globally impinge its activity pattern on the network, making it fractionally dominant as compared to the native representation (Figure 3.1). This is primarily due to the fact that at criticality, when the system has the highest susceptibility to the external input, this input biases the state of the network towards the new representation, and the emergence of long-distance correlations allows it to spread throughout the system. Subsequently, state dependent synaptic plasticity allows for long-term storage (consolidation) of this new representation, even as it competes with a broad range of native configurations (Figure 3.3d). Thus here, similarly to results shown in self-organizing recurrent models [7], presentation of organized input results in a shift in the parametric location of criticality (Figure 3.1b), due to increased stability of native representation or storage of the new representation (Figure 3.3b).

We thus hypothesize that criticality on one hand provides metastability to already stored configurations, so that if a native memory is presented through input, the memory is retrieved via the stabilized attractor, while on the other hand criticality provides a dynamical substrate for storage and consolidation of the new representations.

We find that successful new memory consolidation possibly changes the underlying dynamical state from being near-critical to being slightly sub-critical (Figure 3.3). Previous studies have reported a similar, slightly sub-critical dynamical state of the brain [27] which here seems to be the result of system consolidation to new information. Indeed, we see a similar deviation from critical to sub-critical dynamics in hippocampal recordings of mice successfully consolidating fear memories in vivo (Figure 3.4). This phenomenon can be

explained as follows: before learning, susceptibility to external input is maximal near a critical point but, as learning commences, the system adapts by strengthening the connectivity to consolidate this new information extending the region of dynamical stability.

Our results indicate that the brain operates near criticality, possibly slightly sub-critical, and that plasticity plays an active role in reducing the dynamical state away from criticality during learning and consolidation. This is an agreement with previous work that suggests slightly sub-critical dynamics still impart increased tunability, response to external input, and long-range spatial and temporal correlations [138]. The extension of the Hopfield model we present here suggest that the critical point is indeed shifted (Figure 3b), rather than the phase transition region is widened, what would be indicative of emergence of Griffith phase [139]. However, some of our unpublished results obtained in models of self-organized criticality, which are similar to integrate and fire models, suggest that the critical point may indeed expand suggesting emergence of Griffith phase (data not shown).

This raises an interesting question: how does the brain finally reset to a near-critical state after learning, so that another (new) memory can be consolidated? Our work here does not address this issue, but previous work by others has shown that neurons and networks in the brain have built-in homeostatic mechanisms which serves to recalibrate synaptic efficacies (see e.g. Refs. [51, 140-143]), a process that was proposed to happen also during development [144]. Thus, it could be that homeostatic plasticity together with reduced external input during sleep is sufficient to drive the system towards criticality, as shown by *Zierenberg, J. et al.* [145]. Indeed, our in vivo analysis indicates that the role of sleep is not purely homeostatic [51], but instead involves active learning processes, in line with previous reports [135, 146]. As an additional consideration: both the model system and mice subjected to fear stimuli involve relatively strong inputs to be learned. In processes that occur over longer time periods, the dynamical shift may be weak compared to homeostatic dynamical rescaling, making it hard to detect on such short time scales as we show here. Future work should thus be done to investigate the interplay between homeostatic-based and learning-based changes in system dynamics near criticality.

Chapter 4

Slow-wave sleep facilitates hippocampal consolidation of contextual fear memory through temporal coding of neuronal representations

In this final chapter, I combine modeling and analysis of in vivo hippocampal recordings to elucidate the role of NREM-associated oscillations in driving STDP to consolidate memories. I show that phase-coding of information during sleep following rate-coding during wake may account for the observation of homogenized neuronal firing rates. This work is in preparation to be submitted to the journal of eLife.

How the brain binds features of the physical world into a neural code is a long-standing question in neuroscience. Ongoing evidence across multiple systems indicate that synaptic plasticity between time-synchronized neuronal spike patterns plays a major role in this process [7, 31, 107], selectively strengthening connections among neurons and creating lasting memory traces. A widely-accepted hypothesis, supported by experimental evidence, is that plasticity preferentially potentiates connections between neurons whose activity undergo experience-associated changes (often referred to as "engram neurons") [12]. Such representation of information constitutes a mnemonic "rate code". However, rate coding has limitations for long-term information storage in the brain in vivo.

First, individual neurons have widely divergent baseline firing rates which would constitute background "noise", obscuring firing rate changes in a sparse subset of neurons representing new information. Second, because individual neurons have a fairly limited dynamic range over which their firing rates can vary (e.g. due to homeostatic plasticity), rate coding has a limited capacity for neuronal information integration over time. In other words, neurons' peak firing rates would represent a "ceiling", after which no additional information could be carried by a given neuron by rate-coding alone. Third, changes in firing rate will alter spike timing-dependent plasticity (STDP) rules governing activity-

dependent synaptic strengthening or weakening between constituent network neurons. As neurons increase their firing rates, they will either reduce STDP-mediated synaptic strengthening/weakening (due to differences in firing rates between pre- and postsynaptic neurons), or bias STDP toward potentiation [8]. In either case, the information-carrying capacity of STDP for the network is limited by firing rate increases. Thus, for long-term information storage (consolidation) in a network of heterogeneous neurons, non-rate-based coding mechanisms must be invoked.

Another long-accepted assumption is that following information encoding during learning, memories are consolidated, via either modification of existing synaptic connections or de novo creation of additional synapses. Such synaptic changes constitute a structural network heterogeneity, which subsequently serves as a so-called dynamical attractor in the network [7, 17]. When such an attractor is present, incomplete patterns of spiking among engram-encoding neurons could lead to subsequent recollection of the complete pattern, i.e. a memory. This recollection process possibly consists of augmentation of a memory trace with new populations of cells that were not as active initially [15]. It remains unclear what activity-dependent mechanisms could mediate consolidation leading to such an attractor formation. However, oscillatory patterning, such as that occurring during sleep, has been implicated in promoting synaptic plasticity [15, 48, 53, 147-149], thereby affecting the network thought to be an integral part of memory representations.

Various network oscillations appear preferentially in specific brain circuits (e.g., hippocampal and thalamocortical circuits) during particular behavioral states (i.e. waking, REM and NREM sleep) and have been established to be highly predictive of state-dependent cognitive processes and network plasticity [150]. Further, network oscillations have been implicated in promoting STDP by precisely timing the firing between pairs of neurons [12, 53]. Thus, hypothetically, oscillatory dynamics could mediate a switch between rate coding (present during learning) and timing-dependent coding (in which neurons' respective timing, rather than firing rate, carries information). We have recently shown that in a weakly coupled regime, an activity-dependent shift in theta-band cellular resonance among neurons may play an important role in explaining learning-initiated network reorganization, via STDP-based mechanisms [43, 53]. We hypothesize that, for

the reasons outlined above, this switch from rate to phase coding (i.e. neurons' respective phase of firing on each cycle conveys the information about the network patterning) is essential for long-term information storage in brain networks.

Here, we combined computational modeling of a highly reduced CA1 network with the analysis of in vivo mouse CA1 activity to elucidate the role of coincident firing during theta band (4-12 Hz) oscillations in short- and long-term reorganization of hippocampal network structure and dynamics during contextual fear memory (CFM) consolidation. Contextual fear conditioning (CFC) leads to rapid formation and consolidation of new memories (i.e. after single-trial learning) [110]. CFM consolidation relies on ad lib sleep in the hours immediately following CFC [146], and is associated with augmented theta-frequency activity in CA1 in the hours following CFC [43]. Critically, recent in vivo work has shown CFM is disrupted when theta oscillations are suppressed during post-CFC sleep via optogenetic or pharmacogenetic inhibition of CA1 fast-spiking interneurons [111]. Conversely, CFM consolidation can be rescued from disruption caused by experimental sleep deprivation when theta oscillations are driven optogenetically (via rhythmic activation of fast-spiking interneurons) in CA1 [43].

We show that resonance properties of the CA1 network model, mediated via changes in Ach levels during varying vigilance states, are sufficient to recreate all of these experimental phenomena, including temporal stabilization of network representations after CFC during NREM, the state-dependence of fear memory consolidation, and disruption of consolidation through interneuron inhibition. Furthermore, we demonstrate that STDP-based memory consolidation during a dynamical network state analogous to sleep: (1) causes increased stability of functional network functional connectivity, subsequently leading to (2) an asymmetric change in synaptic connectivity patterns, between fast and slow spiking neurons, and (3) causes dramatic, differential changes in the activity profile of highly active vs. sparsely firing neuronal populations. These same dynamic changes are not observed during the dynamical network state corresponding to waking.

Together, these results show that successful memory consolidation requires the brain to switch between information coding schemes which occur naturally through the sleep-wake cycle. In so doing, the functional network structures associated with engrams become more stable and leads to successful long-term memory storage.

Finally, similar frequency dependent changes in firing frequency mediated by NREM sleep, were observed in other modalities than the hippocampus (i.e. visual cortex [15]). Thus we conclude that even though this study is limited to investigating role of theta frequency network oscillations in fear memory, it may yield insight into potential universal neuronal network mechanisms of information storage in the brain.

4.1 The Introduction of Memory Traces Augments Network Oscillatory Dynamics and Stability of Functional Connectivity Patterns.

We hypothesized that network plasticity in hippocampal area CA1 following single-trial contextual fear conditioning (CFC) [112] would be a plausible biological system to investigate how rapid memory encoding affects the underlying neural network dynamics. Since CA1 network activity is necessary for fear memory consolidation in the hours following CFC [113], we analyzed and simulated recordings of the same population of CA1 neurons over a 24-h baseline and for 24 h following CFC to determine how functional network dynamics were affected by de novo memory formation. For brief description of the experimental protocol see methods, for more details please see [110].

To investigate the mechanisms involved in sleep-dependent memory consolidation, we simulated a reduced CA1 network model composed of two cell types: excitatory pyramidal neurons and inhibitory interneurons that loosely represent parvalbumin positive (PV+) interneurons. In the model, the time-dependent voltage $\boldsymbol{V_i}$ of a single neuron is given by

$$\boldsymbol{C_m \frac{d}{dt} V_i = -I_{Na} - I_K - I_{K_s} - I_{leak} + I_{ext} - I_{Synaptic}}$$

where $\boldsymbol{C_m}$ is the membrane capacitance, $\boldsymbol{I_{ext}}$ is the fixed external input used to elicit spiking, $\boldsymbol{I_{leak} = 0.02(V_i + 60)}$ is the leakage current, and $\boldsymbol{I_{Synaptic}} = \left(\sum_{j \in Excitatory} \boldsymbol{g_{E-X}\, S_{ij}}\right)\boldsymbol{(V_i - V_{Excitatory})} + \left(\sum_{j \in Inhibitory} \boldsymbol{g_{I-X}\, S_{ij}}\right)\boldsymbol{(V_i - V_{Inhibitory})}$ is the total summed synaptic input received by a neuron from its pre-synaptic partners and $\boldsymbol{g_{I-X}}$ and $\boldsymbol{g_{E-X}}$ represent the synaptic conductance for connections from inhibitory and excitatory neurons to their post synaptic targets X (values provided below). The synaptic reversal potentials are $\boldsymbol{V_{Excitatory} = 0\, mV}$ and $\boldsymbol{V_{Inhibitory} = -75\, mV}$. Here, $\boldsymbol{S_{ij}} = \boldsymbol{exp}\left(-\frac{\Delta t_{ji}^{spk}}{\tau_s}\right) - \boldsymbol{exp}\left(-\frac{\Delta t_{ji}^{spk}}{\tau_f}\right)$ represents the shape of the synaptic current, given the

difference in spike timing between the post-synaptic neuron $\boldsymbol{i}$ and the recently fired pre-synaptic neuron $\boldsymbol{j}$, ($\Delta \boldsymbol{t}_{\boldsymbol{ji}}^{\boldsymbol{spk}}$), with $\boldsymbol{\tau_f = 5\,ms}$ and $\boldsymbol{\tau_s = 250\,ms}$ or $\boldsymbol{\tau_s = 30\,ms}$ for excitatory synaptic currents and inhibitory synaptic currents, respectively.

The ionic currents are $\boldsymbol{I_{Na}, I_K, and\ I_{K_s}}$, representing sodium (Na), potassium (K), and muscarinic slow potassium (K_s), respectively. More specifically: $\boldsymbol{I_{Na} = g_{Na} m_\infty^3 h(V_i - V_{Na})}$, with $\boldsymbol{m_\infty = (1 + exp\left(\frac{-V_i - 30}{9.5}\right))^{-1}}$ being the activation of the channel and where $\boldsymbol{h}$, the inactivation, is given by the solution to $\boldsymbol{\frac{d}{dt}h = (h_\infty - h)/\tau_h}$, with $\boldsymbol{h_\infty = (1 + exp\left(\frac{V_i + 53}{7}\right))^{-1}}$ and $\boldsymbol{\tau_h = 0.37 + 2.78(1 + exp\left(\frac{V_i + 40.5}{6}\right))^{-1}}$; $\boldsymbol{I_K = g_K n^4 (V_i - V_K)}$ with $\boldsymbol{\frac{d}{dt}n = (n_\infty - n)/\tau_n}$ where $\boldsymbol{n_\infty = (1 + exp\left(\frac{-V_i - 30}{10}\right))^{-1}}$ and $\boldsymbol{\tau_n = 0.37 + 1.85(1 + exp\left(\frac{V_i + 27}{15}\right))^{-1}}$; and $\boldsymbol{I_{K_s} = g_K s(V_i - V_K)}$ with $\boldsymbol{\frac{d}{dt}s = (s_\infty - s)/75}$ where $\boldsymbol{s_\infty = (1 + exp\left(\frac{-V_i - 39}{5}\right))^{-1}}$. The reversal potentials are $\boldsymbol{V_{Na} = 55\,mV}$ and $\boldsymbol{V_K = -90\,mV}$ and the maximal conductances are $\boldsymbol{g_{Na} = 24\ \frac{mS}{cm^2}}$, $\boldsymbol{g_K = 3.0\ \frac{mS}{cm^2}}$. Each simulation was completed using the RK4 integration method with a step size of h = 0.05 ms.

The slow potassium conductance, $\boldsymbol{g_{K_s}}$, serves as a control parameter for individual neuron dynamics, switching the neuron's response to input from being integrative to resonant [151]. Integrators, also called Type 1 (T1) excitable oscillators, exhibit a decrease in timing between spikes regardless of when excitatory external input is received. Resonators, called Type 2 (T2) excitable oscillators, exhibit an increase in timing between spikes when excitatory input is received soon after an initial spike but a decrease in spike timing when excitatory input is received later in the firing cycle [151]. Studies have shown that T2 oscillators have robust spike-frequency adaptation compared to T1 oscillators [53] but that T1 oscillators are characterized by wider firing-rate distributions and attain overall higher firing-rates compared to T2 oscillators. Further, due to the biphasic response of T2 neurons, they can synchronize much more easily to input compared to T1 oscillators, indicating that synchronization is mediated by both internal neuronal properties as well as how neurons are connected.

In this model, the $\boldsymbol{g_{K_s}}$ conductance controls the action of muscarinic acetylcholine (ACh) receptors [57, 152]: high ACh concentration blocks pyramidal cells' slow-varying potassium currents through these receptors, yielding increased gain in the firing frequency response to varying (excitatory) input and decreased capacity for firing to synchronize with rhythmic input, i.e. T1 dynamics. Low concentrations of ACh allows these slow-varying potassium currents to play a larger role in membrane excitability, switching them to T2 oscillators: firing rate distributions are narrowed [56], neurons experience considerably enhanced spike-frequency adaptation [53], and, as we will see, are able to synchronize their dynamics much more easily than T1 oscillators. Hence, I can control the effective amount of Ach in the system just by changing the values of this slow potassium conductance:

$$\boldsymbol{g_{K_s}} = \begin{cases} \mathbf{0}\ \frac{\boldsymbol{mS}}{\boldsymbol{cm^2}} & \boldsymbol{if\ ACh\ is\ high} \\ \mathbf{1.5}\ \frac{\boldsymbol{mS}}{\boldsymbol{cm^2}} & \boldsymbol{if\ ACh\ is\ low} \end{cases}.$$

As we have discussed, the high concentrations of Ach have been implicated in attentive wakefulness and is thought to play a role in the initial binding of memories into neural codes [55, 153]. During NREM sleep, the level of Ach drops precipitously before rebounding in REM sleep. Conveniently then, modifying the value of $\boldsymbol{g_{K_s}}$ allows us insight into a simple model approximation of the biochemical differences between wake and sleep and how the emergent processes of networks facilitate information encoding.

The network used in these studies consists of N=1000 neurons, with $\boldsymbol{N_e = 800}$ excitatory neurons and $\boldsymbol{N_I = 200}$ inhibitory neurons. Connections are arranged to form a random network with different levels of connectivity dependent on the pairwise pre- and post-synaptic neuron identity: Inhibitory neurons project to 50% of the inhibitory neurons and 30% to the excitatory neurons whereas excitatory neurons project to just 6% of both the inhibitory and excitatory neurons, with self-connections being forbidden in all cases. The initial synaptic weights are $\boldsymbol{g_{I-I} = 0.0013\ mS/cm2}$, $\boldsymbol{g_{I-E} = 0.0005\ mS/cm2}$, $\boldsymbol{g_{E-I} = 0.00046\ \frac{mS}{cm2}}$, and $\boldsymbol{g_{E-E} = 0.00003\ mS/cm2}$ but can be changed (manually or through synaptic interactions) to form an initial memory engram.

Neural correlates of memory are thought to emerge due to the strengthening and weakening of synaptic strengths in an activity-based manner following spike timing-dependent plasticity (STDP)[7]. Here, we use a symmetric learning rule that uniformly

increases or decreases synaptic weights based on the time-ordering of pre- and post-synaptic pair firings, only in excitatory-to-excitatory connections. If a pre-synaptic neuron fires before its post-synaptic partner, the conductance increases by an amount $\boldsymbol{\rho\, Exp}\left(-\frac{t_{pre}^{spk}-t_{post}^{spk}}{10}\right)$; similarly, a weakening of synaptic strength occurs by an amount $\boldsymbol{\rho\, Exp}\left(-\frac{t_{post}^{spk}-t_{pre}^{spk}}{10}\right)$ when a post-synaptic neuron fires before its pre-synaptic partner, roughly following experimental trends [7]. In both cases, if the time difference between spike pairs is too great, the change in synaptic strength will approach zero. On the other hand, highly coincident spike pairs will have a maximal change given by the learning rate $\boldsymbol{\rho = 10^{-3}}$. It should be noted that while the synaptic weight is prohibited from becoming negative, there is no upper-bound set on the synaptic strength, though previous work has shown saturation of synaptic weights given enough time [53].

Within this framework, we first investigated how changes in connectivity to a limited subset of neurons, an analogy for memory encoding by engram neurons (i.e. the end result of a previous learning experience using STDP), affect network activity patterns. To this end, we mimicked the formation of an engram by first artificially increasing the

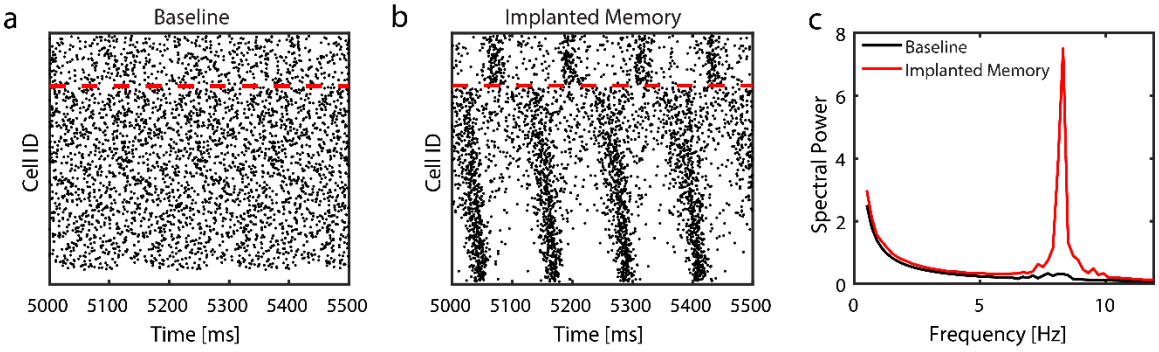

Figure 4.1 Model networks respond to sparse strengthening of excitatiory synapses through emergance of theta rhythm and phase-locking. a-b) Raster plots of the network before (left) and after (right) introduction of a memory of strength 20. Neurons below the red, dashed line are excitatory and those above the line are inhibitory. Inhibitory neurons connected randomly to 50% of the other inhibitory neurons and to 30% of the pyramidal neurons whereas excitatory neurons connected randomly to 6% of both the excitatory and inhibitory neurons; a random subset of excitatory neurons increased their synaptic strength to other excitatory neuron, constituting the memory strength. c) Fourier transform of the excitatory network signal reveals a sharp increase in spectral power only near 8 Hz; black line marks baseline; red line – spectral power after synaptic strenghtening.

outgoing synaptic connections of a randomly chosen subset pyramidal neurons by various degrees, which we refer to as the "memory strength". Comparing raster plots for the

network when pyramidal neurons exhibit type 2 dynamics, before vs. after introduction of the engram (Figure 4.1a-b; corresponding to a memory strength of 20), reveals the emergence of well-defined oscillations and, correspondingly, an increase in theta-band spectral power (4-12 Hz; Figure 4.1c). The emergence of network oscillations with temporally locked neuronal spiking (i.e. characteristic oscillations in the theta-band regime) might be sufficient to generate reliable sequences of spiking patterns thought to play an important role in memory consolidation.

4.2 Hippocampal Network Stabilization in vivo Predicts Effective Memory Consolidation.

To investigate how emergence of oscillations affects formation of the temporal patterning between the neuronal firing patterns, we calculated the stability of network functional connectivity, using the functional network stability metric (FuNS; see Chapter 2.1) [62](Figure 4.2).

The FuNS metric is sensitive to reconfiguration of functional connectivity patterns within the network, but not to fluctuations in the precise spike order between pairs of neurons. High stability can be detected for weak but consistent functional connectivity patterns. Therefore, FuNS qualitatively provides information about temporal evolution of functional connectivity as compared to direct assessment of correlation between spike bouts. We thus used FuNS to measure stability of functional connectivity measured from the experimental recordings as well as in the model simulations.

Experimentally, the mice underwent CFC (n = 5 mice), sham conditioning (n = 3 mice), or CFC followed by 6 h of sleep deprivation (SD; a manipulation known to disrupt fear memory consolidation [43, 129, 146]; n = 5 mice); a full description of the experiments and subsequent recordings performed can be found in Appendix A.1.

We measured changes in FuNS in these recordings after each manipulation by quantifying FuNS on a minute-by-minute basis over the entire pre- and post-training (i.e., 24 h duration) intervals and calculating their respective difference within each animal. Consistent with previous findings [110], we observed a significant increase in FuNS over the 24 h following CFC during NREM sleep (Figure 4.2a). In contrast, no change in NREM FuNS was seen in Sham mice or following CFC (during recovery sleep) in SD mice.

Group differences in NREM FuNS were reflected in the behavior of the mice 24 h post-training, when context-specific fear memory was assessed (inset Figure 4.2c, inset). Mice allowed ad lib sleep following CFC showed significantly greater freezing behavior when returned to the conditioned context than did Sham or SD mice. Moreover, training-induced changes in NREM-specific FuNS for individual mice predicted context-specific freezing during memory assessment (Figure 4.2c). Thus, successful consolidation of a behaviorally-accessible memory trace in vivo is accompanied by increased FuNS in the CA1 network.

In terms of in silico modeling, we allowed synapses between pyramidal neurons to undergo STDP while we monitored changes in FuNS. We simulated networks with differential exposure to learning (i.e. strengthening the connectivity between the subset of "engram" neurons as described in above section) and varying vigilance states (type 1 for wake or type 2 for NREM sleep) to mimic the experiments described above. Specifically, to model sleep deprivation (SD), pyramidal neurons were set to type 1 excitability in the presence of engram neurons (memory strength = 10), whereas to model NREM effects on CFC and Sham the pyramidal cells were set to type 2 excitability, with CFC having a similarly defined engram as SD and Sham having no engram.

We found that only when the engram was present in networks with type 2 excitability was the network capable of successfully consolidating information, maintaining stable network dynamics over time (Figure 4.2b). Conversely, type 2 networks without an engram and type 1 networks with an engram showed no significant increase in stability, even in the context of STDP-based learning. These results led us to hypothesize that oscillatory patterning is critical for successful STDP-based consolidation of a hippocampal memory trace. We focus on this phenomenon in the following sections.

Further, to investigate how initial memory strength (corresponding to just after CFC experimentally) affects FuNS and power of theta band oscillations we measured the two quantities as a function of strengthening of synaptic connections between "engram" neurons (i.e. the memory strength, Figure 4.2d). By increasing the memory strength, we observed a positive linear relationship between increased FuNS and increased theta-band power in the model network (Figure 4.2d). This suggests that both the locking of neuronal firing with network oscillations and the associated stabilization of network functional

connectivity patterns is driven by synaptic potentiation between engram neurons and other neurons in the network. This finding agrees with experimental results, showing that contextual fear memory consolidation is associated with both increased theta-band oscillations, and increased FuNS, during sleep in the hours immediately following conditioning [43, 110, 111]. We proceed to show that it is this phase-locking of neurons to underlying oscillations that is important for memory consolidation.

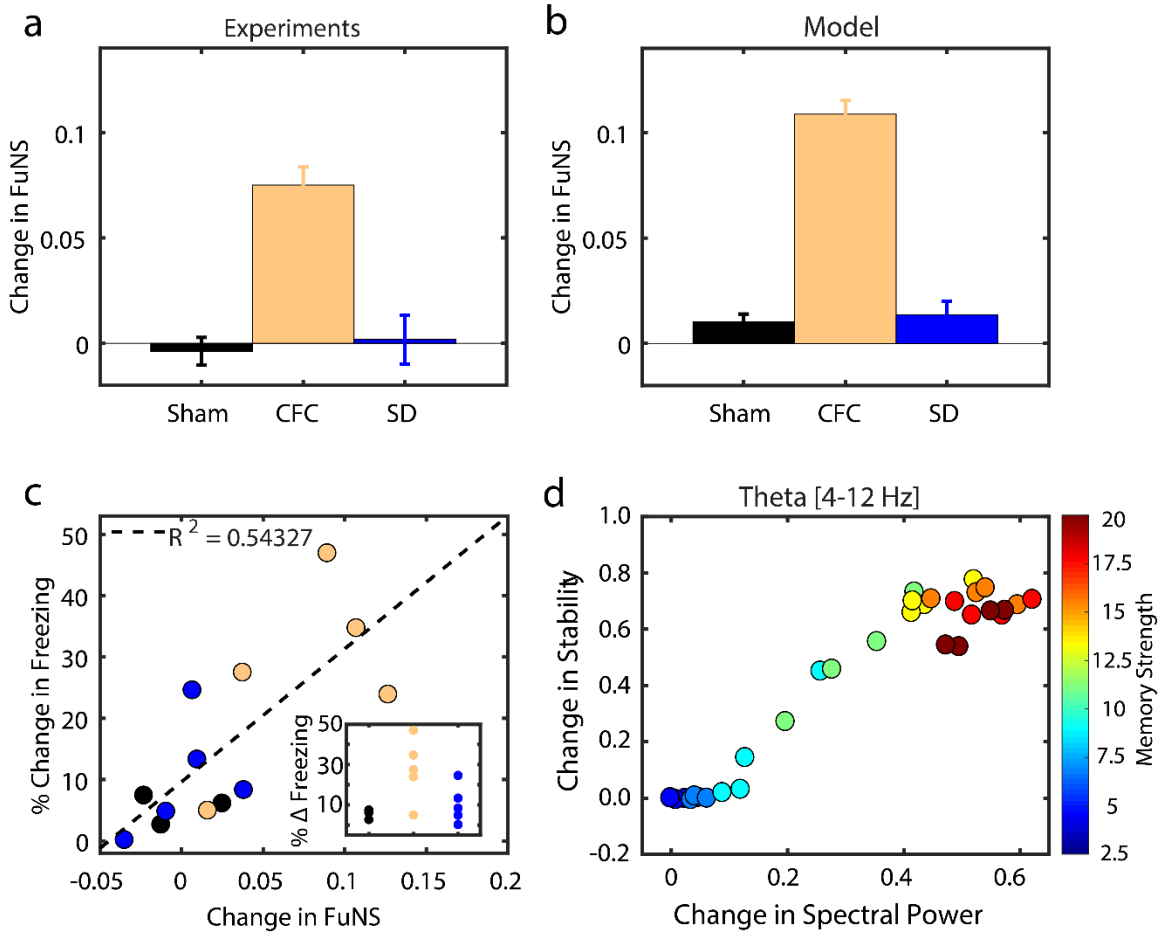

Figure 4.2 Functional network stability predicts future level of memory consolidation. a) Analysis of in vivo recordings in the mouse hippocampus CA1 area following CFC for corresponding behavioral states (sham — black [n = 3] , SD — blue [n = 5], CFC – Gold [n = 5]) reveals that ad lib sleep post conditioning leads to the greatest increase in FuNS. b) Model predictions for the change in FuNS in each simulation group: Sham (learning in NREM states without selective connectivity strenghtening; n= 5) and SD (learning in Wake dynamical state with selective connectivity strenghtening; n = 5) show only marginal changes in FuNS whereas CFC (learning in NREM dynamical state with selective connectivity strenghtening; n = 5) show a maximal increase in FuNS. All error bars represent the standard error of the mean. c) Examining the change in FuNS as a function of change in learning (% freezing; raw values shown as inset) reveals a linear relationship with a goodness of fit $R^2 = 0.54327$. d) Measuring the change in stability as a function of the change in spectral power density (integrated over theta-band frequencies) for increasing the strength of those selected synaptic connections reveals a linear releationship. The color bar represents the magnitude of synaptic strength change from baseline.

4.3 Input-dependent Phase Locking of firing to Network Oscillations Predicts Firing rate Reorganization across a Period of Sleep

A series of recent studies have demonstrated that sleep has heterogeneous effects on neuronal firing rates within neural circuits. Specifically, initially highly active neurons reduce their firing rates across a period of sleep, while sparsely firing neurons simultaneously increase their firing rate [15, 147, 154]. Sleep is essential for this redistribution of firing rates, as firing rate changes do not occur in animals that are experimentally sleep deprived [15].

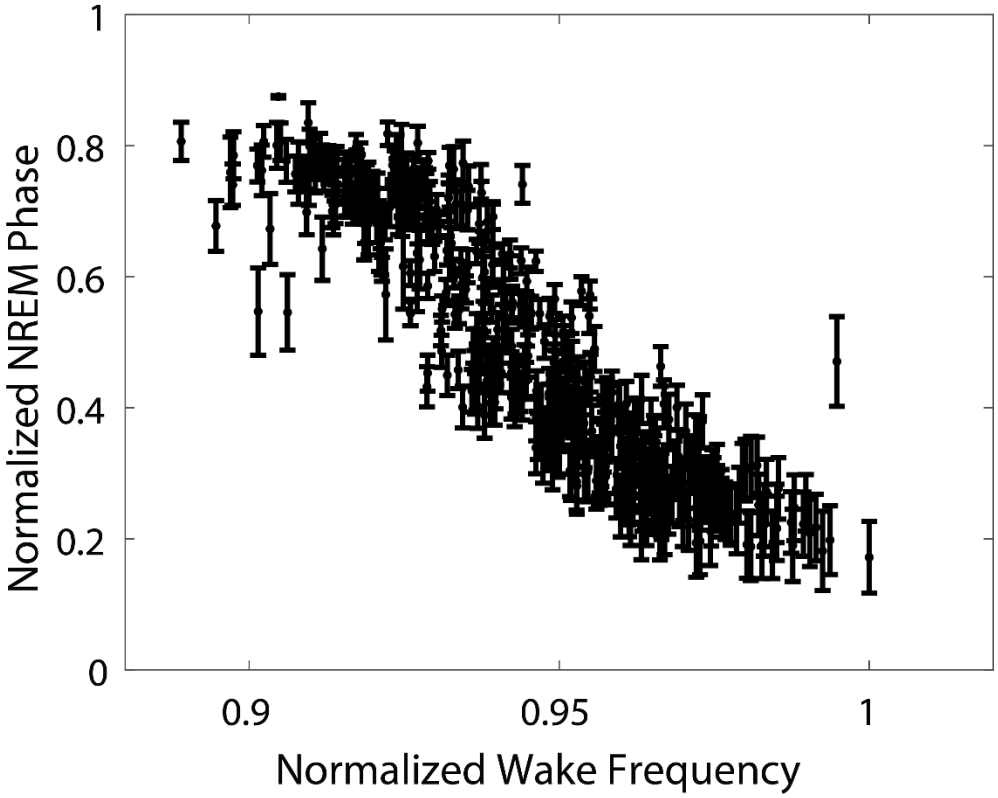

Figure 4.3: Neurons arrange in NREM oscillatory phase by Wake firing rate. Normalized phase of firing in NREM versus normalized wake frequency of pyramidal neurons reveals that the neurons firing with the highest frequency align with an earlier phase of the pyramidal network population whereas slower firing neurons align with later phases.

We hypothesized that this phenomenon could result from an asymmetric plasticity mechanism accompanying an input-dependent phase-locking of neurons' firing to oscillations observed during sleep. Specifically, we predict that neurons that are highly active during wake would fire at an earlier phase than sparsely-firing neurons during subsequent NREM oscillations. Figure 4.3 illustrates the relationship between model

neurons' phase of firing calculated during type 2 dynamics before learning as a function of the normalized frequency during type 1 dynamics pre-learning. Specifically, the phase relationships reported were calculated using a continuous inhibitory signal (inhibitory spikes convolved with a Gaussian function with $\mu = 0$ ms and $\sigma = 5$ ms) from the inhibitory neuron population spike vector. Peaks in the inhibitory signal acted as the start and end of a given phase and non-engram excitatory spike times were used to calculate the phase difference between an individual neuron and the signal. The phase difference was normalized to give values between 0 and 1.

We observed that the activation of fastest firing neurons during waking indeed occurs earlier in the phase of the pyramidal network oscillation during sleep. This suggests that neurons take on a phase-based, temporal coding strategy during strongly coherent network oscillations (i.e., those occurring during sleep) which reflects relationships in the prior rate coding strategy present during wake. Based on this relationship, we hypothesized that excitatory connections from high-firing to low-firing neurons are strengthened via STDP during sleep, while connections from low-firing to high-firing neurons are weakened.

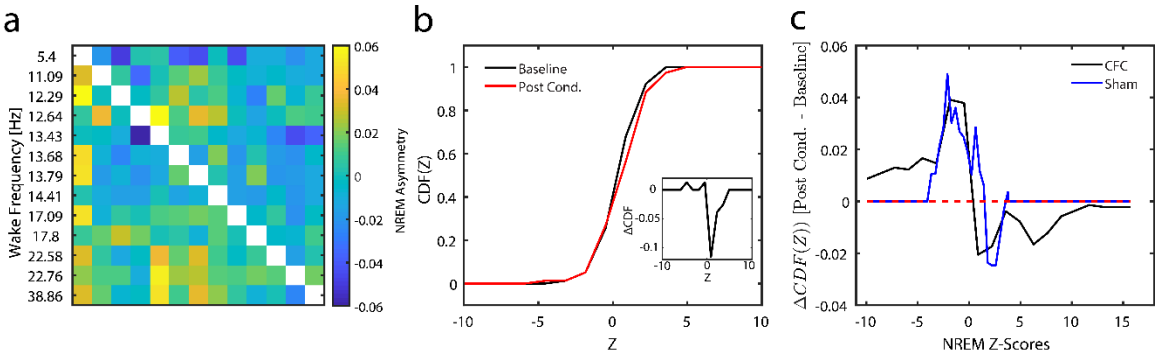

Figure 4.4 Analysis of spiking asymmetry reveals enhanced wake frequency dependent temporal relationships between firing neurons after CFC. a) Pairwise spiking asymmetry recorded in an example CFC mouse during NREM; rows and columns have been arranged by wake frequency (y-axis). b) Example cumulative distribution functions of pairwise asymmetry Z-Scores (Z) [NREM Asymmetry ordered by Wake Frequency] for baseline (Black) and post-conditioning (Red) recordings. Inset: difference between the two distributions. c) Weighted average of differences in CDFs for CFC mice (Black) and Sham mice (blue).

To investigate whether this phase-locking phenomenon can be observed experimentally, we compared in vivo and in silico spiking dynamics using a metric detecting frequency dependent spiking asymmetry between the pairs of firing neurons; for a full description of the following, please refer to Appendix B. Briefly, in CFC and Sham

hippocampal recordings, we first detected bursts of activity in NREM and we subsequently calculated the frequency dependent spiking asymmetry between neurons – i.e. whether, statistically, neurons that fire more frequently during wake tend to lead spiking during NREM sleep. We define the asymmetry matrix A, an $\boldsymbol{N \times N}$ matrix whose rows and columns were ordered by the firing rate of the same neurons within wake (Figure 4.4a), leaving the lower triangular matrix to represent neurons with a positive frequency difference during wake. We subsequently calculate the significance of the results, a Z score for every pair, by performing bootstrapping; by convention, $\boldsymbol{Z_{ij} > 0}$ indicates a positive asymmetry and we take $\boldsymbol{Z_{ij} \geq 2}$ to be significant.

We next compared the cumulative distribution functions (CDFs) of $\boldsymbol{Z_{ij}}$ calculated before and after exposure to CFC (Figure 4.4b) or sham. The inset in Figure 4.4b shows the difference between the two distributions. Finally, we calculated the weighted average of the differences between CDFs of the CFC and sham animals, respectively, and found that that the difference in CFC distributions (with respect to baseline) are skewed towards positive Z values as compared to sham ones (Figure 4.4c), indicating that CFC significantly increases the number of the locked pairs in the frequency dependent manner. This further indicates presence of a phase-coding mechanism.

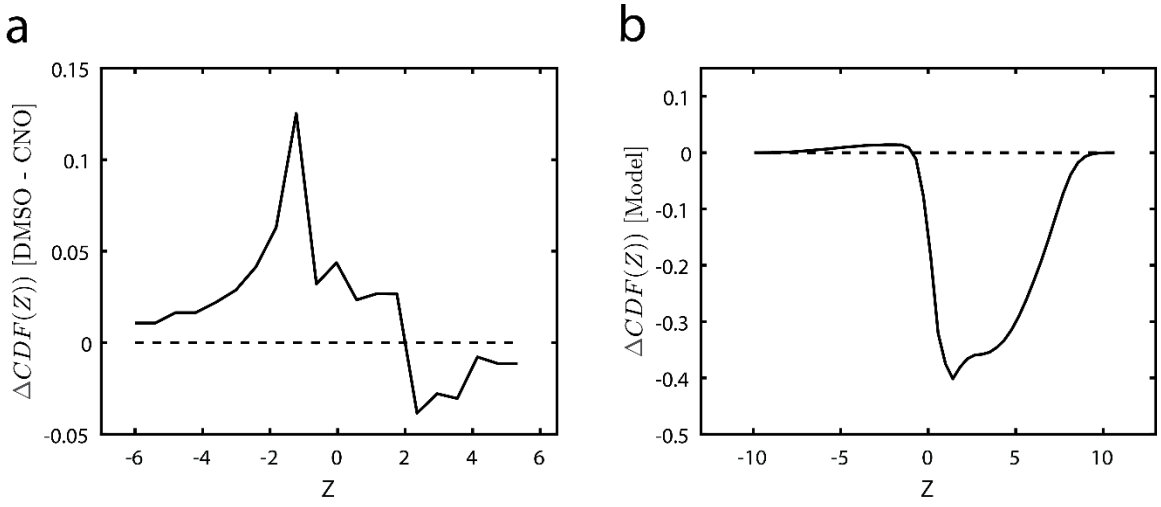

Figure 4.5 Disruptions to network oscillations lead to diminished temporal relationships between firing neurons. Weighted sum of differences in experimental CDFs between DMSO and CNO experimental recordings (a) and between model CDFs, between full networks and networks with blocked inhibitory interneurons (b).

4.4 *Disruption of Oscillatory Rhythms Affects Temporal Coding During Sleep*.

We next wanted to investigate how disruptions in network-wide oscillations affect phase-coding mechanisms. To this end, we analyzed the spiking asymmetry of hippocampal recordings of mice that were transgenically modified so that the presence of an inhibitory block (CNO dissolved in DMSO) shuts down PV+ interneuron activity and subsequently ablates network-wide oscillations and diminishes learning, whereas the presence of DMSO alone does not adversely affect oscillations or learning [43] (an experimental description is found in Appendix A.2). We observed a decrease in positive asymmetries taken from post-learning recordings in CNO threated animals as compared to DMSO only threated ones, indicating a subsequent disruption in frequency dependent phase-coding due to disruptions in network oscillations (Figure 4.5a). We found a similar disruption of phase-coding in our model simulations (Figure 4.5b), when we blocked inhibitory neurons from firing.

4.4.1 *Frequency dependent Firing Asymmetry Affects Network Reorganization in the Presence of STDP*.

As a last step, we examined how individual neuron frequencies change due to STDP-based synaptic reorganization during NREM sleep. In our network model, we compared the effects of this reorganization when networks had type 2 pyramidal cell excitability (low Ach, indicative of NREM sleep) with a condition where network oscillations are abolished by silencing inhibitory neurons (similar to CNO mice in vivo).

We examined the changes in wake-state neuronal firing frequencies after a reorganization interval in these two scenarios. In the model, when STDP-based synaptic reorganization took place when networks had type 2 dynamics, we observed a significant increase in the firing rates of excitatory neurons initially firing at low rates during wake; in contrast, those with the highest initial firing showed a decrease in firing rate (Figure 4.6a). We color-coded neurons based on their relative change in frequency across sleep and identify in the corresponding pre-reorganization raster plot (Figure 4.6a inset) that neurons that fire faster (slower) also fire earlier (later) in the oscillation during type 2 dynamics (consistent with Figure 4.3) and show a decrease (increase) in firing frequency due to learning. Note that some neurons did not fire during NREM and so did not show a significant change in firing frequency due to learning (black points in Figure 4.6a). By

comparison, silencing of inhibitory neurons in the network led to a homogenous increase in firing rates for those neurons active consistently across learning (Figure 4.6b). We compared these results with data recorded from the hippocampus of mice subjected to corresponding experimental manipulations [43]. We measured the change in log firing frequency across a six-hour time interval at the start of the animals' rest phase (i.e., starting at lights on), with recordings occurring either the day before CFC (baseline) and immediately after training (post CFC). Following CFC, mice either were (a) allowed ad lib sleep with administration of a vehicle (DMSO), or (b) were allowed ad lib sleep with pharmacogenetic inhibition of hippocampal PV+ fast-spiking interneurons (CNO). Changes in log firing rate for all stably-recorded neurons were calculated in each condition, as a function of their baseline firing rate. The resulting best-fit lines reveal that all cases show relatively low rescaling before exposure CFC but that initially low firing neurons increase their firing rate in response to fear conditioning, with the greatest increase seen in the control condition, DMSO mice (Figure 4.6c). Indeed, comparing the slopes of firing rate change (Figure 4.6d) for post-shock recordings show significantly weaker reorganization of firing rates in CNO than in in DMSO treated mice. This rescaling of firing rates across the network is an important prediction of the model, as it suggests a possible universal network-level correlate of sleep-based memory consolidation in vivo.

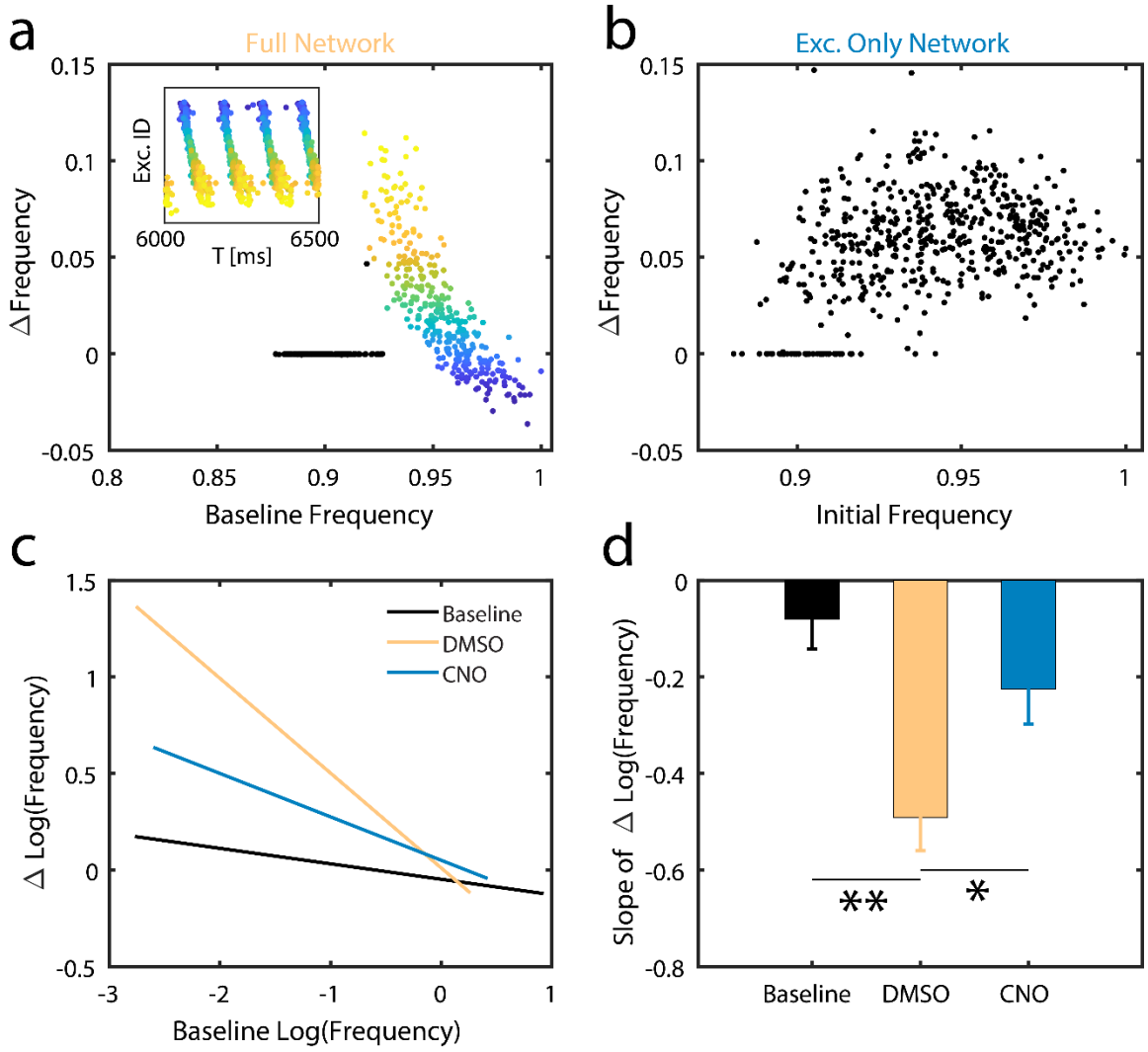

Figure 4.6 Memory consolidation during sleep differentially affects frequency of firing neurons – model prediction and experiment. **MODEL**: a) Changes in individual neuron firing frequency (normalized to baseline) due to memory consolidation during sleep as a function of normalized baseline firing frequency. Neurons are color-coded based on their change in frequency from baseline and the color is conserved in the raster plot; black-colored neurons are those which did not fire during NREM sleep. Of the neurons that are consistently active, those with initially lower frequency increase their frequency whereas neurons with initially higher frequency decrease their frequency. Excitatory-excitatory connectivity strength was increased to $g_{E-E} = 0.00001\ mS/cm2$ to proctect against synaptic depression. Inset:a snapshot of the corresponding raster plot in NREM sleep before learning. b) Change in firing frequency (normalized to baseline) as a function of normalized baseline firing frequency when blocking inhibition diminished oscillations. Unlike in (a), firing frequency changes homogenously across baseline firing rates. **EXPERIMENT**: c) Best fit lines of the change in log firing rates vs the initial log firing rate, comparing baseline recordings (solid lines; composite n = 11) to post recordings (dashed lines) for the first 6 hours post CFC in experimental DMSO (Yellow; n = 3), CNO (Teal; n = 3), and SD (Blue; n = 5). d) Slope comparison of change in log firing rates for DMSO, CNO, and SD baseline and post-shock recordings. Analysis of covariation revealed statistically significant slope differences between DMSO and CNO (*, $p = 0.0114$), and DMSO and Baseline (**, $p < 0.0001$).

4.5 Discussion

Sleep has long been known to be vital for successful memory consolidation. Sleep's requirement for long-term memory storage has been demonstrated across organisms and across different types of memories (e.g., those mediated by network activity in the hippocampus vs. sensory cortex; [150]). Similarly, recent advances have shown that varying oscillatory dynamics accounts for different aspects of memory consolidation [52, 155]. Here, we argue that neuronal phase-locking to network oscillations expressed during NREM sleep promotes feed-forward synaptic plasticity (i.e., STDP) from highly active neurons to less active ones that promotes long-term memory storage.

Our model demonstrates that the initial encoding of memories in the hippocampus during CFC increases theta-band oscillation power (Figure 4.1, 4.2d). The augmentation of these oscillations occurs in a NREM sleep-like, low-ACh, type 2 network state This is due to the fact that the neurons are less responsive to input (i.e. the input-frequency curve flattens), and they exhibit higher propensity to synchronization as their phase response curve becomes type 2. Subsequently, these oscillations mediate input-dependent phase locking of neuronal spiking. This locking to network oscillations leads to more stable firing relationships between neurons. We observe this increase in stability (i.e. increase in FuNS) both in our hippocampal network model after introduction of a synaptically-encoded memory and experimentally after hippocampus-dependent memory encoding in vivo (Figure 4.2).

By assessing the relative change in FuNS recorded in the CA1 network following CFC, we have determined the salient features of network wide dynamics that accompany successful sleep-dependent memory consolidation (Figures 4.2-4.6) [15, 43, 110, 112, 147, 156]. We find that increased FuNS is not only a result of stronger low-frequency oscillatory patterning of the network, but that it also predicts whether experimental conditions will support, disrupt, or rescue fear memory consolidation (Figure 4.2c).

This increased stability in turn mediates mapping between frequency response during wake and relative phase-of-firing patterns during sleep dynamics (Figure 4.3) – neurons that fire with higher frequency during wake tend to fire during sleep before those with lower spiking frequency during wake. We showed that this frequency/timing relationship can be detected in experimental data during sleep phases (Figure 4.4).

Moreover, when the oscillations are abolished by blocking activity of PV+ interneurons (in the experiments and in the model, Figure 4.5) the frequency/timing relationship is reduced, indicating that oscillations do play central role in this process.

Finally, when model networks subsequently undergo structural re-organization through classic STDP mechanisms, we observe renormalization of firing frequencies, with sparsely firing neurons increasing their firing rate substantially, and highly active neurons decreasing their firing rate. These results are consistent with experimental observations in CA1 during CFM consolidation (Figure 4.6). Furthermore, we again observed reduction of these effect when the activity of PV+ neurons is blocked. Similar firing rate changes were also reported in neurons across periods of sleep in other brain structures [15, 147].

Thus, taken together, our results hint at a universal mesoscopic network mechanism underlying what is commonly referred to as systems consolidation. Our results further point to the hypothesis that while the brain may use a rate code to initiate memory encoding during experience, memories may be preferentially consolidated through phase-based information coding, in the form of phase locking of firing network oscillations in subsequent sleep. This is supported by the idea that rate-based information coding in the brain is highly limited (as described in the introduction) and is experimentally supported by our data showing that CFM consolidation does not occur in the absence of sleep (Figure 4.2 inset) – unless a network oscillation is artificially generated in CA1 [43]. A simplified schematic of this idea is shown in Figure 4.7.

While the present study is focused on matching parameters of computational models to data from the hippocampus during fear memory consolidation, we believe that the mechanisms outlined here may be universally true. For example, sleep, and sleep-associated network oscillations, are required for consolidation of experience-dependent sensory plasticity in the visual cortex [52, 157, 158]. Moreover, similar frequency-dependent changes in neuronal firing rates are also observed across periods of sleep in the visual cortex [15]. Based on these and other recent data linking network oscillations in sleep to many forms of memory consolidation, this suggests a unifying principle for sleep effects on cognitive function, and one that could reconcile discrepant findings on how sleep affects synaptic strength [150]. It also paints a more complex but complete role of sleep in memory management that is often proposed [51]. Here we show that sleep may on one

hand facilitate disconnection of high frequency cells that initially participate in memory coding, but at the same time, mediates recruitment of new neuronal populations during consolidation of memory engram.

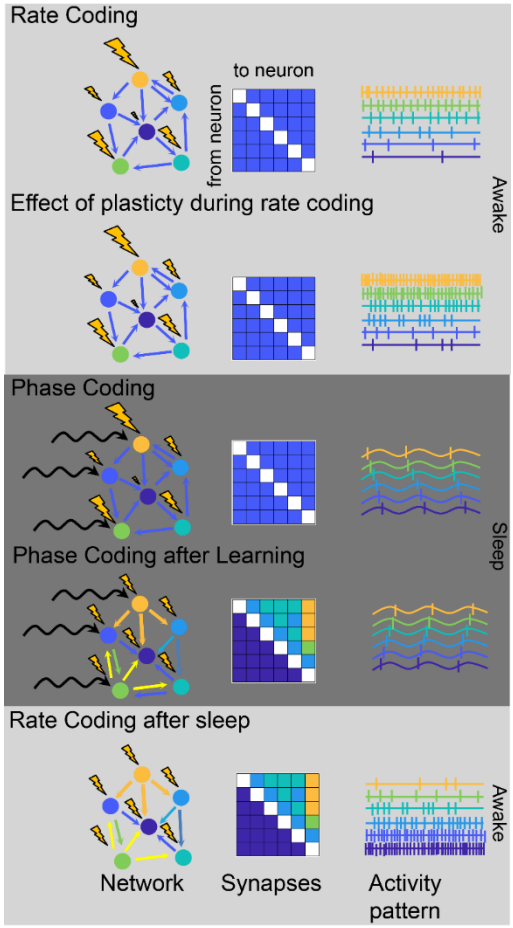

Figure 4.7: Proposed progression of memory consolidation and replay during wake and sleep. Top (wake): Rate coding during wakefulness arranges spiking behavior in order of frequency, where neurons receiving a larger input fire faster. Learning during rate coding effectively speeds up the firing frequency and causes a uniform increase in all synaptic connectivity strengths. Middle (Sleep): Phase coding, on the other hand, occurs when neurons are firing at nearly the same frequency, with those neurons receiving greater input occurring first in a firing sequence. The effect of learning during rate coding is a preferential increase in synaptic connectivity strength, as well as de novo creation of new synapses, following the sequence. The firing phases eventually even out and then switch, so that the neurons firing first in the original sequence fire simultaneously and then last as learning progresses, respectively. Switching back to rate-coding after consolidation occurs during phase-coding effectively switches the order of the fastest firing neurons. Thereby, the process repeats itself and those neurons that are phase locked will see a net potentiation and therefore consolidation of a fear memory.

Chapter 5

Conclusion

Our brains have evolved to become efficient information processing machines, easily converting analog sensory signals into compressed digital codes. Interpreting this code is effortless for the brain itself but remains mysterious for those of us trying to understand it. Steady advancement of technology over the past century has provided a wealth of tools at our disposal to crack the brain's coding scheme and monumental efforts have been enacted to do just that. Studies show that information is contained in the temporal coding of spikes among neurons, which is facilitated by networks of synaptic connections. How this code should be interpreted is an ongoing discussion with many competing and cooperating theories for all manners of neurological processes. In the work presented here, I outlined three interconnected theories relating to the process of memory consolidation: functional network stability, system criticality, and oscillation-driven phase-coding. At first, the interplay between critical dynamics and oscillations and functional network stability seems counterintuitive. How can criticality, a state often defined as an asynchronous and irregular state [138, 144], produce synchronization leading to emergent oscillations and stable networks?

The explanation is that the time scale between these processes is not necessarily the same. Critical dynamics are typically assessed on the order of system-wide interspike intervals [58] while synchronization may utilizes longer time scales and the oscillations produced can vary in frequency from very slow as in NREM sleep (Delta [0.1-4Hz] & Theta [4-12Hz]) to fast as in Wake (Beta[12-30Hz], Gamma [30-80Hz]) and FuNS is limited by features of functional connectivity identification (though AMD is "parameter free" and has no characteristic time scale). At the same time, evidence has shown that critical dynamics and oscillations coexist due to the network structure itself [159], indicating that these feature may be inextricably linked and that the results I present above are indeed present in tandem.

The importance of criticality in information processes has been shown in numerous studies [3, 20-23, 120]. Here, I showed for the first time that criticality may be important for the storage of new memories. Long-range correlations in the model network support a transfer of information about the new configuration throughout the network (Figure 3.2b), allowing the initial encoding of new information near criticality (Figure 3.2a) to be successfully consolidated into a stable attractor through plasticity (Figure 3.2c). Upon storage, critical dynamics are shifted to become more stable and help facilitate the existence of a dynamical attractor representing the new memory (Figure 3.2, 3.3), a feature I found evidence of in NREM hippocampal recordings of mice exposed to fear stimuli (Figure 3.4). This dynamic stabilization was similarly observed at the network level in the same mice (Figure 2.16, 4.2) (further proof of link between these dynamical classifications), indicating that (1) the FuNS algorithm is powerful and (2) that oscillatory activity of NREM is providing a convenient mode of communication between neurons through temporal coding.

The results in Chapter II show many features that may be important in understanding information encoding at the network-level: (1) Average-minimal distance provides an analytical framework to rapidly asses relationships among neurons (Figures 2.3-2.6), which may be important in time-sensitive situations such as in identifying the onset of epileptic seizures [160]; (2) tracking network similarity over time using FuNS reveals potential metastable networks utilized by the brain to represent distinct features; and (3) that AMD and FuNS are robust with respect to how data is generated and can even be used to track the stability in space as well as time. Oscillations are thought to provide this spatiotemporal means of communication.

Oscillatory activity is ongoing in the brain, with different modes of oscillations attributed to different vigilance states and cognitive functions. In Chapter IV, I found that slow-wave oscillations associated with NREM may facilitate a phase-coding mechanism, where neurons that are highly active during wake lead oscillations during NREM, whereas less active neurons in wake lag within oscillations during NREM. In model simulations, I showed that heterogeneity in synaptic connections leads to emergent oscillations among resonator, Type 2 (T2), neurons (Figure 4.1) which was revealed to precisely structure spike timings in a frequency dependent manner (Figure 4.3). Ongoing evidence supports

that sleep may be a period of active [15, 147], and not just passive[51] plasticity, and may then, when combine with this precise spike ordering, preferentially strengthen connections from fast-spiking neurons to slow-spiking neurons (while weakening in the other direction), leading to a homogenization of firing rates among neurons (Figure 4.6; see also [15, 147, 156]). Hence, phase-coding may provide the dynamics to facilitate FuNS in NREM sleep for mice that learn well.

Taken together, I showed that dynamical characterizations of spiking activity help provide mechanistic explanations of how the brain processes, encodes, and consolidates new information. While contention about the critical brain rages on [25, 27], I showed that it can be a unifying mechanism, cooperating with neuronal oscillations to form dynamical attractors as stable network interactions.

5.1 Future Directions

The work I presented here only scratches the surface on the dynamical interplay between these mechanisms and their collective and individual roles in elucidating brain function. I believe that future projects stemming from this work should attempt to answer two central questions: (1) how does the interplay between various plasticity mechanisms shape network structure and dynamics in the context of memory?; and (2) what affects does hippocampal memory recall have on the dynamical state of other cortical areas?

5.1.1 Investigating Plasticity Mechanisms and Their Role in Memory Consolidation

In the work I present here, I investigated activity-based plasticity. In Chapter 3, this took the form of a standard spike-timing dependent plasticity between excitatory neurons. Multiple other forms of synaptic plasticity are known to exist, including plasticity of inhibitory connections and between inhibitory neurons [9] as well as homeostatic plasticity mechanisms, which are thought to be responsible for maintaining a relative stability of firing rates in individual neurons [51, 161]. Homeostatic plasticity in particular is thought to facilitate synaptic downscaling, where connections are reduced and prime networks for subsequent periods of learning. The sleep homeostasis hypothesis claims that this is the main role of plasticity during sleep [51], though our results in Chapter 2 and 3 show that sleep may play a more active role in plasticity. Still, understanding how these plasticity mechanisms interact and the effect they have on networks and spiking dynamics could provide potentially useful information about encoding multiple memories. One simple way

to investigate STDP together with homeostatic plasticity is through an extension of the model outlined in Chapter 3. Instead of plasticity only occurring during short time windows when both pre- and postsynaptic neurons are active, STDP can be set to compete with a constant synaptic change which attempts to fix input and output strength to a (possibly predetermined) level. This homeostatic plasticity would have to be relatively weak overall, lest it adversely shut down STDP. At the same time, evidence exists that STDP may require a bursts of activity between pre- and post-synaptic scales to be switched on [38]. This could also be accounted for in the model by introducing a gating variable (following a standard sigmoidal distribution) for changing synaptic weight. Together, these studies may provide insight into more realistic means of synaptic plasticity that permit both the storage of new memories and the maintenance required to continually encode them.

5.1.2 Multiscale Brain Dynamics during Hippocampal Memory Recall

Each model considered here was confined to a single layer of neuronal networks, but the brain is known to be multilayered and hierarchical [2]. It would be interesting to investigate whether adding new layers with more extensive connectivity motifs adversely affects dynamical states and how information would propagate through the layers. This opens potentially new avenues of tracking dynamics and information transfer across regions of the brain. Are all parts of the brain firing under a universal dynamical state as suggested by the theory of brain criticality, or do different regions have slightly different firing patterns? If so, what are the benefits? Does this help information transfer between regions? Can rate-coding play a larger role in this context?

In the context of contextual fear conditioning, this raises the possibility of tracking how hippocampal fear memory reactivation propagates to other cortical regions. It may be that hippocampal reactivation of the fear memory elicits a similarly stable network response in somatosensory cortex, for example, which processes tactile stimulations; recording from both regions across the training cycle may reveal a hierarchical processing that unifies dynamics across cortical regions. At the same time, output from the hippocampus may be tracked and the signal can be related to oscillatory tone of the brain, further revealing potential temporal-coding processes related to information transfer between regions.

Appendix A

Experimental Methods

A.1 Hippocampal Recordings, Fear Conditioning, and Sleep Deprivation

All procedures were approved by the University of Michigan Animal Care and Use Committee. Male C57BL/6J mice between 2 and 6 months were implanted (described in more detail in [110, 111]) consisted of custom built drivable headstages with two bundles of stereotrodes implanted within right-hemisphere CA1. 3 EMG electrodes were placed in nuchal muscle to monitor activity. The signals from the stereotrodes were split into local field potential (0.5-200 Hz) and spike data (200 Hz-8 kHz). Single neuron data was discriminated as described in Ognjanovski N et al. 2014 [110], and only neurons stably recorded across each 24 hour period were used in the subsequent analyses.

The mice underwent CFC (placement into a novel environmental context, followed 2.5 min later by a 2-s, 0.75 mA foot shock; n = 5 mice), sham conditioning (placement in a novel context without foot shock; Sham; n = 3 mice), or CFC followed by 6 h of sleep deprivation (SD; n = 5 mice). Spike data from individual neurons was discriminated offline using standard methods (consistent waveform shape and amplitude on the two stereotrode wires, relative cluster position of spike waveforms in principle component space, ISI $\geq$ 1 ms) [15, 43, 52, 110, 111]. Only neurons that were stably recorded and reliably discriminated throughout the entire baseline and post-conditioning period were included in subsequent analyses of network dynamics.

Post-shock, animals were either allowed ad lib sleep or 6 hours of sleep deprivation via gentle handling. 24 hours later, freezing behavior to conditioning context was assessed to evaluate the formation of a contextual fear memory.

A.2 Pharmacogenetic Inhibition of Interneurons

2-3-month-old male Pvalb-IRES-CRE mice were bilaterally injected with either the inhibitory receptor hM4Di (rAAV2/Ef1A-DIo-hM4Di-mCherry; UNC Vector Core: Lot #AV4,708) or a control mCherry reporter (raav2/Ef1A-DIo-mCherry; UNC Vector Core:

Lot #AV4375FA) (methods further elaborated in [111]). Using the same implant procedures described above, the animals were implanted with stereotrode bundles.

After allowing 4 weeks for viral expression, the animals underwent contextual fear conditioning (as described above). Post-shock, mice were either given 0.04 mL i.p. injecting of either 0.3 mg/kg clozapine-N-oxide (CNO) dissolved in DMSO (to activate the DREADD) or DMSO alone (as a control). Further detail can be found in Ognjanovski N et al. 2018 [43], where this data set was originally published.

Appendix B

Analytical Methods

B.1 Burst Detection and Spiking Asymmetry

In order to elucidate the presence of a phase-coding mechanism, we analyzed bursts of activity for spiking asymmetry between active neurons. First, recordings of a given interval of length L were segmented into smaller windows of length x (25 ms in CFC, 50 ms in CNO/DMSO (Appendix A.2), and 100ms in model simulations; with times chosen to maximize number of pairwise co-activations occurring) with windows overlapping by 12.5 ms to increase the sampling of the interval L and to reduce effects of windows onset. Then, the total number of active neurons, in each window is determined and used to define a burst-detection threshold: a burst occurs if the activity in a window is significantly greater than the mean background activity, averaged over all intervals of a given vigilance state. More precisely, if a window $\boldsymbol{w_i}$ has a corresponding number of active neurons $\boldsymbol{k_i}$, then the set of windows representing bursts over all intervals is given as $\boldsymbol{B} = \{\boldsymbol{w_i} | \boldsymbol{k_i} \geq \boldsymbol{\mu_k} + \boldsymbol{2\sigma_k}\}$, where $\boldsymbol{\mu_k}$ and $\boldsymbol{\sigma_k}$ are the mean and standard deviation across all w.

Spiking Asymmetry Calculation

Next, the pairwise spiking asymmetry A is calculated across all bursts, where A is an NxN matrix with entries $\boldsymbol{A_{m,n}} > \boldsymbol{0}$ if the spikes of neuron m occur before the spikes of neuron n on average, and $\boldsymbol{A_{m,n}} < \boldsymbol{0}$ in the opposite case. The exact value of an entry $\boldsymbol{A_{m,n}}$ is given as the normalized sum of percent differences between the number of spikes of neuron n occurring after and before each spike of neuron m, across all detected bursts B:

$$\boldsymbol{A_{m,n}} = \left(\frac{\boldsymbol{1}}{\boldsymbol{B_{m,n}}}\right) \sum \sum \frac{(\boldsymbol{T_{n>m}} - \boldsymbol{T_{n<m}})}{(\boldsymbol{T_{n>m}} + \boldsymbol{T_{n<m}})}$$

Where $\boldsymbol{T_{n>m}}$ represents the number of spikes of neuron n occurring after a given spike of neuron m, the inner sum is over the number of spikes of neuron m within a given burst, the outer sum is over all bursts, and the normalization factor is the total number of bursts where neurons n and m are coactive. Given that the firing rate of each neuron is (at least) slightly

different, the asymmetry is typically not symmetric about the $\boldsymbol{A_{m=n}}$ diagonal, i.e. $\boldsymbol{A_{m,n} - A_{n,m} \neq 0}$.

B.2 From Spiking Asymmetry to Phase Asymmetry

The asymmetry matrix A informs on the relative spike timings between different neurons but does not necessarily inform about corresponding phase relationships. Thus, to bridge the gap between spiking asymmetry and phase asymmetry, the rows and columns of A are sorted by neuron spiking rate measured within a given vigilance state. Sorting in this way and working within the framework of our hypothesis, spiking asymmetry of slow firing rate neurons compared to high firing rate neurons will (a) compose the lower triangular matrix of A and (b) will be more positive than the upper triangular matrix of A if faster firing neurons lead slower firing neurons. We thus compared each pairwise entry of $\boldsymbol{A_{m,n}}$ in the lower triangular matrix with its reciprocal $\boldsymbol{A_{n,m}}$ in the upper triangular,

$$\widetilde{\boldsymbol{A}}_{\boldsymbol{m,n}} = \boldsymbol{A_{m,n} - A_{n,m}}$$

under the expectation that if $\widetilde{\boldsymbol{A}}_{\boldsymbol{m,n}} > \boldsymbol{0}$, then the faster firing neuron leads the slower firing neuron on average. This comparison is thus a remapping from the NxN matrix to vector of length 0.5(N^2-N), where we have excluded the $\widetilde{\boldsymbol{A}}_{\boldsymbol{m=n}}$ entries.

We next determined the significance of each $\widetilde{\boldsymbol{A}}_{\boldsymbol{m,n}}$ by randomizing the timing of each neuron's spike within each burst, 100 times. The value of significance is then given by the Z-score,

$$\boldsymbol{Z}\left(\widetilde{\boldsymbol{A}}\right)_{\boldsymbol{m,n}} = \frac{\widetilde{\boldsymbol{A}}_{\boldsymbol{m,n}} - < \widetilde{\boldsymbol{A}}_{\boldsymbol{m,n}} >_{\boldsymbol{randomized}}}{\boldsymbol{\sigma}(\widetilde{\boldsymbol{A}}_{\boldsymbol{m,n}})_{\boldsymbol{randomized}}},$$

where $< \widetilde{\boldsymbol{A}}_{\boldsymbol{m,n}} >_{\boldsymbol{randomized}} \sim \boldsymbol{0}$ in all cases (due to the method of randomization) and thus $\boldsymbol{Z}\left(\widetilde{\boldsymbol{A}}\right)_{\boldsymbol{m,n}} \geq \boldsymbol{2}$ indicates that neuron m leads neuron n in a significant way. Moving forward, it is this distribution of Z-scores Z that we use to assess the phase asymmetry.

We have found that the most concise way of characterizing changes in Z (e.g. due to learning) is by comparing the cumulative distribution functions, CDF(Z). For example, imagine there are two CDFs, $\boldsymbol{CDF_1}$ and $\boldsymbol{CDF_2}$, with the same range and span of Zs. Then, at any particular value $\boldsymbol{Z_x}$, $\boldsymbol{CDF_1(Z_x) > CDF_2(Z_x)}$ indicates that there are more $\boldsymbol{Z > Z_x}$ remaining in $\boldsymbol{CDF_2}$ than in $\boldsymbol{CDF_1}$. We make use of this relationship by taking the weighted averaged of differences in CDFs,

$$\boldsymbol{D} = \frac{\mathbf{1}}{\sum_{g}(\boldsymbol{N_1} + \boldsymbol{N_2})_g} \sum_{g} (\boldsymbol{N_1} + \boldsymbol{N_2})_g (\boldsymbol{CDF_1}(\boldsymbol{Z}) - \boldsymbol{CDF_2}(\boldsymbol{Z}))_g$$

Where N represents the size of each vector Z and the summation is over all recordings belonging to a specific group (e.g. all CFC mice).

BIBLIOGRAPHY

1. L. D. Landau and E. M. Lifshitz, Statistical Physics. (Butterwork-Heinemann, Sain Louis, Missouri, 1984).
2. O. Sporns, Networks of the Brain. (The MIT Press, Cambridge, Massachusetts London, England, 2011).
3. J. M. Beggs and D. Plenz, Journal of Neuroscience 23 (35), 10 (2003).
4. E. R. Kandel, J. H. Schwartz and T. M. Jessell, Principals of Neural Science, 4 ed. (McGraw-Hill, 2000).
5. M. Borczyk, M. A. Sliwinska, A. Caly, T. Bernas and K. Radwanska, Scientific Reports 9 (2019).
6. N. Hardingham, J. Dachtler and K. Fox, Frontiers in Cellular Neuroscience 7 (2013).
7. G.-q. Bi and M.-m. Poo, Annual Review of Neuroscience 24, 27 (2001).
8. G. M. Wittenberg and S. S.-H. Wang, Journal of Neuroscience 26 (24), 7 (2006).
9. G. Hennequin, E. J. Agnes and T. P. Vogels, Annual Review of Neuroscience 40, 22 (2017).
10. D. O. Hebb, Organization of Behavior: A Neuropsychological Theory. (Psychology Press, 1949).
11. K. S. Lashley, in Society for Experimental Biology; Physiological Mechanisms in animal behavior (Academic Press, 1950), pp. 454-482.
12. T. J. Ryan, D. S. Roy, M. Pignatelli, A. Arons and S. Tonegawa, Science 348 (6238), 6 (2015).
13. A. Attardo, J. E. Fitzgerald and M. J. Schnitzer, Nature 523 (7562), 4 (2015).
14. S. Trouche, B. Bontempi, P. Roullet and C. Rampon, Proceedings of the National Academy of Sciences of the United States of America 106 (14), 5 (2009).
15. B. C. Clawson, J. Durkin, A. K. Suresh, E. J. Pickup, C. G. Broussard and S. J. Aton, Frontiers in Systems Neuroscience 12 (40) (2018).
16. N. Rodriguez-Iglesias, A. Sierra and J. Valero, Frontiers in Cell and Devlopmental Biology 7 (24) (2019).
17. J. J. Hopfield, Proceedings of the National Academy of Sciences of the United States of America 79 (8), 4 (1982).
18. P. Bak, C. Tang and K. Wiesenfeld, Physical Review Letters 59 (381) (1987).
19. A. Sornette and D. Sornette, Europhysics Letters 9 (3), 5 (1989).
20. C. Haldemann and J. M. Beggs, Physical Review Letters 94 (058101) (2005).
21. S. H. Gautam, T. T. Hoang, K. McClanahan, S. K. Grady and W. L. Shew, PLOS Computational Biology 11 (12) (2015).
22. W. L. Shew, H. Yang, S. Yu, R. Roy and D. Plenz, Journal of Neuroscience 31 (1), 8 (2011).
23. N. M. Timme, N. J. Marshall, N. Bennett, M. Ripp, E. Lautzenhiser and J. M. Beggs, Frontiers in physiology 7 (425), 19 (2016).
24. C. Rossert, P. Dean and J. Porrill, PLOS Computational Biology 11 (10) (2015).
25. J. Touboul and A. Destexhe, Physical Review E 95 (1), 14 (2017).
26. B. Cessac, A. L. Ny and E. Locherbach, Neural Computation 29 (1), 24 (2017).
27. V. Priesemann, M. Wibral, M. Valderrama, R. Propper, M. L. V. Quyen, T. Geisel, J. Triesch, D. Nikolic and M. H. J. Munk, Frontiers in Systems Neuroscience 8 (2014).
28. J. Wilting and V. Priesemann, Nature Communications 9 (2325) (2018).

29. D. R. Chialvo, Nature Physics 6 (10), 6 (2010).
30. W. L. Shew and D. Plenz, The Neuroscientist 9 (1) (2013).
31. C. V. D. Malsburg, Neuron 24 (1), 10 (1999).
32. W. Singer, Neuron 24 (1), 17 (1999).
33. M. Riesenhuber and T. Poggio, Neuron 24 (1), 7 (1999).
34. E. D. Adrian and Y. Zotterman, The Journal of Physiology 61 (4), 18 (1926).
35. A. Thompson, A. Gribizis, C. Chen and M. C. Crair, Current Opinion in Neurobiology 42, 7 (2017).
36. L. Gao, K. Kostlan, Y. Wang and X. Wang, Neuron 91, 14 (2016).
37. E. Ahissar, R. Sosnik and S. Haidarliu, Nature 406 (6793) (2000).
38. F. Zeldenrust, W. J. Wadman and B. Englitz, Frontiers in computational neuroscience 12, 14 (2018).
39. K. Karbowski, Journal of Neurology 249, 2 (2002).
40. D. Senkowski, T. R. Schneider, J. J. Foxe and A. K. Engel, Trends in Neurosciences 31 (8), 10 (2008).
41. P. Fries, Trends in Cognitive Sciences 9 (10) (2005).
42. G. Buzsaki and B. O. Watson, Dialogues in Clinical Neuroscience 14 (4), 23 (2012).
43. N. Ognjanovski, C. Broussard, M. Zochowski and S. J. Aton, Cerebral Cortex 28 (10) (2018).
44. R. T. Canolty and R. T. Knight, Trends in Cognitive Sciences 14 (11), 12 (2010).
45. M.-B. Moser, D. C. Rowland and E. I. Moser, Cold Spring Harbor Perspectives in Biology 7 (2) (2015).
46. C. Lindemann, J. Ahlbeck, S. H. Bitzenhofer and I. L. Hanganu-Opatz, Neural Plasticity 2016, 14 (2016).
47. B. O. Watson and G. Buzsaki, Daedalus 144 (1), 15 (2015).
48. B. C. Clawson, J. Durkin and S. J. Aton, Neural Plasticity 2016 (2016).
49. J. G. Jenkins and K. M. Dallenbach, The American Journal of Psychology 35 (4), 8 (1924).
50. S. Diekelmann and J. Born, Nature Review Neuroscience 11, 12 (2010).
51. G. Tononi and C. Cirelli, Neuron 81 (1), 22 (2014).
52. J. Durkin, A. K. Suresh, J. Colbath, C. Broussard, J. Wu, M. Zochowski and S. J. Aton, Proceedings of the National Academy of Sciences of the United States of America 114 (39), 5 (2017).
53. J. P. Roach, A. Pidde, E. Katz, J. Wu, N. Ognjanovski, S. J. Aton and M. R. Zochowski, Proceedings of the National Academy of Sciences of the United States of America 115 (13), 8 (2018).
54. G. Buzsaki, Dialogues in Clinical Neuroscience 14 (4), 22 (2012).
55. I. Klinkenberg, A. Sambeth and A. Blokland, Behavioural Brain Research 221 (2), 12 (2011).
56. C. G. Fink, G. G. Murphy, M. Zochowski and V. Booth, PLOS Computational Biology 9 (3) (2013).
57. K. M. Stiefel, B. S. Gutkin and T. J. Sejnowski, PLOS One 3 (12) (2008).
58. V. Priesemann, M. Valderrama, M. Wibral and M. L. V. Quyen, PLOS Computational Biology 9 (3) (2013).

59. C. Meisel, E. Olbrich, O. Shriki and P. Achermann, Journal of Neuroscience 33 (44), 9 (2013).
60. R. Havekes, A. J. Park, J. C. Tudor, V. G. Luczak, R. T. Hansen, S. L. Ferri, V. M. Bruinenberg, S. G. Poplawski, J. P. Day, S. J. Aton, K. Radwanska, P. Meerlo, M. D. Houslay, G. S. Baillie and T. Abel, eLife 5 (2016).
61. W. Nicola and C. Clopath, Nature Neuroscience 22, 13 (2019).
62. J. Wu, Q. M. Skilling, D. Maruyama, C. Li, N. Ognjanovski, S. Aton and M. Zochowski, Journal of Neuroscience Methods 296, 15 (2018).
63. Q. M. Skilling, N. Ognjanovski, S. J. Aton and M. Zochowski, Entropy 21 (11) (2019).
64. G. Buzsaki, Nature Neuroscience 7 (5), 7 (2004).
65. J. Livet, J. W. Lichtman and J. R. Sanes, Nature Review Neuroscience 9 (6) (2008).
66. L. Luo, E. M. Callaway and K. Svoboda, Neuron 57 (5), 26 (2008).
67. E. Chorev, J. Epsztein, A. R. Houweling, A. K. Lee and M. Brecht, Current Opinion in Neurobiology 19 (5), 7 (2009).
68. K. Friston, R. Moran and A. K. Seth, Current Opinion in Neurobiology 23 (2), 7 (2012).
69. A. M. Bastos and J. M. Schoffelen, Frontiers in Systems Neuroscience 9 (2016), 175 (2016).
70. A. Cimenser, P. L. Purdon, E. T. Pierce, J. L. Walsh, A. F. Salazar-Gomez, P. G. Harrell, C. Tavares-Stoeckel, K. Habeeb and E. N. Brown, Proceedings of the National Academy of Sciences of the United States of America 108 (21), 6 (2011).
71. R. Cestnik and M. Rosenblum, Physical Review E 96 (1), 012209 (2017).
72. Y. V. Zaytsev, A. Morrison and M. Deger, Journal of Computational Neuroscience 39 (1) (2015).
73. D. Poli, V. P. Pastore, S. Martinoia and P. Massobrio, Journal of Neural Engineering 13 (2) (2016).
74. K. Shen, R. M. Hutchison, G. Bezgin, S. Everling and A. R. McIntosh, Journal of Neuroscience 35 (14) (2015).
75. H. E. Wang, C. G. Benar, P. P. Quilichini, K. J. Friston, V. K. Jirsa and C. Bernard, Frontiers in Neuroscience 8 (405) (2014).
76. M. E. J. Newman, Physical Review E 69 (6) (2004).
77. M. E. J. Newman, Physical Review E 74 (3), 036104 (2006).
78. M. E. J. Newman, Networks: An Introduction. (Oxford University Press, Oxford, United Kingdom, 2010).
79. S. C. Ponten, A. Daffertshofer, A. Hillebrand and C. J. Stam, NeuroImage 52 (3) (2010).
80. M. Rubinov and O. Sporns, NeuroImage 52 (3) (2010).
81. F. D. V. Fallani, J. Richiardi, M. Chavez and S. Achard, Philosophical transactions of the Royal Society B 369 (1653), 20130521 (2014).
82. K. Supekar, V. Menon, D. Rubin, M. Musen and M. D. Greicius, PLOS Computational Biology 4 (6) (2008).
83. O. Sporns, G. Tononi and G. M. Edelman, Cerebral Cortex 10 (2) (2000).
84. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez and D.-U. Hwang, Physical Reports 424 (4), 34 (2006).

85. J. M. Stafford, B. R. Jarrett, O. Miranda-Dominguez, B. D. Mills, N. Cain, S. Mihalas, G. P. Lahvis, K. M. Lttal, S. H. Mitchell, S. V. David, J. D. Fryer, J. T. Nigg and D. A. Fair, Proceedings of the National Academy of Sciences of the United States of America 111 (52) (2014).
86. S. E. Petersen and O. Sporns, Neuron 88 (1) (2015).
87. B. Misic and O. Sporns, Current Opinion in Neurobiology 40, 8 (2016).
88. H.-J. Park and K. Friston, Science 342 (6158) (2013).
89. D. S. Bassett, D. L. Greenfield, A. Meyer-Lindenberg, D. R. Weinberger, S. W. Moore and E. T. Bullmore, PLOS Computational Biology 6 (4), e1000748 (2010).
90. S. Feldt, P. Bonifazi and R. Cossart, Trends in Neurosciences 34 (5), 12 (2011).
91. S. Gu, F. Pasqualetti, M. Cieslak, Q. K. Telesford, A. B. Yu, A. E. Kahn, J. D. Medaglia, J. M. Vettel, M. B. Miller, S. T. Grafton and D. S. Bassett, Nature Communications 6 (1) (2015).
92. J. D. Medaglia, F. Pasqualetti, R. H. Hamilton, S. L. Thopson-Schill and D. S. Bassett, Neuroscience and Biobehavioral Reviews 75, 12 (2017).
93. E. N. Davison, K. J. Schlesinger, D. S. Bassett, M.-E. Lynall, M. B. Miller, S. T. Grafton and J. M. Carlson, PLOS Computational Biology 11 (1) (2015).
94. A. M. Hermundstad, K. S. Brown, D. S. Bassett and J. M. Carlson, PLOS Computational Biology 7 (6), e1002063 (2011).
95. D. S. Bassett, N. F. Wymbs, M. A. Porter, P. J. Mucha, J. M. Carlson, S. T. Grafton and M. E. Raichle, Proceedings of the National Academy of Sciences of the United States of America 108 (18), 6 (2011).
96. M. Shimono and J. M. Beggs, Cerebral Cortex 25 (10) (2015).
97. S. Nigam, M. Shimono, S. Ito, F.-C. Yeh, N. Timme, M. Myroshnychenko, C. C. Lapish, Z. Tosi, P. Hottowy, W. C. Smith, S. C. Masmanidis, A. M. Litke, O. Sporns and J. M. Beggs, Journal of Neuroscience 36 (3), 15 (2016).
98. S. Pajevic and D. Plenz, PLOS Computational Biology 5 (1), e1000271 (2009).
99. A. Nakhnikian, G. V. Rebec, L. M. Grasse, L. L. Dwiel, M. Shimono and J. M. Beggs, PLOS ONE 9 (3), e89443 (2014).
100. S. Feldt, J. Waddell, V. L. Hetrick, J. D. Berke and M. Zochowski, Physical Review E 79 (5) (2009).
101. R. C. Froemke, Annual Review of Neuroscience 38 (1), 25 (2015).
102. J. Barral and A. D'Reyes, Nature Neuroscience 19 (12), 7 (2016).
103. S. S. Poil, R. E. Hardstone, H. D. Mansvelder and K. Linkenkaer-Hansen, Journal of Neuroscience 32 (29), 7 (2012).
104. R. W. Berg, A. Alaburda and J. Hounsgaard, Science 315 (5810), 4 (2007).
105. R. Ran, L. F. Abbott and H. Sompolinsky, Proceedings of the National Academy of Sciences of the United States of America 114 (44) (2017).
106. D. E. Feldman, Neuron 75 (4), 14 (2012).
107. S. Song, L. F. Abbott and K. D. Miller, Nature Neuroscience 3 (9), 7 (2000).
108. C. A. Ghiani, L. Beltran-Parrazal, D. M. Sforza, J. S. Malvar, A. Seksenyan, R. Cole, D. J. Smith, A. Charles, P. A. Ferchmin and J. d. Vellis, Neurochemical Research 32 (2), 14 (2007).
109. P. Vanderhaeghen and H.-J. Cheng, Cold Spring Harbor Perspectives in Biology 2 (6) (2010).

110. N. Ognjanovski, D. Maruyama, N. Lashner, M. Zochowski and S. J. Aton, Frontiers in Systems Neuroscience 8 (1) (2014).
111. N. Ognjanovski, S. Schaeffer, J. Wu, S. Mofakham, D. Maruyama, M. Zochowski and S. J. Aton, Nature Communications 8, 14 (2017).
112. N. C. Tronson, C. Schrick, Y. F. Guzman, K. H. Huh, D. P. Srivastava, P. Penzes, A. L. Guedea, C. Gao and J. Radulovic, Journal of Neuroscience 29 (11) (2009).
113. S. Daumas, H. Halley, B. Frances and J.-M. Lassalle, Learning and Memory 12 (4), 8 (2005).
114. C. v. Vreeswijk and H. Sompolinsky, Science 274 (5293) (1996).
115. H. Jiang, S. Wang, Y. Huang, X. He, H. Cui, X. Zhu and Y. Zheng, Cell 163 (1), 15 (2015).
116. W. M. Jacobs and D. Frenkel, Biophysical Journal 112 (4), 9 (2017).
117. M. H. T. Bonhoeffer, Neuron 67 (3), 8 (2010).
118. E. H. Baeg, Y. B. Kim, J. Kim, J.-W. Ghim, J. J. Kim and M. W. Jung, Journal of Neuroscience 27 (4), 9 (2007).
119. I. Goshen, M. Brodsky, R. Prakash, J. Wallace, V. Gradinaru, C. Ramakrishnan and K. Deisseroth, Cell 147 (3) (2011).
120. J. M. Beggs, Philosophical transactions of the Royal Society A 366 (1864) (2008).
121. J. Hesse and T. Gross, Frontiers in Systems Neuroscience 8 (2014).
122. T. Petermann, T. C. Thiagarajan, M. A. Lebedev, M. A. L. Nicolelis, D. R. Chialvo, D. Plenz and E. Marder, Proceedings of the National Academy of Sciences of the United States of America 106 (37) (2009).
123. V. Pasquale, P. Massobrio, L. L. Bologna, M. Chiappalone and S. Martinoia, Neuroscience and Biobehavioral Reviews 153 (4) (2008).
124. E. D. Gireesh and D. Plenz, Proceedings of the National Academy of Sciences of the United States of America 105 (21) (2008).
125. C. Tetzlaff, S. Okujeni, U. Egert, F. Worgotter and M. Butz, PLOS Computational Biology 6 (12) (2010).
126. N. Friedman, S. Ito, B. A. W. Brinkman, M. Shimono, R. E. L. Deville, K. A. Dahmen, J. M. Beggs and T. C. Butler, Physical Review Letters 108 (20) (2012).
127. A. J. Fontenele, N. A. P. D. Vasconcelos, T. Feliciano, L. A. A. Aguiar, C. Soares-Cunha, B. Coimbra, L. D. Porta, S. Ribeiro, A. J. Rodrigues, N. Sousa, P. V. Carelli and M. Copelli, Physical Review Letters 122 (20) (2019).
128. H. Yang, W. L. Shew, R. Roy and D. Plenz, Journal of Neuroscience 32 (3) (2012).
129. L. A. Graves, E. A. Heller, A. I. Pack and T. Abel, Learning and Memory 10 (3) (2003).
130. B. Rasch and J. Born, Physiologial Reviews 93 (2), 86 (2013).
131. D. J. Amit, H. Gutfreund and H. Sompolinsky, Annals of Physics 173 (1) (1987).
132. D. J. Amit, Modeling Brain Function: The world of attractor neural networks. (Cambridge University Press, Cambridge, 1989).
133. D. J. Watts and S. H. Strogatz, Nature 393 (6684) (1998).
134. B. Giri, H. Miyawaki, K. Mizuseki, S. Cheng and K. Diba, Journal of Neuroscience 39 (5) (2019).
135. M. A. Wilson and B. L. McNaughton, Science 265 (5172) (1994).
136. B. D. Papa, V. Priesemann and J. Triesch, PLOS ONE 12 (5), e0178683 (2017).

137. P. Bak, C. Tang and K. Wiesenfeld, 59 4 (5), 5 (1987).
138. J. Wilting, J. Dehning, J. P. Neto, L. Rudelt, M. Wibral, J. Zierenberg and V. Priesemann, Frontiers in Systems Neuroscience 12, 55 (2018).
139. M. A. Munoz, R. Juhasz, C. Castellano and G. Odor, Physical Review Letters 105 (12), 128701 (2010).
140. D. Hsu and J. M. Beggs, Neurocomputing 69 (10), 3 (2006).
141. A. D. A. Costa, M. Copelli and O. Kinouchi, Journal of Statistical Mechanics: Theory and Experiment 2015 (6), P06004 (2015).
142. M. Uhlig, A. Levina, T. Geisel and J. M. Herrmann, Frontiers in computational neuroscience 7, 87 (2013).
143. K. B. Hengen, A. T. Pacheco, J. N. McGregor, S. D. V. Hooser and G. G. Turrigiano, Cell 165 (1), 12 (2016).
144. F. Y. K. Kossio, S. Goedeke, B. v. d. Akker and B. Ibarz, Physical Review Letters 121 (5), 6 (2018).
145. J. Zierenberg, J. Wilting and V. Priesemann, Physical Review X 8 (3), 031018 (2018).
146. T.-M. Prince, M. Wimmer, J. Choi, R. Havekes, S. Aton and T. Abel, Neurobiology of Learning and Memory 109, 9 (2014).
147. D. Levenstein, B. O. Watson, J. Rinzel and G. Buzsaki, Current Opinion in Neurobiology 44 (9), 34 (2017).
148. G. Girardeau, I. Inema and G. Buzsaki, Nature Neuroscience 20 (11), 9 (2017).
149. D. Khodagholy, J. N. Gelinas and G. Buzsaki, Science 358 (6361), 4 (2017).
150. C. Puentes-Mestril and S. J. Aton, Frontiers in Neural Circuits 11, 61 (2017).
151. R. M. Smeal, G. B. Ermentrout and J. A. White, Philosophical transactions of the Royal Society B 365, 15 (2010).
152. K. M. Stiefel, B. S. Gutkin and T. J. Sejnowski, Journal of Computational Neuroscience 26 (2), 12 (2009).
153. M. E. Hasselmo, Current Opinion in Neurobiology 16 (6), 6 (2006).
154. B. O. Watson, D. Levenstein, J. P. Greene, J. N. Gelinas and G. Buzsaki, Neuron 90 (4), 14 (2016).
155. J. W. Antony, M. Schonauer, B. P. Staresina and S. A. Cairney, Trends in Neurosciences 42 (1), 3 (2019).
156. H. Miyawaki, B. O. Watson and K. Diba, Scientific Reports 9 (1), 14 (2019).
157. J. Durkin and S. J. Aton, Sleep 39 (1), 5 (2016).
158. S. J. Aton, A. Suresh, C. Broussard and M. G. Frank, Sleep 37 (7), 7 (2014).
159. S.-J. Wang, C. C. Hilgetag and C. Zhou, Frontiers in computational neuroscience 5, 30 (2011).
160. N. Moghim and D. W. Corne, PLOS ONE 9 (6), e99334 (2014).
161. H.-K. Lee and A. Kirkwood, Frontiers in Cellular Neuroscience 13 (2019).