



ADVERTIMENT. L'accés als continguts d'aquesta tesi queda condicionat a l'acceptació de les condicions d'ús establertes per la següent llicència Creative Commons:  <https://creativecommons.org/licenses/?lang=ca>

ADVERTENCIA. El acceso a los contenidos de esta tesis queda condicionado a la aceptación de las condiciones de uso establecidas por la siguiente licencia Creative Commons:  <https://creativecommons.org/licenses/?lang=es>

WARNING. The access to the contents of this doctoral thesis it is limited to the acceptance of the use conditions set by the following Creative Commons license:  <https://creativecommons.org/licenses/?lang=en>

The Role of Behavioral Timescale Synaptic Plasticity for Memory Storage in Neural Networks



Pan Ye Li

Supervised by

Alex Roxin

Department of Mathematics
Autonomous University of Barcelona

A thesis submitted for the degree of
Doctor of Philosophy

September 2024

Abstract

Episodic memory depends crucially on the capacity of neuronal circuits to store information in a one-shot fashion about events that unfold over a time-scale of seconds. Standard Hebbian plasticity rules, such as STDP that require repeated pairing of pre- and post-synaptic activation, are inadequate as physiological mechanisms underlying this type of rapid learning.

Contrary to this, Behavioral Timescale Synaptic Plasticity (BTSP), a newly discovered form of plasticity in the hippocampus, operates on a timescale of seconds. This mechanism induces long-lasting synaptic changes after a single experience, driven by dendritic plateau potentials, making it ideally suited for encoding episodic memories. After just one trial, BTSP's ability to rapidly form place fields in CA1 neurons underscores its critical role in memory formation.

This thesis investigates the role of BTSP in memory storage within the hippocampal network. We derive a simplified BTSP model that lends itself to rigorous mathematical analysis, extending this framework to recurrent networks such as the CA3 region of the hippocampus to explore its memory storage properties. Through a detailed examination of recall dynamics, our results demonstrate that BTSP facilitates the encoding and retrieval of a large number of memories, with variability enhancing both storage and recall. Additionally, we explore the non-Hebbian aspect of BTSP, showing that it supports homogeneous representations in CA3. Consequently, we conclude that BTSP is a viable candidate mechanism underlying episodic memory.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my thesis supervisor, Alex Roxin, for offering me this incredible opportunity in computational neuroscience. His constant support, encouragement, and advice have been the backbone of this work. I am also profoundly thankful to Sandro Romani, Senior Group Leader at Janelia Research Campus, HHMI, for the excellent opportunity to work with him during my stay. Our joint efforts in studying homogeneous representation, featured in one of the chapters, made for an exceptionally enriching experience.

I sincerely appreciate my colleagues and the administrative staff at the Computational Neuroscience Group at CRM. Their support and assistance with the necessary paperwork have been invaluable. I could not have navigated this journey without them.

Additionally, I extend heartfelt thanks to my beloved pets, Suger and Yogurt, for their companionship and emotional support. During the challenging Cronocovid period, Suger was my source of strength, taking care of me through both my switch from mathematics to neuroscience and all the troubles of quarantine. When Suger left for China, Yogurt became my steady companion. She has also been a source of day-to-day joy and emotional steadiness, whereby I have made good progress and comfort during this journey.

Lastly, I express my deepest gratitude to my family and friends for their unwavering support and encouragement. I want to extend special thanks to myself for the courageous decision to resume my studies at the age of 22 — a choice that has significantly influenced the past 16 years and

ultimately enabled me to attain my goal of becoming a doctor. This decision and the support I received are a testament to the power of personal determination and its impact on one's life.

Contents

1	Introduction	1
1.1	The hippocampus: anatomical and functional foundations for episodic memory	2
1.2	Theory of synaptic plasticity	4
1.2.1	Computational models of Hebbian learning	7
1.3	Attractor networks	9
1.3.1	Hopfield Network	9
1.3.2	Attractor networks with bounded synapses	10
1.4	A candidate mechanism for episodic memory	12
2	Behavioral timescale synaptic plasticity	15
2.1	A brief review of behavioral timescale synaptic plasticity	15
2.2	Weight-dependent model of BTSP (Milstein et al., 2021)	19
3	Mathematical modeling of BTSP: one-dimensional map	22
3.1	Constructing spatially dependent plasticity functions	23
3.2	CA3-CA1 feedforward networks	25
3.3	Simulations confirm that 1D map reproduces results of full model . .	26
4	BTSP and memory storage in recurrent networks	30
4.1	1D map for BTSP in recurrent networks	30
4.1.1	Steady-state statistics: Mean and variance	35
4.1.2	Spatial statistics: mean and variance of the memory trace . .	37
4.2	Sparse coding with BTSP in recurrent networks	40
4.2.1	Sparse coding enhances the storage capacity	41
4.2.2	Detailed calculation of 1D map	43

4.2.3	Memory capacity for balanced potentiation and depression . .	45
4.3	Discussion	48
5	Neuronal networks endowed with BTSP can encode a large number of spatial memories as bump attractors	50
5.1	Network dynamics	50
5.2	Ring-model approximation	54
5.3	Impact of variability and network parameters on Turing bifurcation in sparse coding networks	61
5.3.1	Stationary uniform solutions of the ring model	62
5.3.2	Linear stability analysis and Turing bifurcation of the ring model in the absence of noise	65
5.3.3	The role of system size on quenched variability in the ring model	67
5.4	Discussion	71
6	Maintenance of uniform representation in CA3 through BTSP	74
6.1	Fluctuating coding levels in learning n memories	75
6.1.1	Model for BTSP	75
6.1.2	Hebbian learning	77
6.2	Results and discussion	78
7	Conclusion	83
	Bibliography	86
A	Detailed calculation of memory traces and quenched variability in the BTSP rule	97
B	Calculation of the Spatial Fourier Spectrum for Quenched White Noise	102
C	Figure: Bifurcation diagram for $s = 0.2$	107

List of Figures

1.1	The anatomy and circuitry of the hippocampal formation	3
1.2	Scheme of three binary-synapse models	12
2.1	Experimental finding in Bittner et al. 2017	16
3.1	Place field emergence after plateau potential in CA1	23
3.2	A temporal plasticity kernel from the biophysical model can be converted to a spatial kernel for a 1D map	24
3.3	Fitting of spatial kernels captures the degree of plasticity	27
3.4	Consistency of models across various induction protocols	28
3.5	The animal's speed modulates the plasticity kernel	29
4.1	Schematic of BTSP-based learning in a recurrent network	31
4.2	Spatial modulation of synaptic weights decreases with age	33
4.3	The value of P and D can change the shape of spatial statistics	34
4.4	BTSP-based learning in a sparse recurrent network with network size M	42
4.5	Memory capacity improves with sparse coding and increased population size	46
4.6	Absolute difference between $\langle F \rangle^2$ and $\langle F^2 \rangle$ when $P = D$	46
4.7	Storage capacity and optimal plasticity parameter in neural networks with different sizes and sparseness levels	47
5.1	Choice of F-I curve	52
5.2	Network dynamics can be approximated by the ring model in the sparse limit	53
5.3	Comparison of bifurcation diagram for full network model and ring-model approximation	56

5.4	The zoomed-in bifurcation plot distinctly shows a supercritical Turing bifurcation	58
5.5	Calculation of bump amplitude over a range of η show the bump state is constrained to the desired environment	60
5.6	Phase diagram as a function of W_0 and W_1	60
5.7	Bifurcation diagram for the stationary solutions of the ring model . .	64
5.8	Bifurcation diagram for $W_1 = 3$	64
5.9	The memory capacity of the network is shown to depend non-monotonically on the resolution of spatial tiling of place cell	70
5.10	Deviations of the network model from the ring-model approximation are due to the emergence of mixed attractors	72
5.11	Memory capacity in the network model outperforms ring model approximation and scales optimally with system size	73
6.1	Coding levels of external signals and frequencies of plateau potentials	79
6.2	Mean and variance of memory traces	80
6.3	The relationship between f_A and f_{PP}	81
B.1	Comparison of the theory (lines) with numerical simulations for different values of B	104
B.2	Comparison of the theory (lines) with numerical simulations for different values of C.	105
C.1	Bifurcation diagram for $s = 0.2$	107

Chapter 1

Introduction

Among the most remarkable capacities of the human brain is its ability to remember past experiences. These recollections, called episodic memories, are not just echoes of the past but constitute some of the building blocks that shape and define who we are today [1]. Episodic memory is a memory system that encodes, stores, and retrieves the consciously experienced events of an individual's life. It allows one to “relive” one's past and envision possible futures. This ability to mental time travel that is associated with episodic memory highlighted as an essential feature of the system by Tulving, who defined it as “the capacity to have conscious recollection of previous episodes in one's life” [2,3].

It is further characterized by the type of information it encodes, encapsulated by the three W's: “what” happened, “where” it occurred, and “when” it took place [4]. This definition of episodic memory is not only confined to humans but also applied to animals. For instance, in the experiment by Fugazza et al. [5], dogs successfully reproduced a sequence of actions that had previously been shown to them once by a human experimenter. The results indicated that dogs could remember and imitate these actions after a one-minute or one-hour delay. Less cognitively complex animals, such as scrub jays, have also been shown to recall the type of food they cached, the location, and the time of storage. After a short delay, they prefer worms over peanuts, but after a longer delay, they select peanuts because the worms are no longer edible [6]. Evidence of episodic-like memory in other species further supports the view that such memory is not unique to humans [7].

1.1 The hippocampus: anatomical and functional foundations for episodic memory

A considerable body of evidence has established the hippocampus as one core structure for episodic memory. Patients with hippocampal lesions are observed to suffer from hippocampal amnesia, a disorder characterized by a loss in the acquisition of new episodic memories and retrieval of the most recent ones, whereas their distant memories and other intellectual functions are relatively preserved. This condition was first fully documented in the case of Henry Molaison (H.M.), who became one of the most highly studied patients with amnesia in the history of neuroscience [8]. It was later shown that a similar deficit in the episodic memory function as H.M. had could be replicated in monkeys with hippocampi removed bilaterally, thus underscoring the importance of the hippocampus in the encoding and retrieving memories of particular events [9, 10]. fMRI studies in human subjects have extended these findings, showing a reliable activation of the hippocampus during the retrieval of episodic memories and increased activity explicitly related to tasks that require the recall of personal events, a finding that further supports the more central role of the hippocampus within the network supporting episodic memory [11, 12].

Anatomically, the hippocampus is a complex structure located deep within the medial temporal lobe and consists of three distinct subregions: the dentate gyrus (DG), the hippocampus proper, and the subiculum. The hippocampus proper can be further divided into CA3, CA2, and CA1; these subfields are organized in proximal-to-distal order with respect to the DG, with CA3 being more proximal and CA1 being more distal, illustrated in Fig 1.1. The critical distinction between different regions of CA is based on the size of their pyramidal neurons and their synaptic connectivity with the DG. Pyramidal neurons in CA3 are larger in size compared to CA1 and receive direct inputs from DG via mossy fibers, which is lacking in CA1. Between the CA1 and CA3 regions lies a more minor subfield known as CA2. This area is characterized by pyramidal neurons similar in size to those found in CA3. However, unlike CA3, CA2 does not receive input from the mossy fibers originating in the DG; instead, its connectivity resembles that of CA1. Despite the increasing interest in the CA2 (related to social memory, see reviews [13]), this thesis will limit its interest in the CA1 and CA3 hippocampal subregions.

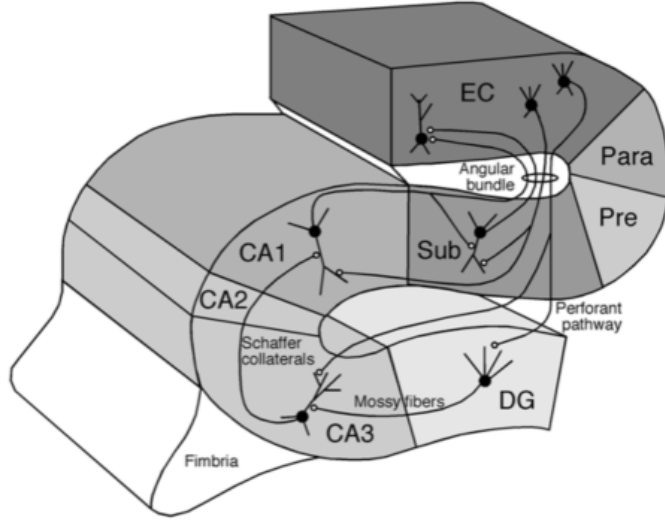


Figure 1.1: **The anatomy and circuitry of the hippocampal formation.** See in-text description. Figure adapted from Amaral and Lavenex 2007 [14].

Unlike the neocortex, where interconnections between areas are reciprocal, the hippocampal formation¹ is predominantly characterized by an unidirectional flow of information. The entorhinal cortex (EC) plays a pivotal role as the main entrance of neocortical inputs into the hippocampus, which receives two types of information in parallel. The medial entorhinal cortex (MEC) is primarily responsible for receiving spatial input, while the lateral entorhinal cortex (LEC) is more focused on nonspatial information, such as objects and items [15]. After inputs arrive at the EC, they are sent to the hippocampus. Layer II neurons from both MEC and LEC project via the perforant pathway to the DG and CA3, making integrating different information types possible. However, there are apparent differences in layer III: LEC layer III neurons project to the distal part of CA1, whereas MEC layer III neurons project to the proximal part of CA1, thereby making functional differences within CA1 [16]. The projection from the EC to the DG initiates the hippocampal processing sequence, where DG granule cells receive input and project solely to the CA3 region via mossy fibers. In CA3, neurons extend axons to CA1 through the Schaffer collaterals but also send collateral branches to other CA3 neurons, supporting intra-regional processing; this unique recurrent connection in CA3 is essential for maintaining memory,

¹The hippocampal formation includes the dentate gyrus, hippocampus proper, subiculum, pre-subiculum, parasubiculum, and entorhinal cortex. The entorhinal cortex can be distinguished into six layers: four cellular layers (II, III, V, VI) and two acellular or plexiform layers (I, IV). [14]

especially spatial memory [17, 18]. CA1 neurons then project back to the entorhinal cortex, particularly in its deeper layers V and VI, either directly or indirectly via the subiculum. Finally, the EC relays the processed information back to the neocortex, thus closing the extensive loop of the information process in the hippocampal formation. (See [14, 16, 19] for a detailed description.)

Within the hippocampus, separate neuronal populations encode different aspects of episodic memory. Of these, place cells, discovered by John O’Keefe in 1971, provide a significant contribution to spatial processing [20]. These neurons become active when rats occupy a specific location and, surprisingly, do not turn off, even in complete darkness — indicating that they rely on some internal cognitive map rather than on external visual inputs [21, 22]. The region where a place cell is active is called its “place field”, and different place cells may have different firing fields. Collectively, these cells build up a dynamic, flexible internal representation of the environment, which is necessary for effective navigation and spatial memory [23].

Beyond spatial encoding, it has been shown that nonspatial factors can modulate place cell activity. A specific subset of place cells, called splitter cells, has revealed this in a way through different patterns of firing based on the particular context or goal of the animal within the same environment, pointing out that it is not only spatial information encoded by the hippocampus but also contextual [24, 25]. The identification of time cells furthers the role of the hippocampus in expanding beyond spatial representation. It comprises neurons that fire at specific moments during an experience, thus representing a timeline of events for temporal sequences of activities and providing a timeframe for memories [26]. These neurons together enable the hippocampus to integrate the “what,” “where,” and “when” of experiences, forming a unique episode for episodic memory.

1.2 Theory of synaptic plasticity

Neurons are the basic computational units of the brain; in the hippocampus, various populations of cells selectively respond to one or another dimension of episodic memory. Following years of theoretical and experimental research, a model of how memories form may be conceptualized as the following: synchronized firing evoked by an event or experience leads to long-lasting changes in neural circuits and forms

a stable assembly of neurons. These assemblies, sometimes referred to as “memory traces,” have been considered the neuronal substrates for information storage in the brain.

Richard Semon first established this theoretical idea of memory traces in 1921. He proposed that an engram was a permanent change within the brain resulting from a specific event or experience. Once formed, the engram remains inactive until a retrieval cue, usually part of the original experience, elicits a response similar to the experience of the original event. Notably, such engrams are not static entities; re-exciting an engram creates a new one associated with the original to enhance the memory [27–29]. However, one central question that remained to be answered in this framework was how these engrams form. Donald Hebb filled this gap by proposing that repeated and simultaneous activation of neurons strengthens the synapses among them [30]. Hebb’s theory posits that external stimuli stimulate a specific set of neurons during memory encoding. By synchronously activating these neurons, their synaptic links are strengthened, forming a stable and interconnected network, which Hebb called a “cell assembly”, akin to Semon’s engram, essential for sustaining the memory in the brain.

Since then, Hebb’s concept of synaptic plasticity, activity-dependent modifications in synaptic strengths, has been believed to be the basis of learning and memory. The first compelling evidence showing that changes in the efficacy of communication between neurons modify the behavioral output came from pioneering work done on *Aplysia californica* [31,32]. This work showed that behavioral habituation, a reduction in response after repeated stimulation of the gill and siphon withdrawal reflex, is associated with reduced neurotransmitter release at specific synapses.

Furthermore, the discovery of long-term potentiation in mammals gave substantial evidence for the theory of synaptic plasticity. Bliss and Lømo [33] first observed it at the DG-CA3 synapses in the rabbit hippocampus, where repeated activation of excitatory synapses resulted in sustained enhancements of postsynaptic responses. It is usually induced through high-frequency stimulation, whereby massive glutamate releases activate AMPA receptors (AMPA) by depolarizing postsynaptic neurons. A sufficiently strong depolarization [34], in turn, removes the magnesium block of NMDA receptors (NMDARs), allowing calcium influx and the intracellular signaling

cascades necessary for LTP induction [35–37]. The voltage dependence of NMDARs further enables the LTP induction through the pairing of low-frequency stimulation with direct depolarization of postsynaptic neurons [38]. Studies show that blocking LTP induction by NMDA antagonists [39] and that saturating LTP during learning [40] disrupt hippocampus-dependent tasks such as spatial learning, pointing to an essential correlation between LTP and memory [41, 42].

As the counterpart of LTP, LTD weakens synaptic strength, thereby reversing the potentiating effect of LTP [43–45]. LTD can be induced through low-frequency stimulation protocols and, similar to LTP, often involves the activation of NMDA receptors but engages distinct calcium signaling pathways [36, 37]. Research extending into the temporal domain revealed that the relative timing between pre- and postsynaptic activity is fundamental to determining the direction and polarity magnitude of the synaptic change [46–48]. Synaptic potentiation is favored when the presynaptic spikes precede the postsynaptic spikes in a time window of tens of milliseconds. Conversely, synaptic depression is induced when the postsynaptic spike precedes presynaptic activity. This precise timing relationship is captured in models, such as spike-time-dependent plasticity (STDP), a Hebbian variant of synaptic plasticity that explicitly relates synaptic changes to the timing of neural events and provides a temporal context of how experience sculpts neural circuits [49, 50].

On the other hand, the engram hypothesis has been extensively tested in various types of memory, especially associative memory. In auditory Pavlovian fear conditioning, researchers have identified the lateral amygdala as a critical site for storing fear engrams [51]. Research using molecular markers like CREB has shown that neurons with high levels of CREB are more likely to be included in an engram and, therefore, play a more significant role in forming fear memory [52]. Manipulating these identified engrams has demonstrated that these neurons are neuronal substrates for memory. For example, selective ablation of neurons believed to be part of the engram disrupts the associated memory, confirming that these neurons are constituent parts of the memory trace [53]. Optogenetic re-activation of the engram cells shows a successful recall of that memory, providing direct evidence that stimulating an engram-specific population is enough to trigger memory recall [54].

These results have emphasized the mutual roles of synaptic plasticity and engram

theory in explaining mechanisms that control memory storage and retrieval in the brain. Memory is retained not only by activating a particular set of neurons but by the distribution of plastic synaptic weights across a neural network [30, 55]. Experiences initiate changes in the synaptic links between neurons, which, in turn, influence the patterns of neuronal firing across the network that are intimately linked with remembered experiences. Such interplay between synaptic plasticity and network dynamics provides the basis whereby learned experiences are stored and retrieved by the brain.

1.2.1 Computational models of Hebbian learning

Hebbian learning is especially central in the study of synaptic plasticity for determining mechanisms of memory and learning. Several mathematical formulations have been proposed [56–58] to describe the change of synaptic efficacy w within the framework of Hebbian plasticity. These models can generally be divided into two main paradigms: the first involves the correlation of neuronal firing rates, which results in synaptic strengths being changed by the coactivation of neurons, and the second paradigm is based on spike-timing-dependent plasticity (STDP), where synaptic changes depend on the precise timing of presynaptic and postsynaptic spikes. Here, we provide a brief description of these models.

The simplest version of Hebb’s rule can be expressed as simultaneous firing driving changes in synaptic strength, represented as :

$$\tau \frac{d}{dt} w = vu,$$

where u and v denote the firing rate of the presynaptic and postsynaptic neurons, respectively. Both take positive value, and their relationship is given by $v = wu$. However, this formulation accounts for only long-term potentiation (LTP). To implement long-term depression (LTD) into the model, it would be necessary to introduce a postsynaptic threshold, θ_v , such that:

$$\tau \frac{d}{dt} w = (v - \theta_v)u.$$

This modification aligns with experimental induction protocol, where LTD occurs when presynaptic activity is paired with postsynaptic activity below a given threshold

(i.e., $v - \theta_v < 0$), and LTP is induced when the postsynaptic activity exceeds the threshold (i.e., $v - \theta_v > 0$). Such modulation allows for the effective inclusion of both LTP and LTD within this model, depending on the postsynaptic response relative to the threshold.

Furthermore, an alternative model can be formulated by applying the threshold to the presynaptic activity instead of the postsynaptic activity:

$$\tau \frac{d}{dt} w = v(u - \theta_u).$$

When considering the averaged inputs and assuming that the threshold represents the average activity in each case, both approaches converge to the following general model:

$$\tau \frac{d}{dt} w = Cw,$$

with $C = \langle (u - \langle u \rangle)u \rangle = \langle (u - \langle u \rangle)(u - \langle u \rangle) \rangle$ is the input covariance matrix. Due to the involvement of C , this model is called the covariance rule.

On the other hand, spike-timing-dependent plasticity (STDP) incorporates a temporal element into Hebbian learning by taking advantage of the precise timing of spikes to determine synaptic changes. In STDP, synapse change depends on the time difference between presynaptic and postsynaptic spikes, given as $\Delta t = t_{pre} - t_{post}$. LTP occurs when a presynaptic spike happens before a postsynaptic spike ($\Delta t > 0$). Conversely, if the postsynaptic spike precedes the presynaptic spike ($\Delta t < 0$), LTD is induced. Therefore, weight update in STDP can be mathematically described by:

$$\Delta w = \begin{cases} A_+ e^{-\Delta t / \tau_+} & \text{provided that } \Delta t > 0 \\ A_- e^{\Delta t / \tau_-} & \text{for } \Delta t < 0 \end{cases}$$

where A_+ and A_- represent amplitudes for LTP and LTD, respectively, and τ_+ and τ_- are time constants that define the time window over which LTP and LTD occur.

At the end, our goal is to understand how plasticity shapes the connectivity within neuronal networks. This leads us to explore the dynamic behavior of these networks, particularly attractor networks, which store memories as stable fixed-point attractors. In the following section, we delve deeper into how attractor networks function and their role in memory storage.

1.3 Attractor networks

One of the central questions in the field of computational neuroscience is how the brain encodes, stores, and retrieves memories. To address this, Donald Hebb proposed that activity-dependent synaptic plasticity underlies memory formation, with changes in synaptic efficacy serving as the cellular substrate for memory. Hebb's model presumes that neural transmission becomes progressively more effective with repeated stimulation, allowing neural networks to maintain representations after removing a stimulus. Building on this, Hopfield developed a mathematical formalism with attractor networks that store memories as stable states and allow recall even from partially or noisily degraded inputs, a hallmark of associative memory processes in the brain. These mathematical models allow for an invaluable investigation of the capacity of memories and, therefore, are especially relevant in studies related to the dynamics of neural memory systems using attractor networks. In this section, we will consider a few different types of attractor networks, focusing on the Hopfield model and its more biologically plausible variants.

1.3.1 Hopfield Network

The Hopfield network was one of the earliest neural models proposed with the Hebbian learning rule [59]. The network consists of N binary neurons, each of which can be in either an active state $s_i = +1$ or an inactive state $s_i = -1$. The synaptic matrix $W = \{w_{ij}\}$ governs the neurons' communication. Under Hebbian learning, the synaptic weight from neuron j to neuron i is designed in the following form to store p memories, each encoded by a pattern $\xi^\mu = \{\xi_1^\mu, \xi_2^\mu, \dots, \xi_N^\mu\}$ for $\mu \in \{1, \dots, p\}$:

$$w_{ij} = \frac{1}{N} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu. \quad (1.1)$$

When a new memory is introduced, the synaptic weight is updated by the rule: $w_{ij}^{new} = w_{ij}^{old} + \frac{1}{N} \xi_i^{new} \xi_j^{new}$. These memories are stored as stable states within an energy landscape: the network updates the neuron states asynchronously, lowering the energy function with every iteration until it converges into a stable state. If the dynamics start near one of the memories, the system evolves to that memory. The initial states that lead to this convergence define the basin of attraction for that memory.

In such a configuration, the network exhibits a vast memory capacity. The number of memory that the network can remember and correctly recall depends on the number of independent synapses, which scales with the size of the network even for random and uncorrelated memories [60]. Using the replica method from statistical physics [61, 62], Amit, Gutfreund, and Sompolinsky determined that the memory capacity scaled linearly with the system size, and specifically $p_{\max} \approx 0.138N$ [63, 64]. Below this threshold, the system can accurately recall all stored patterns, but exceeding this critical value leads to memory retrieval failure due to increased interference between the stored patterns.

Sparse coding can significantly increase memory capacity [64, 65]. In the Hopfield network, memory capacity is limited mainly by interference between stored patterns because the signal, which is the contribution of the target memory, remains constant regardless of the number of patterns. Suppose memories are encoded by a smaller subset of neurons such that only a fraction f of neurons is active for each memory. In that case, overlap between patterns is significantly reduced, lowering noise by a factor of $f < 1$ relative to the signal and substantially increasing capacity. The capacity can reach N^2 when $f \sim 1/N$.

As the number of stored memories grows too large, the network eventually loses the ability to recall any specific memory, a phenomenon termed “memory blackout” [66]; this arises because all memories contribute equally to the weight matrix (see Eq 1.1). As more memories are stored, the weight grows unconstrained, ultimately saturating the system. In the energy landscape, each memory has an identically shaped basin of attraction whose size is proportional to the number of stored patterns, p . When the network is under critical capacity, these basins are well-separated and uniformly distributed, assuming random and uncorrelated memories. However, basins start to overlap for $p > p_{\max}$, and spurious states — unwanted local minima — emerge. If too many memories are stored, all basins collapse into one, causing the network to lose the ability to recall any previously stored memory.

1.3.2 Attractor networks with bounded synapses

Under the assumption that synaptic weights can grow indefinitely by linearly integrating new memories, the Hopfield network exhibits significant memory capacity. How-

ever, in biological systems, synapses are constrained to a finite range. When bounded synapses are introduced into networks, even though the modification looks minor, it completely changes the network dynamics. In such systems, synaptic strengths can reach their limits; once they hit the limits, the older memories are lost due to encoding new ones. The slow erasure of older patterns gives rise to a palimpsest-like system where the previous information is progressively replaced [66,67]. This type of network significantly differs from the Hopfield net, where all memories are stored in the synaptic structure, eventually leading to memory catastrophe.

Incorporating biological constraints into attractor networks drastically reduces memory capacity, shifting from linear scaling with network size to a logarithmic dependence. While the original Hopfield model preserves the memory traces indefinitely in time, the memory traces decay exponentially in models proposed by Amit and Fusi [68], where the synapses undergo transitions between discrete states. This exponential decay results in a capacity that scales with $\log N$ rather than network size. In subsequent work, Amit and Fusi [69] then showed that sparse coding, if implemented, would considerably improve the capacity to $N^2/\log^2 N$ given that the coding level scales as $f \sim \log N/N$ and synaptic transitions are controlled by f .

Empirical data, however, suggests that the loss of memory traces is better described by a power law rather than an exponential function [70,71]. In theoretical models with simple binary synapses, synaptic states can be either potentiated or depotentiated, and the transitions are controlled by the learning rate q , see the left panel in Fig 1.2. The synapses can be classified into fast or slow depending on the relative value of q . For $q \rightarrow 1$, synapses switch states fast to encode new memories but forget them very quickly, which makes them fast synapses. On the other hand, for $q \rightarrow 0$, synapses resist the change of state; therefore, memories survive longer. However, memory decays exponentially for homogeneous models with a single learning rate for all synapses.

Various models have been proposed to implement power-law memory decay, two of which are the cascade model by Fusi, Drew, and Abbott [72] and the heterogeneous synapse model by Roxin and Fusi [73]. In the cascade model (see the right panel in Fig 1.2), synapses have the potential for further potentiation or depotentiation, with each synapse having multiple hidden states. These states either switch (ongoing

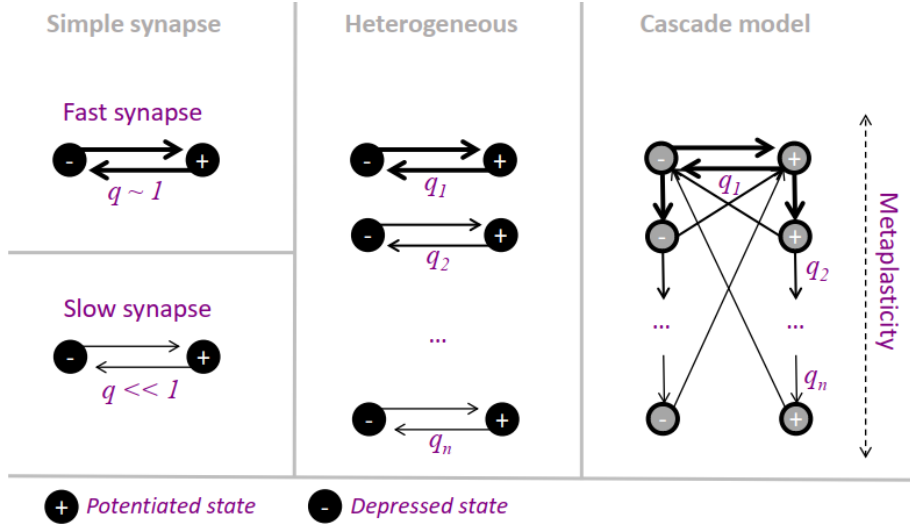


Figure 1.2: **Scheme of three binary-synapse models.** See in-text description. Figure adapted from Fusi 2021 [60].

plasticity) or transition to a more stable metaplastic state, which is more resistant to switching. By integrating different timescales, the network achieves a power-law forgetting phenomenon. Similarly, in the heterogeneous synapse model (see the middle panel in Fig 1.2), synapses are divided into partitions with distinct learning rates, represented by a set of qs , giving rise to a network exhibiting power-law decay. The memory capacity shifts from logarithmic scaling to \sqrt{N} in both cases.

In summary, the Hopfield network was the first transparent setting in which the memory capacity of neural networks could be investigated, but its memory storage declines when overloaded. Incorporation of biologically relevant constraints, such as bounded synapses, yields gradual forgetting in Fusi networks [68, 69], and this reduces capacity from scaling linearly with network size to $\log N$. Other models, like those by Fusi et al. [72] and Roxin and Fusi [73], increase capacity to \sqrt{N} by adding complexity, which shifts the distribution of the forgetting rate from exponential to power-law, in line with experimental findings [70, 71]. Finally, both models find that sparse coding dramatically enhances memory capacity.

1.4 A candidate mechanism for episodic memory

One peculiarity of episodic memory is that many experiences are unique, requiring a fast-learning mechanism that links distinct events over extended timescales.

Experimental evidence indicates that the induction of LTP and LTD relies on repeated stimulation, and STDP only operates on a timescale of milliseconds, making the classical Hebbian rule unsuitable for episodic memory formation. Recently, a novel plasticity rule has been discovered in the hippocampus of mice, dubbed Behavioral Timescale Synaptic Plasticity (BTSP), that could shape the place cell dynamics within one single trial. This plasticity mechanism has a window of several seconds and operates independently of postsynaptic activity, making it a non-Hebbian type of learning. Given its derivation from *in vivo* data, BTSP aligns well with Fusi-type models of bounded synapses. However, the memory storage properties of networks endowed with BTSP have not yet been rigorously analyzed.

This thesis presents a rigorous mathematical analysis of BTSP, focusing on its impact on memory storage and retrieval in recurrent networks, with the main results published in [74]. To this end, the thesis is structured as follow:

- Chapter 2 reviews BTSP and introduces a computational model that successfully explains the observed behaviors found in the hippocampal CA1 region.
- Chapter 3 reduces this model to a one-dimensional (1D) map that can reproduce the results of the full model using more intricate simulation protocols.
- Chapter 4 extends the 1D map to a recurrent network framework to study the storage properties and capacity of BTSP during exploration of a large number of distinct environments. Using the signal-to-noise ratio technique, we find that the memory of networks with BTSP scales logarithmically with the network size and that sparse coding greatly improves the memory capacity.
- Chapter 5 explores network dynamics by applying the synaptic weight matrix, obtained after freezing the learning process, to firing rate equations. A ring-model approximation is employed to evaluate the dynamics of networks with a low proportion of active neurons, offering a more in-depth understanding of the network’s underlying dynamics. Studying network dynamics shows that real memory capacity has the same scaling properties as in the previous section.
- Chapter 6 studies the role of BTSP in maintaining homogeneous attractors in CA3. We show that BTSP can sustain homogeneous spatial maps in CA3

by dynamically adjusting the frequency of plateau potential for fluctuations in coding levels due to external sensory input, such as rewards.

- Chapter 7 summarizes the work by concluding that BTSP is a strong candidate for episodic memory formation and maintaining robust spatial maps in CA3.

Chapter 2

Behavioral timescale synaptic plasticity

2.1 A brief review of behavioral timescale synaptic plasticity

In 2017, Bittner et al. [75] discovered another new form of synaptic plasticity termed behavioral timescale synaptic plasticity, or BTSP. Their findings provide new insights into how place fields in hippocampal CA1 neurons emerge rapidly during active behavior. Unlike STDP plasticity, BTSP acts within a behavioral timescale of seconds and can induce long-lasting changes after a single trial. This property makes BTSP a primer candidate mechanism underlying one-shot learning - a key feature of the formation of episodic memories.

In this seminal study by Bittner et al., mice were head-fixed, running on a linear treadmill in a virtually enriched environment with visual cues. During runs, water rewards were provided to the mice. At the end of each run, the mice were instantly teleported back to the beginning of the track, equivalent to a circular track. Moreover, the experimental setting allowed for the manipulation and measurement of electrical currents in CA1 neurons through *in vivo* whole-cell patch-clamp recording. Despite the technical challenges of *in vivo* intracellular recordings, they found from recorded data that previously silent neurons became place cells and place fields persisted to the end of the session (Fig 2.1A).

Trial-by-trial analysis of membrane potential (V_m) showed that during the formation of place fields there was, at some point along the environment, a significant ramp-like depolarization of V_m , resulting in a complex spike burst (lap 11 in Fig 2.1B).

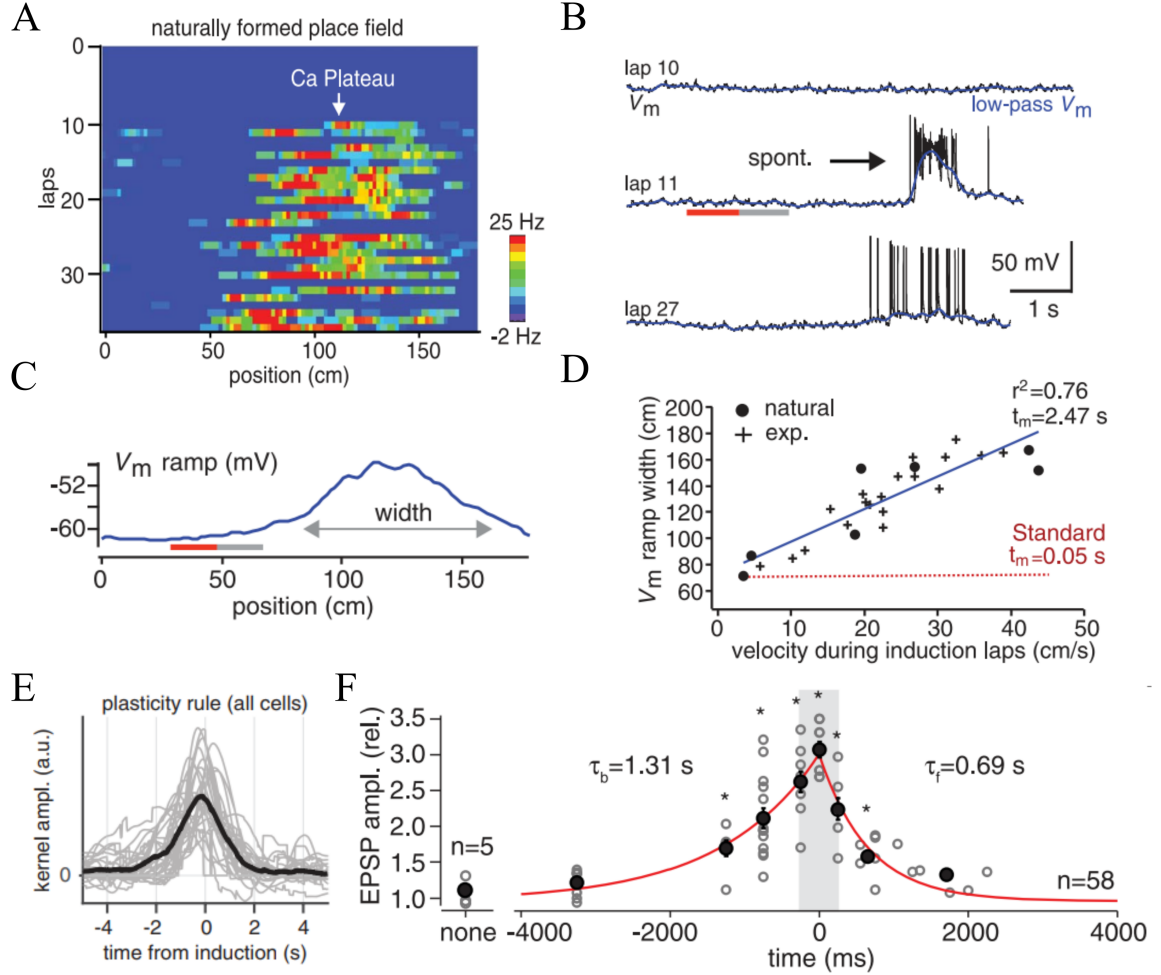


Figure 2.1: **Experimental finding in Bittner et al. 2017** (A) Spatial firing rates of a CA1 neuron across sequential laps. (B) Membrane potential (V_m , black) and its low-pass filtered counterpart (blue) shown for laps before (lap 10), during (lap 11), and after plateau potential induction (lap 27). (C) Averaged subthreshold V_m ramp, with the gray and red lines indicating the initial 20 cm and the preceding 20 cm of the ramp, respectively. (D) The relationship between the width of the V_m ramp and the average running velocity during induction trials, with data fitted using a linear equation (blue line). The red dashed line represents the linear relationship predicted by the standard synaptic plasticity rule. (E) The inferred synaptic plasticity rule across all CA1 place cells, with the black trace representing the mean. (F) Normalized post-induction EPSP amplitude plotted against the induction interval for the full population of neurons, showing a plasticity kernel similar to that observed in hippocampal slice preparations. **Note:** Figure adapted from Bittner et al. 2017 [75].

This significant depolarization was due to the initiation of dendritic calcium plateau potentials (PPs) in the distal dendrites of silent CA1 pyramidal neurons [76], causing a widespread calcium influx and prolonged somatic depolarization. Both *in vitro* [77] and *in vivo* [78] studies demonstrated that inputs from the entorhinal cortex (EC3) - one of the major inputs to CA1 - contribute to the generation of PPs. Optogenetic silencing of EC3 inputs [78, 79] strongly suppressed plateau probability highlighting the distinct contribution of EC3 to the generation of dendritic PPs. Importantly, these PPs are global and instructive signals, similar to error signals in supervised learning [80], and allow the potentiation of excitatory inputs arriving around the plateau [78]. They can further be modulated by novelty, surprise, and reward [75, 81].

Post-induction, neurons exhibited increased activity before the induction location due to plateau-induced potentiation (lap 27 in Fig 2.1B). Examination of the sub-threshold V_m revealed a slow ramping that extended back to locations where no action potential firing was observed during the induction trials (Fig 2.1C). This potentiation did not follow the standard Hebbian rule. If Hebbian learning were responsible, the ramp width would not vary substantially because, with a short timescale, CA1 neurons always integrate a similar amount of spatially tuned input from CA3 neurons, resulting in a comparable ramp. However, the data indicated a positive correlation between the ramp width and animals' speed with a slope of the order of seconds (Fig 2.1D). This slope ruled out the possibility of Hebbian learning and confirmed the existence of a plasticity mechanism operating on a timescale of seconds.

To quantify synaptic plasticity, they hypothesized that CA1 neurons received inputs from a population of CA3 uniformly distributed along the track, each with a narrow Gaussian place field. The variation in subthreshold membrane potential (ΔV_m) due to the plateau was modeled as a linear combination of convolved presynaptic firings modulated by plasticity-induced synaptic weight changes (ΔW). The plasticity ($\sim \Delta W$) inferred from the data exhibited an asymmetric shape with the peak occurring ahead of the plateau, within a time window of ± 4 seconds centered on the plateau onset (Fig 2.1E).

A biologically inspired model proposed by authors could also reproduce the same phenomenon. In this model, they suggested that the synaptic strength change was determined by overlapping local signals and a global signal triggered by plateau po-

tentials. These local signals, originating from glutamate release due to presynaptic firings, create so-called eligibility traces (see review in [82]), indicating that the synapses targeting the presynaptic neurons are ready to change. This asymmetric kernel inferred from data could also be observed using hippocampal slice preparation (Fig 2.1F) and was later confirmed by directly measuring synaptic transmission [83] and spine calcium dynamics [84] through an all-optical approach.

Hippocampal place cells can shift their place fields in response to stimuli, known as place cell remapping. Juxtacellular stimulation, similar to intracellular current injection, can induce place fields in previously silent neurons and trigger place field remapping [85]. Studies by Bittner et al. demonstrated that BTSP could induce place fields, but its role in remapping remained unclear. To address this, Milstein and colleagues [86] conducted a series of experiments to explore the role of BTSP in neurons that already exhibit place fields, whether formed naturally or experimentally. They discovered that inducing a plateau at different locations in behaving mice can shift place fields toward the new induction site. This shift is achieved by potentiating the synaptic weights of presynaptic inputs around the new induction site and depotentiating the efficacy of previously induced plateaus, making BTSP bidirectional.

Moreover, the direction and magnitude of synaptic weight change were dependent on the initial weight rather than the postsynaptic state: potentiation dominated in synapses with initially weak weights, whereas strong synapses underwent depression. This finding was tested using computational modeling, which could explain the data. The following section briefly provides an overview of this biophysical model.

This novel plasticity mechanism, characterized by an asymmetric kernel, can induce place field formation and trigger remapping. The resulting place fields, which encode future locations ahead of the animal, align with prior researches demonstrating a backward shift in CA1 place fields [87, 88]; therefore, it provides predictive power about future positions [75, 80]. It operates on a timescale of seconds induced by plateau potentials originating from the EC3. It requires synaptic activity from CA3 [83], but not the level of inhibition, which was found to be homogeneous in CA1 neurons [89], nor the postsynaptic activity. On a molecular level, Xiao et al. [90] demonstrated that while α -calcium-calmodulin-dependent protein kinase II (α CaMKII) is not involved in the generation of plateau potentials or eligibility traces,

it plays a crucial role in the expression mechanism of BTSP. Mice with a T286A point mutation in α CaMKII do not express BTSP after plateau induction compared to control groups.

Studies indicate that BTSP may contribute to the over-representation of CA1 neurons and the formation of context-dependent splitter cells [81,91]. It is crucial for forming and consolidating place cell activity at the network level [81,83,92]. Recent research also demonstrates that BTSP is present in CA3 with a symmetric kernel [93], which supports memory storage as stationary attractors [74]. These findings suggest that BTSP is a potent mechanism for shaping network-wide connectivity patterns in the hippocampus in one shot, likely playing a significant role in the formation of episodic memories.

2.2 Weight-dependent model of BTSP (Milstein et al., 2021)

The computational model proposed in [86] has successfully described the changes in membrane potential observed in CA1 cells after the induction of a plateau potential (PP) at a given location along a virtual track. The model keeps track of both a synaptic eligibility trace related to the activity of CA3 cells presynaptic to the CA1 cell of interest and an instructive signal associated with the occurrence of the PP, originating from EC3. The instructive dendritic signal is global, allowing for the possibility of all activated synapses to be updated simultaneously. The resulting plasticity at a given synapse depends on the convolution of these two signals, which are passed through a non-linearity and integrated over the lap. There is a different eligibility trace for potentiation and depression. This section will briefly overview this computational model. For more details, please refer to that manuscript.

They modeled behavioral time scale synaptic plasticity by considering a CA1 place cell that receives inputs from N excitatory CA3 place cells uniformly distributed on a circular track of length L . It was assumed that a virtual animal ran at a constant velocity v and, as it crossed a given location denoted by x_{PP} , a PP occurred either naturally or artificially induced through intracellular current injection.

For each CA3 cell i , they modeled the firing rate R_i using a Gaussian function:

$$R_i = R_i(x(t)) = R_{max} \cdot e^{-\frac{1}{2} \left(\frac{y_i - x(t)}{\sigma} \right)^2},$$

where y_i was the peak firing position, the animal's trajectory was $x(t)$ and the parameters $R_{max} = 1$ and $\sigma = 90/(3\sqrt{2})$. A postsynaptic dendritic PP during each lap k was defined by a binary function:

$$P(x(t)) = \begin{cases} 1 & \text{during a plateau} \\ 0 & \text{otherwise} \end{cases},$$

with a duration of 300 *ms*.

The presynaptic activity, R_i of CA3 cells and the PP activated two distinct biochemical signals: an eligibility trace and an instructive signal with exponential decay rate τ_{ET} and τ_{IS} , respectively, and the overlap of the signals drove distinct potentiating and depressing plasticity processes. They modeled these processes using sigmoidal gain functions:

$$\begin{aligned} q^+(ET_i \cdot IS) &= s(ET_i \cdot IS, \alpha^+, \beta^+) \\ q^-(ET_i \cdot IS) &= s(ET_i \cdot IS, \alpha^-, \beta^-) \\ s(x, \alpha, \beta) &= \frac{\hat{s}(x, \alpha, \beta) - \hat{s}(0, \alpha, \beta)}{\hat{s}(1, \alpha, \beta) - \hat{s}(0, \alpha, \beta)} \\ \hat{s}(x, \alpha, \beta) &= (1 + e^{-\beta(x-\alpha)})^{-1}, \end{aligned}$$

where ET_i is the eligibility signal activated by presynaptic neuron i and IS is the instructive signal propagating to all synapses.

The change in weight at each synapse depended on the current value of synaptic weight w_i and the plasticity processes, q^+ and q^- with corresponding learning constants k^+ and k^- :

$$\frac{dw_i}{dt} = (1 - w_i)k^+q^+(ET_i \cdot IS) - w_ik^-q^-(ET_i \cdot IS), \quad 0 \leq w_i \leq 1.$$

Although the plasticity rule is continuous in time, the total net change in synaptic weight Δw_i was computed once per lap, integrating the updating from initial time to end time of the track, t_0 and t_1 :

$$\Delta w_i = (1 - w_i)k^+\Delta Q^+ - w_ik^-\Delta Q^-, \quad 0 \leq w_i \leq 1, \quad (2.1)$$

where $\Delta Q^* = \int_{t_0}^{t_1} q^*(ET_i \cdot IS)dt$. Parameters of this model are τ_{ET} , τ_{IS} , α^+ , β^+ , α^- , β^- , k^+ , k^- .

Although this computational model successfully captures the essential features of BTSP and explains the translocation of place fields observed in vivo, its complexity

limits further analytical exploration. Cone and Shouval [94] proposed an extension of this model, representing the instructive and eligibility traces using differential equations. However, the resulting place fields can only be analytically calculated if the presynaptic place fields are assumed to be rectangular. Recent work has also demonstrated that a simplified BTSP rule using binary synapses can store and retrieve large numbers of binary inputs in a feedforward architecture reminiscent of CA1 [95].

This doctoral work aims to develop a mathematical framework for BTSP that replicates the results of the biophysical model in CA1 and can be extended to recurrent networks, similar to the CA3 structure, to study memory formation and recall. This framework seeks to overcome the limitations of existing models by providing a more flexible and analytically tractable approach, potentially offering new insights into the mechanisms underlying hippocampal function and memory processes.

Chapter 3

Mathematical modeling of BTSP: one-dimensional map

Previously silent CA1 pyramidal cells can suddenly become place cells after the occurrence of a plateau potential (PP). Experimental evidence suggests that the PP effectively “switches on” synapses from spatially tuned CA3 inputs, leading to a tuned subthreshold membrane potential in the CA1 cell [75, 83]. This phenomenon is illustrated in Fig 3.1. As the animal runs along a linear track, some CA1 cells receive little or no CA3 input due to the ineffectiveness of the synaptic connections, see Fig 3.1A. Therefore, the resulting membrane potential of the CA1 cell is initially spatially untuned, Fig 3.1B. When a PP occurs in the CA1 cell at a given location along the track (see PP symbol in Fig 3.1A), synapses from CA3 place cells that are active within a window of a few seconds around the PP become potentiated, leading to spatial tuning, Fig 3.1C, 3.1D.

The hippocampal place cell activities are rapidly shaped by a novel synaptic plasticity known as behavioral timescale synaptic plasticity, which was reviewed in the previous section. The computational model proposed by Milstein et al. [86] shows that synaptic weight change inversely depends on the initial weight: weaker initial weight facilitates potentiation while strong weights favor depression. Here, we show that the total plasticity occurring over a lap from this model can be described straightforwardly as a one-dimensional map. Namely, the synaptic weight from a presynaptic cell j to a postsynaptic cell i on a lap k can be written

$$w_{ij}^k = w_{ij}^{k-1} + \Delta w_{ij}^k, \quad (3.1)$$

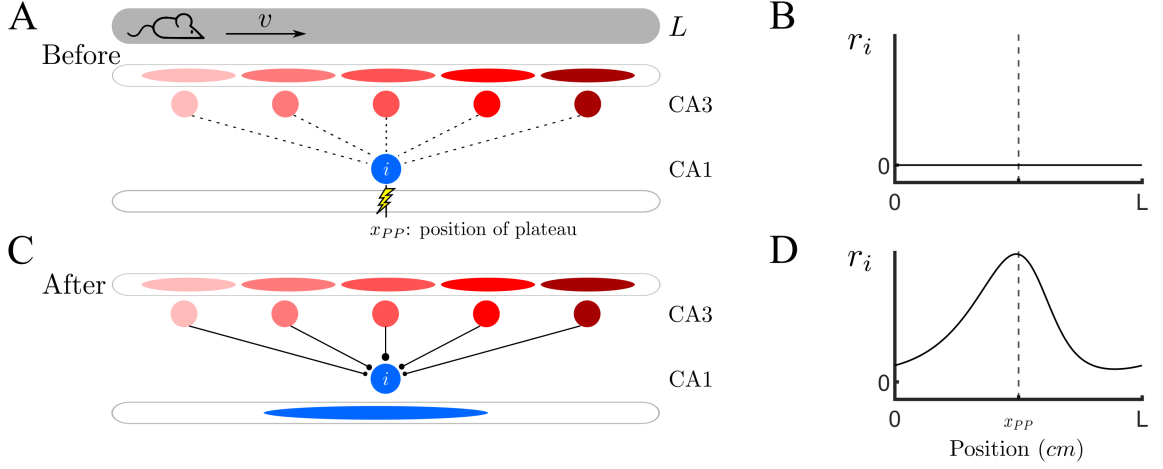


Figure 3.1: **Place field emergence after plateau potential in CA1** (A) A mouse runs at constant velocity v on a virtual linear track of length L . Before the trial, N CA3 place cells are weakly connected to the postsynaptic CA1 cell i , i.e., the synapses are ineffective. On a subsequent trial, a dendritic plateau potential (PP) occurs inside the CA1 cell when the animal reaches a position x_{PP} . (B) The firing rate of cell i before the PP is zero over the length of the track. (The dashed line indicates the position of the mouse when the PP occurs.) (C-D) After the occurrence of the PP, synapses from CA3 place cells are potentiated. As a result, the CA1 cell i develops spatial tuning.

where the change in the weight due to the occurrence of a PP on lap k does not explicitly depend on continuous time but only on the identity of the presynaptic and postsynaptic neurons.

3.1 Constructing spatially dependent plasticity functions

Previous studies [75, 86] have shown that the kernel of the BTSP has an asymmetric shape that spans for seconds around plateau potential onset. Modeling of the BTSP, described in the section 2.2, reveals that such plasticity can be decomposed into two different plasticity processes, potentiation and depression, resulting from the convolution of the synaptic eligibility trace and the instructive signal. Integrating them over the induction lap is essential in explaining the synaptic weight change. In the biophysical model, both (integrated) potentiation and depression are asymmetric and skewed with respect to the plateau onset. These processes are well fit by functions proportional to wrapped skew- t distributions.

To do so, we assume a sufficiently large number of discretized positions, which allows us first to define the plasticity rules in terms of continuous functions. We do this using the probability density function (PDF) of a skew- t distribution [96] with location μ , scale σ , skewness λ and ν degrees of freedom:

$$\bar{f}(t) = \frac{2}{\sigma} t_\nu \left(\frac{t - \mu}{\sigma} \right) T_{\nu+1} \left(\lambda \frac{t - \mu}{\sigma} \sqrt{\frac{\nu + 1}{\nu + \left(\frac{t - \mu}{\sigma} \right)^2}} \right), \quad (3.2)$$

where t_ν and $T_{\nu+1}$ denote the PDF and cumulative distribution function of the standardised t -distribution, and $t = 0$ is the onset time of plateau, see Fig 3.2 left. As the animal runs along the track and presynaptic place cells are activated in order, we can assign a time difference to each cell index j .

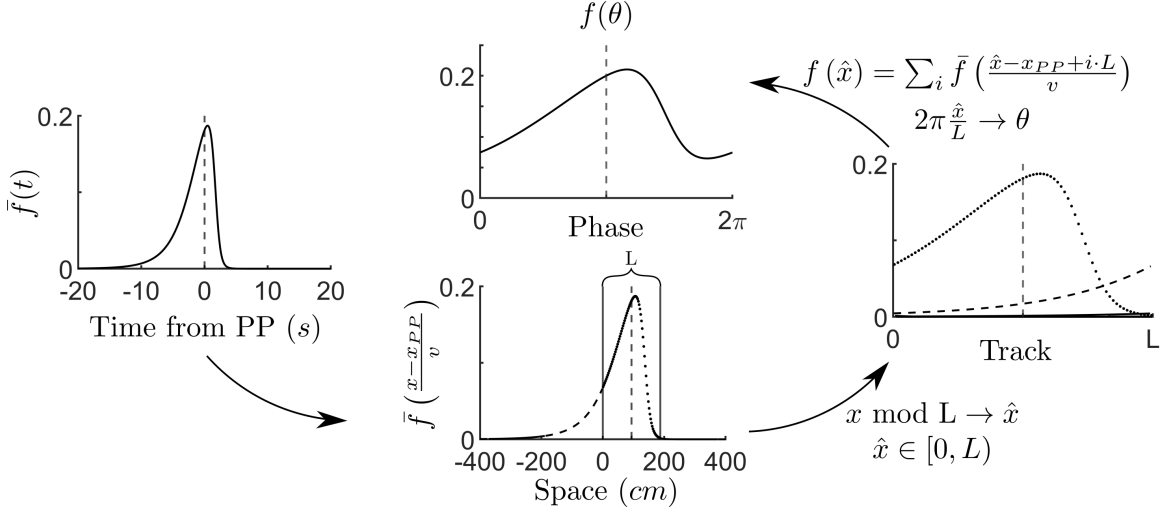


Figure 3.2: A temporal plasticity kernel from the biophysical model can be converted to a spatial kernel for a 1D map. Left: The biophysical model can be fit to a temporal plasticity kernel extracted from the experiment. Bottom: For an animal running at a constant velocity, the temporal kernel can be expressed as a spatial kernel with a linear transformation. Right: For running on a virtual track with teleportation, the spatial kernel must be “wrapped”. Top: A phase can be defined by normalizing the spatial kernel by the track length.

If animals run at a constant velocity, we can directly convert time to space $x = v \cdot t + x_{PP}$, where x_{PP} is the position of the PP. Solving this equation for t , we have $t = (x - x_{PP})/v$. Replacing t by $(x - x_{PP})/v$ in the Eq 3.2, the resulting plasticity will be defined in space, $\bar{f}_X(x) = \bar{f}((x - x_{pp})/v)$, see Fig 3.2 bottom.

One confound when converting from time to space is that while time extends along the whole real axis, space is restricted to lying between 0 and the total track length

L (in the virtual reality set-up, the animal is teleported back to the beginning of the track, which is equivalent to a circular track). In practical terms, this means that the plasticity of the synapse from a given presynaptic cell may have several contributions at different points in time. Specifically, portions of the spatial plasticity rule outside the track region (below 0 and greater than L) must be wrapped around to fit within the domain. In mathematical terms, we wrap $\bar{f}_X(x)$ on a circle with length L :

$$f(\hat{x}) = \sum_{i \in \mathbb{Z}} \bar{f} \left(\frac{\hat{x} - x_{pp} + i \cdot L}{v} \right), \quad \hat{x} \in [0, L), \quad (3.3)$$

see Fig 3.2 bottom and right.

Finally, we convert \hat{x} to $\theta = 2\pi\hat{x}/L$ to define a phase. The resulting spatial plasticity function $f(\theta)$ is served to define potentiation $f_P(\theta)$ and depression $f_D(\theta)$, see Fig 3.2 top. Importantly, these functions depend on the onset of PP. In a more general situation, let $\bar{\theta}$ the plateau onset phase, $f_P(\theta)$ and $f_D(\theta)$ can be represented as $f_P(\theta, \bar{\theta})$ and $f_D(\theta, \bar{\theta})$, respectively.

3.2 CA3-CA1 feedforward networks

Once we have defined spatial dependent plasticity functions, we can replace potentiation ΔQ^+ and depression ΔQ^- in Eq 2.1 by f_P and f_D respectively. We define $w(\theta)$ as the strength of connection from presynaptic CA3 cell that peaks at phase $\theta \in [0, \pi)$ to CA1 cell. Therefore, we can rewrite the Eq 2.1 as

$$\Delta w^k(\theta, \bar{\theta}) = (P \cdot (1 - w^{k-1}(\theta)) \cdot f_P(\theta, \bar{\theta}) - D \cdot w^{k-1}(\theta) \cdot f_D(\theta, \bar{\theta})) \cdot \mathbb{I}(\bar{\theta}), \quad (3.4)$$

where

1. $\bar{\theta}$ is plateau onset phase,
2. $f_P(\theta, \bar{\theta})$ and $f_D(\theta, \bar{\theta})$ are spatial dependent plasticity for potentiation and depression with corresponding learning constants P and D respectively,
3. and $\mathbb{I}(\bar{\theta})$ takes a value of 1 if there is plateau potential during the lap and 0 otherwise.

Then, the synaptic weights update for each lap k , $w^k(\theta)$, according to the 1D map:

$$w^k(\theta) = w^{k-1}(\theta) + \Delta w^k(\theta, \bar{\theta}), \quad w(\theta) \in [0, 1]. \quad (3.5)$$

And parameters of the 1D map are $P, \mu^P, \sigma^P, \lambda^P, \nu^P, D, \mu^D, \sigma^D, \lambda^D, \nu^D$.

3.3 Simulations confirm that 1D map reproduces results of full model

To evaluate the suitability of the proposed 1D map as a mathematical model for the full model presented in [86], we conducted simulations with a virtual mouse trained to run at a constant velocity of $v = 25 \text{ cm/s}$ along a circular track with a length of $L = 187 \text{ cm}$. We simulate the synaptic efficacy of $N = 100$ excitatory CA3 cells using two different models and varying induction protocols, subsequently comparing the results. The source code for this study is implemented in MATLAB and is available at [74], which includes a re-implementation of the biophysical model [97].

To achieve this, we first fit the spatial-dependent plasticity functions, f_P and f_D , to the biophysical model with parameters: $\tau_{ET} = 1664.1 \text{ ms}$, $\tau_{IS} = 737 \text{ ms}$, $\alpha^+ = 0.415$, $\beta^+ = 3.609$, $\alpha^- = 0.026$, $\beta^- = 13.815$, $k^+ = 0.9$, and $k^- = 0.275$. By manually adjusting the skew-t distributions, we determine the parameter set $\{P = 2.365, \mu^P = 0.685, \sigma^P = 1.65, \lambda^P = -1.61, \nu^P = 3.5, D = 2.57, \mu^D = 1.75, \sigma^D = 3.65, \lambda^D = -5.35, \text{ and } \nu^D = 5\}$ that quantitatively reproduces the kernels, see Fig 3.3A. As a result, the synaptic weight change derived from the map, plotted as a function of the presynaptic cells' position and the initial synaptic weight value, closely matches that of the biophysical mode, Fig 3.3B.

Second, we employ these plasticity functions, designed for a single PP, to determine the synaptic weights after a series of inductions. We tested different induction protocols.

1. Two single inductions: We induce two single PPs at different positions during two consecutive trials. The first PP is at 30 cm from the origin of the track in a trial, and the second is at 90 in another trial cm . $x_{PP}^1 = 30$, $x_{PP}^2 = 90$.
2. Repeating single inductions: A single PP is triggered in the middle of the track and lasts five consecutive trials. $x_{PP}^i = 93.5 \text{ cm}$, for $i \in \{1, 2, 3, 4, 5\}$.
3. Multiple inductions at once: Two inductions are simultaneously driven at two distinct locations in the same trial, followed by a third induction in the second trial. $x_{PP}^1 = 30, 120$ and $x_{PP}^2 = 93.5$.

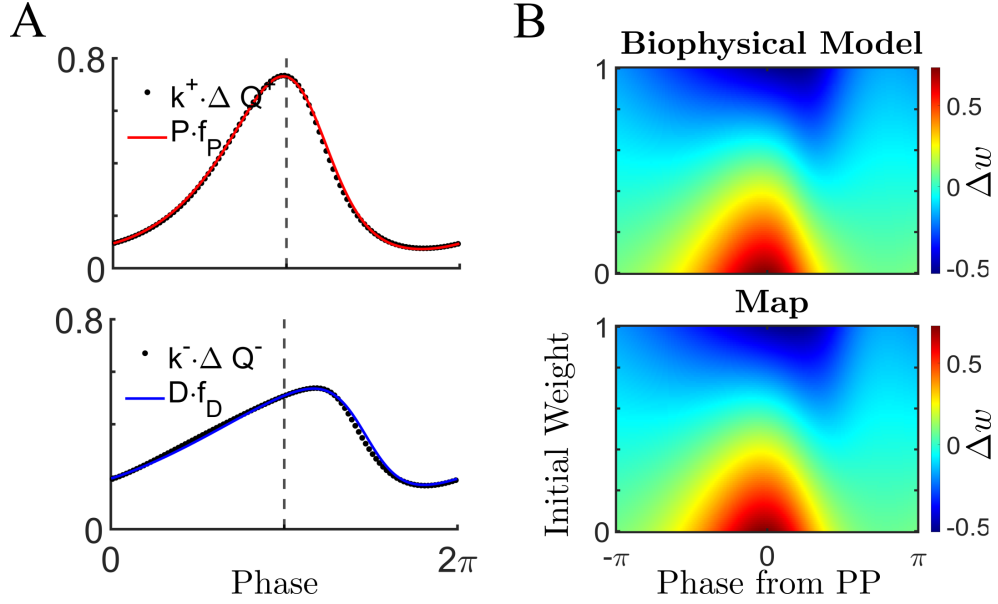


Figure 3.3: **Fitting of spatial kernels captures the degree of plasticity.** (A) We define f_P (red line) and f_D (blue line) as the normalized spatial plasticity functions. Here, they are compared to results from simulations of the biophysical model, which are also expressed as a function of phase along the track. (B) The degree of plasticity inferred from the experiment depends not only on the plasticity functions but also on the previous synaptic strength, normalized here to lie between 0 and 1.

We initialized the synaptic weights to zero and conducted simulations using these protocols. The numerical results demonstrate that after a single traversal with a plateau, the synaptic weights in the 1D map perfectly aligned with those in the biophysical model, see Fig 3.4A and 3.4B, top. Furthermore, this alignment remains accurate after a second (or fifth) trial at a different (or same) position, see Fig 3.4A and 3.4B, bottom. Interestingly, although the plasticity functions are fitted to a single PP during a trial, the results suggest that the superposition of two PPs within the same trial also worked, see Fig 3.4C. These results indicate that a direct connection between the biophysical model and the 1D map can be fully established once the underlying plasticity kernels are determined.

So far, numerical simulations have shown no difference between the two models. However, one question remains: Can the 1D map with wrapped skew-t distributions still fit the biophysical model when the animal's speed varies? Experimentally, it has been observed that running speed modulates the width of the emergent place field, which is a hallmark of BTSP [75]. More precisely, the BTSP time window

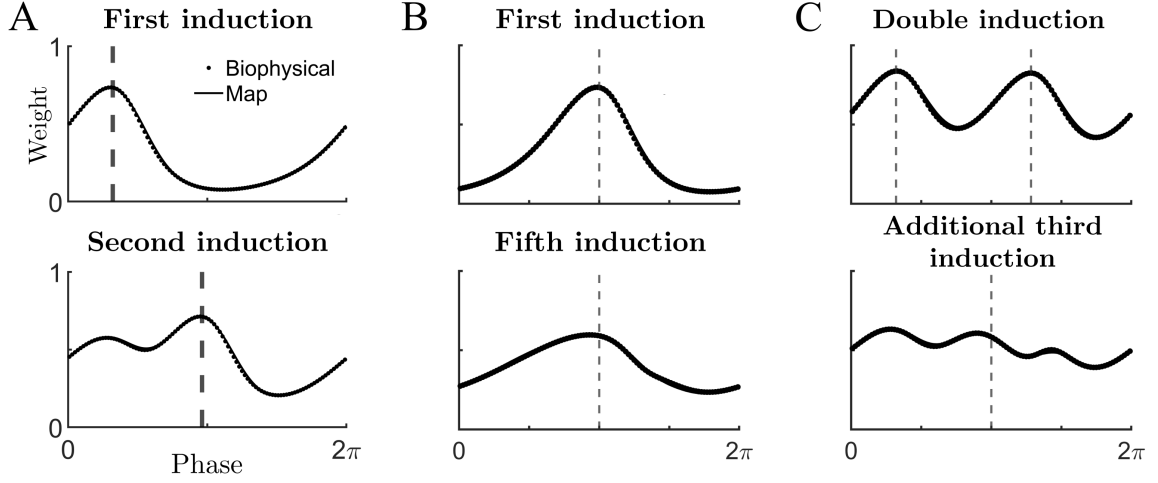


Figure 3.4: **Consistency of models across various induction protocols.** (A) Synaptic weights after a single induction (top) and after two inductions at a different position on consecutive trials (bottom). (B) After five inductions at the same position. (C) The results of two inductions at distinct locations on the same trial (top), and the subsequent change in weights after a third induction on the second trial (bottom).

spans several seconds of the plateau phase. If the animal’s speed is relatively slow, the synaptic inputs from different neurons are relatively few, resulting in a narrower plasticity kernel. Conversely, when the animal’s speed is faster, the plasticity kernel becomes wider because more synaptic inputs from neurons are received within the same time frame. In the biophysical model, varying the animal’s speeds while keeping the rest of the model unchanged results in changes in the shape of the plasticity kernels. For each speed, we could fit the 1D map with different sets of parameters (see the figure caption), as shown in Fig 3.5.

Putting all the previous results together, this analysis shows that the 1D map can reproduce all the relevant features of the biophysical model. However, one aspect we did not focus on was identifying the homotopy of plasticity kernels when the animal’s speed varies - in other words, determining the transformation of skew-t distributions that fit different velocities. Here, we propose a mathematical model of BTSP by constructing the plasticity functions at a constant speed. Nonetheless, if the animal’s running speed is not constant, the mapping of the plasticity rule from time to space still exists but will be nonlinear.

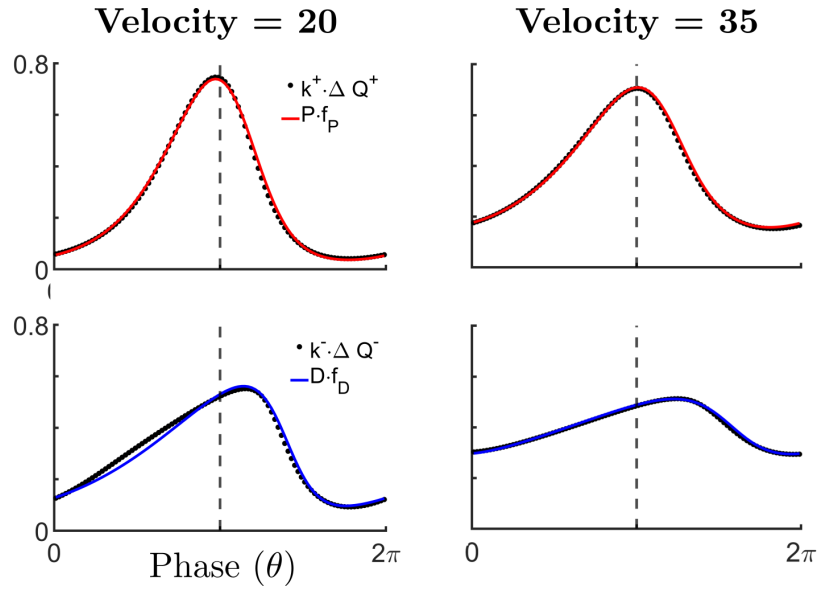


Figure 3.5: **The animal's speed modulates the plasticity kernel.** Potentiation (top) and depression (bottom) plasticity functions for different velocities. Parameters: **Velocity = 20**: $P = 2.625$, $\mu^P = 0.75$, $\sigma^P = 1.825$, $\lambda^P = -1.55$, $\nu^P = 5$, $D = 2.8$, $\mu^D = 1.875$, $\sigma^D = 3.45$, $\lambda^D = -4.55$, and $\nu^D = 5$; **velocity = 35**: $P = 2$, $\mu^P = 0.6$, $\sigma^P = 1.42$, $\lambda^P = -1.625$, $\nu^P = 2$, $D = 2.15$, $\mu^D = 1.54$, $\sigma^D = 3.85$, $\lambda^D = -5.75$, and $\nu^D = 3$.

Chapter 4

BTSP and memory storage in recurrent networks

Now that we have shown that the 1D map is suitable to describe BTSP in CA1, we want to extend it to recurrent networks of place cells to study how the BTSP shapes the recurrent structures for the storage and recall of memories. Specifically, we are interested in how the plasticity rule forms stable internal representations of different spatial environments. Doing so requires two distinct steps. First, we must determine how plasticity shapes the matrix of recurrent connections. We will do this through direct analysis of the 1D map. In doing so, we will calculate the average correlation of the synaptic weight matrix with any given environment and the degree of quenched variability. This will allow us to perform a signal-to-noise ratio calculation and determine how the memory capacity scales qualitatively with network size, coding sparseness, and the learning rates P and D [68, 72]. Secondly, and importantly, we need to study how the connectivity shapes the activity in a network model and determine the true memory capacity of the network in terms of stable attractor states [69, 98], which will be studied in the next section.

4.1 1D map for BTSP in recurrent networks

We first consider the most straightforward scenario in which all cells in the network have place fields in a given environment. We assume that when the animal first explores a novel environment, place cells in CA3 are either already present or quickly form due to plasticity in afferent inputs. Thus, we can arrange the cells along a linear track according to their place field locations, even before any plasticity occurs in

the recurrent connectivity, as shown in Fig 4.1A (before). As the animal runs along the track, we assume that, over several traversals, plateau potentials (PPs) occur in all cells, with their occurrence timing coinciding with the maximal firing rate of the post-synaptic cell. This way, strong potentiation of recurrent excitatory inputs is expected between cells with adjacent place fields. In contrast, cells with distant place fields may experience no potentiation or even depression, depending on the specifics of the plasticity functions and the previous state of the synapse. Consequently, the recurrent connectivity becomes correlated with the ordering of the cells' place fields in the novel environment, as shown in Fig 4.1A (after).

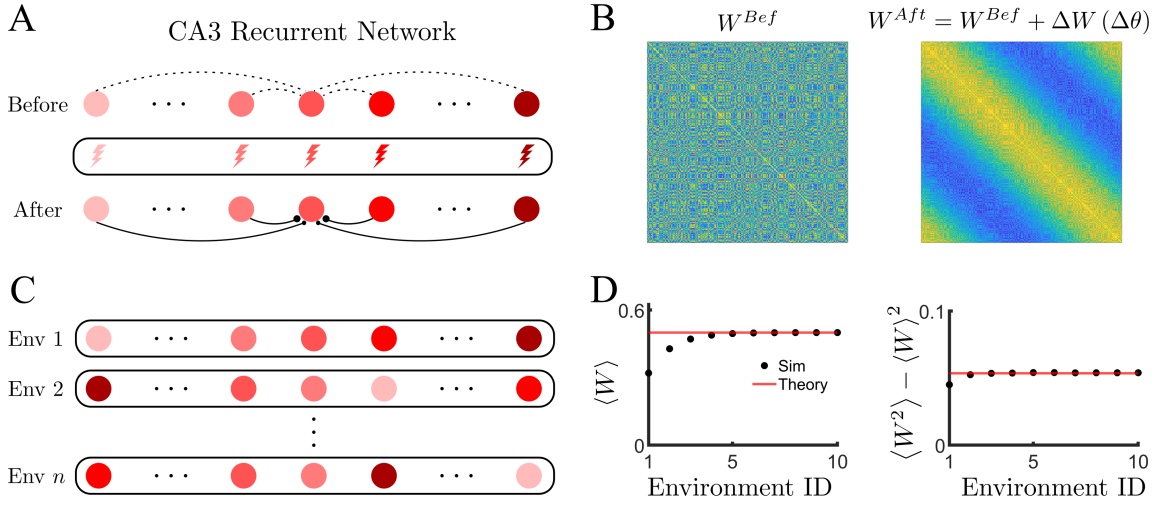


Figure 4.1: **Schematic of BTSP-based learning in a recurrent network.** (A) Before the learning: the recurrent connectivity in a population of N CA3 place cells is initially uncorrelated with the novel environment. During the exploration of the novel track: a dendritic plateau potential occurs inside each CA3 cell, driving plasticity. After: the synaptic weights are modulated according to the phase difference. (B) Simulations show that ordering cells w.r.t. the novel environment before (left) and after (right) the BTSP-based learning reveals the emergence of correlation. (C) Schematic of plasticity across n distinct environments. Global remapping for each novel environment, representing in a random permutation of place field locations. (D) Global statistics of the connectivity matrix, showing the average (left) and variance (right) as a function of the number of explored environments

We extend the previous 1D map to recurrent networks, such as those in area CA3 of the hippocampus. In this scenario, we assume that BTSP exclusively shapes the strength of the recurrent excitatory connections between place cells. Therefore, we consider that cells are already place cells due to spatially-tuned feedforward inputs.

The synaptic weight change due to BTSP of a connection from cell j with a place field centered at phase θ_j to cell i with a place field centered at phase θ_i in environment k can be expressed as $\Delta w_{ij}(\Delta\theta_{ij}^k)$. The resulting 1D map for the recurrent networks is

$$w_{ij}^k = w_{ij}^{k-1} + \Delta w_{ij}(\Delta\theta_{ij}^k) \quad (4.1)$$

$$= w_{ij}^{k-1} + P(1 - w_{ij}^{k-1})f_P(\Delta\theta_{ij}^k) - Dw_{ij}^{k-1}f_D(\Delta\theta_{ij}^k), \quad (4.2)$$

where $\Delta\theta_{ij}^k = \theta_i^k - \theta_j^k$ is the phase difference in the place field positions of cells i and j in environment k . Plasticity functions are considered to be symmetric in storing memories as stationary attractors. An asymmetric rule would lead to the formation of dynamic attractors rather than stable spatial maps [99, 100]. Therefore, any symmetric functions on the phase $[-\pi, \pi]$ can be used to define f_P and f_D , ensuring that the resulting plasticity peaks at $\Delta\theta_{ij}^k = 0$ and decreases as $|\Delta\theta_{ij}^k|$ increases. Recent experimental data in CA3 show that the plasticity kernel is indeed symmetrical, confirming our hypothesis [93].

To understand how this learning rule shapes recurrent connectivity, we specify the plasticity functions and run simulations. In practice, $f_P(\theta) = 1 + \cos \theta$ and $f_D(\theta) = 1 - \cos \theta$ are used. We model the plasticity due to the exploration of n distinct linear tracks, Fig 4.1C. The ordering of the place cells uniquely determines each environment. In the simulation, a random permutation of cells' order represents a new environment. Running simulations using this rule (with $N = 256$, $n = 50$, and $P = D = 0.3$), we observe that the hypothesis in Fig 4.1A holds true; after a single traversal, spatial correlations to the novel environments emerge (Fig 4.1B right) in the recurrent connections of place cells that were initially uncorrelated with the novel environments (Fig 4.1B left). In Fig 4.1B, the weight matrix is sorted w.r.t. the novel environment before and after the learning. As synaptic weights are updated for each new environment explored, the statistics of the weight matrix eventually reach a steady state after a certain number of environments, see Fig 4.1D, depending on the learning constants. The steady-state statistics can be calculated directly from the plasticity rule (Fig 4.1D red lines), e.g. the mean $\mu = \langle w_{ij} \rangle$ and variance, $\sigma^2 = \langle w_{ij}^2 \rangle - \mu^2$, where the brackets indicate averages over the phases of the network; see later section.

As the learning process becomes stationary, we investigate how the memory traces of previously explored environments are stored in the synaptic weight matrix. After

exploring a sufficient number of environments, the synaptic weight matrix reaches a steady state. To analyze this, we sort the matrix according to the previous environments. Since the ordering of cells represents each environment, we can rearrange the matrix ($\{w_{ij}\}$) in both indices and then align each row of the matrix to the null phase difference, $\Delta\theta = 0$. Fig 4.2 shows the synaptic weight after this sorting for four previously explored environments, with $\Delta\theta^n$ being the most recently explored one.

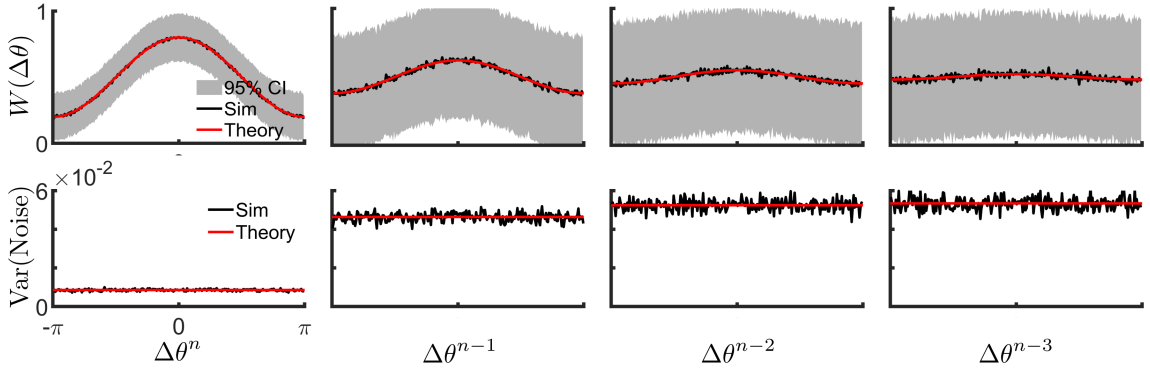


Figure 4.2: **Spatial modulation of synaptic weights decreases with age.** Spatially ordered statistics of weight matrix (sorted and centered w.r.t previously explored environments) in the steady state. Top: Mean (black line) and 95% confidence intervals (grey lines) of simulated weight as a function of phase difference at each environment, as well as the theoretical prediction (red line). Bottom: Variance of the ordered weights: simulation (black) and theoretical curve (red)

Simulations show a rapid decay in spatial modulation as memory age increases. Here, we define the memory trace as the mean spatial connectivity in each environment, represented by the black lines in the top panel. The amplitude of this trace indicates the spatial modulation. Although the rule is deterministic, the process is stochastic due to the random shuffling of cells for each environment. This stochasticity results in variability of the synaptic weight that increases as the memory ages. The shaded region in the top panel of Fig 4.2 represents the 95% confidence interval of simulated weights, which is quantified in the bottom panel, shown by the black lines. Interestingly, these spatial statistics depend on the choice of P and D , see Fig 4.3. For a balanced network, when $P = D$, the variability around the mean remains constant. Conversely, a non-homogeneous shape emerges when $P \neq D$.

In the remaining section, we analyze this simple 1D map in the recurrent network to compute the steady-state and spatial statistics using general functions defined on

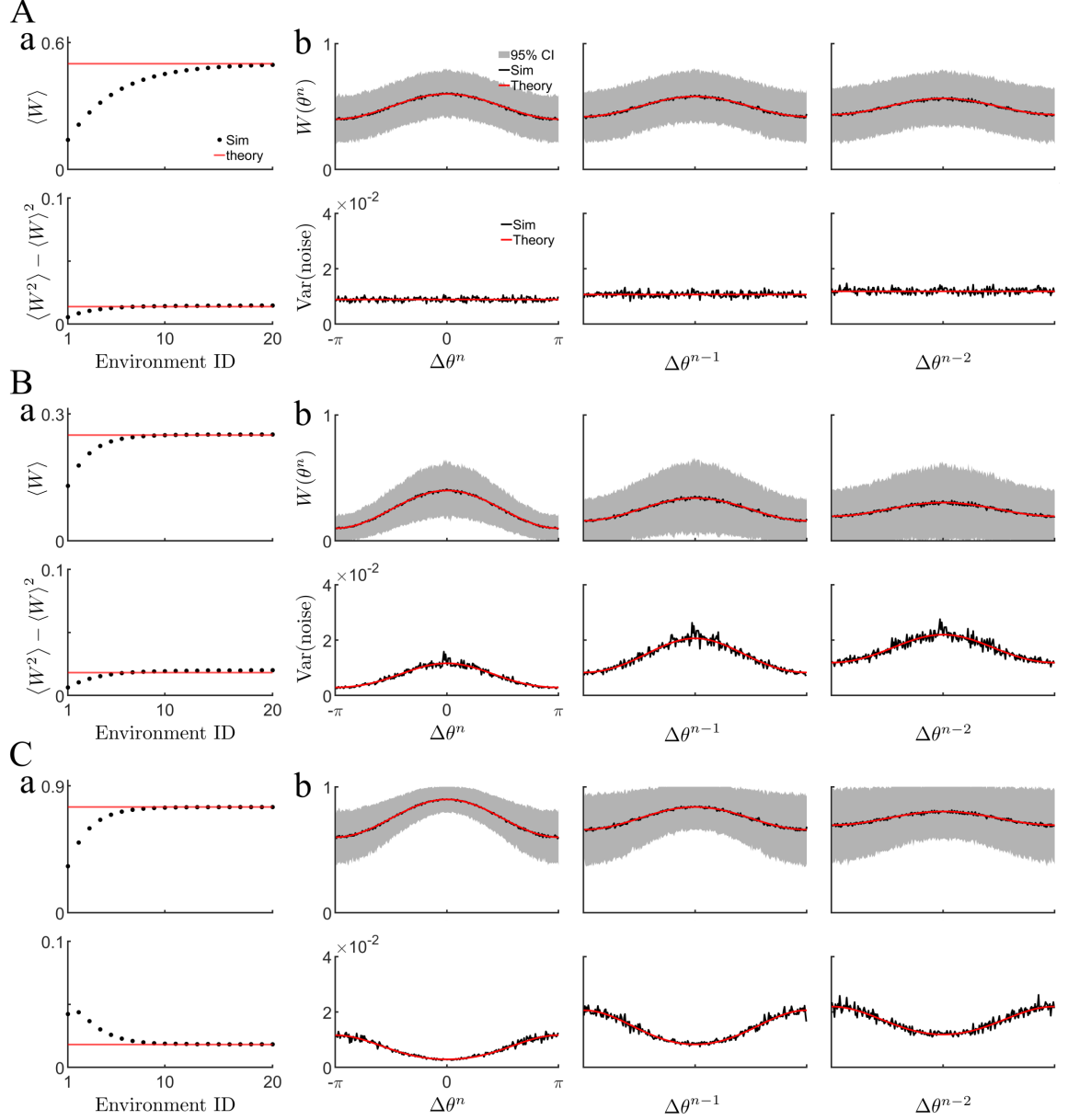


Figure 4.3: **The value of P and D can change the shape of spatial statistics.** (A-C) a. Global statistics, average (top) and variance (bottom), as a function of a number of explored environments. b. Spatial statistics of weight matrix at steady state: mean curve with 95% confidence interval (top) and variance of noise (bottom) as a function of phase difference. **Parameters:** (A) $P = D = 0.1$; (B) $P = 0.1$ and $D = 0.3$; (C) $P = 0.3$ and $D = 0.1$.

a circle. Additionally, we derive the closed-form expressions for the specific choices of f_P and f_D .

4.1.1 Steady-state statistics: Mean and variance

After plasticity in a sufficiently large number of environments, the synaptic weight matrix will reach a statistically steady state. This steady state is reached more quickly with larger learning rates P and D , while convergence takes longer for small learning rates. This effect can be observed by comparing Fig 4.1D ($P = D = 0.3$) and panel a of Fig 4.3A ($P = D = 0.1$).

Assuming that steady-state has been reached, the mean weight across the matrix upon an exploration of environment k can be written $\langle w_{ij}^k \rangle = \langle w_{ij}^{k-1} \rangle = \mu_w$, where the brackets indicate an average over the entire matrix without any particular ordering. Applying this average to Eq 4.2,

$$\begin{aligned}
\langle w_{ij}^k \rangle &= \langle w_{ij}^{k-1} + P(1 - w_{ij}^{k-1})f_P(\Delta\theta_{ij}^k) - Dw_{ij}^{k-1}f_D(\Delta\theta_{ij}^k) \rangle \\
&= \langle w_{ij}^{k-1} \rangle + \langle P(1 - w_{ij}^{k-1})f_P(\Delta\theta_{ij}^k) \rangle - \langle Dw_{ij}^{k-1}f_D(\Delta\theta_{ij}^k) \rangle \\
&= \langle w_{ij}^{k-1} \rangle + P\langle f_P(\Delta\theta_{ij}^k) \rangle - P\langle w_{ij}^{k-1} \rangle \langle f_P(\Delta\theta_{ij}^k) \rangle - D\langle w_{ij}^{k-1} \rangle \langle f_D(\Delta\theta_{ij}^k) \rangle.
\end{aligned} \tag{4.3}$$

In the learning procedure, the w_{ij}^{k-1} is shaped by environments up to $k - 1$, which indicates no correlation between it and the environment k , on f_P and f_D rely. Consequently, we can compute the average of the product as the product of the averages in the third equality in Eq 4.3. Replacing the $\langle w_{ij}^k \rangle$ and $\langle w_{ij}^{k-1} \rangle$ by μ_w and simplifying the notation for f_P and f_D by ignoring the variable, the previous equation yield to

$$\mu_w = \mu_w + P\langle f_P \rangle - P\mu_w\langle f_P \rangle - D\mu_w\langle f_D \rangle,$$

which leads to

$$\mu_w = \frac{P\langle f_P \rangle}{P\langle f_P \rangle + D\langle f_D \rangle}, \tag{4.4}$$

consistent with previous modeling studies [86, 94]. Given that the plasticity functions f_P and f_D are defined as periodic in space, in practice, the averages can be taken by ordering the neurons and integrating

$$\langle f_\alpha \rangle = \langle f_\alpha(\Delta\theta_{ij}^k) \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_*(\Delta\theta_{ij}^k) d(\Delta\theta_{ij}^k), \quad \alpha \in \{P, D\}.$$

Using the definition of the variance of a random variable X : $\text{Var}(X) = \text{E}[X^2] - \text{E}[X]^2$, we can compute the steady-state variance of the weight matrix $\sigma_w^2 = \langle w^2 \rangle - \langle w \rangle^2$. For simplicity, we have removed the super- and subscripts from the weights here. Using Eq 4.2 we have

$$\begin{aligned}\langle w^2 \rangle &= \langle (w + P(1 - w)f_P - Dw f_D)^2 \rangle \\ &= \langle (Pf_P + wF)^2 \rangle \\ &= \langle P^2 f_P^2 + 2wP f_P F + w^2 F^2 \rangle \\ &= P^2 \langle f_P^2 \rangle + 2P \langle w \rangle \langle f_P F \rangle + \langle w^2 \rangle \langle F^2 \rangle,\end{aligned}$$

where $F = 1 - Pf_P - Df_D$. Solving for $\langle w^2 \rangle$ we find that

$$\sigma_w^2 = \frac{P^2 \langle f_P^2 \rangle + 2P \mu_w \langle f_P F \rangle}{1 - \langle F^2 \rangle} - \mu_w^2. \quad (4.5)$$

The red solid lines in Fig 4.1D and panel a of Fig 4.3 are calculated using Eqs 4.4 and 4.5 respectively, where the plasticity functions were chosen.

The remaining section gives a closed-form formula for the steady-state statistics for $f_P(\theta) = 1 + \cos \theta$ and $f_D(\theta) = 1 - \cos \theta$. Before that, we will first compute all the terms:

$$\begin{aligned}\langle f_P(\theta) \rangle &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (1 + \cos \theta) d\theta = 1, \\ \langle f_D(\theta) \rangle &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (1 - \cos \theta) d\theta = 1, \\ \langle f_P^2(\theta) \rangle &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (1 + \cos \theta)^2 d\theta \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (1 + 2 \cos \theta + \cos^2 \theta) d\theta = \frac{3}{2}, \\ \langle f_P(\theta) f_D(\theta) \rangle &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (1 + \cos \theta)(1 - \cos \theta) d\theta \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (1 - \cos^2 \theta) d\theta = \frac{1}{2}.\end{aligned}$$

Using previous results, we have

$$\langle f_P F \rangle = 1 - \frac{3}{2}P - \frac{1}{2}D.$$

With these relations, the mean and variance of the weight matrix are

$$\mu_w = \frac{P}{P + D}, \quad (4.6)$$

$$\sigma_w^2 = \frac{2P^2 D^2}{(P + D)^2 \cdot (2(PD + P + D) - 3/2(P + D)^2)}, \quad (4.7)$$

4.1.2 Spatial statistics: mean and variance of the memory trace

Once a statistical steady state has been reached, we can calculate the memory trace of past environments. When examining the weight matrix after environment n and looking back at environment $n - \eta$, we find that the statistical properties of the memory trace depend solely on the memory age, η , rather than the absolute order $n - \eta$. This holds as long as the matrix is already in a steady state after environment $n - \eta - 1$.

To quantify the memory trace in environment $n - \eta$, we use Eq 4.2 and iteratively expand it with previous environments. Specifically, we write, for the weights after plasticity in environment n

$$w_{ij}^n = Pf_P(\Delta\theta_{ij}^n) + w_{ij}^{n-1}F(\Delta\theta_{ij}^n), \quad (4.8)$$

where we now indicate the ordering of the phases explicitly in each environment. This equation is valid for any n , and hence we can expand w_{ij}^{n-1} to express it in terms of w_{ij}^{n-2} , and so on. Here, we show how to expand it to the environment $n - 3$. First, we replace the w_{ij}^{n-1} using the definition of rule in environment $n - 1$:

$$\begin{aligned} w_{ij}^n &= Pf_P(\Delta\theta_{ij}^n) + w_{ij}^{n-1}F(\Delta\theta_{ij}^n) \\ &= Pf_P(\Delta\theta_{ij}^n) + (Pf_P(\Delta\theta_{ij}^{n-1}) + w_{ij}^{n-2}F(\Delta\theta_{ij}^{n-1}))F(\Delta\theta_{ij}^n) \\ &= Pf_P(\Delta\theta_{ij}^n) + Pf_P(\Delta\theta_{ij}^{n-1})F(\Delta\theta_{ij}^n) + w_{ij}^{n-2}F(\Delta\theta_{ij}^{n-1})F(\Delta\theta_{ij}^n). \end{aligned}$$

Same for w_{ij}^{n-2} , we get

$$\begin{aligned} w_{ij}^n &= Pf_P(\Delta\theta_{ij}^n) + Pf_P(\Delta\theta_{ij}^{n-1})F(\Delta\theta_{ij}^n) \\ &\quad + (Pf_P(\Delta\theta_{ij}^{n-2}) + w_{ij}^{n-3}F(\Delta\theta_{ij}^{n-2}))F(\Delta\theta_{ij}^{n-1})F(\Delta\theta_{ij}^n) \\ &= Pf_P(\Delta\theta_{ij}^n) + Pf_P(\Delta\theta_{ij}^{n-1})F(\Delta\theta_{ij}^n) + Pf_P(\Delta\theta_{ij}^{n-2})F(\Delta\theta_{ij}^{n-1})F(\Delta\theta_{ij}^n) \\ &\quad + w_{ij}^{n-3}F(\Delta\theta_{ij}^{n-2})F(\Delta\theta_{ij}^{n-1})F(\Delta\theta_{ij}^n) \\ &= Pf_P(\Delta\theta_{ij}^n) + Pf_P(\Delta\theta_{ij}^{n-1})F(\Delta\theta_{ij}^n) + Pf_P(\Delta\theta_{ij}^{n-2})F(\Delta\theta_{ij}^{n-1})F(\Delta\theta_{ij}^n) \\ &\quad + (Pf_P(\Delta\theta_{ij}^{n-3}) + w_{ij}^{n-4}F(\Delta\theta_{ij}^{n-3}))F(\Delta\theta_{ij}^{n-2})F(\Delta\theta_{ij}^{n-1})F(\Delta\theta_{ij}^n). \end{aligned}$$

In a compact way:

$$w_{ij}^n = \left(Pf_P(\Delta\theta_{ij}^{n-3}) + w_{ij}^{n-3-1}F(\Delta\theta_{ij}^{n-3}) \right) \prod_{k=0}^2 F(\Delta\theta_{ij}^{n-k}) + P \sum_{l=0}^2 f_P(\Delta\theta_{ij}^{n-l}) \prod_{k=0}^{l-1} F(\Delta\theta_{ij}^{n-k}).$$

Finally, we extend the analysis to the environment $n - \eta$:

$$w_{ij}^n = \left(P f_P(\Delta\theta_{ij}^{n-\eta}) + w_{ij}^{n-\eta-1} F(\Delta\theta_{ij}^{n-\eta}) \right) \prod_{k=0}^{\eta-1} F(\Delta\theta_{ij}^{n-k}) + P \sum_{l=0}^{\eta-1} f_P(\Delta\theta_{ij}^{n-l}) \prod_{k=0}^{l-1} F(\Delta\theta_{ij}^{n-k}), \quad (4.9)$$

where the first term completely contains the memory trace from environment $n - \eta$. Within the parentheses, we observe a potentiation term combined with a mixed potentiation and depression term, which is premultiplied by the synaptic weight before plasticity. This trace is then degraded in a multiplicative way by all subsequent learning, from environment $n - (\eta - 1)$ to environment n . The second term represents an additive noise due to the interference between more recently learned environments, again spanning from $n - (\eta - 1)$ to n .

We calculate the strength of the memory trace by extracting the first Fourier mode of the spatial modulation in the relevant environment. Here, we assume plasticity functions f_P and f_D are even, which means that it is sufficient to consider the cosine Fourier component alone. Specifically, if we order the neurons according to their place field location in environment $n - \eta$, we can approximate the connectivity between neurons through

$$M_\eta = \mu_w + a_\eta \cdot \cos(\Delta\theta^{n-\eta}), \quad (4.10)$$

where a_η is the amplitude of the memory trace. In the case of pure cosine plasticity functions, this formula is exact, while for other shapes, it can be extended to higher-order Fourier modes. Note that because the matrix reached a steady state, this connectivity profile depends only on the age of the memory η . We determine the amplitude a_η through integration

$$\begin{aligned} a_\eta &= \mathbb{R} \left(\langle e^{i\Delta\theta_{ij}^{n-\eta}}, w_{ij}^n \rangle \right) \\ &= 2 \langle \cos(\Delta\theta_{ij}^{n-\eta}), w_{ij}^n \rangle, \end{aligned} \quad (4.11)$$

where \mathbb{R} indicates the real part. Plugging in the formula for the weight given by Eq 4.9 and using the fact that the phases in different environments are uncorrelated, i.e., $\forall l \neq n - \eta$

$$\langle e^{i\Delta\theta_{ij}^{n-\eta}}, f_P(\Delta\theta_{ij}^l) \rangle = 0,$$

we find that the amplitude is

$$a_\eta = 2 \left(P \langle \cos(\theta), f_P(\theta) \rangle + \mu_w \langle \cos(\theta), F(\theta) \rangle \right) \langle F \rangle^\eta.$$

The amplitude a_η is a mean across the network. That is, if we consider the connectivity profile for a given neuron i after learning and then average over all i , we obtain a_η in the limit of many neurons. However, for any given i , there will be deviations from this mean in the form of quenched variability. This variability will impact the ability of the network to retrieve a given memory. We can calculate this variability by subtracting the mean connectivity from the weight matrix:

$$V_\eta = \text{Var}(w_{ij}^n - M_\eta) = \langle (w_{ij}^n - a_\eta \cos(\Delta\theta_{ij}^{n-\eta}))^2 \rangle - \mu_w^2. \quad (4.12)$$

A general expression for V_η for arbitrary f_P and f_D could be derived but would be lengthy. On the other hand, below, we provide the explicit formula for a simple choice of f_P and f_D , which we use in this study. A more detailed derivation can be found in the Appendix A.

For $f_P(\theta) = 1 + \cos \theta$ and $f_D(\theta) = 1 - \cos \theta$, we find

$$\begin{aligned} \langle \cos(\theta), f_P(\theta) \rangle &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos \theta (1 + \cos \theta) d\theta = \frac{1}{2}, \\ \langle \cos(\theta), F(\theta) \rangle &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \cos \theta (1 - P(1 + \cos \theta) - D(1 - \cos \theta)) d\theta \\ &= \frac{D - P}{2}, \\ \langle F^2(\theta) \rangle &= \frac{1}{2\pi} \int_{-\pi}^{\pi} (1 - P(1 + \cos \theta) - D(1 - \cos \theta))^2 d\theta \\ &= 1 + \frac{3P^2 + 3D^2 + 2PD - 4P - 4D}{2}. \end{aligned} \quad (4.13)$$

Therefore, the mean and variance of the memory trace, derived from the Appendix A considering $s = 1$, are

$$a_\eta = \frac{2PD}{P + D} (1 - P - D)^\eta, \quad (4.14)$$

$$V_\eta = A_\eta + B_\eta \cos(\Delta\theta^{n-\eta}) + C_\eta \cos^2(\Delta\theta^{n-\eta}) \quad (4.15)$$

The coefficients of the variance curve are

$$\begin{aligned}
A_\eta &= A_0 \langle F^2 \rangle^\eta + \mu^2 (\langle F^2 \rangle^\eta - 1) + \frac{3}{2} P^2 \frac{1 - \langle F^2 \rangle^\eta}{1 - \langle F^2 \rangle} \\
&\quad + 2P^2 \left(1 - \frac{3}{2}P - \frac{1}{2}D\right) \frac{1}{\langle F \rangle - \langle F^2 \rangle} \left(\frac{1 - \langle F \rangle^\eta}{1 - \langle F \rangle} - \frac{1 - \langle F^2 \rangle^\eta}{1 - \langle F^2 \rangle} \right) \\
&\quad + 2\mu P \left(1 - \frac{3}{2}P - \frac{1}{2}D\right) \frac{\langle F \rangle^\eta - \langle F^2 \rangle^\eta}{\langle F \rangle - \langle F^2 \rangle}, \\
B_\eta &= B_0 \langle F^2 \rangle^\eta + 2a_0 \mu (\langle F^2 \rangle^\eta - \langle F \rangle^{2\eta}) \\
&\quad + 2a_0 P \left(\left(1 - \frac{3}{2}P - \frac{1}{2}D\right) \frac{\langle F \rangle^\eta - \langle F^2 \rangle^\eta}{\langle F \rangle - \langle F^2 \rangle} - \frac{\langle F \rangle^\eta - \langle F \rangle^{2\eta}}{1 - \langle F \rangle} \right), \\
C_\eta &= C_0 \langle F^2 \rangle^\eta + a_0^2 (\langle F^2 \rangle^\eta - \langle F \rangle^{2\eta}).
\end{aligned}$$

where

$$\begin{aligned}
A_0 &= P^2 + 2P(1 - P - D)\mu + (1 - P - D)^2 \langle w^2 \rangle - \mu^2, \\
B_0 &= 2 \frac{P - D}{P + D} \left(P^2 + P(1 - 2P - 2D)\mu - (P + D)(1 - P - D) \langle w^2 \rangle \right), \\
C_0 &= \left(\frac{P - D}{P + D} \right)^2 \left(P^2 - 2P(P + D)\mu + (P + D)^2 \langle w^2 \rangle \right).
\end{aligned} \tag{4.16}$$

4.2 Sparse coding with BTSP in recurrent networks

In the previous section, we examined a network where every neuron has a place field in every environment, and each cell encoded a distinct phase in any given environment. This configuration resulted in low storage capacity due to maximal interference between plasticity events from different environments and a low signal-to-noise ratio (SNR) for any given location. Interference between memories can be reduced by considering sparse coding, where only a fraction s of neurons are place cells in any given environment. The SNR at a given location can be increased by considering a population of M cells for each location. Specifically, we consider a network of NM cells, where N is the number of uniformly distributed spatial phases, and M is the number of cells available to encode each phase. We had $M = 1$ and $s = 1$ in the previous section.

We model BTSP in the recurrent connections as before, with the difference that, given a sparseness s , there are now sM neurons encoding each spatial phase, see Fig 4.4A. It is important to note that neurons with overlapping place fields belonging to the same group of sM neurons in one environment generally do not have overlapping fields in another one. Specifically, the individual cells undergo global remapping, not the cell populations. We now consider the sparseness in the 1D map by keeping track of which neurons are active in any given environment and, hence, which synapses get updated. The rule is now

$$w_{ij}^n = w_{ij}^{n-1} + \left(P \cdot (1 - w_{ij}^{n-1}) f_P(\Delta\theta_{ij}^n) - D \cdot w_{ij}^{n-1} \cdot f_D(\Delta\theta_{ij}^n) \right) S_i^n S_j^n, \quad (4.17)$$

where 1) i and j go from 1 to MN and 2) S_i^n takes a value of 1 or 0, indicating whether the neuron is active in environment n . We assume the activation of neurons is a Bernoulli process with probability s , i.e., $P(S_\alpha^n = 1) = s$, $\alpha \in \{i, j\}$.

4.2.1 Sparse coding enhances the storage capacity

The structure of the connectivity matrix after plasticity varies significantly with different values of M and s used in the simulations. As M increases, a block structure emerges in the matrix, as illustrated in Fig 4.4B upper row, which enhances the SNR. Specifically, suppose each post-synaptic neuron receives the average synaptic weight from a pre-synaptic population (block). In that case, the amplitude of the memory trace remains constant regardless of the population size (Fig 4.4C red left), while the variance decreases inversely with the population size (Fig 4.4C red right). Consequently, the SNR, defined as the amplitude of the memory trace at age η divided by the square root of the constant component of the variance:

$$\text{SNR}_\eta = \frac{a_\eta}{\sqrt{\langle V_\eta \rangle}}$$

increases with larger M for a given s . This increase in SNR directly translates to an enhanced storage capacity. The capacity of a network is the maximum memory age at which the SNR remains above a specified threshold T ; for late analysis, we set $T = 1$:

$$\eta_{max} = \max_\eta \{ \text{SNR}_\eta \geq T \}. \quad (4.18)$$

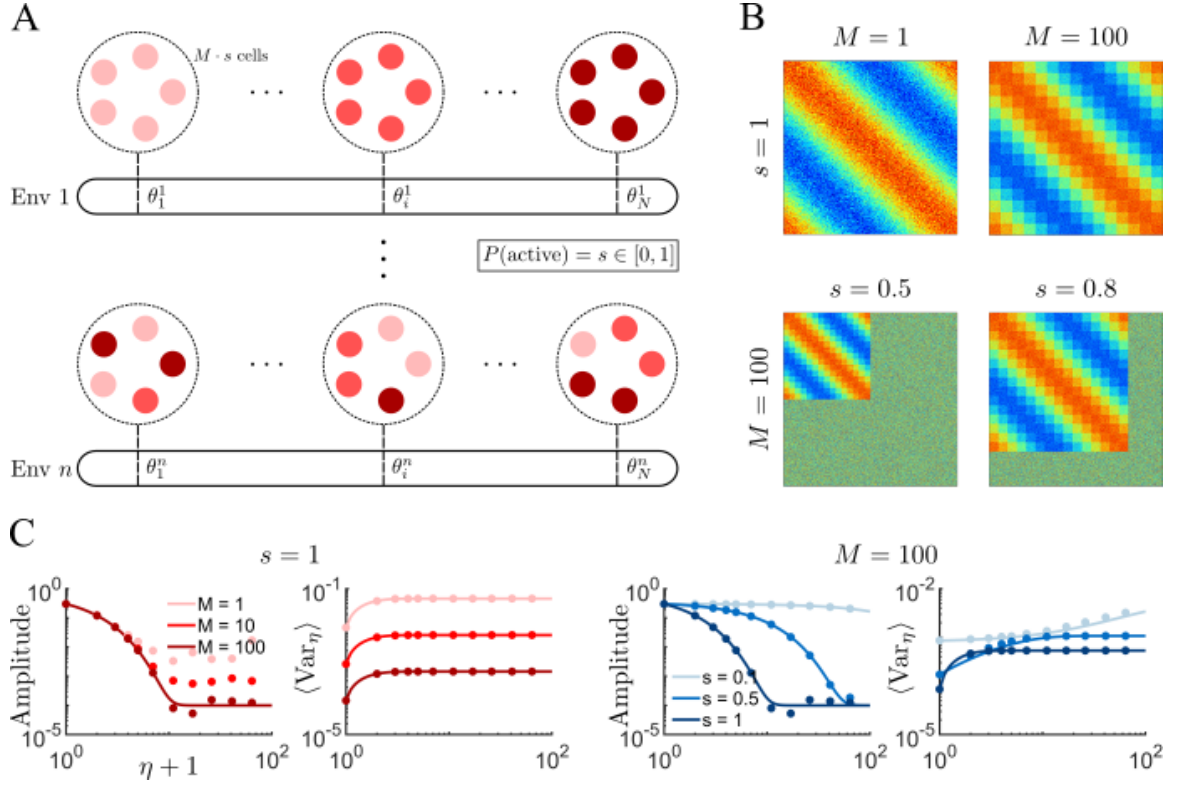


Figure 4.4: **BTSP-based learning in a sparse recurrent network with network size M .** (A) Schematic representation of the recurrent network with a population size M and sparseness s . In any given environment, each cell has a probability s of being a place cell. A population of sM cells with place fields centered at position θ_i . For each environment, cell positions undergo global remapping (reshuffling). (B) The weight matrix is shown for different values of M and s . (C) Amplitude of the mean memory traces a_η (Fig 4.2 top) and of the variance V_η (Fig 4.2 bottom) as a function of number of past explored environments in log-log scale. To avoid the memory age 0, the x-axis is represented as memory age + 1. Also, a lower bound is applied to y-axes to restrict the range.

This indicates that higher values of M allow for longer-lasting memory traces, thereby increasing the effective storage capacity of the network.

When $s < 1$, a portion of the synaptic weights remain unchanged during plasticity in a novel environment and, therefore, uncorrelated with the environment, which is depicted in Fig 4.4B bottom row. The reduction of memory interference leads to a more gradual decay in the amplitude, specifically for lower values of s , as shown in Fig 4.4C blue left. While the impact on variance is nontrivial, the theoretical framework well accounts for it, as illustrated in Fig 4.4C blue right. The slower decay in amplitude for sparse networks will enhance the SNR over extended periods, compared to networks with the same population size but higher s . Consequently, the memory capacity markedly increases in the sparse-network limit.

4.2.2 Detailed calculation of 1D map

This section provides a comprehensive analysis of the 1D map for sparse networks for the choice of

$$f_P(\theta) = 1 + \cos \theta, \quad f_D(\theta) = 1 - \cos \theta,$$

encompassing the analysis from the previous section as a special case. In this context, the mean and variance of the weight matrix, μ_w and σ_w^2 , remain the same, see Eqs 4.6-4.7. However, the mean memory trace and its variance will differ slightly in their calculations.

To do so, we first rewrite the weight equation, Eq 4.17, as a function of the environment $n - \eta$. Following the same derivation in the Section 4.1.2, we get:

$$\begin{aligned} w_{ij}^n = & \left(P f_P(\theta_{ij}^{n-\eta}) S_i^{n-\eta} S_j^{n-\eta} + w_{ij}^{n-\eta-1} F(\theta_{ij}^{n-\eta}) \right) \prod_{k=0}^{\eta-1} F(\Delta \theta_{ij}^{n-k}) \\ & + P \sum_{l=0}^{\eta-1} f_P(\theta_{ij}^{n-k}) S_i^{n-k} S_j^{n-k} \prod_{k=0}^{l-1} F(\Delta \theta_{ij}^{n-k}) \end{aligned} \quad (4.19)$$

where $F(\Delta \theta_{ij}^k) = 1 - S_i^k S_j^k (P f_P(\Delta \theta_{ij}^k) + D f_D(\Delta \theta_{ij}^k))$. When we calculate the mean and variance of the memory trace in environment $n - \eta$, we now only consider the subset of active cells, and hence take $S_i^{n-\eta} = 1$, while for all other environments k , S_i^k

is treated as a Bernoulli random variable. Therefore, the amplitude can be calculated as

$$\begin{aligned}
a_\eta &= 2\langle \cos(\Delta\theta_{ij}^{n-\eta}), w_{ij}^n \rangle|_{\{S_i^{n-\eta}=1, S_j^{n-\eta}=1\}} \\
&= \left(P \frac{1}{\pi} \int_{-\pi}^{\pi} f_P(\theta_{ij}^{n-\eta}) \cos(\Delta\theta_{ij}^{n-\eta}) d(\Delta\theta_{ij}^{n-\eta}) \right. \\
&\quad \left. + \mu_w \int_{-\pi}^{\pi} (1 - P f_P(\Delta\theta_{ij}^{n-\eta}) - D f_D(\Delta\theta_{ij}^{n-\eta})) \cos(\Delta\theta_{ij}^{n-\eta}) d(\Delta\theta_{ij}^{n-\eta}) \right) \langle F \rangle^\eta.
\end{aligned}$$

Similar for the variance

$$V_\eta = \langle (w_{ij}^n - a_\eta \cos(\Delta\theta_{ij}^{n-\eta}))^2 \rangle|_{\{S_i^{n-\eta}=1, S_j^{n-\eta}=1\}} - \mu_w^2.$$

See Appendix A for a detailed calculation of the variance.

Since the probability of activation of cells is uncorrelated, we have

$$\begin{aligned}
\langle F \rangle &= \langle F(\Delta\theta_{ij}^k) \rangle \\
&= \langle 1 - S_i^k S_j^k (P f_P(\Delta\theta_{ij}^k) + D f_D(\Delta\theta_{ij}^k)) \rangle \\
&= 1 - s^2(P + D). \\
\langle F^2 \rangle &= \langle (1 - S_i^k S_j^k (P f_P(\Delta\theta_{ij}^k) + D f_D(\Delta\theta_{ij}^k)))^2 \rangle \\
&= 1 + s^2(3P^2 + 3D^2 + 2PD - 4P - 4D)/2,
\end{aligned}$$

where $\langle (S_i^k)^2, (S_j^k)^2 \rangle = \langle S_i^k, S_j^k \rangle = \langle S_i^k \rangle \langle S_j^k \rangle = s^2$. Putting all the calculations together, including previous results (Eq 4.13), the amplitude is

$$a_\eta = \frac{2PD}{P+D}(1 - s^2(P+D))^\eta. \quad (4.20)$$

And the variance is

$$\begin{aligned}
V_\eta &= A_\eta + B_\eta \cos(\Delta\theta^{n-\eta}) + C_\eta \cos^2(\Delta\theta^{n-\eta}) \\
&= (A_\eta + \frac{C_\eta}{2}) + B_\eta \cos(\Delta\theta^{n-\eta}) + \frac{C_\eta}{2} \cos(2\Delta\theta^{n-\eta})
\end{aligned} \quad (4.21)$$

where

$$\begin{aligned}
A_\eta &= A_0 \langle F^2 \rangle^\eta + \mu^2 (\langle F^2 \rangle^\eta - 1) + \frac{3}{2} P^2 s^2 \frac{1 - \langle F^2 \rangle^\eta}{1 - \langle F^2 \rangle} \\
&\quad + 2P^2 s^4 \left(1 - \frac{3}{2}P - \frac{1}{2}D\right) \frac{1}{\langle F \rangle - \langle F^2 \rangle} \left(\frac{1 - \langle F \rangle^\eta}{1 - \langle F \rangle} - \frac{1 - \langle F^2 \rangle^\eta}{1 - \langle F^2 \rangle} \right) \\
&\quad + 2\mu P s^2 \left(1 - \frac{3}{2}P - \frac{1}{2}D\right) \frac{\langle F \rangle^\eta - \langle F^2 \rangle^\eta}{\langle F \rangle - \langle F^2 \rangle}, \\
B_\eta &= B_0 \langle F^2 \rangle^\eta + 2a_0 \mu (\langle F^2 \rangle^\eta - \langle F \rangle^{2\eta}) \\
&\quad + 2a_0 P s^2 \left(\left(1 - \frac{3}{2}P - \frac{1}{2}D\right) \frac{\langle F \rangle^\eta - \langle F^2 \rangle^\eta}{\langle F \rangle - \langle F^2 \rangle} - \frac{\langle F \rangle^\eta - \langle F \rangle^{2\eta}}{1 - \langle F \rangle} \right), \\
C_\eta &= C_0 \langle F^2 \rangle^\eta + a_0^2 (\langle F^2 \rangle^\eta - \langle F \rangle^{2\eta}).
\end{aligned}$$

and A_0 , B_0 , and C_0 remain the same, see Eq 4.16.

As shown in Fig 4.4, the theoretical predictions (solid lines) for the mean and variance of memory traces align closely with the simulated data (dots) across different values of M and s , thus providing compelling validation of the theory. For panel C of Fig 4.4, we apply a lower bound to prevent the amplitude goes to 0 if the memory age gets larger.

4.2.3 Memory capacity for balanced potentiation and depression

Once the mean and variance of memory traces are determined, the signal-to-noise ratio (SNR) provides an approximation of the storage capacity in a purely structural context. Figure 4.5A illustrates how SNR varies with memory age for different values of s and M , maintaining consistency between simulation and theory. We assume that a memory can be retrievable if the SNR associated with that memory exceeds a threshold value T . Here, we set $T = 1$, indicated by the dashed line in Fig 4.5A. The maximum age at which SNR remains above this threshold defines the network's storage capacity:

$$\eta_{max} = \max_{\eta} \{ \text{SNR}_{\eta} \geq 1 \} = \max_{\eta} \left\{ \frac{a_{\eta}}{\sqrt{\langle V_{\eta} \rangle}} \geq 1 \right\}.$$

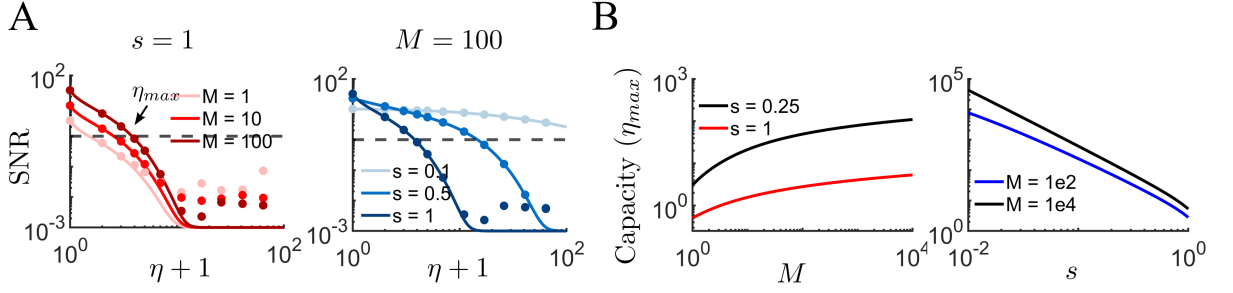


Figure 4.5: **Memory capacity improves with sparse coding and increased population size.** (A) Signal-to-noise ratio using different value of M and s for $P = D = 0.3$. Dashed line indicates $\text{SNR} = 1$. (B) The memory capacity of the system, η_{\max} as a function of population size (left) and sparsity (right).

Deriving a closed-form expression for the network's storage capacity is generally challenging due to the complexity of the variance, as described in Eq 4.21. However, when potentiation and depression are balanced ($P = D$), $\langle F^2 \rangle$ can be approximated by $\langle F \rangle^2$ in the sparse limit, or they are the same when $s = 1$. Figure 4.6 shows the absolute difference between $\langle F \rangle^2$ and $\langle F^2 \rangle$ across different values of P as s varies. In this case, Eqs 4.20 and 4.21 can be simplified to

$$\begin{aligned}
 a_\eta &= P(1 - 2s^2P)^\eta \\
 A_\eta &= \frac{P}{8(1-P)} - \frac{P^2}{2}(1 - 4s^2P + 4s^2P^2)^\eta \quad \text{and} \\
 C_\eta &= P^2(1 - 4s^2P + 4s^2P^2)^\eta - P^2(1 - 2s^2P)^{2\eta},
 \end{aligned}$$

where we only consider components of the variance that are constant in space.

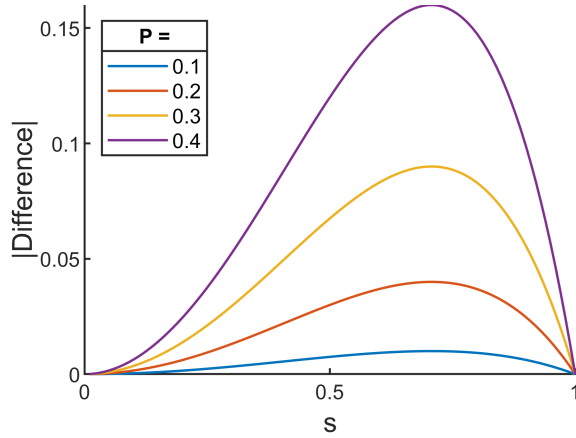


Figure 4.6: **Absolute difference between $\langle F \rangle^2$ and $\langle F^2 \rangle$ when $P = D$.**

Therefore, the SNR for the environment $n - \eta$ is

$$\text{SNR}_\eta = \frac{a_\eta}{\sqrt{\langle V_\eta \rangle}} = \frac{a_\eta}{\sqrt{(A_\eta + \frac{1}{2}C_\eta)/sM}} = \frac{\sqrt{sM}P(1 - 2s^2P)^\eta}{\sqrt{P/(8 - 8P) - P^2(1 - 2s^2P)^{2\eta}/2}}. \quad (4.22)$$

Setting the SNR to 1 and squaring both sides yields:

$$\frac{sMP^2(1 - 2s^2P)^{2\eta}}{P/(8 - 8P) - P^2(1 - 2s^2P)^{2\eta}/2} = 1.$$

Solving this equation by η , we get the memory capacity:

$$\eta_{max} = -\ln(8P(1 - P)(sM + 1/2))/(2\ln(1 - 2s^2P)). \quad (4.23)$$

Analyzing this equation, as illustrated in Fig 4.7A, We find that the range of potentiation values for which real solutions exist depends on the value of network size and sparseness. Increasing the network size extends the range of potentiation values for which real solutions exist and enhances the maximum storage capacity. As the potentiation value varies, the storage capacity exhibits a non-monotonic behavior. Furthermore, when transitioning from a dense network (Left panel: $s = 1$) to a sparse network (Right panel: $s = 0.1$), the storage capacity increases while the range of potentiation values decreases.

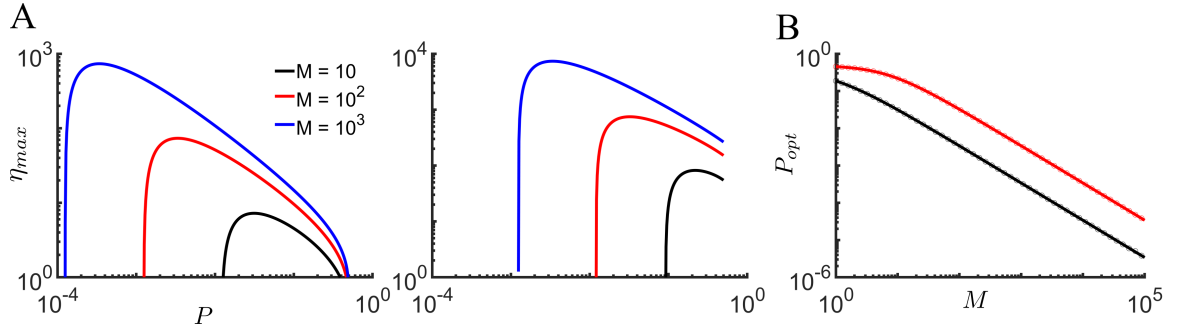


Figure 4.7: Storage capacity and optimal plasticity parameter in neural networks with different sizes and sparseness levels. (A) Storage capacity, η_{max} , as a function of network size, sparseness, and values of P . Lines represent different network sizes. The left panel illustrates the case for $s = 1$, while the right panel shows the scenario for $s = 0.1$. The x-axis denotes different values of P . (B) The optimal value of P that maximizes memory capacity for each M . Circles represent the optimal values found in (A), while lines are the numerical solutions of Eq 4.24. Red denotes $s = 0.1$ and black denotes $s = 1$.

Leveraging the equation for η_{max} , we can determine the optimal value of P that maximizes memory capacity, as shown in panel A. To do so, we differentiate the equation with respect to P and set the derivative to zero:

$$(1 - 2P)(1 - 2s^2P) \ln(1 - 2s^2P) + 2s^2P(1 - P) \ln\left(8P(1 - P)\left(sM + \frac{1}{2}\right)\right) = 0, \quad (4.24)$$

where we only take into account the numerator of the derivative. Fig 4.7B illustrates a good matching of the optimal value of P found using the graph or the above equation for different s . Let P_{opt} be the numerical solution to the above equation. Substituting P_{opt} into Eq 4.23, we obtain the optimal value of η_{max} , which depends exclusively on the network size and sparseness.

Finally, if we assume that $s^2 \ll 1$ and that $sM \gg 1/2$, which is the case if we wish to obtain a large capacity, then the formula further simplifies to

$$\eta_{max} = \frac{1}{4s^2P} \ln\left(sM \cdot 8P(1 - P)\right).$$

Furthermore, taking only the contribution of the system size sM in the logarithm leads to

$$\eta_{max} \sim \frac{1}{4s^2P} \ln(sM). \quad (4.25)$$

4.3 Discussion

The logarithmic scaling of system size is a well-established result for networks with bounded synapses [68,69]. Similarly, the substantial enhancement in capacity through sparse coding is also known [65,69,98,101]. In essence, if additional synaptic resources are available to boost the SNR, the optimal strategy would be to maintain sM while reducing s , effectively making the representation sparser. This implies that $s \sim 1/M$ and consequently, the memory capacity $\eta_{max} \sim M^2$. This scaling is characteristic of several memory systems with sparse coding [60]. Further strategies for enhancing capacity, such as incorporating multiple interacting discrete synaptic states, have been thoroughly investigated in previous studies [72,73,102,103].

It is important to note that the SNR scales with the population size at each position along the track, sM , rather than with the number of positions N . Our use of continuous spatial plasticity functions $f_P(\theta)$ and $f_D(\theta)$ implicitly assumes that N is sufficiently large for this approximation to be valid. This condition essentially

establishes a lower bound on N , governed by the sampling theorem. Beyond this constraint, N does not impact the SNR, as the SNR is determined by the degree of redundancy in the input from cells with similar or identical tuning, which is sM . However, when considering the recall of attractor states in neuronal networks, the number of encoded positions N becomes crucial.

Chapter 5

Neuronal networks endowed with BTSP can encode a large number of spatial memories as bump attractors

In the preceding chapter, we systematically analyzed the statistical properties of recurrent synaptic weights in networks with BTSP. We determined how the capacity for encoding multiple spatial environments scales with system size and sparseness. This analysis implicitly assumed the activation of a subset of cells in any given environment without explicitly modeling the firing rate dynamics.

However, true memory capacity is determined by the recovery of intrinsic dynamical states that correlate with neuronal activity within specific environments. To explore this, we simulate networks of firing rate neurons with connectivity matrices derived directly from the 1D map for BTSP from the preceding section. We specifically focus on the firing rate dynamics in a network where plasticity has already been established, thus no longer allowing for additional plasticity.

5.1 Network dynamics

The one-dimensional map for BTSP enables us to determine the strength of the memory trace corresponding to all past environments within the connectivity matrix itself. While a sufficiently strong SNR of the memory trace is necessary for memory recall, it is not sufficient. To investigate the dynamics of memory recall, we consider a

network of rate neurons with recurrent excitatory connectivity defined by the weight matrix resulting from the plasticity rule.

For simplicity, we model the effect of inhibitory interneurons by assuming a global inhibitory feedback proportional to the mean excitatory activity. This approximation is valid if the dynamics of inhibitory interneurons are significantly faster than those of excitatory neurons. Additionally, we account for the normalization of weights by the plasticity rule and rescale the weights using an overall maximum weight factor. Specifically, we define the rescaled synaptic weight from neuron j to neuron i as

$$\bar{w}_{ij} = W_0 + W_{max}(w_{ij} - \mu_w), \quad (5.1)$$

where w_{ij} is derived directly from the 1D map after the learning process, W_{max} is the maximum synaptic weight (with w_{ij} normalized to be between 0 and 1), and $\mu_w = \langle w_{ij} \rangle$. The mean offset μ_w and the parameter W_0 control the inhibitory effect.

The plasticity rule assumes the presence of place cells with place fields located at N uniformly distributed positions in any given environment. At each position, there are M cells, but only a fraction s of them are active. Consequently, the total number of neurons in the network is MN , while the number of active neurons in any given environment is sMN .

To study the network dynamics, we use the following firing rate equations:

$$\left\{ \begin{array}{l} \tau \frac{d}{dt} r_1 = -r_1 + \phi \left(\frac{1}{\kappa N} \sum_{j=1}^{MN} \bar{w}_{1j} r_j S_j^k + I_0 \right) S_1^k, \\ \tau \frac{d}{dt} r_2 = -r_2 + \phi \left(\frac{1}{\kappa N} \sum_{j=1}^{MN} \bar{w}_{2j} r_j S_j^k + I_0 \right) S_2^k, \\ \vdots \\ \tau \frac{d}{dt} r_i = -r_i + \phi \left(\frac{1}{\kappa N} \sum_{j=1}^{MN} \bar{w}_{ij} r_j S_j^k + I_0 \right) S_i^k, \\ \vdots \end{array} \right., \quad (5.2)$$

where

- r_i represents the firing rate of cell i ,
- ϕ is a nonlinear F-I curve or transfer function,

- I_0 is external input that is uniform across all neurons (this study only considers constant I_0),
- and $S_i^k = 1$ if cell i is active in environment k , otherwise $S_i^k = 0$.

The parameter κ determines the connectivity scaling concerning the neuronal population's size encoding each position. As we shall see in the later section, setting $\kappa = \sqrt{sM}$ allows us to compare directly the retrieval capacity (network dynamic) with the storage capacity (SNR calculation). With this scaling, the mean input increases with population size, while the variance remains on the order of one. Alternatively, setting $\kappa = sM$ keeps the mean constant, but the variance decreases with increasing population size.

For simulations, we define $\phi(x)$ as follows:

$$\phi(x) = \begin{cases} 0, & x < 0 \\ x^2, & x \in [0, 1] \\ 2\sqrt{x - 3/4}, & x > 1 \end{cases} \quad (5.3)$$

This function, illustrated in Fig 5.1, is continuous and smooth, effectively capturing the expansive nonlinearity at low rates typical for irregularly firing neurons in the fluctuation-driven regime.

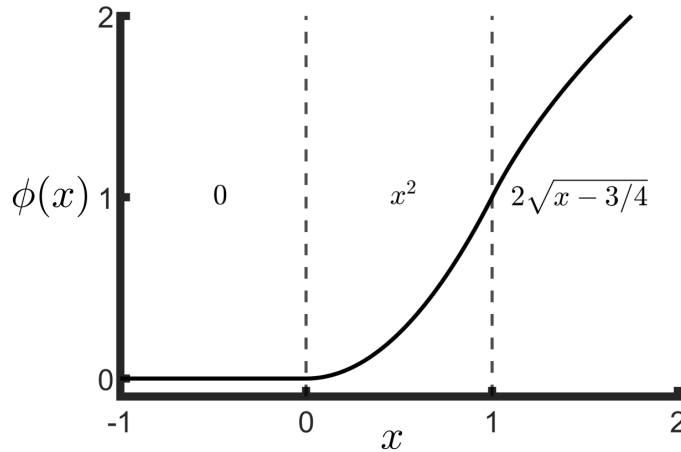


Figure 5.1: **Choice of F-I curve.**

We perform three separate simulations to illustrate the evolution of firing rate dynamics of the same network during the recall of memories associated with the last environments explored n , second to last $n - 1$, and ten environments ago $n - 10$. In

each environment, the firing rate of active neurons, initially set to zero, is updated every 0.5 milliseconds according to Eq 5.2. We assume inactive neurons receive a substantial external inhibitory input ($I_0 \ll 0$), resulting in a null firing rate. For practical purposes, the firing rate of these inactive neurons is set to zero during the simulation. Upon completing the simulation, the firing rates are averaged within each group of sM neurons and then visualized by ordering (averaged) neurons according to their phases in each environment.

Figure 5.2A illustrates the network dynamics for different initial conditions over one second. When the visualized environment (rows) matches the initial condition (columns), bump dynamics rapidly emerge post-simulation, with the overall dynamics evolving to a steady state, as shown in the diagonal panels. In this simulation, the sparseness is set to 0.1, meaning there is a 1% chance that any neuron can be active in two different environments. Consequently, 10% of neurons active in the initial environment are also active in another environment. As a result, when visualizing an environment that does not match the initial condition, some firing activity is still observed, albeit without any structured pattern, as shown in the off-diagonal panels.

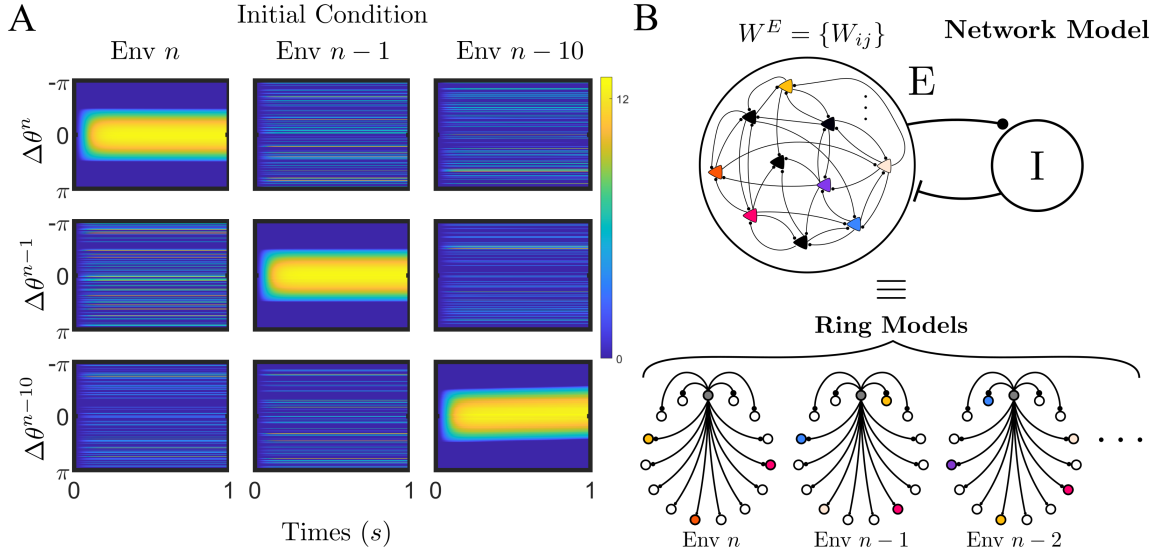


Figure 5.2: **Network dynamics can be approximated by the ring model in the sparse limit.** (A) Illustration of the bump dynamics for a single network with three distinct initial conditions over a one-second period. (B) The topology of the recurrent network shaped by BTSP, and with sparse coding, can be approximated by a series of rings, one for each environment.

These simulations suggest that approximating the full network dynamics by a ring model for each environment is feasible by constraining the dynamics to the manifold corresponding to each environment, as shown in Fig 5.2B. However, there is no guarantee that the whole dynamics will remain confined to this manifold. We hypothesize that this approximation holds if the sparseness is sufficiently low, ensuring distinct representations of environments, but it may break down as $s \rightarrow 1$. To test this hypothesis, we conduct additional simulations of firing rates using the weight matrix derived from the learning process and compare these results with those obtained from the corresponding ring model in each environment, as described in the subsequent section.

5.2 Ring-model approximation

Simulations with 10% sparseness suggest the dynamics in a network of firing neurons with full recurrent connectivity might be approximated by considering a set of distinct ring models, one for each past environment. To test whether this approximation holds for $s = 0.1$, we first generate the connectivity matrix using the statistical properties of the original weight matrix in each environment. These quantities were precisely calculated analytically in the previous chapter. Specifically, in the ring model for an environment k , the connectivity between neurons at a phase difference of $\Delta\theta^{n-\eta}$ can be expressed as:

$$W(\Delta\theta^{n-\eta}) = W_0 + W_1^\eta \cos(\Delta\theta^{n-\eta}) + \Delta W z(\Delta\theta^{n-\eta}), \quad (5.4)$$

where $W_1^\eta = W_{\max} \frac{sM}{\kappa} a_\eta$, $\Delta W = W_{\max} \frac{\sqrt{sM}}{\kappa} \sqrt{V_\eta}$, and z is a zero-mean Gaussian random variable with unit variance. Here, the statistics are rescaled to match those in Eq 5.1. The first two terms determine the mean memory trace in environment $n - \eta$, and the last one quantifies the variability of memory trace around the mean. Additionally, in the simulation, the ΔW depends on the phase difference to reproduce the complex shape of the variance curve, as shown in Figure 4.3.

The advantage of the ring-model formulation lies in its low dimensionality, which significantly simplifies analysis compared to the full network described by Eq 5.2. For instance, by choosing the scaling $\kappa = \sqrt{sM}$ in the network (with $P = D$), it can be demonstrated through the analysis of the ring model that the memory capacity,

estimated via a linear stability analysis of the spatially uniform steady state, scales exactly as in Eq 4.25 (see later section). However, we use $\kappa = sM$ in our numerical simulations.

Equation 5.4 thus establishes a mapping between the statistics of the full network’s connectivity and the ring model’s coupling parameters. To ensure a robust comparison of both models, we generate ten weight matrices using the plasticity rule (with parameters: $N = 256$, $n = 1500$, $M = 60$, $s = 0.1$, $P = D = 0.3$) described in Eq 4.17, varying the random seed for each instance. The seed controls the generation of environments by randomly selecting active neurons and assigning phases in each environment, thereby determining the stochastic component of the learning process. As a result, all generated weight matrices are statistically equivalent.

Subsequently, all weight matrices undergo the firing rate simulation, described in Eq 5.2, with parameters $W_0 = -0.25$, $W_{max} = 40$, and $I_0 = 0.2$. To maintain clarity and avoid excessive simulations, we select a specific set of environments for comparison. Also, our primary focus shifts to emphasizing η rather than $n - \eta$, as we aim to quantify the maximum number of memories related to previously explored environments that can be retrieved through the firing rate, where η indicates the memory age. A memory is considered retrievable if the firing rates of neurons active in that environment present a bump structure at the steady state.

However, a homogenous solution is also possible. To capture both solutions simultaneously, we apply two initial perturbations to the firing rate for each memory age $\eta \geq 0$. For all the active neurons active in environment $n - \eta$, we consider following perturbation: $FR_0(\theta) = C_0(1 + \cos \theta)$. One simulation employs a small perturbation, $C_0 = I_0^2$, and the other uses a large perturbation, $C_0 = 1.5$. This dual approach allows us to capture both solutions simultaneously.

As in the previous simulation, the firing rates are updated every 0.5 milliseconds until a steady state is reached, defined by an absolute difference in the mean firing rate between two consecutive steps of less than 10^{-12} . Here, we refer to the steady state as the stabilization of the bump amplitude. Occasionally, the center of the bump solution may shift after the amplitude has stabilized, which is not captured by this stopping criterion. Once the steady state is achieved, we average the firing rates within each group and then extract the first Fourier mode. The amplitude of

the bump for memory age η is the twice the value of the first Fourier mode for that memory. We also conduct simulations with ring models under the same conditions.

Fig 5.3 shows a comparison of the bifurcation diagram generated using the full network model (red symbols) and the low-dimensional approximation given by the ring model (black lines). The red symbols represent the mean and 95% confidence interval of the final bump amplitude for ten weight matrices using two initial conditions for different memory ages. Notably, there is a direct link between memory age (η) and spatial modulation (W_1). Changing η has two primary effects: first, the spatial modulation in the recurrent connectivity decreases with increasing age; second, the strength and shape of quenched variability also change, as described in Eq 4.21. Therefore, these two terms can be used interchangeably.

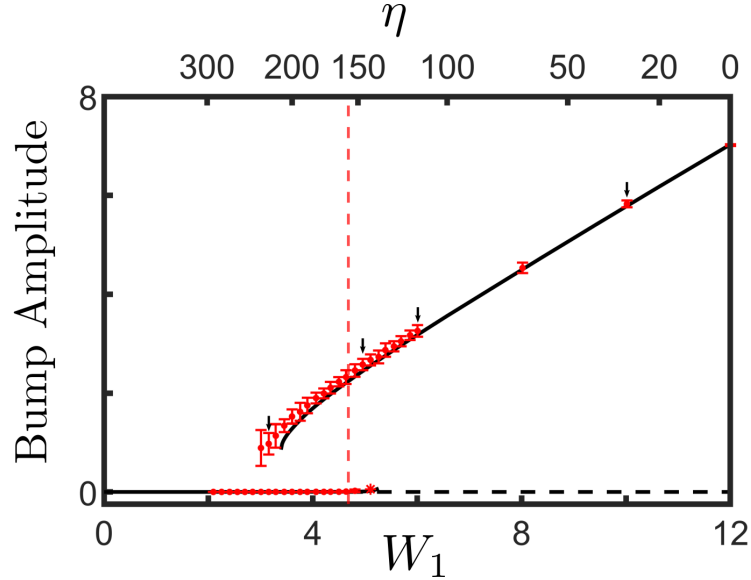


Figure 5.3: **Comparison of bifurcation diagram for full network model and ring-model approximation.** Bifurcation diagram as a function of W_1 (bottom x-axis) or η (top x-axis) for the ring model and network model, respectively. Solid lines represent stable solutions, while dashed lines represent unstable solutions of the ring model. Red dots indicate the mean amplitude of bump solutions from the network model, with error bars representing the 95% confidence interval. The dotted line denotes the analytically calculated critical value of W_1 at the Turing bifurcation for the ring model.

Simulation results show a good agreement between the network simulation and the ring approximation for $s = 0.1$. For large W_1 , both initial conditions converge to a bump solution (large amplitude), rendering the flat solution (null amplitude)

an unstable fixed point. As spatial modulation decreases, the bump amplitude is also reduced. For a small value of W_1 , the bump solution does not exist, and the flat one becomes the unique attractor. For intermediate values of W_1 , both solutions coexist: the small perturbation condition converges to the flat solution, while the large perturbation condition converges to the bump solution.

For the ring model, we can easily identify two distinct bifurcations at the boundaries of this bi-stable region, where flat and bump solutions coexist. There are the saddle-node bifurcation of Turing patterns and Turing bifurcations. To identify the saddle-node bifurcation, we run simulations using large perturbation conditions, starting from a large value of W_1 (indicating the most recent memory) and gradually decreasing to a small W_1 (a large memory age). Each memory age corresponds to one simulation. Under these conditions, the large perturbation continues to converge to the bump solution until it reaches the saddle-node bifurcation. Interestingly, this bifurcation is shifted even further left in the network model compared to the ring model, shown in Fig 5.3 and 5.4A. When examining the network simulations individually, almost all outperform the ring approximation. Plotting the resulting firing rate for one simulation, a clear bump is observed, as shown in Fig 5.4C.

Conversely, to determine the Turing bifurcation, we perform simulations under small perturbation conditions, starting with a low W_1 value and gradually increasing it to a higher W_1 value. In this scenario, the system consistently converges to the flat solution until the Turing bifurcation is reached. Upon examining the flat branch in detail, as shown in Fig 5.4B, we observe that this Turing bifurcation is supercritical, meaning it is continuous and results in very small amplitude but stable bumps just to the right of the bifurcation. The critical value of W_1 , marked by the red dotted line, at which a Turing instability occurs, will be systematically analyzed in the next section. Network simulations also confirm the existence of small bump amplitude solutions, as depicted in Fig 5.4D.

The discrepancies observed at the bifurcations are likely attributable to additional sources of variability inherent in the network model. Firstly, the connectivity matrix for the ring model considers only the spatial statistics, assuming no interaction or interference between environments beyond the Gaussian quenched variability. Secondly, in the ring model, the firing rate at each position represents the average activity

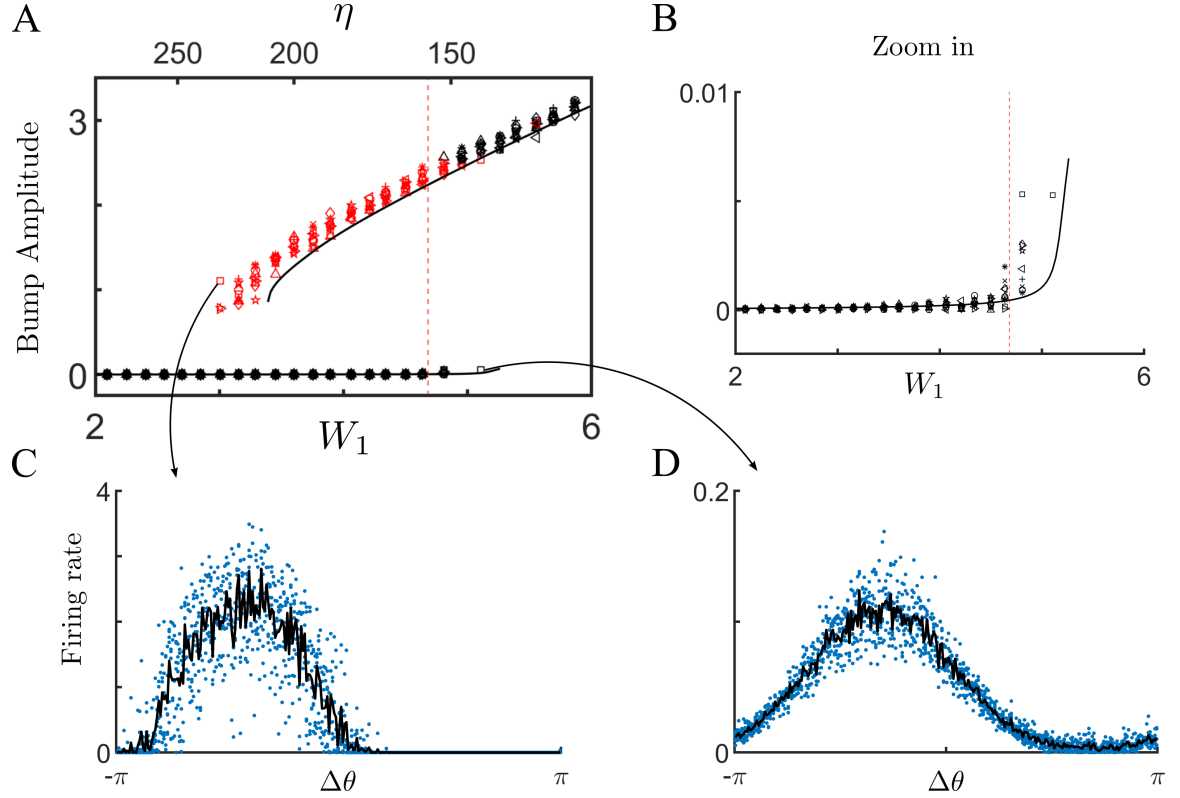


Figure 5.4: **The zoomed-in bifurcation plot distinctly shows a supercritical Turing bifurcation.** (A) Zoomed plot of Fig 5.3 showing bump amplitude in an individual fashion. Black represents the small perturbation condition, and red represents the large one. (B) Detailed bifurcation diagram showing a supercritical bifurcation from the unpatterned state to the small amplitude bump state. (C-D) Sample profile of large and small bump amplitude indicated by the arrows. The black line represents the mean firing rate, while the blue dots represent the individual firing rates from the simulation.

across populations of cells with that preferred location, thereby simplifying the dynamics. In contrast, the network model captures a richer dynamic landscape due to the distribution of firing rates within each population.

It is important to emphasize that the x-axis in the bifurcation diagram in Fig 5.3 represents the memory age, η , and thus each point corresponds to a different low-dimensional manifold. For the ring model, the solution is inherently confined to this manifold by definition, whereas this restriction does not necessarily apply to the full network model. In other words, the y-axis measures the amplitude of a bump corresponding to a specific η , but other orderings are possible, allowing bump amplitudes to be measured for all explored environments. To confirm that the bump solutions depicted in Fig 5.4 are indeed bumps in the desired environment η , we measured the bump amplitude across a range of η values for several example cases. Fig 5.5A demonstrates that for the four chosen values of η , indicated by arrows in Fig 5.3, the bump exists only on a single manifold, while projections of the activity onto other manifolds yield disordered activity (bump amplitude zero). Fig 5.5B provides an example of the steady-state bump solution when a small amplitude bump is seeded according to the ordering of neurons for $\eta = 30$. As shown, the bump remains constrained to the manifold for $\eta = 30$.

Finally, the network's capacity can be determined from the phase diagram of the dynamical states shown in Fig 5.6, which is derived from the ring model. For each value of W_0 , the firing rate simulation is conducted for the ring model to identify the saddle-node bifurcation for bump states, indicated by the dashed line: no bump states exist to the left of this line. The Turing bifurcation, indicated by the solid line, is computed using the analytical formula (detailed in the next section). Between these two lines, any existent bump state is bi-stable with the flat state, provided the ring-model approximation is accurate. Thus, network simulation results (represented by dots) can be overlaid, and the maximum capacity is defined by the greatest age that remains to the right of the saddle-node bifurcation. In this scenario, the network capacity is 210.

These numerical simulations suggest that the ring approximation is valid in the limit of sparse coding. For the case studied, where $s = 0.1$, the minimal overlap between environments causes the network simulation to outperform the ring model,

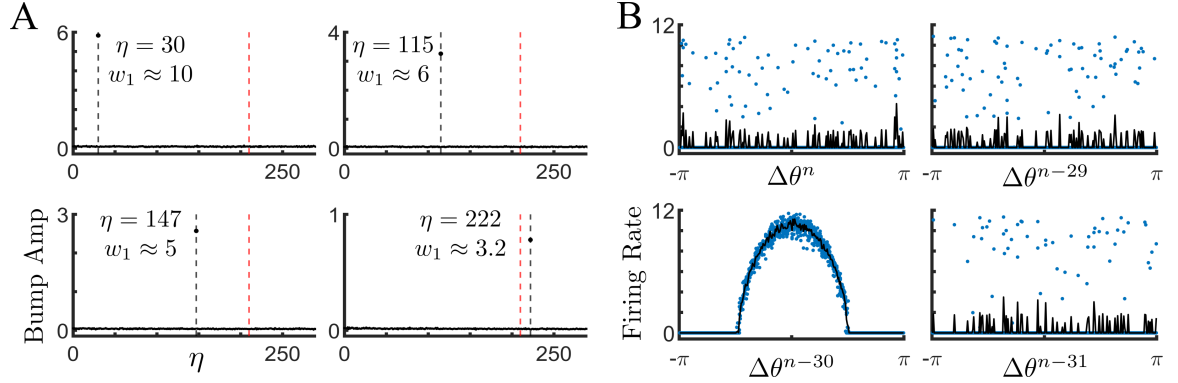


Figure 5.5: **Calculation of bump amplitude over a range of η show the bump state is constrained to the desired environment.** (A) Bump amplitudes calculated by ordering cells according to their preferred positions across a range of environments. These amplitudes are averaged over ten simulations, each using distinct connectivity matrices. The four panels correspond to specific values of η indicated by arrows in Fig 5.3. The black vertical dashed line represents the retrieved environment, while the red dashed line indicates the saddle-node bifurcation of the ring-model approximation. (B) The steady-state neuronal activity in the full network, ordered according to four different environments, for the retrieval of environment $n - 30$. This specific case is marked with an asterisk in panel D.

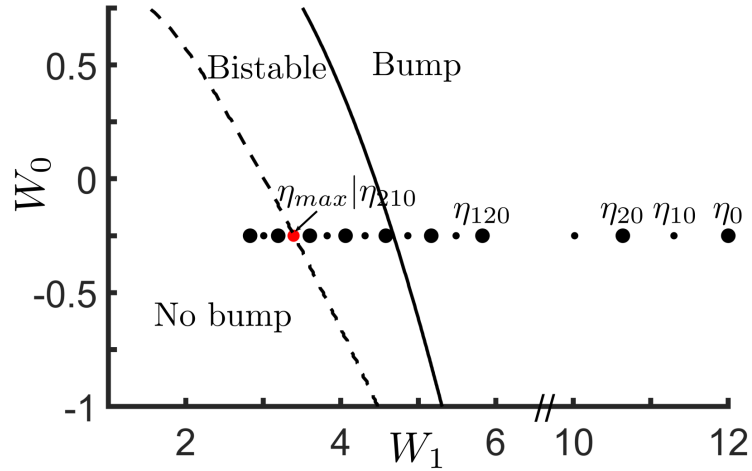


Figure 5.6: **Phase diagram as a function of W_0 and W_1 .** The phase diagram illustrates the transitions between different dynamical states. The dashed line denotes the saddle-node bifurcation for bump states, indicating that no bump states exist to the left of this line. The solid line marks the Turing bifurcation. Between these two lines, bump states are bi-stable with the flat state. Black dots represent the system's state at specific memory ages (η), as analyzed in this study. The red dot indicates the critical point, the maximum memory age where bump states can still exist.

leading to inconsistencies near bifurcation points. These inconsistencies arise from increased variability, indicating that such variability plays a role in enhancing network capacity. However, further analysis of variability in this type of network is necessary to understand its implications fully.

5.3 Impact of variability and network parameters on Turing bifurcation in sparse coding networks

The numerical simulations for $s = 0.1$ suggest that the ring-model approximation is valid in sparse coding. This validation highlights the ring model's effectiveness in capturing the network's critical dynamics. In this section, we seek to understand how the variability, as described by Eq 4.21, influences the Turing bifurcation, a critical transition from a homogeneous state to a patterned state characterized by bump solutions. Understanding this influence is crucial for identifying conditions allowing the network to maintain stable memory despite inherent fluctuations and noise.

For sparse networks, restricting the dynamics to the manifold corresponding to spatial modulations in environment $n - \eta$ allows the ring model to be a good approximation. When the number of spatial positions N is large enough, the summation of the network's contribution to the firing rate, in Eq 5.2, can be approximated by an integral, assuming a circular track. This results in the following equation for the ring model in environment $n - \eta$:

$$\tau \frac{\partial}{\partial t} r(\theta^{n-\eta}, t) = -r(\theta^{n-\eta}, t) + \phi \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\theta^{n-\eta} - \theta') r(\theta', t) d\theta' + I_0 \right), \quad (5.5)$$

where

$$W(\Delta\theta^{n-\eta}) = W_0 + W_1^\eta \cos(\Delta\theta^{n-\eta}) + \Delta W z(\Delta\theta^{n-\eta}),$$

as detailed in Eq 5.4. Consequently, we have a set of ring models, one for each

environment:

$$\begin{cases} \tau \frac{\partial}{\partial t} r(\theta^n, t) = -r(\theta^n, t) + \phi \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\theta^n - \theta') r(\theta', t) d\theta' + I_0 \right) \\ \tau \frac{\partial}{\partial t} r(\theta^{n-1}, t) = -r(\theta^{n-1}, t) + \phi \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\theta^{n-1} - \theta') r(\theta', t) d\theta' + I_0 \right) \\ \tau \frac{\partial}{\partial t} r(\theta^{n-2}, t) = -r(\theta^{n-2}, t) + \phi \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\theta^{n-2} - \theta') r(\theta', t) d\theta' + I_0 \right) \\ \vdots \end{cases}$$

By definition, the high dimensionality of the ring models makes it impractical to analyze the Turing bifurcation directly. To address this, we simplify them by ignoring the dependency of the spatial modulation and noise strength on memory age. This simplification results in the following equation:

$$\tau \frac{\partial}{\partial t} r(\theta, t) = -r(\theta, t) + \phi \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\theta - \theta') r(\theta', t) d\theta' + I_0 \right). \quad (5.6)$$

The connectivity matrix is now given by:

$$W(\Delta\theta) = W_0 + W_1 \cos(\Delta\theta) + \Delta W z(\Delta\theta),$$

where $W_0 \in \mathbb{R}$, W_1 ranges between 0 and 12 (considering $W_{max} = 40$), and ΔW can be either scale value or depending on the phase difference. Leveraging this ring model, we first analyze the stationary uniform solutions and then determine the Turing bifurcation through linear stability analysis.

5.3.1 Stationary uniform solutions of the ring model

For the simplified ring model, the stationary uniform (SU) solutions, in the absence of the noise term, can be derived by setting $\frac{\partial}{\partial t} r(\theta, t) = 0$. This yields:

$$\begin{aligned} r_0 &= \phi \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} W(\theta - \theta') r_0 d\theta' + I_0 \right) \\ &= \phi \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} (W_0 + W_1(\theta - \theta')) r_0 d\theta' + I_0 \right) \\ &= \phi(W_0 r_0 + I_0), \end{aligned} \quad (5.7)$$

where r_0 is a constant and spatially uniform firing rate. Given the non-linear function used in this study, as outlined in Eq 5.3, we can identify a total of four solutions based on the value of $W_0 r_0 + I_0$.

For the range $0 \leq W_0 r_0 + I_0 \leq 1$, the transfer function is $\phi(x) = x^2$. Then, the SU solutions are derived from the equation $r_0 = (W_0 r_0 + I_0)^2$. Solving this equation for r_0 gives us

$$r_{0,T}^{\pm} = \frac{1 - 2W_0 I_0 \pm \sqrt{1 - 4W_0 I_0}}{2W_0^2}.$$

To ensure real solutions, the term under the square root, $\sqrt{1 - 4W_0 I_0}$ must be non-negative, implying that W_0 must be less than or equal to $(4I_0)^{-1}$. On the other hand, if $W_0 r_0 + I_0 \geq 1$, we must solve the equation $r_0 = 2\sqrt{W_0 r_0 + I_0} - 3/4$. The SU solutions in this case are:

$$r_{0,U}^{\pm} = 2W_0 \pm \sqrt{(4W_0^2 + 4I_0 - 3)}.$$

Therefore, these solutions only exist when $W_0 \geq \sqrt{3/4 - I_0}$.

Moreover, the transfer function is continuous at $x = 1$, which indicates that two of the four solutions must be connected if they exist. Specifically, $r_{0,T}^+$ and $r_{0,U}^-$ meet this condition, which occurs when $W_0 r_0 + I_0 = 1$. This condition implies that r_0 must be 1. Solving for W_0 , we find that $W_0 = 1 - I_0$. Taking into account the constraints on W_0 when calculating the solutions, we find that $r_{0,T}^+$ exists within the range $W_0 \in [1 - I_0, (4I_0)^{-1}]$, while $r_{0,U}^-$ exists when $W_0 \in [1 - I_0, \sqrt{3/4 - I_0}]$. Furthermore, these two regions remain in existence when I_0 is less than 0.5, making $I_0 = 0.5$ a bifurcation point where four solutions collapse into two.

Numerical simulations conducted with two different initial conditions reveal that the bounded solutions previously analyzed are unstable when I_0 is below the bifurcation points, as shown in Fig 5.7A. In this case, this unstable solutions region is also bi-stable, limited by dashed lines, where different stable states coexist depending on the initial conditions. When the external input reaches the bifurcation points, Fig 5.7B, the four solutions collapse into two, and the two stable solutions become directly connected, eliminating the bi-stable regions.

The close alignment between numerical and analytical results is disrupted as W_1 increases. When W_1 is shifted from the value 1 (used in Fig 5.7) to 3, discrepancies begin to appear within the bi-stable region, as illustrated in the left panel of Fig 5.8. This divergence arises due to the emergence of bump solutions, which is confirmed by the significant bump amplitude observed in the right panel. This indicates that a bifurcation occurs with changes in W_1 , impacting partially the bi-stable region.

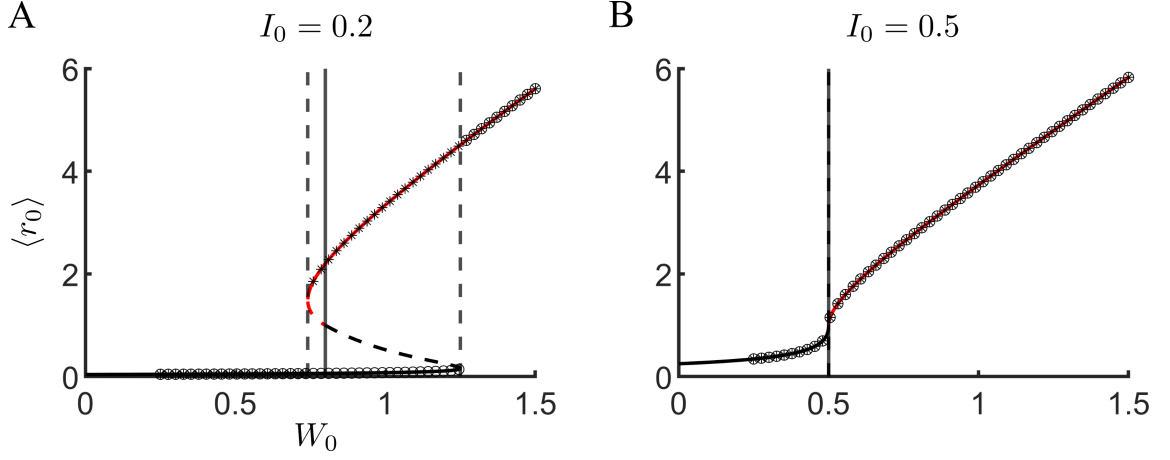


Figure 5.7: **Bifurcation diagram for the stationary solutions of the ring model.** The solid lines represent analytical solutions, with black lines indicating $r_{0,T}$ and red lines indicating $r_{0,U}$. Circles denote numerical solutions initiated from a small initial condition, simulated from a small value of W_0 . Asterisks represent numerical solutions starting from a large initial mean firing rate, simulated from a large W_0 . **(A)** The bi-stable region, marked by the vertical dashed lines, is evident when $I_0 = 0.2$, below the bifurcation point. This region also encompasses the unstable solutions (dashed lines) that converge at $W_0 = 1 - I_0$ (vertical solid line). **(B)** When I_0 reaches the bifurcation point, the bi-stable region vanishes.

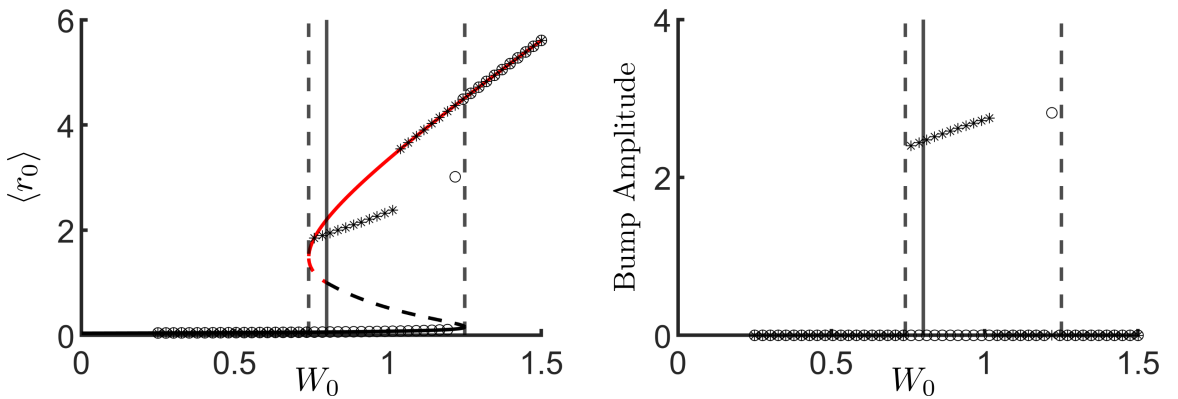


Figure 5.8: **Bifurcation diagram for $W_1 = 3$.** The left panel shows a discrepancy between numerical results and analytical solutions within the bi-stable region. The right panel reveals that this mismatch arises due to the emergence of bump solutions, which affect the amplitude of the numerical results.

By analyzing the impact of parameters on the ring model's stationary uniform solutions, we can identify potential bifurcations based on numerical simulations. However, we do not go further since our primary interest lies in understanding the emergence of the Turing bifurcation. The Turing bifurcation is significant as it marks the transition from a homogeneous state to a patterned state within the network. To this end, we focus on the branch where the average firing rate is nearly zero, denoted as $r_{0,T}^-$. This condition is indicative of scenarios where W_0 takes small or negative values.

This choice is deliberate, as small or negative values of W_0 are associated with low baseline firing rates, which are critical for the emergence of Turing patterns. By restricting our attention to this branch, we aim to isolate and study the conditions under which spatial patterns, or bumps, form in the network. These bumps are essential for memory retrieval, as they represent localized regions of heightened neural activity corresponding to stored memories. Through this focused analysis, we can gain deeper insights into the mechanisms underlying the formation and stability of these memory-related patterns in sparsely coded neural networks.

5.3.2 Linear stability analysis and Turing bifurcation of the ring model in the absence of noise

To determine whether bump attractors can emerge spontaneously, we consider two distinct types of perturbations to the system: spatially homogeneous and spatially inhomogeneous. Given that these perturbations are weak, the response will be linear, allowing us to linearize the equations. Consequently, we can decompose the spatially inhomogeneous perturbation into a series of Fourier modes, and the total solution will be a sum of these modes. Thus, the perturbation ansatz is given by:

$$r(\theta, t) = r_0 + \delta r_0 e^{\lambda_0 t} + \delta r_1 \cos(\theta) e^{\lambda_1 t}, \quad (5.8)$$

where $\delta r \ll 1$ and r_0 is stationary uniform solution previously analyzed.

By differentiating both sides of the ansatz, we obtain:

$$\frac{\partial}{\partial t} r(\theta, t) = \delta r_0 \lambda_0 e^{\lambda_0 t} + \delta r_1 \lambda_1 \cos(\theta) e^{\lambda_1 t}. \quad (5.9)$$

Substituting the ansatz and its derivative into the ring model equation, we have:

$$\begin{aligned} \delta r_0 \lambda_0 e^{\lambda_0 t} + \delta r_1 \lambda_1 \cos(\theta) e^{\lambda_1 t} = & -(r_0 + \delta r_0 e^{\lambda_0 t} + \delta r_1 \cos(\theta) e^{\lambda_1 t}) \\ & + \phi \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} (W_0 + W_1 \cos(\theta - \theta')) (r_0 + \delta r_0 e^{\lambda_0 t} + \delta r_1 \cos(\theta') e^{\lambda_1 t}) d\theta' + I_0 \right) \end{aligned}$$

Using integral properties and the trigonometric identity:

$$\cos(\theta - \theta') \cos(\theta') = \cos \theta \frac{1 - \cos(2\theta')}{2} - \sin \theta \frac{\sin(2\theta')}{2},$$

it can be simplified further to

$$\begin{aligned} \delta r_0 \lambda_0 e^{\lambda_0 t} + \delta r_1 \lambda_1 \cos(\theta) e^{\lambda_1 t} = & -(r_0 + \delta r_0 e^{\lambda_0 t} + \delta r_1 \cos(\theta) e^{\lambda_1 t}) \\ & + \phi \left(W_0 r_0 + W_0 \delta r_0 e^{\lambda_0 t} + \frac{1}{2} W_1 \delta r_1 \cos(\theta) e^{\lambda_1 t} + I_0 \right). \end{aligned}$$

Applying the Taylor expansion to the ϕ around the $W_0 r_0 + I_0$ yields to

$$\begin{aligned} \delta r_0 \lambda_0 e^{\lambda_0 t} + \delta r_1 \lambda_1 \cos(\theta) e^{\lambda_1 t} = & -(\delta r_0 e^{\lambda_0 t} + \delta r_1 \cos(\theta) e^{\lambda_1 t}) \\ & + \phi'(W_0 r_0 + I_0) W_0 \delta r_0 e^{\lambda_0 t} + \phi'(W_0 r_0 + I_0) \frac{1}{2} W_1 \delta r_1 \cos(\theta) e^{\lambda_1 t}, \quad (5.10) \end{aligned}$$

where we use the fact that r_0 is SU solution, $r_0 = \phi(W_0 r_0 + I_0)$, of the ring model. Here, ϕ' is the slope of the transfer function evaluated at the steady state.

Solving Eq 5.10 separately for both eigenvalues yields:

$$\begin{aligned} \delta r_0 \lambda_0 e^{\lambda_0 t} &= -\delta r_0 e^{\lambda_0 t} + \phi'(W_0 r_0 + I_0) W_0 \delta r_0 e^{\lambda_0 t}, \\ \delta r_1 \lambda_1 \cos(\theta) e^{\lambda_1 t} &= -\delta r_1 \cos(\theta) e^{\lambda_1 t} + \phi'(W_0 r_0 + I_0) \frac{1}{2} W_1 \delta r_1 \cos(\theta) e^{\lambda_1 t}. \end{aligned}$$

This leads to the following expressions for the eigenvalues:

$$\lambda_0 = -1 + \phi'_0 W_0, \quad (5.11)$$

$$\lambda_1 = -1 + \phi'_0 \frac{W_1}{2}, \quad (5.12)$$

where $\phi'_0 = \phi'(W_0 r_0 + I_0)$. These equations for the eigenvalues λ describe the growth rates of perturbations. If λ is negative, the perturbation decays; if positive, the perturbation grows. Specifically, a positive growth rate for the cosine perturbation indicates the onset of a bump, known as a Turing instability. Therefore, to avoid a

uniform instability of the steady state, W_0 must satisfy $W_0 < 1/\phi'_0$. The critical value of the spatially modulated connectivity is then given by:

$$W_1^{cr} = \frac{2}{\phi'_0}. \quad (5.13)$$

To connect this result to the network, we substitute the value of W_1 for environment $n - \eta$, which allows us to compute the memory capacity restricted to the Turing branch. Specifically, we set up the equation for W_1 and solve for η , the memory age:

$$W_{max} \frac{sM}{\kappa} \frac{2PD}{P+D} (1 - s^2(P+D))^\eta = \frac{2}{\phi'_0}.$$

Solving for η , we find:

$$\eta_{cr} = \frac{1}{s^2(P+D)} \ln \left(\frac{sM}{\kappa} W_{max} \phi'_0 \frac{PD}{P+D} \right), \quad (5.14)$$

where we assume that $s \rightarrow 0$. Taking $P = D$ and $\kappa = \sqrt{sM}$ and making appropriate approximations, particular large sM , we derive:

$$\eta_{cr} = \frac{1}{2s^2P} \left(\ln(\sqrt{sM}) + \ln \left(W_{max} \phi'_0 \frac{P}{2} \right) \right).$$

This approximation reveals that the leading order term for large sM aligns with the result from the signal-to-noise analysis (Eq 4.25), indicating that the memory capacity derived from the Turing bifurcation analysis confirms the findings from the SNR analysis.

5.3.3 The role of system size on quenched variability in the ring model

Analyzing the Turing bifurcation without noise lays the foundation for understanding network dynamics. However, in the context of the full network, quenched variability — calculated based on the plasticity rule — takes the form $\Delta W(\theta) = \frac{\sqrt{sM}}{\kappa} \sqrt{V(\theta)} z(\theta)$, where θ represents the phase difference between two neurons within the environment of interest, and z is a spatially uncorrelated Gaussian random variable with zero mean and unit variance. Consequently, the strength of this variability is determined by the population size at each position, sM . For example, if we choose $\kappa = sM$, the quenched variability will vanish with increasing population size. The number of encoded positions N does not play a role here.

However, this situation changes significantly near an instability of a spatially modulated mode, such as at a Turing bifurcation. To incorporate the effects of quenched variability, we represent the Gaussian noise process, ignoring the prefactor of $\frac{\sqrt{sM}}{\kappa}$, $\Delta W(\theta) = \sqrt{V(\theta)}z$ using its Fourier series:

$$\begin{aligned}\Delta W(\theta) &= \sum_{j=1}^N \left(c_j e^{ij\theta} + \bar{c}_j e^{-ij\theta} \right), \\ &= 2 \sum_{j=1}^N \alpha_j \cos(j\theta) + 2 \sum_{j=1}^N \beta_j \sin(j\theta),\end{aligned}\tag{5.15}$$

where N denotes the number of spatial positions, equal to the number of place cell populations. Given that the variability is a zero-mean Gaussian process, the coefficients α_j and β_j are also zero-mean Gaussian random variables whose variances and covariances must be determined self-consistently through averaging. Specifically, we calculate $V(\theta) = \langle \Delta W(\theta)^2 \rangle$, leading to:

$$\begin{aligned}V(\theta) &= 2 \sum_{j=1}^N \sum_{l=1}^N \left(\langle \alpha_j \alpha_l \rangle + \langle \beta_j \beta_l \rangle \right) \cos((j-l)\theta) \\ &\quad + 2 \sum_{j=1}^N \sum_{l=1}^N \left(\langle \alpha_j \alpha_l \rangle - \langle \beta_j \beta_l \rangle \right) \cos((j+l)\theta) \\ &\quad + 4 \sum_{j=1}^N \sum_{l=1}^N \left(\langle \alpha_j \beta_l \rangle \sin((j-l)\theta) - \langle \alpha_j \beta_l \rangle \sin((j+l)\theta) \right).\end{aligned}\tag{5.16}$$

We recall that $V(\theta) = A + \frac{C}{2} + B \cos(\theta) + \frac{C}{2} \cos 2(\theta)$.

To determine the statistics of the coefficients in the Fourier series, we apply Parseval's theorem, which states that the variance, or power, of a signal must be conserved in Fourier space. Additionally, since z is a white-noise process in space, the power is not only conserved but also evenly distributed across all modes. Therefore, the variance of the noise process is the sum of the variances of its Fourier coefficients

with each of term contributes evenly:

$$\langle \alpha_1^2 \rangle = \frac{1}{2N} \left(A + \frac{3C}{4} \right), \quad (5.17)$$

$$\langle \beta_1^2 \rangle = \frac{1}{2N} \left(A + \frac{C}{4} \right), \quad (5.18)$$

$$\langle \alpha_{j \neq 1}^2 \rangle = \langle \beta_{j \neq 1}^2 \rangle = \frac{1}{2N} \left(A + \frac{C}{2} \right), \quad (5.19)$$

$$\langle \alpha_j \alpha_{j+1} \rangle = \langle \beta_j \beta_{j+1} \rangle = \frac{B}{4N}, \quad (5.20)$$

$$\langle \alpha_j \alpha_{j+2} \rangle = \langle \beta_j \beta_{j+2} \rangle = \frac{C}{8N}, \quad (5.21)$$

see Appendix B for a detailed derivation. These results show that for finite N , the quenched variability generates power in the mode relevant for a bump instability, which is the first Fourier mode ($j = 1$).

Each place cell in the network has a pair of Fourier coefficients, α_1 and β_1 , contributing a term $R_1 \cos(\theta - \phi_1)$ to the connectivity, where $R_1 = 2\sqrt{\alpha_1^2 + \beta_1^2}$ and $\phi_1 = \arctan(\beta_1/\alpha_1)$. These amplitudes and phases are also random variables, unique to each neuron in the network. Consequently, at some network locations, the quenched variability will facilitate the instability leading to a bump solution (where $\phi_1 \sim 0$), while at others, it will hinder this instability (where $\phi_1 \sim \pi$).

The net contribution is computed by averaging the amplitude over the distribution of α and β , yielding $\langle R_1 \rangle = \sqrt{\frac{2\pi}{N} \left(A + \frac{C}{2} \right) \cdot \frac{\sqrt{sM}}{\kappa}}$. Therefore, the critical value of the connectivity becomes:

$$W_1^{cr} = \frac{2}{\phi'_0} - \langle R_1 \rangle, \quad (5.22)$$

indicating that the bifurcation occurs at lower values of W_1 compared to the case without quenched variability. For the simulations, the Turing bifurcation, indicated by the red vertical dashed lines in Fig 5.3 and 5.4A, is adjusted by considering this correction on the spatial modulation. The critical age is determined by finding the value of η that satisfies the following condition:

$$\eta_{cr} = \max_{\eta} \left\{ a_{\eta} + \sqrt{\frac{2\pi}{N} \left(A_{\eta} + \frac{C_{\eta}}{2} \right) \frac{\sqrt{sM}}{\kappa}} \leq \frac{1}{W_0 r_0 + I_0} \right\},$$

while W_1^{cr} is computed using the amplitude equation, Eq 4.14, by substituting η_{cr} into the equation.

In any case, we can see that the shift in the bifurcation is inversely proportional to the number of positions N . Specifically, smaller N values result in a greater

memory capacity. However, if N is too small, the continuum approximation fails, and no Turing instability would be expected to occur. This indicates that there is an optimal number of encoded positions N that maximizes memory capacity, given that all other parameters remain constant. Simulations using the same parameters as in section 5.2 but varying the number of spatial position N confirm our hypothesis, as shown in Fig 5.9.

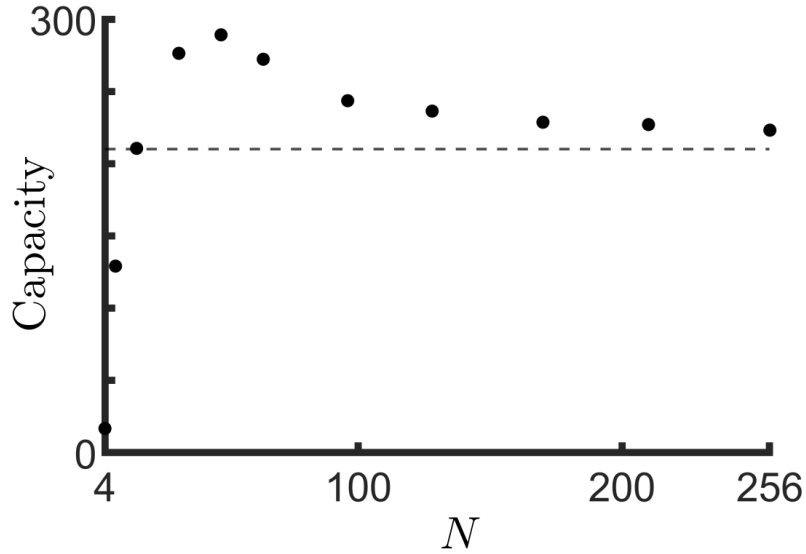


Figure 5.9: **The memory capacity of the network is shown to depend non-monotonically on the resolution of spatial tiling of place cell.** Memory capacity, measured as the number of retrievable bump solutions in the network model, varies with the number of encoded positions, which corresponds to the spatial resolution. In the large N limit, the capacity approaches that of the ring model without quenched variability, represented by the dashed line. As N decreases, the amplitude of the quenched variability increases, which facilitates the generation of bump solutions. However, when N becomes too small, bump solutions no longer exist as the ring model approximation breaks down. This results in an optimal value of N for which memory capacity is maximized.

We have seen that in the sparse coding limit, the interference from other stored memories takes the form of quenched noise in the synaptic weight matrix. Counter-intuitively, this variability actually enhances the memory capacity by driving bump solutions below the noiseless bifurcation point, much like how intermittent oscillations emerge when a damped oscillator is driven by dynamic noise. However, such “fluctuation-driven” bumps likely represent imperfect spatial memories as the variability will drive the bump toward one or more hotspots around the ring. Namely, the

representation will not be spatially homogeneous as one would expect of an unbiased spatial map.

5.4 Discussion

The agreement between network simulations and the ring model approximation holds in the limit of sparse coding. In this scenario, the overlap between sub-populations of place cells encoding distinct environments vanishes, causing the dynamics of each memory to evolve on distinct, weakly-interacting manifolds. This weak interaction manifests as quenched variability. For $s = 0.1$, this approximation holds across all values of η , as shown in section 5.2. However, increasing the value of s introduces nontrivial effects that are not captured by the ring model, such as the emergence of mixed attractors. This phenomenon is illustrated in Fig 5.10. To make a fair comparison, we keep the sM constant and set it to 6 in each of simulation.

Bifurcation diagrams for $s = 0.3$ and $s = 0.5$ are shown in panels A and D, respectively. Here, the lines represent the ring model simulations, while the red circles denote the full network simulations. For recent memories, the attractor states remain confined to a manifold corresponding to a single environment, as seen in Fig 5.10B for $\eta = 3, 12$, and 16 , and in Fig 5.10E for $\eta = 1$. However, this is not the case for more remote memories, which tend to form mixed states when s is sufficiently large, as depicted in Fig 5.10B for $\eta = 26$ and Fig 5.10E for $\eta = 4 - 6$. Mixed states exhibit spatial modulation across several, often many, environments simultaneously. For instance, for $s = 0.3$ and $\eta = 26$, the steady-state solution shows significant spatial modulation in the environment η , as well as in at least the ten most recently explored environments, as shown in Fig 5.10B and C. Generally, in the regime of mixed attractors, there is a high degree of variability between simulations with different realizations of the connectivity matrix, as indicated by the error bars in Fig 5.10A,D.

We conclude that a recurrent network with Behavioral Timescale Synaptic Plasticity can encode a large number of attractors through one-shot learning, provided that the coding is sufficiently sparse. In this sparse coding limit, the interaction between attractors manifests as quenched variability. Interestingly, this variability enhances memory capacity by stabilizing bump attractors in parameter regimes where they

$s = 0.3$

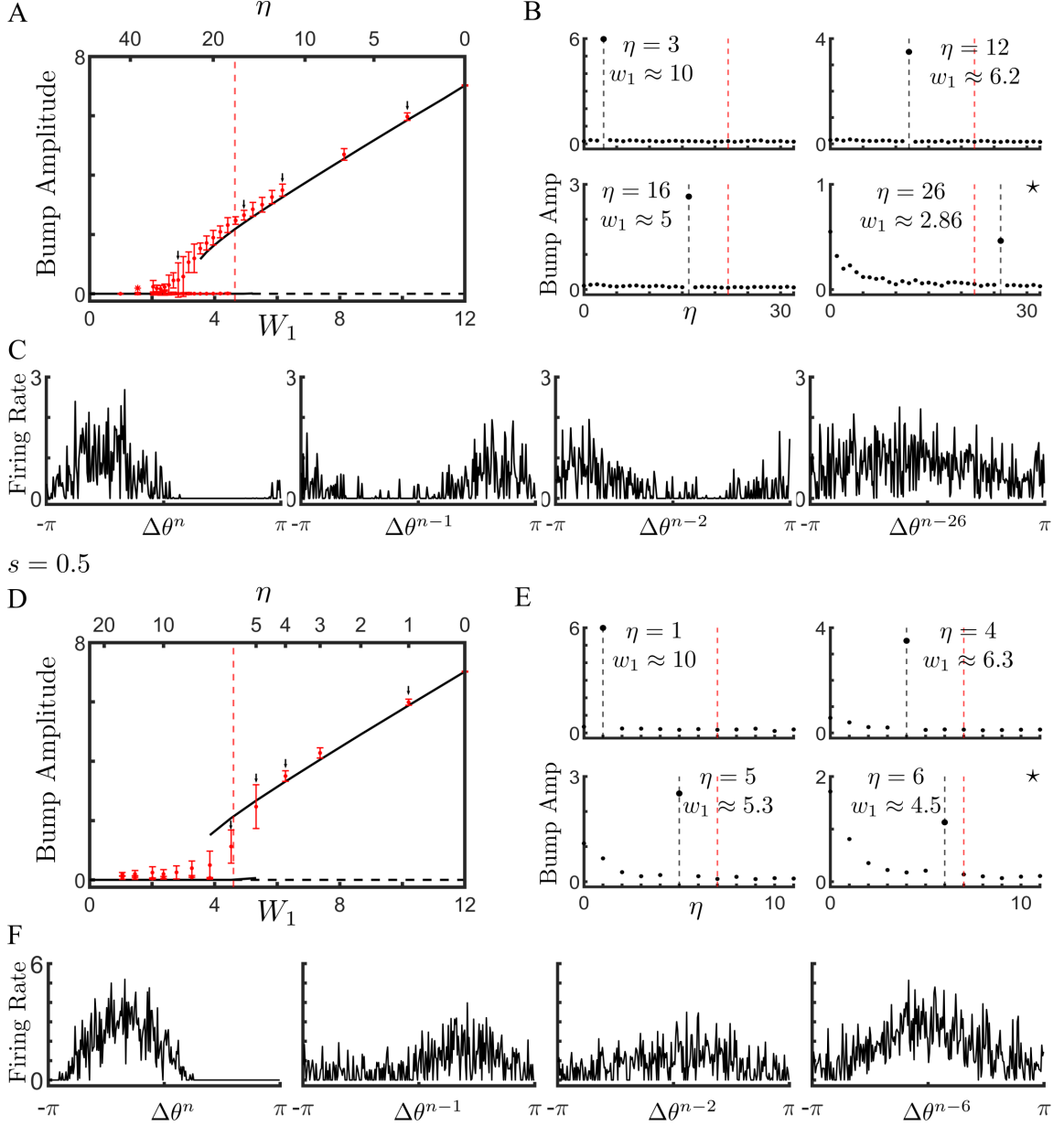


Figure 5.10: **Deviations of the network model from the ring-model approximation are due to the emergence of mixed attractors.** (A) Bifurcation diagram for $s = 0.3$. Lines are from simulation of the ring model, while symbols indicate the mean and 95% confidence intervals from simulations of the network model. (B) Amplitude of the bump as measured with different orderings. Note that for $\eta = 26$ the steady-state solution is a bump in many different environments at the same time. (C) Sample profiles of the same steady-state solution but with distinct neuronal orderings corresponding to different environments. (D-F) Same as A-C but with $s = 0.5$.

would not exist in the absence of noise. In this regime, memory capacity scales as M^2 , where M represents the number of place cells available for encoding each position (with only a fraction s being active), as shown in Fig 5.11. This aligns with the theoretical optimum [60]. Furthermore, the full network with BTSP surpasses the ring-model approximation in this regime, likely due to the additional quenched variability in the firing rates within each population of place cells, as depicted in the inset of Fig 5.11.

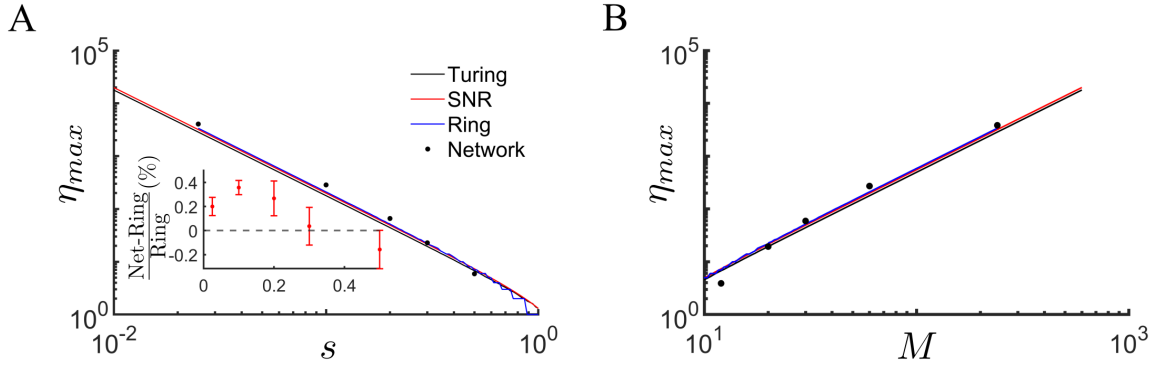


Figure 5.11: Memory capacity in the network model outperforms ring model approximation and scales optimally with system size. (A) The memory capacity η_{max} as a function of the coding sparseness s . Black line: Instability to Turing Bifurcation, Eq 5.14. Red line: SNR calculation, Eq 4.23. Blue line: Numerical estimation of the saddle-node of bump attractors from the ring model. Squares: Capacity of full network. Inset: The fractional difference between the memory capacity of the full network and that of the ring, $\Delta \eta_{max} = (\eta_{max}^{network} - \eta_{max}^{ring}) / \eta_{max}^{ring}$. The network outperforms the ring model in the sparse-coding limit. (B) The memory capacity as a function of the number of place cells encoding each position, M .

Variability plays a crucial role in enhancing memory capacity. For one, interference from other memories can amplify Turing instability in the context of the ring model formulation. Secondly, additional fluctuations in firing rates in the full network lead to a memory capacity surpassing the ring model approximation.

Chapter 6

Maintenance of uniform representation in CA3 through BTSP

While place cells, and hence spatial maps, are present in both areas CA1 and CA3 regions of the hippocampus, there is a crucial qualitative difference between them. Maps in CA1 are strongly modulated by sensory cues, including rewarded locations, which are overrepresented compared to non-rewarded locations [104, 105]. Maps in CA3, on the other hand, are largely homogeneous [93] and unmodulated by reward [104]. Given that salient sensory cues and reward lead to modulation in neuronal firing rates, a purely Hebbian plasticity rule would be expected to lead to modulated spatial maps, as seen in CA1. It, therefore, remains unclear what physiological mechanism conspires to maintain homogeneous maps in CA3.

We propose that BTSP is ideally suited for this task by virtue of the fact that modulation of the frequency of plateau potentials (PPs) can be used to compensate for fluctuations in neuronal activity levels. In short, when there is a spike in neuronal activity which would lead to an overrepresentation of place cells given a purely Hebbian rule, a reduction in PP frequency can nonetheless maintain constant rates of plasticity. More generally, even in the absence of specific cues or rewards, neuronal activity levels naturally fluctuate; maintaining homogeneous representations is, therefore, a very general computational challenge. In this section, we will study how BTSP can maintain such homogeneous representations in the face of fluctuating activity levels.

6.1 Fluctuating coding levels in learning n memories

We adopt the learning paradigm outlined in the introductory chapter to investigate this hypothesis. More precisely, we face the task of learning a sequence of n memories in a network of N neurons by a set of patterns $\xi^\mu = \{\xi_1^\mu, \xi_2^\mu, \dots, \xi_N^\mu\}$ labeled by an index $\mu = 1, \dots, n$. Once the network has reached a stationary state, for $n_0 < n$, the memories learned over this period should have similar lifetimes, as the memory traces are expected to decay at comparable rates. If this is not the case, it would suggest that certain memories are retained more robustly than others within the network, a situation we do not expect.

Ideally, when the coding level remains constant, the only source of stochasticity arises from the random selection of the patterns representing the memories, assuming all memories are random and uncorrelated. This randomness introduces variability in the memory traces. When all of the memories after n_0 have been stacked together, we can measure the variability of the memory trace relative to the mean, and this variation should decrease over the lifetime t . However, if the coding level fluctuates over time, it would increase significantly relative to the constant level. Nevertheless, when introducing the BTSP into the system, adjustment of plateau potential must compensate for such fluctuations in the coding level, improving the variability.

6.1.1 Model for BTSP

In the previous chapters, the 1D plasticity map for BTSP was developed to accumulate synaptic plasticity, allowing for passage along a circuit track. The model may readily be generalized by discretizing time in steps of Δt and allowing all combinations of pre- and postsynaptic activity or plateau potentials at all time lags. However, one critical problem arises from this generalization: the synaptic weight at time $t + 1$ depends on all previous time steps $\{t, t - 1, t - 2, \dots\}$ within the time window of BTSP. As a consequence, some theoretical statistics would be hard to derive. To treat this, we use the framework proposed by Briguglio and Romani in Li et al. [93], simplifying the generalization to a single time step.

In a network of N neurons with binary synapses, neurons receive two distinct inputs: one from external activity-driven signal - let us denote it as A (for example,

due to rewards) - and another from the plateau potentials, denoted as PP . These inputs, however, serve different roles in the context of BTSP. A plateau potential must occur within the postsynaptic neuron to trigger BTSP. On the other hand, presynaptic spike or activity may be generated either by external signals or by plateau potentials; from the perspective of the postsynaptic neuron, there is no functional distinction between a spike that is driven by an external signal and one that is driven by a complex plateau event. To lighten the notation, let us define q as the presynaptic activity, encompassing both A and PP , and p as postsynaptic activity, represented only by PP . Both q and p are Bernoulli variables, with probability of success $f_q = P(A = 1) + P(PP = 1) - P(A = 1 \ \& \ PP = 1)$ and $f_p = P(PP = 1)$, respectively. The coding levels for the activities and the frequency of plateau potentials are defined by $P_A = P(A = 1)$ and $P_{PP} = P(PP = 1)$, respectively.

The synaptic weight at time $t + 1$ is described by the following equation:

$$\begin{aligned} W(t+1) &= W(t) + (1 - W(t))q(t)p(t) \\ &- W(t)(q(t)p(t-1) + q(t-1)p(t) - q(t)q(t-1)p(t)p(t-1)). \end{aligned} \quad (6.1)$$

In the depotentiated state, when $W(t) = 0$, potentiation occurs only when the pre and postsynaptic activities coincide at time t , i.e., $q(t) = p(t) = 1$. When the synapse is at potentiated state $W(t) = 1$, depotentiation occurs in case of timing mismatch, with either the presynaptic activity preceding or following the postsynaptic one.

This process is non-Markovian as that depends on both t and $t - 1$. However, it can be made Markovian by introducing two auxiliary variables to store the earlier memories: $\alpha(t) = W(t)q(t-1)$ and $\beta(t) = W(t)p(t-1)$. The system then becomes:

$$\begin{cases} W(t+1) = W(t) + (1 - W(t-1))qp - W(t)(\beta q + \alpha p - \alpha\beta qp), \\ \alpha(t+1) = W(t)q + (1 - W(t-1))qp - W(t)(\beta q + \alpha qp - \alpha\beta qp), \\ \beta(t+1) = W(t)p + (1 - W(t-1))qp - W(t)(\beta qp + \alpha p - \alpha\beta qp). \end{cases} \quad (6.2)$$

For ease of notation, we suppress the dependence on t for q , p , α , and β on the right-hand side of equations. This system yields five possible states:

$$(W, \alpha, \beta) \in \{(0, 0, 0), (1, 0, 0), (1, 1, 0), (1, 0, 1), (1, 1, 1)\},$$

with the following transition probability matrix between these states:

$$M = \begin{pmatrix} 1 - f_q f_p & 0 & f_p & f_q & 1 - (1 - f_q)(1 - f_p) \\ 0 & (1 - f_q)(1 - f_p) & (1 - f_q)(1 - f_p) & (1 - f_q)(1 - f_p) & (1 - f_q)(1 - f_p) \\ 0 & f_q(1 - f_p) & f_q(1 - f_p) & 0 & 0 \\ 0 & (1 - f_q)f_p & 0 & (1 - f_q)f_p & 0 \\ f_q f_p & f_q f_p & 0 & 0 & 0 \end{pmatrix},$$

where for each column, the sum is 1.

This transition matrix allows us now to calculate the desired memory traces for any time: we consider, as already mentioned, all the memories we want to track after the system reaches a stationary state. Here, the stationary probability vector given by $\vec{\pi} = \{\pi_i\}_{i=1:5}$, which is the left eigenvector corresponding to the eigenvalue $\lambda_1 = 1$. We are interested in following the assembly of neurons that participate in the plateau potential, so we take $f_q = f_p = 1$. The initial distribution of the memory trace will therefore be $\vec{x}(0) = M_{|f_q=f_p=1} \vec{\pi} = \{\pi_3 + \pi_4 + \pi_5, 0, 0, 0, \pi_1 + \pi_2\}$. The mean strength of the memory trace is taken as the portion of synapses that are potentiated in the assembly, and it will be given by $\pi_1 + \pi_2$. Moreover, the long-term behavior of the memory traces is controlled by the decay rate $1 - \lambda_2$, where λ_2 is the second-leading eigenvalue, depending on the f_A and f_{PP} .

To compensate for fluctuations in A , we adjust the frequency of PP so that the second-leading eigenvalue is kept constant. The hypothesis proposed is tested by running simulations under the following conditions:

- When both the coding levels of A and frequency of PP are constant, denoted as $BSTP(A, PP)$;
- when the coding level of A fluctuates but the frequency of PP remains the same, referred to as $BTSP(A_{per}, PP)$;
- When the coding level of A fluctuates and is regulated by the frequency of PP , named to as $BTSP(A_{per}, PP_{per})$.

6.1.2 Hebbian learning

In Hebbian learning, co-activating two neurons at time t leads to synaptic potentiation. In contrast, a mismatch in the activity at time t leads to synaptic depotentiation

with a probability dependent on the coding level. Mathematically, this is described by:

$$W(t+1) = W(t) + (1 - W(t))q(t)p(t) - cW(t)((1 - q(t))p(t) + q(t)(1 - p(t))), \quad (6.3)$$

where c is the depression rate.

In the simulation, we apply variable plateau potentials as an external source of activity to maintain the consistent assembly size under both BTSP and Hebbian learning. We run two simulations: one in which the frequency of plateau potential was fixed, so-called *Heb(PP)* model, and another one where the frequency of this potential was allowed to vary - *Heb(PP_{per})*. Since the Hebbian model contains only one variable, we consider an additional condition of BTSP in the absence of external activity A with fluctuating PP , we call it *BTSP(PP_{per})*.

6.2 Results and discussion

In the simulation, we set both f_A and f_{PP} to 0.1. To introduce perturbations in the coding level of activity, we utilize the beta distribution, $\text{Beta}(\alpha, \beta)$, due to its flexibility in shaping the probability distribution. Specifically, we select three beta distributions by change the parameter β : $\text{Beta}(1.5, 1.5)$, $\text{Beta}(1.5, 4)$, and $\text{Beta}(1.5, 15)$. To ensure that the coding level remains between 0 and 1, and is centered around the desired value, we rescale the generated values by dividing by the mean, $\alpha/(\alpha + \beta)$, and then multiply by f_A . This results in the following three distributions for the coding level, we name these conditions as low, middle, and high variance, see the top row in Fig 6.1. Using the second-leading eigenvalue of the transition matrix M , we obtain numerically the corresponding value of f_{PP} , see the bottom row in Fig 6.1.

We initialized the synaptic weights randomly between 0 and 1 and allowed the system to run for 1000 iterations to stabilize the weight matrix. Once the system reached a steady state, we tracked the storage of 1000 additional memories throughout 150 iterations for each model discussed in the previous section under each condition. The results, shown in Fig 6.2, reveal some interesting patterns. Notably, the initial strength of the averaged memory traces varies, particularly for the Hebbian model, as the tracked assembly consists of neurons receiving external stimuli. Consequently,

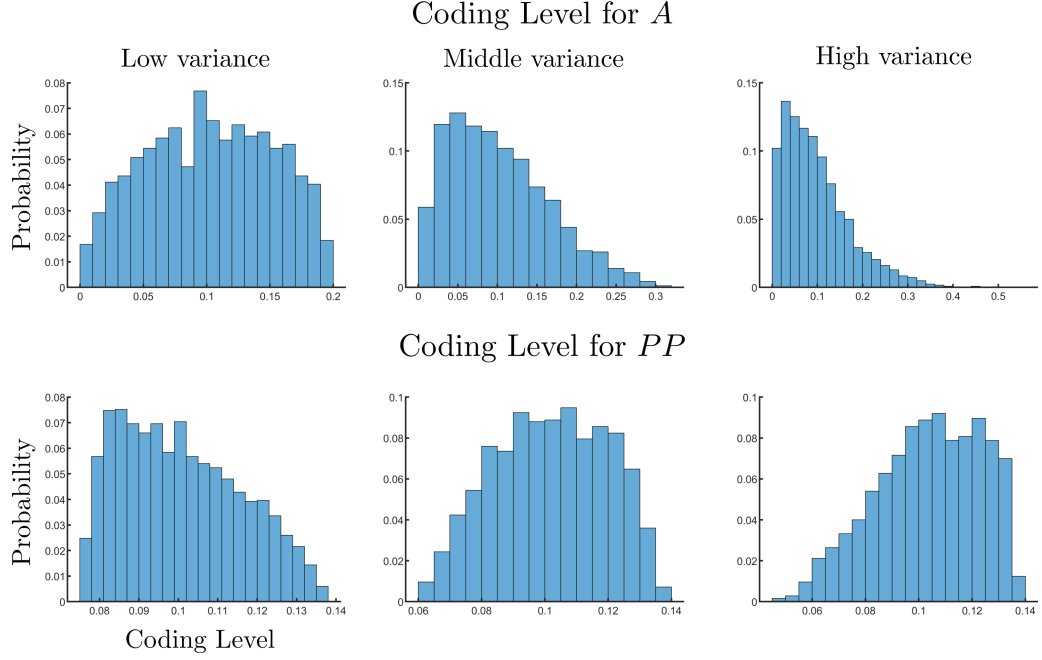


Figure 6.1: **Coding levels of external signals and frequencies of plateau potentials.**

the synapses between them undergo potentiation, and the plot begins immediately after the induction phase.

The overall decay rates across models appear similar at first glance, yet the variance reveals a more profound complexity, as illustrated in Fig 6.2B. In the Hebbian model, variance spikes sharply following induction before gradually returning to baseline levels. In contrast, the BTSP model, depicted in Fig 6.2C, demonstrates a steady decline in variance post-induction, indicating a more controlled and stable synaptic adaptation process.

Furthermore, as the coding level f_A fluctuates over time, the average memory strength remains relatively similar, but the variance undergoes a notable increase depending on the conditions. In cases where the frequency of plateau potential f_{PP} is kept constant (dashed line in Fig 6.2C), the variance rises by nearly 1.5 times, indicating a significant decrease in memory trace stability compared to the optimal scenario where both f_A and f_{PP} are maintained at constant levels. When the plateau potential is adjusted to account for fluctuations in f_A (denoted by the * symbol in Fig 6.2C), the variance is considerably reduced, approaching levels seen in the optimal case. This effect becomes even more pronounced as the variance in f_A increases, high-

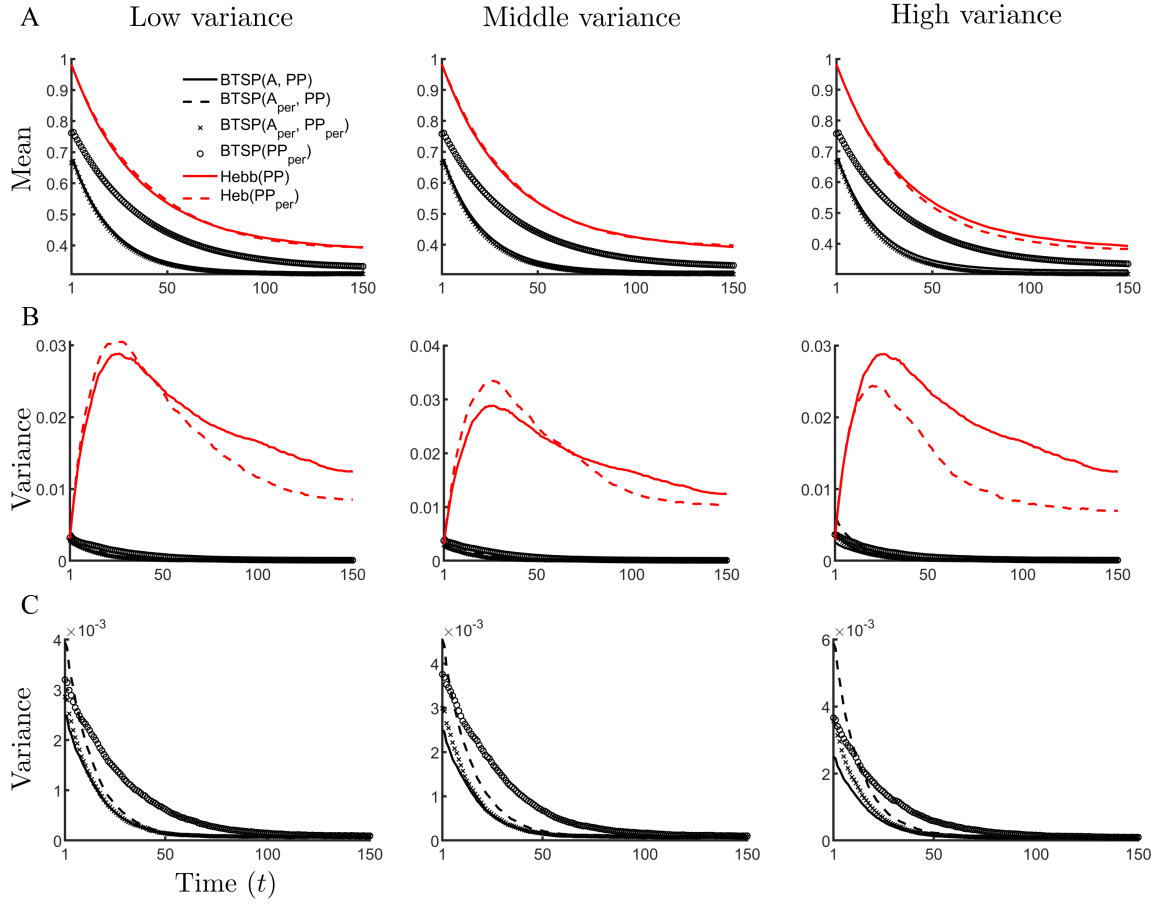


Figure 6.2: **Mean and variance of memory traces.** (A) The average memory trace. (B) The variance of memory traces around the mean trace. (C) The variance excluded those for the Hebbian learning to compare the BTSP. **Note:** Columns correspond to different conditions.

lighting that modulating the plateau potential effectively counters the destabilizing impact of fluctuating activity and preserves the stability of memory traces.

Next, we explore the intricate relationship between the coding level of A and the frequency of PP , as illustrated in Fig 6.3. We initialized f_A and f_{PP} at equal values, then systematically varied f_A to observe the effects. Using the definition $\lambda_2 = \lambda_2(f_A, f_{PP})$, we computed the corresponding value of f_{PP} , revealing the dynamic interplay between these two parameters. Each line in Fig 6.3 represents a distinct scenario, with the initial value indicated by a star symbol. The results clearly demonstrate that as the coding level of the external signal increases, the frequency of the plateau potential decreases, compensating for this change and maintaining balance within the system.

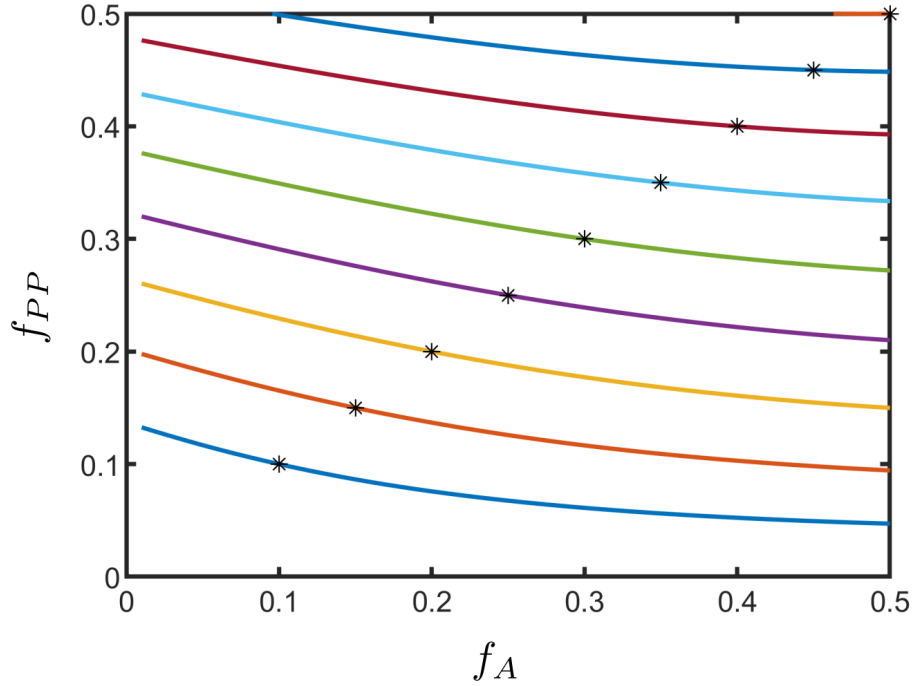


Figure 6.3: **The relationship between f_A and f_{PP} .**

The inverse relationship between f_A and f_{PP} presents a compelling mechanism within hippocampal dynamics. In the CA3 region, neurons primarily receive input from entorhinal cortex layer III (EC3), where plateau potentials are generated. Notably, optogenetic studies [78,79] have shown that inhibition of EC3 can significantly reduce the probability of plateau potentials. This suggests the existence of a regulatory circuit involving CA3-somatostatin (SOM) interneurons-EC3-CA3. In this

model, external inputs increase the firing rate of CA3 pyramidal neurons, which in turn activate SOM interneurons. Once active, SOM interneurons inhibit EC3, the primary source of plateau potentials. As this inhibition strengthens, the likelihood of plateau potential generation in EC3 diminishes, reducing f_{PP} . This f_{PP} decline stabilizes CA3 excitatory activity, preventing overexcitation and ensuring a balanced network. Thus, this feedback loop dynamically modulates excitatory-inhibitory interactions, maintaining the functional equilibrium of the hippocampal circuit.

In conclusion, our simulations demonstrate a robust feedback mechanism in the hippocampus, where fluctuations in coding level f_A inversely impact the frequency of plateau potentials f_{PP} . This dynamic balance ensures stable excitatory activity in CA3, as increased external input is compensated by reduced f_{PP} . Modulating plateau potentials in response to coding fluctuations effectively stabilizes memory traces, with the BTSP model showing more controlled synaptic adaptation than Hebbian learning, especially in high-variance conditions. This mechanism highlights the hippocampus's ability to maintain functional equilibrium under variable neural activity.

Chapter 7

Conclusion

Episodic memory, a system for recalling past experiences based on their temporal and spatial contexts, necessitates rapid or one-shot learning, frequently occurring in everyday life. Hebbian plasticity theory, which relies on repetitive stimuli, falls short of explaining the formation of episodic memories. However, a novel plasticity mechanism discovered in hippocampal regions demonstrates the ability to shape place field activity with just a few or even a single induction, offering a plausible candidate for episodic memory formation. Computational models have successfully replicated these phenomena observed *in vivo*, providing an initial step toward understanding behavioral timescale synaptic plasticity (BTSP).

In this study, we demonstrate that the biophysical model of BTSP [86], which replicates *in-vivo* plasticity protocols in the CA1 region of awake, behaving mice [75], can be effectively reduced to a one-dimensional (1D) map. The map, which describes synaptic weight updates following a plateau-potential-driven plasticity event, quantitatively fits the biophysical model. This reduction to a 1D map offers a powerful mathematical tool for analytically calculating the statistics of the learning process.

Leveraging the 1D map with a symmetric BTSP rule, we investigated plasticity in a recurrent network, akin to the CA3 region of the hippocampus, as an animal explores various environments. For simplicity, we assumed statistically equivalent environments, each with a ring topology. Although this assumption is somewhat unrealistic, it facilitates a straightforward characterization of the network’s memory capacity. It is reasonable to presume that storing memories of environments with different topologies would result in a similar scaling of memory capacity with system parameters. The key determinant of memory capacity was the sparseness of neuronal

coding, i.e., the fraction of place cells active in a given environment. For sparse coding, the dynamics of memory recall in the network can be approximated by its projection onto the low-dimensional manifold corresponding to the desired environment. In our network model, this approximation holds when $s \leq 0.2$. Recent large-scale in-vivo calcium imaging studies of the hippocampus indicate that the fraction of active place cells in any given environment is about 5% [106], fitting well within our model’s sparse coding regime.

In the sparse coding regime, interference between memories manifests as quenched variability in the connectivity matrix, uncorrelated with the analyzed environment. Surprisingly, this variability enhances memory capacity by promoting the formation of bump attractors in parameter regions that would not otherwise support them. This phenomenon, where quenched variability increases the robustness of intrinsically generated patterns and expands the hysteresis regime near a bifurcation, has been noted in spatially extended systems [107].

In this regime, we can approximate the memory capacity of the recurrent network by analyzing the dynamics projected onto the low-dimensional manifolds corresponding to past environments. This approach equates to a series of ring models, each with a connectivity profile analytically derived from the 1D map. We calculated memory capacity by determining the age of the environment where the recurrent connectivity’s spatial modulation could generate a bump attractor, standing at a Turing pattern bifurcation. This calculation considered the effect of quenched variability, which shifted the bifurcation to lower spatial modulation values and older memories. This calculation provided a lower bound on capacity, as a region of multi-stability typically existed below the bifurcation, where large amplitude bump attractors co-existed with the unpatterned state. Thus, the corresponding ring-model projections predicted the recurrent network dynamics well as long as the coding was sufficiently sparse. This approximation failed when $s \rightarrow 1$, and mixed attractor states formed, partially correlated with multiple environments.

Our analysis suggests that BTSP, with a temporally symmetric plasticity window, is well-suited for one-shot encoding of memories as steady-state attractors in recurrent networks. Such a symmetric form of BTSP has recently been found in CA3 recordings [93], making it a plausible mechanism for episodic memory formation, a function

attributed to the hippocampus. For simplicity, we considered the encoding of spatial memory for distinct environments. The global remapping of place cells ensures largely uncorrelated representations between environments. We also assumed that the time between encoding memories is much longer than the plasticity timescale, suitable for exploring distinct environments. Recent theoretical work suggests BTSP may be particularly efficient in storing correlated patterns [93]. It remains to be determined to what extent BTSP generally underlies episodic memory formation.

Furthermore, our exploration into the role of BTSP in maintaining homogeneous spatial representations in the CA3 region of the hippocampus has unveiled some intriguing findings. Unlike the modulation of firing rates and spatial maps in CA1 by sensory cues and reward signals, the CA3 network preserves its homogeneity despite these influences. Our analysis and simulations have revealed that BTSP is a key player in this process, compensating for fluctuations by dynamically regulating the frequency of plateau potentials. Unlike Hebbian plasticity, BTSP consistently sustains uniform spatial maps despite variable activity levels. This dynamic balance between external input fluctuations and plateau potential adjustment provides a compelling mechanism for CA3 network stability, supporting consistent memory representation. These findings, which shed new light on the role of BTSP, emphasize its critical role in promoting resilient and adaptable hippocampal function, extending its importance beyond episodic memory.

Bibliography

- [1] Nabil Bouizegarene and Frederick L Philippe. Episodic memories as building blocks of identity processing styles and life domains satisfaction: Examining need satisfaction and need for cognitive closure in memories. *Memory*, 24(5):616–628, 2016.
- [2] Endel Tulving et al. Episodic and semantic memory. *Organization of memory*, 1(381-403):1, 1972.
- [3] Endel Tulving. Episodic memory: From mind to brain. *Annual review of psychology*, 53(1):1–25, 2002.
- [4] Endel Tulving. *Elements of episodic memory*. Oxford University Press, 1983.
- [5] Claudia Fugazza, Ákos Pogány, and Ádám Miklósi. Recall of others’ actions after incidental encoding reveals episodic-like memory in dogs. *Current Biology*, 26(23):3209–3213, 2016.
- [6] Nicola S Clayton and Anthony Dickinson. Episodic-like memory during cache recovery by scrub jays. *Nature*, 395(6699):272–274, 1998.
- [7] Jonathon D Crystal. Evaluating evidence from animal models of episodic memory. *Journal of Experimental Psychology: Animal Learning and Cognition*, 47(3):337, 2021.
- [8] William Beecher Scoville and Brenda Milner. Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, 20(1):11, 1957.
- [9] Larry R Squire and Stuart Zola-Morgan. The medial temporal lobe memory system. *Science*, 253(5026):1380–1386, 1991.

- [10] Hugo J Spiers, Eleanor A Maguire, and Neil Burgess. Hippocampal amnesia. *Neurocase*, 7(5):357–382, 2001.
- [11] Eleanor A Maguire. Hippocampal involvement in human topographical memory: evidence from functional imaging. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1360):1475–1480, 1997.
- [12] Laura L Eldridge, Barbara J Knowlton, Christopher S Furmanski, Susan Y Bookheimer, and Stephen A Engel. Remembering episodes: a selective role for the hippocampus during retrieval. *Nature neuroscience*, 3(11):1149–1152, 2000.
- [13] Steven J Middleton and Thomas J McHugh. Ca2: A highly connected intrahippocampal relay. *Annual Review of Neuroscience*, 43(1):55–72, 2020.
- [14] David Amaral and Pierre Lavenex. Hippocampal neuroanatomy. In *The Hippocampus Book*. Oxford University Press, 2007.
- [15] Timothy A Allen and Norbert J Fortin. The evolution of episodic memory. *Proceedings of the National Academy of Sciences*, 110(supplement_2):10379–10386, 2013.
- [16] Sachin S Deshmukh and James J Knierim. Hippocampus. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(2):231–251, 2012.
- [17] MR Bennett, WG Gibson, and J Robinson. Dynamics of the ca3 pyramidal neuron autoassociative memory network in the hippocampus. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 343(1304):167–187, 1994.
- [18] Misha Tsodyks and Terrence Sejnowski. Associative memory and hippocampal place cells. *International journal of neural systems*, 6:81–86, 1995.
- [19] Menno P Witter. Connectivity of the hippocampus. In *The Hippocampal microcircuits: a computational modeler’s resource book*. Springer, 2018.
- [20] John O’Keefe and Jonathan Dostrovsky. The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat. *Brain research*, 1971.

- [21] John O’Keefe and A 1987 Speakman. Single unit activity in the rat hippocampus during a spatial memory task. *Experimental brain research*, 68:1–27, 1987.
- [22] Gregory J Quirk, Robert U Muller, and John L Kubie. The firing of hippocampal place cells in the dark depends on the rat’s recent experience. *Journal of Neuroscience*, 10(6):2008–2017, 1990.
- [23] John O’keefe and Lynn Nadel. *The hippocampus as a cognitive map*. Oxford university press, 1978.
- [24] Emma R Wood, Paul A Dudchenko, R Jonathan Robitsek, and Howard Eichenbaum. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron*, 27(3):623–633, 2000.
- [25] Loren M Frank, Emery N Brown, and Matthew Wilson. Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron*, 27(1):169–178, 2000.
- [26] Howard Eichenbaum. Time cells in the hippocampus: a new dimension for mapping memories. *Nature Reviews Neuroscience*, 15(11):732–744, 2014.
- [27] Sheena A Josselyn, Stefan Köhler, and Paul W Frankland. Finding the engram. *Nature Reviews Neuroscience*, 16(9):521–534, 2015.
- [28] Sheena A Josselyn, Stefan Köhler, and Paul W Frankland. Heroes of the engram. *Journal of Neuroscience*, 37(18):4647–4657, 2017.
- [29] Sheena A Josselyn and Susumu Tonegawa. Memory engrams: Recalling the past and imagining the future. *Science*, 367(6473):eaaw4325, 2020.
- [30] Donald O. Hebb. *The Organization of Behavior: A Neuropsychological Theory*. Wiley, 1949.
- [31] Vincent Castellucci, Harold Pinsker, Irving Kupfermann, and Eric R Kandel. Neuronal mechanisms of habituation and dishabituation of the gill-withdrawal reflex in aplysia. *Science*, 167(3926):1745–1748, 1970.
- [32] Vincent F Castellucci and Eric R Kandel. A quantal analysis of the synaptic depression underlying habituation of the gill-withdrawal reflex in aplysia. *Proceedings of the National Academy of Sciences*, 71(12):5004–5008, 1974.

- [33] Tim VP Bliss and Terje Lømo. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of physiology*, 232(2):331–356, 1973.
- [34] Bruce L McNaughton, RM Douglas, and Go Vo Goddard. Synaptic enhancement in fascia dentata: cooperativity among coactive afferents. *Brain research*, 157(2):277–293, 1978.
- [35] Tim VP Bliss and Graham L Collingridge. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature*, 361(6407):31–39, 1993.
- [36] Robert C Malenka and Mark F Bear. Ltp and ltd: an embarrassment of riches. *Neuron*, 44(1):5–21, 2004.
- [37] Ami Citri and Robert C Malenka. Synaptic plasticity: multiple forms, functions, and mechanisms. *Neuropsychopharmacology*, 33(1):18–41, 2008.
- [38] H Wigström, B Gustafsson, Y-Y Huang, and WC Abraham. Hippocampal long-term potentiation is induced by pairing single afferent volleys with intracellular[^] injected depolarizing current pulses. *Acta Physiologica Scandinavica*, 126(2):317–319, 1986.
- [39] Richard GM Morris, Elizabeth Anderson, Gary S Lynch, and Michel Baudry. Selective impairment of learning and blockade of long-term potentiation by an n-methyl-d-aspartate receptor antagonist, ap5. *Nature*, 319(6056):774–776, 1986.
- [40] Edvard I Moser, Kurt A Krobert, May-Britt Moser, and Richard GM Morris. Impaired spatial learning after saturation of long-term potentiation. *Science*, 281(5385):2038–2042, 1998.
- [41] Stephen J Martin, Paul D Grimwood, and Richard GM Morris. Synaptic plasticity and memory: an evaluation of the hypothesis. *Annual review of neuroscience*, 23(1):649–711, 2000.

- [42] Tomonori Takeuchi, Adrian J Duzskiewicz, and Richard GM Morris. The synaptic plasticity and memory hypothesis: encoding, storage and persistence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1633):20130288, 2014.
- [43] G Barrionuevo, F Schottler, and G Lynch. The effects of repetitive low frequency stimulation on control and “potentiated” synaptic responses in the hippocampus. *Life sciences*, 27(24):2385–2391, 1980.
- [44] Satoshi Fujii, Kazuo Saito, Hiroyoshi Miyakawa, Ken-ichi Ito, and Hiroshi Kato. Reversal of long-term potentiation (depotentiation) induced by tetanus stimulation of the input to ca1 neurons of guinea pig hippocampal slices. *Brain research*, 555(1):112–122, 1991.
- [45] Ursula Staubli and Gary Lynch. Stable depression of potentiated synaptic responses in the hippocampus with 1–5 hz stimulation. *Brain research*, 513(1):113–118, 1990.
- [46] WB Levy and O Steward. Temporal contiguity requirements for long-term associative potentiation/depression in the hippocampus. *Neuroscience*, 8(4):791–797, 1983.
- [47] Guo-qiang Bi and Mu-ming Poo. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of neuroscience*, 18(24):10464–10472, 1998.
- [48] Henry Markram, Joachim Lübke, Michael Frotscher, and Bert Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275(5297):213–215, 1997.
- [49] Larry F Abbott and Sacha B Nelson. Synaptic plasticity: taming the beast. *Nature neuroscience*, 3(11):1178–1183, 2000.
- [50] Natalia Caporale and Yang Dan. Spike timing–dependent plasticity: a hebbian learning rule. *Annu. Rev. Neurosci.*, 31(1):25–46, 2008.

- [51] Glenn E Schafe, Valérie Doyère, and Joseph E LeDoux. Tracking the fear engram: the lateral amygdala is an essential locus of fear memory storage. *Journal of Neuroscience*, 25(43):10010–10014, 2005.
- [52] Jin-Hee Han, Steven A Kushner, Adelaide P Yiu, Christy J Cole, Anna Matyenia, Robert A Brown, Rachael L Neve, John F Guzowski, Alcino J Silva, and Sheena A Josselyn. Neuronal competition and selection during memory formation. *science*, 316(5823):457–460, 2007.
- [53] Jin-Hee Han, Steven A Kushner, Adelaide P Yiu, Hwa-Lin Hsiang, Thorsten Buch, Ari Waisman, Bruno Bontempi, Rachael L Neve, Paul W Frankland, and Sheena A Josselyn. Selective erasure of a fear memory. *Science*, 323(5920):1492–1496, 2009.
- [54] Xu Liu, Steve Ramirez, Petti T Pang, Corey B Puryear, Arvind Govindarajan, Karl Deisseroth, and Susumu Tonegawa. Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature*, 484(7394):381–385, 2012.
- [55] David Marr. *Simple memory: a theory for archicortex*. Springer, 1991.
- [56] Wulfram Gerstner and Werner M Kistler. Mathematical formulations of hebbian learning. *Biological cybernetics*, 87(5):404–415, 2002.
- [57] Peter Dayan and Laurence F Abbott. *Theoretical neuroscience: computational and mathematical modeling of neural systems*. MIT press, 2005.
- [58] Wulfram Gerstner, Werner M Kistler, Richard Naud, and Liam Paninski. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.
- [59] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.
- [60] Stefano Fusi. *Oxford Handbook of Human Memory. vol. 1. 1st ed.*, chapter Memory capacity of neural network models. Oxford University Press, 2021.

- [61] Samuel Frederick Edwards and Phil W Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965, 1975.
- [62] VJ Emery. Critical properties of many-component systems. *Physical Review B*, 11(1):239, 1975.
- [63] Daniel J Amit, Hanoch Gutfreund, and Haim Sompolinsky. Spin-glass models of neural networks. *Physical Review A*, 32(2):1007, 1985.
- [64] Daniel J Amit, Hanoch Gutfreund, and Haim Sompolinsky. Information storage in neural networks with low levels of activity. *Physical Review A*, 35(5):2293, 1987.
- [65] Mikhail V Tsodyks and Mikhail V Feigel'man. The enhanced storage capacity in neural networks with low activity level. *Europhysics Letters*, 6(2):101, 1988.
- [66] Daniel J Amit. Storage capacity of ann's. In *Modeling brain function: The world of attractor neural networks*. Cambridge university press, 1989.
- [67] JP Nadal, G Toulouse, JP Changeux, and S Dehaene. Networks of formal neurons and memory palimpsests. *Europhysics letters*, 1(10):535, 1986.
- [68] Daniel J Amit and Stefano Fusi. Constraints on learning in dynamic synapses. *Network: Computation in Neural Systems*, 3(4):443–464, 1992.
- [69] Daniel J Amit and Stefano Fusi. Learning in neural networks with material synapses. *Neural Computation*, 6(5):957–982, 1994.
- [70] John T Wixted and Ebbe B Ebbesen. On the form of forgetting. *Psychological science*, 2(6):409–415, 1991.
- [71] John T Wixted and Ebbe B Ebbesen. Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & cognition*, 25:731–739, 1997.
- [72] Stefano Fusi, Patrick J Drew, and Larry F Abbott. Cascade models of synaptically stored memories. *Neuron*, 45(4):599–611, 2005.

- [73] Alex Roxin and Stefano Fusi. Efficient partitioning of memory systems and its importance for memory consolidation. *PLoS computational biology*, 9(7):e1003146, 2013.
- [74] Pan Ye Li. Code repository for the mathematical modeling of behavioral timescale synaptic plasticity. *Github*, <https://github.com/PanYe87/BTSP-Based-Learning>.
- [75] Katie C Bittner, Aaron D Milstein, Christine Grienberger, Sandro Romani, and Jeffrey C Magee. Behavioral time scale synaptic plasticity underlies ca1 place fields. *Science*, 357(6355):1033–1036, 2017.
- [76] Christine Grienberger, Xiaowei Chen, and Arthur Konnerth. Nmda receptor-dependent multidendrite ca2+ spikes required for hippocampal burst firing in vivo. *Neuron*, 81(6):1274–1281, 2014.
- [77] Hiroto Takahashi and Jeffrey C Magee. Pathway interactions and synaptic plasticity in the dendritic tuft regions of ca1 pyramidal neurons. *Neuron*, 62(1):102–111, 2009.
- [78] Katie C Bittner, Christine Grienberger, Sachin P Vaidya, Aaron D Milstein, John J Macklin, Junghyup Suh, Susumu Tonegawa, and Jeffrey C Magee. Conjunctive input processing drives feature selectivity in hippocampal ca1 neurons. *Nature neuroscience*, 18(8):1133–1142, 2015.
- [79] Christine Grienberger and Jeffrey C Magee. Entorhinal cortex directs learning-related changes in ca1 representations. *Nature*, 611(7936):554–562, 2022.
- [80] Jeffrey C Magee and Christine Grienberger. Synaptic plasticity forms and functions. *Annual review of neuroscience*, 43(1):95–117, 2020.
- [81] James B Priestley, John C Bowler, Sebi V Rolotti, Stefano Fusi, and Attila Losonczy. Signatures of rapid plasticity in hippocampal ca1 representations during novel experiences. *Neuron*, 110(12):1978–1992, 2022.

- [82] Wulfram Gerstner, Marco Lehmann, Vasiliki Liakoni, Dane Corneil, and Johanni Brea. Eligibility traces and plasticity on behavioral time scales: experimental support of neohebbian three-factor learning rules. *Frontiers in neural circuits*, 12:53, 2018.
- [83] Linlin Z Fan, Doo Kyung Kim, Joshua H Jennings, He Tian, Peter Y Wang, Charu Ramakrishnan, Sawyer Randles, Yanjun Sun, Elina Thadhani, Yoon Seok Kim, et al. All-optical physiology resolves a synaptic basis for behavioral timescale plasticity. *Cell*, 186(3):543–559, 2023.
- [84] Kevin C Gonzalez, Adrian Negrean, Zhenrui Liao, Franck Polleux, and Attila Losonczy. Synaptic basis of behavioral timescale plasticity. *bioRxiv*, pages 2023–10, 2023.
- [85] Maria Diamantaki, Stefano Coletta, Khaled Nasr, Roxana Zeraati, Sophie Laturus, Philipp Berens, Patricia Preston-Ferrer, and Andrea Burgalossi. Manipulating hippocampal place cell activity by single-cell stimulation in freely moving mice. *Cell reports*, 23(1):32–38, 2018.
- [86] Aaron D Milstein, Yiding Li, Katie C Bittner, Christine Grienberger, Ivan Soltesz, Jeffrey C Magee, and Sandro Romani. Bidirectional synaptic plasticity rapidly modifies hippocampal representations. *Elife*, 10:e73046, 2021.
- [87] Mayank R Mehta, Carol A Barnes, and Bruce L McNaughton. Experience-dependent, asymmetric expansion of hippocampal place fields. *Proceedings of the National Academy of Sciences*, 94(16):8918–8921, 1997.
- [88] Inah Lee, Geeta Rao, and James J Knierim. A double dissociation between hippocampal subfields: differential time course of ca3 and ca1 place cells for processing changed environments. *Neuron*, 42(5):803–815, 2004.
- [89] Christine Grienberger, Aaron D Milstein, Katie C Bittner, Sandro Romani, and Jeffrey C Magee. Inhibitory suppression of heterogeneously tuned excitation enhances spatial coding in ca1 place cells. *Nature neuroscience*, 20(3):417–426, 2017.

- [90] Kuo Xiao, Yiding Li, Raymond A Chitwood, and Jeffrey C Magee. A critical role for camkii in behavioral timescale synaptic plasticity in hippocampal ca1 pyramidal neurons. *Science Advances*, 9(36):eadi3088, 2023.
- [91] Xinyu Zhao, Ching-Lung Hsu, and Nelson Spruston. Rapid synaptic plasticity contributes to a learned conjunctive code of position and choice-related information in the hippocampus. *Neuron*, 110(1):96–108, 2022.
- [92] Sachin P Vaidya, Raymond A Chitwood, and Jeffrey C Magee. The formation of an expanding memory representation in the hippocampus. *biorxiv*, pages 2023–02, 2023.
- [93] Yiding Li, John J Briguglio, Sandro Romani, and Jeffrey C Magee. Mechanisms of memory storage and retrieval in hippocampal area ca3. *biorxiv*, pages 2023–05, 2023.
- [94] Ian Cone and Harel Z Shouval. Behavioral time scale plasticity of place fields: mathematical analysis. *Frontiers in computational neuroscience*, 15:640235, 2021.
- [95] Yujie Wu and Wolfgang Maass. Memory structure created through behavioral time scale synaptic plasticity. *biorxiv*, pages 2023–04, 2023.
- [96] Adelchi Azzalini. *The skew-normal and related families*, volume 3. Cambridge University Press, 2013.
- [97] Aaron D Milstein. Code repository for the mathematical modeling of behavioral timescale synaptic plasticity. *Github*, <https://github.com/PanYe87/BTSP-Based-Learning>.
- [98] Francesco P Battaglia and Alessandro Treves. Attractor neural networks storing multiple space representations: a model for hippocampal place fields. *Physical Review E*, 58(6):7738, 1998.
- [99] Maxwell Gillett, Ulises Pereira, and Nicolas Brunel. Characteristics of sequential activity in networks with temporally asymmetric hebbian learning. *Proceedings of the National Academy of Sciences*, 117(47):29948–29958, 2020.

- [100] Davide Spalla, Isabel Maria Cornacchia, and Alessandro Treves. Continuous attractors for dynamic memories. *Elife*, 10:e69499, 2021.
- [101] Daniel D Ben Dayan Rubin and Stefano Fusi. Long memory lifetimes require complex synapses and limited sparseness. *Frontiers in computational neuroscience*, 1:117, 2007.
- [102] Stefano Fusi and LF Abbott. Limits on the memory storage capacity of bounded synapses. *Nature neuroscience*, 10(4):485–493, 2007.
- [103] Marcus K Benna and Stefano Fusi. Computational principles of synaptic memory consolidation. *Nature neuroscience*, 19(12):1697–1706, 2016.
- [104] David Dupret, Joseph O’neill, Barty Pleydell-Bouverie, and Jozsef Csicsvari. The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nature neuroscience*, 13(8):995–1002, 2010.
- [105] Stig A Hollup, Sturla Molden, James G Donnett, May-Britt Moser, and Edvard I Moser. Accumulation of hippocampal place fields at the goal location in an annular watermaze task. *Journal of Neuroscience*, 21(5):1635–1644, 2001.
- [106] Thomas Hainmueller and Marlene Bartos. Parallel emergence of stable and dynamic memory engrams in the hippocampus. *Nature*, 558(7709):292–296, 2018.
- [107] Hezi Yizhaq and Golan Bel. Effects of quenched disorder on critical transitions in pattern-forming systems. *New Journal of Physics*, 18(2):023004, 2016.

Appendix A

Detailed calculation of memory traces and quenched variability in the BTSP rule

We start with the 1D map for BTSP in a recurrent network with sparse coding, Eq 4.17. We rewrite the equation as

$$\begin{aligned} w_{ij}^n &= Pf_P(\Delta\theta_{ij}^n)S_i^n S_j^n + w_{ij}^{n-1}(1 - (Pf_p(\Delta\theta_{ij}^n) + Df_D(\Delta\theta_{ij}^n))S_i^n S_j^n), \\ &= Pf_P(\Delta\theta_{ij}^n)S_i^n S_j^n + w_{ij}^{n-1}F(\Delta\theta_{ij}^n), \end{aligned} \quad (\text{A.1})$$

where $S_i^n = 1$ if cell i is a place cell in environment n and otherwise is zero. We use Eq A.1 iteratively, to solve for w_{ij}^n in terms of the weights for environment $n - 2$, $n - 3$ and so on, until we arrive at a formulation in terms of the plasticity occurred in environment $n - \eta$. This is precisely Eq 4.19. In a slight abuse of notation, which however will make the calculation easier to follow, we will write $f_{P,\eta} = f_P(\Delta\theta_{ij}^{n-\eta})$, which indicates that the plasticity function for potentiation should be ordered in the environment with age η , i.e. environment $n - \eta$. Using similar notation (indicating the age and not the exact identity of the memory) for F , S and w we arrive at

$$w_{ij}^n = (Pf_{P,\eta}S_i^\eta S_j^\eta + w_{ij}^{\eta+1}F_\eta) \cdot \prod_{l=0}^{\eta-1} F_l + P \sum_{k=0}^{\eta-1} f_{P,k}S_i^k S_j^k \prod_{l=0}^{k-1} F_l. \quad (\text{A.2})$$

We recall that the mean and variance of the weights are written μ_w and σ_w^2 , and are given by Eqs 4.6 and 4.5. Using these two statistics and Eq A.2 we can calculate the correlation of the weight matrix with environment $n - \eta$. Specifically, we calculate the amplitude $a_\eta = 2\langle \cos(\Delta\theta_{ij}^{n-\eta}), w_{ij}^n \rangle$, where the brackets indicate an

average over all neurons. We note that in Eq A.2, the functions $f_{P,\eta}$ and F_η both have phases ordered in environment $n - \eta$, while all other terms have orderings which are random and uncorrelated with this space. We also note that we have defined $f_p(\theta) = 1 + \cos \theta$ and $f_D(\theta) = 1 - \cos \theta$, which means that $F(\theta) = 1 - [(P + D) - (D - P) \cos \theta] S_i S_j$. The averages of these functions over neurons are $\langle f_p \rangle = \langle f_D \rangle = 1$, and $\langle F \rangle = 1 - s^2(P + D)$, where s is the coding sparseness. We will also need the fact that $\langle F^2 \rangle = 1 + s^2 \left(\frac{3}{2}(P^2 + D^2) + PD - 2(P + D) \right)$. These last two relations come about due to the fact that $\langle S_i S_j \rangle = \langle S_i \rangle \langle S_j \rangle = s^2$, i.e. the expected value of S_i is one times the probability of a neuron being active s , and also $\langle (S_i)^2 \rangle = \langle S_i \rangle = s$.

Therefore

$$\begin{aligned}
a_\eta &= 2 \langle \cos(\Delta\theta_{ij}^{n-\eta}) \cdot (P f_{P,\eta} S_i^\eta S_j^\eta + w_{ij}^{\eta+1} F_\eta) \cdot \prod_{l=0}^{\eta-1} F_l \rangle, \\
&= \frac{1}{\pi} \int_{-\pi}^{\pi} (P(1 + \cos \theta) + \mu_w(1 - D - P + (D - P) \cos \theta)) \cos \theta d\theta \cdot \langle F \rangle^\eta, \\
&= \frac{2PD}{P + D} (1 - s^2(P + D))^\eta,
\end{aligned} \tag{A.3}$$

where we have used the fact that $\mu_w = P/(P + D)$ and $S_i^\eta = 1$ as we are considering only active place cells in that environment. The amplitude a_η is the amplitude of the first Fourier mode of the spatial connectivity in space η . For the simple plasticity functions considered here it provides a complete characterization of the spatial modulation. Now, the mean synaptic connectivity ordered in space η is

$$M_\eta = \mu_w + a_\eta \cos(\Delta\theta^{n-\eta}).$$

However, if we consider the connectivity profile for a specific post-synaptic cell i , it will deviate from this mean due to the quenched variability caused by global remapping.

The variance of the connectivity about this *ordered* mean is

$$\begin{aligned}
V_\eta &= \langle (w_{ij}^n - M_\eta)^2 \rangle \\
&= \langle (w_{ij}^n - a_\eta \cos(\Delta\theta^{n-\eta}))^2 \rangle - \mu_w^2 \\
&= A_\eta + B_\eta \cos(\Delta\theta^{n-\eta}) + C_\eta \cos^2(\Delta\theta^{n-\eta}),
\end{aligned} \tag{A.4}$$

where the brackets now indicate an average over neurons, but carried out for each $\Delta\theta^{n-\eta}$ between neuronal pairs, i.e. there is no averaging over $\Delta\theta^{n-\eta}$. Using equation

Eq A.2, we find that

$$\begin{aligned}
A_\eta &= \langle (P + w(1 - (P + D)))^2 \prod_{l=0}^{\eta-1} F_l^2 \rangle - \mu_w^2 \\
&\quad + 2P \langle (P + w(1 - (P + D))) \prod_{l=0}^{\eta-1} F_l \cdot \sum_{k=0}^{\eta-1} f_{p,k} S_i^k S_j^k \prod_{l=0}^{k-1} F_l \rangle \\
&\quad + P^2 \langle \sum_{k=0}^{\eta-1} f_{p,k} S_i^k S_j^k \prod_{l=0}^{k-1} F_l \sum_{m=0}^{\eta-1} f_{p,m} S_i^m S_j^m \prod_{r=0}^{m-1} F_r \rangle \\
&= \langle (P^2 + 2Pw(1 - P - D) + w^2(1 - P - D)^2) \prod_{l=0}^{\eta-1} \langle F_l^2 \rangle - \mu_w^2 \\
&\quad + 2P \langle (P + w(1 - P - D)) \sum_{k=0}^{\eta-1} \langle f_{p,k} F_k S_i^k S_j^k \rangle \prod_{l=0}^{k-1} \langle F_l^2 \rangle \prod_{r=k+1}^{\eta-1} \langle F_r \rangle \\
&\quad + P^2 \sum_{k=0}^{\eta-1} \langle (f_{p,k})^2 \rangle \langle (S_i^k)^2 \rangle \langle (S_j^k)^2 \rangle \prod_{l=0}^{k-1} \langle F_l^2 \rangle \\
&\quad + 2P^2 \sum_{l=1}^{\eta-1} \sum_{k=0}^{l-1} \langle f_{p,l} \rangle \langle f_{p,k} F_k \rangle \langle S_i^l \rangle \langle S_j^l \rangle \langle S_i^k \rangle \langle S_j^k \rangle \prod_{m=0}^{k-1} \langle F_m^2 \rangle \prod_{r=k+1}^{l-1} \langle F_r \rangle \\
&= (P^2 + 2P\mu_w(1 - P - D) + \langle w^2 \rangle (1 - P - D)^2) \langle F^2 \rangle^\eta \\
&\quad + 2P(P + \mu_w(1 - P - D)) \langle f_p F \rangle s^2 \cdot \frac{\langle F \rangle^\eta - \langle F^2 \rangle}{\langle F \rangle - \langle F^2 \rangle} \\
&\quad + P^2 s^2 \langle f_p^2 \rangle \frac{1 - \langle F^2 \rangle^\eta}{1 - \langle F^2 \rangle} \\
&\quad + 2P^2 s^4 \frac{\langle f_p F \rangle}{\langle F \rangle - \langle F^2 \rangle} \left(\frac{1 - \langle F \rangle^\eta}{1 - \langle F \rangle} - \frac{1 - \langle F^2 \rangle^\eta}{1 - \langle F^2 \rangle} \right), \tag{A.5}
\end{aligned}$$

where we have used the fact that $S_k = \sum_{i=0}^{k-1} z^i = \frac{1-z^k}{1-z}$.

$$\begin{aligned}
B_\eta &= 2\langle [(P + w(1 - P - D)) \prod_{l=0}^{\eta-1} F_l + P \sum_{k=0}^{\eta-1} f_{p,k} S_i^k S_j^k \prod_{l=0}^{k-1} F_l] \\
&\quad \cdot [(P - w(P - D)) \prod_{l=0}^{\eta-1} -a_0 \langle F \rangle^\eta] \rangle \\
&= 2\langle (P + w(1 - P - D)) (P - w(P - D)) \prod_{l=0}^{\eta-1} F_l^2 \rangle \\
&\quad + 2P \langle (P - w(P - D)) \sum_{k=0}^{\eta-1} f_{p,k} F_k S_i^k S_j^k \prod_{l=0}^{k-1} F_l^2 \prod_{j=k+1}^{\eta-1} F_j \rangle \\
&\quad - 2a_0 \langle F \rangle^\eta \left(\langle (P + w(1 - P - D)) \prod_{l=0}^{\eta-1} F_l \rangle + P \langle \sum_{k=0}^{\eta-1} f_{p,k} S_i^k S_j^k \prod_{l=0}^{k-1} F_l \rangle \right) \\
&= 2(P^2 + \mu_w P(1 - 2P) - \langle w^2 \rangle (P - D)(1 - P - D)) \prod_{l=0}^{\eta-1} \langle F \rangle \\
&\quad + 2a_0 s^2 P \sum_{k=0}^{\eta-1} \langle f_p F \rangle \prod_{l=0}^{k-1} \langle F^2 \rangle \prod_{j=k+1}^{\eta-1} \langle F \rangle \\
&\quad - 2a_0 \mu_w \langle F \rangle^\eta \prod_{l=0}^{\eta-1} \langle F \rangle - 2a_0 s^2 P \langle F \rangle^{\eta-1} \sum_{k=0}^{\eta-1} \prod_{l=0}^{k-1} \langle F \rangle \\
&= 2(P^2 + P\mu_w(1 - 2P) - \langle w^2 \rangle (P - D)(1 - P - D)) \langle F^2 \rangle - 2a_0 \mu_w \langle F \rangle^{2\eta} \\
&\quad + 2a_0 s^2 P \left(\langle f_p F \rangle \frac{\langle F \rangle^\eta - \langle F^2 \rangle^\eta}{\langle F \rangle - \langle F^2 \rangle} - \frac{\langle F \rangle^\eta - \langle F \rangle^{2\eta}}{1 - \langle F \rangle} \right) \\
&= 2 \frac{(P - D)}{(P + D)} (P^2 + P\mu_w(1 - 2P - 2D) - \langle w^2 \rangle (P + D)(1 - P - D)) \langle F^2 \rangle^\eta \\
&\quad + 2a_0 \mu_w (\langle F^2 \rangle^\eta - \langle F \rangle^{2\eta}) \\
&\quad + 2a_0 s^2 P \left(\langle f_p F \rangle \frac{\langle F \rangle^\eta - \langle F^2 \rangle^\eta}{\langle F \rangle - \langle F^2 \rangle} - \frac{\langle F \rangle^\eta - \langle F \rangle^{2\eta}}{1 - \langle F \rangle} \right), \tag{A.6}
\end{aligned}$$

where we have used the fact that $P - \mu_w(P - D) = 2PD/(P + D) = a_0$. Finally we

have

$$\begin{aligned}
C_\eta &= \langle ((P - w(P - D)) \prod_{l=0}^{\eta-1} F_l - a_0 \langle F \rangle^\eta)^2 \rangle \\
&= \langle (P^2 - 2wP(P - D) + w^2(P - D)^2 \prod_{l=0}^{\eta-1} F_l^2) \\
&\quad - 2a_0 \langle F \rangle^\eta \langle (P - w(P - D)) \prod_{l=0}^{\eta-1} F_l \rangle + a_0^2 \langle F \rangle^{2\eta} \rangle \\
&= \langle (P^2 - 2\mu_w P(P - D) + \langle w^2 \rangle (P - D)^2) \langle F^2 \rangle^\eta \\
&\quad - 2a_0 (P - \mu_w(P - D)) \langle F \rangle^{2\eta} + a_0^2 \langle F \rangle^{2\eta} \rangle \\
&= \left(\frac{P - D}{P + D} \right)^2 (P^2 - 2P\mu_w(P + D) + \langle w^2 \rangle (P + D)^2) \\
&\quad + a_0^2 (\langle F^2 \rangle^\eta - \langle F \rangle^{2\eta}). \tag{A.7}
\end{aligned}$$

Appendix B

Calculation of the Spatial Fourier Spectrum for Quenched White Noise

The quenched variability which arises through the global remapping of place cells when BTSP shapes recurrent weights, takes the form $\Delta W(\theta) = \sqrt{V(\theta)}z(\theta)$, where $V(\theta) = A + B \cos \theta + C \cos^2 \theta$ is the variance calculated in the Appendix A, and z is a zero-mean Gaussian random variable with unit variance. We understand θ here to be the phase difference between the centroid of the place fields of a pair of neurons. We write $\Delta W(\theta)$ in terms of the finite Fourier series

$$\Delta W(\theta) = 2 \sum_{j=1}^N \alpha_j \cos(j\theta) - 2 \sum_{j=1}^N \beta_j \sin(j\theta). \quad (\text{B.1})$$

The coefficients α_j and β_j are zero-mean Gaussian random variables with variances and covariances which must be determined self-consistently. Specifically, we average both sides over the variability at each position θ : averaging over neurons, or equivalently, averaging over the distribution of the α_j s and β_j s. Specifically we have $V(\theta) = \langle \Delta W(\theta)^2 \rangle$, where

$$\begin{aligned} V(\theta) &= 2 \sum_{j=1}^N \sum_{l=1}^N (\langle \alpha_j \alpha_l \rangle + \langle \beta_j \beta_l \rangle) \cos((j-l)\theta) \\ &\quad + 2 \sum_{j=1}^N \sum_{l=1}^N (\langle \alpha_j \alpha_l \rangle - \langle \beta_j \beta_l \rangle) \cos((j+l)\theta) \\ &\quad + 4 \sum_{j=1}^N \sum_{l=1}^N (\langle \alpha_j \beta_l \rangle \sin((j-l)\theta) - \langle \alpha_l \beta_j \rangle \sin((j+l)\theta)). \end{aligned} \quad (\text{B.2})$$

We then calculate the Fourier coefficients explicitly for the only non-zero modes. For this, we rewrite the variance as $V(\theta) = A + \frac{C}{2} + B \cos(\theta) + \frac{C}{2} \cos(2\theta)$. We now use Parseval's theorem, which roughly states that the signal's power is conserved in Fourier space, leading to the following relations

$$\frac{1}{2} \int_{-\pi}^{\pi} V(\theta) d\theta = A + \frac{C}{2} = \sum_{j=1}^N (\langle \alpha_j^2 \rangle + \langle \beta_j^2 \rangle), \quad (\text{B.3})$$

$$\frac{1}{2} \int_{-\pi}^{\pi} V(\theta) \cos \theta d\theta = B = \sum_{j=1}^{N-1} (\langle \alpha_j \alpha_{j+1} \rangle + \langle \beta_j \beta_{j+1} \rangle), \quad (\text{B.4})$$

$$\frac{1}{2} \int_{-\pi}^{\pi} V(\theta) \cos 2\theta d\theta = \frac{C}{2} = \sum_{j=1}^{N-2} (\langle \alpha_j \alpha_{j+2} \rangle + \langle \beta_j \beta_{j+2} \rangle) + \langle \alpha_1^2 \rangle - \langle \beta_1^2 \rangle. \quad (\text{B.5})$$

Eqs B.3-B.5 provide only three constraints for many unknowns of order N . However, we can use the fact that the power spectrum of a white noise process is flat, meaning that the power is spread evenly amongst all modes, which leads to Eqs 5.17-5.21. Note that the variance of the coefficients of the first mode is a special case because we have two constraints

$$\begin{aligned} \langle \alpha_1^2 \rangle + \langle \beta_1^2 \rangle &= \frac{1}{N} \left(A + \frac{C}{2} \right), \\ \langle \alpha_1^2 \rangle - \langle \beta_1^2 \rangle &= \frac{C}{4N}, \end{aligned}$$

where the second relation arises due to the spatial inhomogeneity of the quenched variability. A comparison of this theory with numerical simulation of a quenched, Gaussian white-noise process is shown in Figs B.1 and B.2.

Here we can see the role of N , the number of encoded positions around the ring. Although the power (variance) of the quenched variability is entirely independent of N , the power in each Fourier mode decreases as $1/N$. Therefore, in the vicinity of instability in a spatial modulation mode, the value of N will be important. In fact, the noiseless case is recovered in the limit as $N \rightarrow \infty$, even though the total power in the quenched variability is unaffected. On the other hand, for finite N , any spatial bifurcation will be shifted.

In the case of the Turing bifurcation, we can calculate this effect analytically. Specifically, the connectivity profile for a cell i is $W(\theta_i - \theta_j) = W_0 + W_1 \cos(\theta_i - \theta_j) + \sqrt{V(\theta_i - \theta_j)} z(\theta_i - \theta_j)$. It is important to note that the value of z is different for each

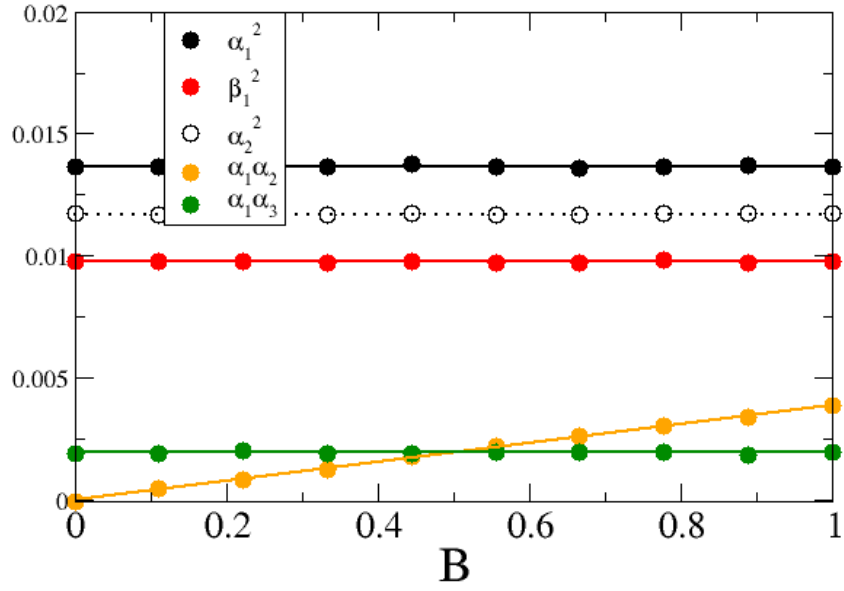


Figure B.1: Comparison of the theory (lines) with numerical simulations for different values of B . Parameters are: $A = 1$, $C = 1$, $N = 64$. Symbols are averages of 10,000 realizations of a quenched, Gaussian white noise process on a ring with variance V_θ .

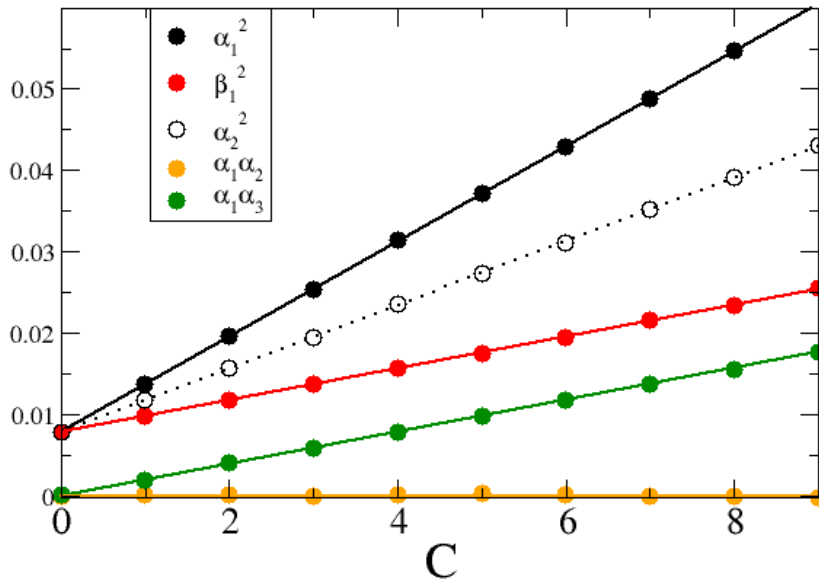


Figure B.2: Comparison of the theory (lines) with numerical simulations for different values of C . Parameters are: $A = 1$, $C = 0$, $N = 64$. Symbols are averages of 10,000 realizations of a quenched, Gaussian white noise process on a ring with variance V_θ .

neuronal pair (i, j) . To know how the bifurcation is shifted, We should extract the component of the variability commensurate with the cosine mode. This has the form $R \cos(\theta - \psi)$, where $R = 2\sqrt{\alpha^2 + \beta^2}$ and $\psi = \tan^{-1}(\beta/\alpha)$. Therefore, amplitude and phase R and ψ are also random variables. We can calculate the mean amplitude by integrating over the distributions of α and β

$$\begin{aligned}
\langle R \rangle &= \frac{1}{\pi\sigma^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} d\bar{x}d\bar{y} \sqrt{\bar{x}^2 + \bar{y}^2} e^{-\frac{\bar{x}^2}{2\sigma^2}} e^{-\frac{\bar{y}^2}{2\sigma^2}} \\
&= \frac{2\sqrt{2}\sigma}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} dxdy \sqrt{x^2 + y^2} e^{-x^2 - y^2} \\
&= 4\sqrt{2}\sigma \int_0^{\infty} dr \cdot r^2 e^{-r^2} \\
&= 4\sqrt{2}\sigma \left[\frac{\sqrt{\pi}}{4} \text{erf}(r) - \frac{1}{2} e^{-r^2} \right]_0^{\infty} \\
&= \sqrt{2\pi}\sigma,
\end{aligned} \tag{B.6}$$

where $\sigma^2 = \frac{1}{N}(A + \frac{C}{2})$, and the last integral is obtained by assuming polar coordinates, for which $dxdy \rightarrow r dr d\theta$. To solve the integral analytically, we have assumed that the variances of α_j and β_j are the same, which is not the case for $j = 1$ unless $C = 0$. In that case, the integral can be solved numerically.

On average, the quenched variability will contribute an amplitude $\langle R \rangle$ to the connectivity. However, the phase is also important. If $\psi = 0$, the quenched variability will add to W_1 and make instability more likely. However, if $\psi = \pi$, the quenched variability will have the opposite effect and make instability less likely. Given that each neuron in the network has a distinct phase ψ if N is large enough, we can expect that for some of them $\psi \sim 0$. The quenched variability, therefore, always makes the instability more likely, shifting the bifurcation to lower values of W_1 , although the emergent bump will be biased to certain locations. Note that the fact that $\alpha_1 > \beta_1$ due to the spatial inhomogeneity of the noise through the coefficient C means that ψ will be biased to values near 0, reducing the effect of so-called ‘‘hot spots’’. This effect requires further study.

Appendix C

Figure: Bifurcation diagram for $s = 0.2$

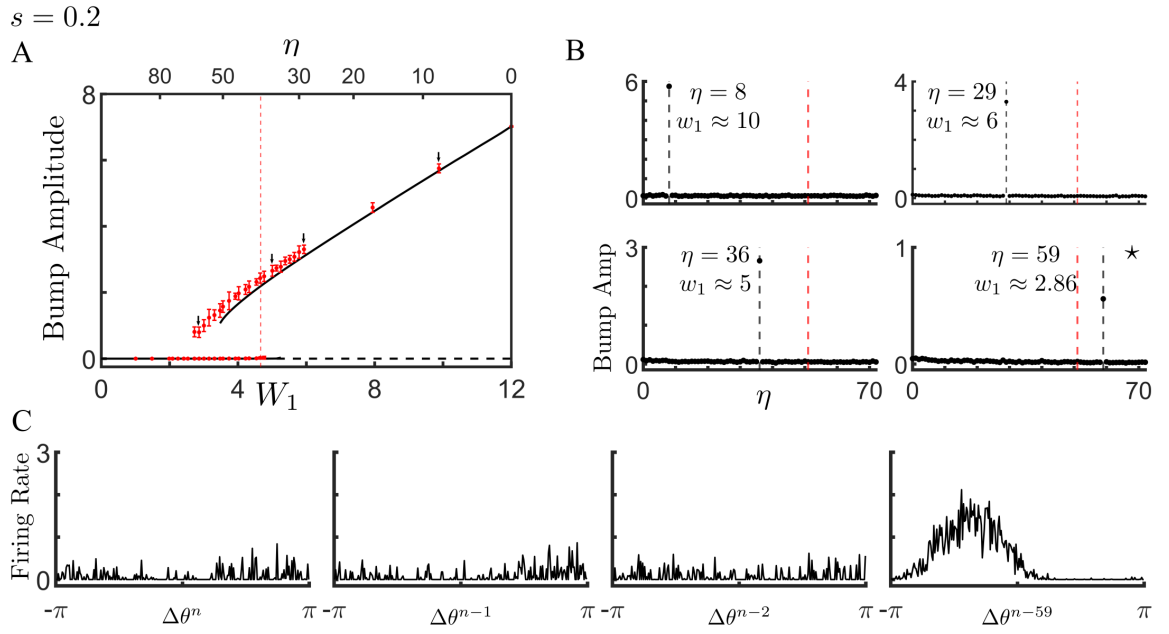


Figure C.1: Bifurcation diagram for $s = 0.2$. Similar to the Fig 5.10.