Modeling Basic Aspects of Working Memory via Dendritic Bistability

By

JIACHENG XU

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Physics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Daniel L. Cox, Chair

Mark S. Goldman

Steve J. Luck

Committee in Charge

2024

Abstract

Working memory refers to information actively held in the brain and is essential for advanced functions like thinking, decision-making, or learning. In this thesis, I focus on two key problems in working memory: 1) how to maintain a graded amplitude of a local memory in a classical ring architecture, and 2) how to maintain a novel pattern of graded neural activity in an unstructured network. To address these problems, I start with reexamining previous methodologies to limit modeling possibilities. Then, I propose two neural circuit models with dendritic bistability, each of which is treated analytically and which are robust against various perturbations. This work links physiological properties to functionality, contributing to a deeper understanding of the mechanisms underlying working memory.

Acknowledgments

First of all, I would like to express my sincere gratitude to my advisor, Daniel. From the very beginning, Daniel has been consistently supportive of my work in an insightful and resourceful way.

I would also like to express my sincere gratitude to my two co-advisors, Mark and Steve. Mark's critical suggestions have greatly helped me sharpen my work. Steve has helped me initialize my work and is always a great communicator.

Without their consistent encouragement, insightful suggestions, and generous time, it would have been impossible for me to accomplish this dissertation. Thank you!

This dissertation was supported by the Air Force Office of Scientific Research (AFOSR) under award number FA9550-22-1-0532. I would like to express my special thanks for funding my work over the years.

To my parents, Xu Gang and Zhang Xiaojun,

I could not have accomplished this without your support along the way.

List of Figures

Contents

CHAPTER 1

Introduction

1.1. What is working memory

Working memory refers to information actively maintained in the brain without external input. The duration of working memory is flexible. The information can be readily available for processing for seconds or minutes. Once it is no longer being actively maintained, it disappears immediately.

Working memory can store diverse contents. It can store a familiar item such as a red apple, or a novel item such as a new anime character. It can store representations of an abstract logical relation, an excerpted sentence, a piece of music, or a human face. It can hold one item or multiple items, until it hits its capacity. It can exist in the mammalian neocortex, and functionally similar systems also exist more broadly, including prominent work in the vertebrate oculomotor system and the head direction system.

Working memory is of fundamental importance for the brain. It provides readily available information for many advanced cognitive tasks such as thinking, planning, or decision making. In addition, a new input may disappear too quickly to induce any valid training. Holding a novel memory can provide a prolonged time to help the gradual creation of long-term memory representations by training [50]. The capacity of working memory is reported to be positively related to intelligence and cognitive performance [41,53]. Malfunction of working memory is associated with diseases, such as schizophrenia [43]. Overall, the importance of working memory in cognition and behavior calls for a more fundamental understanding of the underlying mechanisms.

1.2. How to understand working memory mechanisms

Given the versatility of working memory content, formal models usually focus on one type of information, and multiple separated working memory systems may coexist in the brain [7]. In general, there are two approaches to modeling working memory.

One approach starts with a simple problem: how information, such as a binary value, can be maintained by a neural circuit. People seek an analytical understanding of the core mechanism by taking some mathematical simplification. Some of them are briefly introduced below. Comprehensive overviews of such models are provided in previous reviews [10, 16, 21, 22, 33, 54, 69, 94, 103]. In this thesis, I stick with this approach.

The other approach focuses on the functional role of working memory, that is, how it is used in information processing. Such models are more high-level and descriptive [7, 28], instead of starting with neural circuit building blocks. Usually, there are a huge amount of parameters in such models, making mathematical analysis infeasible. Historically, these models are intensively studied in the language system, while recently being extended to visual-spatial working memory [1, 8, 29, 31, 70, 102].

1.3. A brief introduction to neural modeling

1.3.1. Rate model of a neuron

A neuron is commonly modeled as a node with multiple weighted inputs and an output. Each input is a firing rate (positive) that may vary over time, and it enters the node through a synapse with a corresponding weight value, which can be either excitatory (positive) or inhibitory (negative). The node sums over the products of inputs and corresponding synaptic weights, and then puts these through a nonlinear input-output function to give an output firing rate.

A network of mutually connected node neurons can maintain a memory. In this case, the inputs to a node consist of a sensory input from an external signal and recurrent inputs provided by other neurons in the same network. During a memory task, the network starts with all neurons at a baseline level of firing. An external signal is provided for a period of time and then turned off to zero. If the weights of the recurrent connections are zero, the neurons no longer receive input activity and therefore drop back to silence. If all the connections among the neurons in the network are strong, the excitation of some of the neurons by the sensory input will recurrently excite the other neurons, and this activity will reverberate among the neurons, maintaining persistent activity after the sensory input has terminated. When only a few neurons are strongly connected and the other recurrent connections are near zero, local activity will persist among the interconnected neurons while the rest remain silent, allowing the network to store a local memory in a specific subset of neurons.

The weight values for each recurrent connection can be plastic, slowly changing by training. The plasticity rule governing a synaptic weight is typically determined by the firing rates of the presynaptic neuron (the neuron projecting to the synapse) and the postsynaptic neuron (the neuron where the synapse is located). A common form of plasticity is Hebbian plasticity, where neurons that fire together strengthen their mutual connection weights, while other connections weaken. This associative nature of Hebbian plasticity groups strongly firing neurons together, providing a connectivity pattern that can support local memory.

1.3.2. Spiking model of a neuron

The rate model is a popular choice to capture the key dynamics of a neuron. However, it can be a bit far from a real neuron, where more complex dynamics exists and may provide a more powerful tool for modeling. Here we introduce a second model which incorporates more biophysical details

A real neuron is more complex than a simple node. It consists of a soma, which sums inputs, and branching dendrites from the soma, which can receive inputs. In addition, an axon stretches out of the soma and projects to other neurons. Different ion channels are distributed in the surface (membrane) of the neuron. Ion channels may be gated (opened and closed) by various factors, most commonly by membrane potential or by neurotransmitters. Neurotransmitters are the messager one neuron talks to another, and there are excitatory and inhibitory receptors. Two common classes of excitatory receptors are α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptors and N-methyl-D-aspartate (NMDA) receptors. There also exist common inhibitory receptors like gamma-aminobutyric acid (GABA) receptors. These receptors may open upon the arrival of certain neurotransmitters, allowing electric current to flow into the neuron. Neural models also include hypothetical "leak" channels that model an approximately constant component of conductance with a reversal potential below the neuronal firing threshold.

The neuron undergoes a series of dynamics after receiving an input. The neuronal input is a temporal train of impulses, with the frequency specified as the firing rate (impulses per second). The arrival of an impulse leads to the release of neurotransmitter from the presynaptic neuron, which bind to receptors on the postsynaptic neuron, opening some ion channels in the postsynaptic neuron and eventually causing changes in the electrical potential across the cell membrane at the site of the synapse. These potential changes propagate along the dendrites and are summed at the soma. If the somatic (or, more precisely, axon hillock) voltage exceeds a threshold, an action potential (voltage impulse) is generated and the voltage is reset below the threshold. This action potential travels in two directions, propagating through the neuron's axon to other neurons and also, to a greater or lesser extent depending on the neuron type [87], propagating back into its own dendrites, exerting a postsynaptic effect on these dendrites.

1.3.3. How working memory is maintained

Neural circuit models in working memory can be divided into four main categories:

(1) *Network models with stationary memory.* The canonical memory mechanism is based on mutual excitation of neurons, which leads to self-sustained and stationary memory activity when the input signal is removed. In a state space manifold, with each point representing a particular activity pattern across the neurons in the network, the memory can be represented geometrically as a point attractor. For instance, a double-well attractor, which has two stable states, can retain binary information and can be used for decision making. [42, 86, 93]. A line attractor, which can stabilize at any scalar value along a one-dimensional region in the state space, can be used to encode a one-dimensional variable such as occurs for encoding eye position in the oculomotor system [20, 46, 59, 84]. A line attractor can be periodic, the so-called ring attractor, to represent a feature in a circular space of values, such as head direction [5, 23, 26].

(2) *Network models with dynamic memory.* Network models of memory with non-stationary neural activity patterns have also been proposed, motivated by the observation that the pattern of activity across neurons can vary over time during the memory maintenance period. For example, memory activity may ramp up [16] over time, or neurons may change selectivity [71, 72]. Even when memory activity is non-stationary, a stable memory representation may exist. Such models can exhibit stochastic oscillation [23], seemingly random activity [32], sequential activity [45], or chaotic activity [9]. Overall, the mechanisms by which non-stationary activity may maintain memories is less well defined, but is the subject of active study [25,65].

(3) *Cellular-based models.* Persistent memory activity may also exist in a single neuron, without any recurrent excitation between neurons. This has been observed *in vitro* for graded persistent activity [34] and *in vivo* for excitatory hilar mossy cells [96]. Zylberberg and Strowbridge, 2017 [103], discuss three potential mechanisms: (i) voltage-gated Ca^{2+} and Na^{+} currents, (ii) inward currents that track intracellular Ca^{2+}, and (iii) Ca^{2+}-triggered long-term changes in neuronal excitability. Such models may be extended to the network level to obtain richer circuit dynamics.

(4) *Activity-silent models.* Unlike the models described above, which emphasize neuronal firing, activity-silent models rely on transient changes in synaptic strength [10]. The locally stored information gradually decays if there is no subsequent activity. This mechanism can collaborate with certain activity-based maintenance mechanisms to store working memory [73], with less activity and thus higher energy efficiency. It is also claimed to be more consistent with an experiment [11] showing reduced activity during the memory storage period [10].

1.4. Working memory beyond binary information

Most models focus on how binary or categorical information is stored in working memory. However, the content of working memory is far more complex than binary representations. Here, I will introduce two additional questions about the storage of more complex information.

1.4.1. Encoding a scalar value in a ring

In everyday life, it is often necessary to memorize an object with two dimensions such that one represents a continuous feature and the other represents the amplitude of the feature. For instance, one might need to remember the hue of a given colored object (the feature dimension) along with its intensity (the amplitude dimension). A central goal of this dissertation is therefore to develop a model that can encode two dimensions, both a feature and the related scalar.

Traditionally, feature and scalar working memory have been separately modeled despite they are both one-dimensional. Feature models often use a one-dimensional ring attractor architecture, in which each neuron represents a location along the ring. The network maintains a local persistent activity as memory, with the feature value encoded by the location of the peak firing neuron. This activity has a fixed amplitude, even if input strength varies a lot. On the other hand, scalar values are typically represented using different amplitudes of activity. In this case, stable memory activity requires that the excitatory input perfectly balances the inhibitory input. If the excitation is even slightly too strong, the activity will spread across the entire network; if the inhibition is even slightly too strong, the activity will die out rather than being maintained. Therefore, as activity amplitude increases, the corresponding inhibition must also increase by the amount that perfectly balances the excitation. This need for fine-tuned parameters makes the system fragile, so that it will fail as a result of any slight perturbation by background noise or any minor parameter mistuning.

To encode a feature along with its intensity, we could have a model that merges the two approaches. Indeed, such models exist [19, 20, 60, 98]. They are based on a ring attractor, with an additional line attractor at each location for amplitude encoding, instead of encoding a binary amplitude. However, these models also inherit the fine-tuning problem from line attractor models.

1.4.2. Encoding a novel pattern

Novel encoding differs mechanistically from the encoding of familiar items. In the encoding of familiar items, the signal of an item must be repeatedly delivered to the system, training the system to become familiar with it over hours or days, by forming an attractor. Once this attractor is established, the system can memorize the item whenever it is presented again. In contrast, in novel encoding, a new item can be encoded without any specific training beforehand. For example, we can immediately memorize a new math rule, a new concept that integrates several established concepts, or a new smell which triggers a distributed neural activity pattern. In real life, familiar and novel encoding can work together. For example, seeing a real red apple, I can memorize both the familiar information, a red apple, and novel information, the surface dents and spots unique to each apple. For the sake of theoretical modeling, I focus on the novel encoding part. Further, I narrowly deal with a novel graded input pattern of neural activity.

It is important to distinguish between novel encoding and generalization based on the encoding of familiar items [51]. When a network is trained to memorize, say, pictures of lions, and then a picture of clouds is provided as the input, the network can represent the cloud only with respect to its similarity to the lions. In contrast, novel encoding can memorize a cloud of any shape. In practice, if the network's generalization ability is strong enough to cover many untrained inputs, the performance of the two can be similar.

Novel encoding is important for several reasons. 1) In everyday life, we often encounter novel information. Novel encoding without a slow training process is necessary for quick decision-making. 2) Novel encoding keeps a new pattern active for an extended time, during which slower but longer-lasting learning processes can occur [50].

However, the majority of working memory models only work with familiar inputs. Indeed, memorization in these models relies on attractors. Almost all of these attractors are formed, or presumably formed, through slow training, making inputs no longer novel. However, exceptions can exist, where attractors may be gene-encoded without pretraining, especially for working memory related to evolutionarily important intuitions.

1.5. Complex neurons for computation: channel and morphological properties

The modeling approaches introduced so far have been driven by the question of ”how to maintain information”. In this section, I consider a different, yet overlapping, perspective, which asks why single-neuron morphology and biophysics is so complex. How does this complexity contribute to computational capabilities? In the next section, I will merge these two perspectives and introduce my dendrite-based network models for working memory.

Neurons have complex structures. Neurons can differ in their morphologies and channel types, which affects their functions [66, 75]. At the network level, complex neurons can be connected in various ways, such as the interplay of excitatory and inhibitory neurons through dendritic or somatic connections [26,99]. In general, it is still an open question how neural complexity contributes to network level functionalities.

In the following, I will selectively introduce certain aspects of single-neuron biophysics that are relevant to the working memory models proposed later. For more comprehensive overviews of neuronal biophysics and computation, see [6,62,67,78,87,88].

1.5.1. The role of NMDA receptors

NMDA receptors play a key role in working memory, as supported by experiments [90,91,100]. Mechanistically, NMDA receptors exhibit several useful properties:

1. NMDA receptors have slow dynamics. Therefore, even though spikes may be received at random times, NMDA receptors can smooth out this randomness and maintain a high level of openness throughout. This allows for more stable, self-sustained neural activity, as suggested previously [92].

2. NMDA receptors can perform coincidence detection. The opening of NMDA receptors requires not just presynaptic input but also postsynaptic voltage. The opening requires a short time window where both presynaptic input and elevated postsynaptic voltage approximately coincide. In some cases, such as in neurons that have strong backpropagation of action potentials, the source of the elevated voltage can be due to the postsynaptic neuron spiking so that the NMDA detects approximate coincidences of pre- and postsynaptic spikes.

3. NMDA receptors may not merely passively add presynaptic inputs linearly but can also provide rich dynamics that are beneficial for working memory. For instance, nonlinearities induced by NMDA receptors may help sustain working memory [75]. Additionally, NMDA receptors can help to mediate local dendritic voltage bistability due to the dependence of the

conductance of NMDA receptors on postsynaptic voltage, so that the NMDA conductance can be either high or low for a given amount of bound transmitter. Evidence has been observed experimentally [68, 83, 95], where NMDA receptors induce voltage plateaus for milliseconds, and models have been proposed [61, 82].

1.5.2. Separated computational units

The branching dendritic morphology of a neuron suggests the presence of multiple separated computational units. This idea has been implemented in models [77, 88] and has been supported by some experiments [63, 79]. Others suggest that even individual spines can function as separate computational units [6, 27].

However, intracellular interactions may make these components interdependent rather than independent. These interactions are common and may be helpful for information processing, as in the coincidence detection mentioned above. However, these interactions could potentially make dendrites less independent, thereby weakening the benefit of having multiple dendrites. This worry is mitigated by a model [12], which demonstrates that even with strong back-propagated somatic action potentials, dendrites still maintain a high level of independence.

1.6. My methodology towards open questions - adding more constraints

In this thesis, I model working memory using neural circuit models by incorporating several constraints. Traditionally, research has focused on how simple information, such as a color, a location, or a binary choice, can be maintained. It is a relatively easy problem that does not strongly constrain models, as many existing models mentioned can address it. Therefore, I consider additional requirements that are important for working memory, which turns out to better constrain the set of possible models. The subsequent chapters will show that models of working memory using dendritic computation can satisfy this additional set of constraints.

1.6.1. Encoding a scalar value in a ring - broader considerations

As mentioned above, there is a trade-off in sensitivity and robustness for encoding a scalar value in the amplitude of neural activity. To ensure sensitivity under small input signals, models require fine-tuning of parameters, making them not robust to even slight perturbations. To ensure robustness to perturbations, models need to form discrete attractors, but this has the downside of reducing their sensitivity to small signals.

Here, I consider this trade-off between robustness and sensitivity in more detail. A working memory system needs to be sufficiently robust, given the pervasive noise (or task-irrelevant activity) in the brain. This robustness makes the system insensitive to small activity changes, but may not reduce the system's sensitivity to an external stimulus. That is, a small stimulus change might trigger a sharp increase in activity, allowing the memory to transition from one attractor state to another, thereby updating the memory. This may happen for encoding the hue of a color or the intensity of a concept, such as friendliness. However, it seems not the case in integrators, such as the oculomotor system or head direction system, where experiments have shown that small stimulus changes result in small activity changes.

In addition, fine-tuning-based models [20, 46, 59, 84] may struggle when multiple items are encoded and compete with each other. Each memory receives not only local excitation to maintain itself but recurrent inhibition. The magnitude of received recurrent inhibition can vary significantly depending on factors such as the presence or absence of other inputs/memories. Consequently, it becomes challenging to balance a relatively fixed level of excitation with the varying levels of inhibition. This difficulty in achieving balance may render stable amplitude encoding impossible in these models.

Due to these two considerations, I prioritize robustness at the price of some insensitivity in my model. The model, which is introduced in detail below, is motivated by merging two previous models: the integrator model with multiple bistable dendrites for robust amplitude encoding and the classical ring model for feature encoding.

1.6.2. Encoding a novel pattern - broader considerations

Even though fast Hebbian plasticity is not well supported, the widely distributed NDMA receptors can produce effects that are analogous to fast Hebbian plasticity. That is, the coincidence detection nature of NMDA receptors makes NMDA receptors' activation associative. This property can be utilized to encode information for a brief period of time [89]. If the dynamics of NMDA receptors are tuned to be bistable [61,82], the activation can be maintained for an extended period of time just like plasticity. It is not yet clear how NMDA receptors can be utilized in novel encoding.

CHAPTER 2

Robust encoding of both stimulus location and amplitude in a working memory model based on dendritic bistability

Working memory can hold basic features like orientation or location. Canonically, the memory is envisioned as a point within an attractor on a neural activity manifold, which represents a stationary local activity pattern in a neuronal network. Implementing this concept, the classical ring model has achieved great success. However, this model struggles with encoding activity amplitude, which can be an important quantity for memory intensity or precision. To address this, we propose a modified network model based on the classical ring, where each neuron has multiple conductance-based bistable dendrites. This allows robust encoding of memory amplitude in neural spiking by activating different dendrites. By simplifying the spiking network to a firing rate one, we found an approximate analytical mapping between input and memory amplitude. Different from most previous models, our model requires no fine-tuning of parameters. Memory is robust against background noise, while amplitude drift or location diffusion may happen. Our model provides an example of how dendritic properties can contribute to working memory.

2.1. Introduction

Working memory refers to memory actively held in the mind. It contains information readily available for advanced cognitive tasks such as recall, reasoning, and decision-making.

While working memory content is versatile, a popular focus is simple one-dimensional features, including orientation, hue, stimulus luminance or vibration frequency, leading to many experiments and computational models [16, 46, 54, 56, 64, 94]. However, the underlying mechanisms of working memory remain not fully understood.

A ring model is a canonical model to encode a simple local feature. This includes a hue in a color wheel or an orientation within 360 degrees. The model arranges neurons in a one-dimensional ring, with each location corresponding to a feature value. The neurons are connected via translationally-invariant local recurrent excitation and lateral inhibition [5,23, 26], as shown in Fig. 2.1. The strong local excitation allows for self-sustained memory in a particular feature location. This memory stabilizes when inhibitory/decay inputs balance out excitatory inputs in each neuron, as shown in Fig. 2.1B. Over the years, many experimental works have found support for such a model across different species and brain regions [55,94]. However, the memory amplitude has a single fixed non-zero value and cannot encode different amplitudes, as shown in Fig. 2.1C.

In addition to the ring model, a second class of model to encode a one-dimensional feature is by activity amplitude. Thiscan represent useful information such as mechanical vibration frequency [15], stimulus luminance [24] or eye position [3]. In computational modeling studies of parametric working memory [16,46,56], most previous studies have modeled the activity by a line attractor, where each point along the line, representing activity amplitude, is stable and maps to a certain value. Previous models for parametric working memory usually construct the line attractor through fine-tuning [20,46,59,84]. That is, the excitation, inhibition, and intrinsic or synaptic decay processes in such models are tuned to be perfectly balanced at any given memory firing level. Unfortunately, such fine-tuned models are quite fragile. Even small background noise drives memory along the line attractor randomly. In addition, slight mistuning of model parameters breaks the perfect balance, which completely ruins the line attractor, making the memory either saturated or reduced to baseline activity.

However, exceptions exist [47,56,74], where neuronal or dendritic bistability makes perfect balancing possible without fine-tuning.

Therefore, it is preferred if the ring model can encode activity amplitude by including an additional parametric working memory model. In this way, the model can encode two-dimensional information, viz., a location code (e.g., a hue on the color wheel) and an amplitude code (e.g., the intensity of the hue), which is also suggested by neural recording [2,52]. However, previous modeling efforts in this direction need fine-tuning of parameters as in the parametric line attractor case [19,20,60,98].

In addition, although firing rates seem like a natural way to encode the intensity of a stimulus, they might also be used to represent the precision or uncertainty about a given feature value. Theoretical works [57,85,97] have demonstrated how activity amplitude can be used to represent precision of locally stored memory. In the present work, the stable firing rates could represent either the intensity or the precision of a given feature.

Here we show that dendritic bistability makes the joint coding of amplitude and location possible without fine-tuning. We propose a ring model with each neuron having multiple bistable dendrites. This model merges the local feature encoding and amplitude encoding of that feature. We begin by examining the core mechanism in a simplified autapse case for parametric working memory and demonstrate how different weight functions affect memory performance. We show that this mechanism also enables amplitude encoding in the ring model. Next, we show the robustness of this local memory under background noise, characterizing the amplitude drift and location diffusion, and how the effect of noise depends on dendritic bistability and network connectivity. Finally, we discuss the biological plausibility, interpretation and limitations of our model.

2.2. Materials and Methods

In the following, we present three models based on neurons containing multiple bistable dendritic compartments. First, we consider an ‘autapse’ network consisting of a single neuron with excitatory synapses onto its own conductance-based bistable dendrites. Second, we abstract this model to a simplified rate-based model that is amenable to analytical study. Third, we construct a ‘ring network’ of such rate-based multicompartment neurons to demonstrate how a network can simultaneously and robustly encode both the location and the amplitude of a stimulus. Below, we describe the equations governing each of these models, followed by the simulation details.

2.2.1. Autapse model with multiple conductance-based, bistable dendrites

The conductance-based model neuron consists of an integrate-and-fire soma electrically connected to N=10 conductance-based dendrites with voltages $\mathrm{V_{d,i}}$. The somatic voltage $\mathrm{V_s}$ is governed by:

$$\mathrm{C}\frac{\mathrm{dV_s}}{\mathrm{dt}} = -\mathrm{I_L} - \mathrm{I_d} - \mathrm{I_e} - \mathrm{I_{ton}} \tag{2.1}$$

$$= -\mathrm{g_L}(\mathrm{V_s} - \mathrm{E_L}) - \sum_{\mathrm{i}}^{\mathrm{N}} \mathrm{g_{ds}}(\mathrm{V_s} - \mathrm{V_{d,i}}) - \mathrm{g_e s_e}(\mathrm{V_s} - \mathrm{E_{AMPA}}) - \mathrm{I_{ton}}, \tag{2.2}$$

where $\mathrm{I_L}$ is the leak current, $\mathrm{I_d}$ is the summed current from all dendrites, $\mathrm{I_e}$ is the current from external excitatory synaptic inputs through AMPA receptors, and $\mathrm{I_{ton}}$ is the tonic background current. We set the membrane capacitance to C = 10 nF/mm^2, the maximal conductance and equilibrium potentials for the currents to $\mathrm{g_L}$ = 0.5 µS/mm^2, $\mathrm{E_L}$ = –80 mV, $\mathrm{g_e}$ = 5.2 µS/mm^2, $\mathrm{E_{AMPA}}$ = 0 mV, the conductance mediating the flow of currents from dendrite to soma as $g_{ds} = 0.007$ μS/mm^2, and $I_{ton} = –17.2$ nA/mm^2. The synaptic activation variable s_e obeyed dynamics described below. The firing threshold for the soma is $V_{th} = –50$ mV. Each spike sets the somatic voltage to 30 mV for 3 ms before it resets to $V_{reset} = –80$ mV.

The voltage dynamics of the dendritic compartments were adapted from [82], but with multiple dendritic compartments, modified maximum conductance values and the inward rectifier potassium (KIR) conductance simplified to not include a synaptically driven component:

$$C\frac{dV_{d,i}}{dt} = –I_{L,i} - I_{s,i} - I_{AMPA,i} - I_{NMDA,i} - I_{KIR,i}, \tag{2.3}$$

where $I_{L,i}$ gives the leak current, I_s gives the current arriving from the soma, and the final three terms give the currents arriving through AMPA receptors, NMDA receptors, and inwardly rectifying potassium (KIR) channels, respectively. These currents are described by the following equations:

$$I_{L,i} = g_L(V_{d,i} - E_L), \tag{2.4}$$

$$I_{s,i} = \varkappa g_{ds}(V_{d,i} - V_s), \tag{2.5}$$

$$I_{AMPA,i} = g_{AMPA}s_{AMPA}(V_{d,i} - E_{AMPA}), \tag{2.6}$$

$$I_{NMDA,i} = g_{NMDA,i}s_{NMDA}\frac{V_{d,i} - E_{NMDA}}{1 + 0.15e^{–0.08V_{d,i}}}, \tag{2.7}$$

$$I_{KIR,i} = g_{KIR}\frac{V_{d,i} - E_{KIR}}{1 + e^{0.1(V_{d,i}–E_{KIR}+10)}}, \tag{2.8}$$

where $\varkappa = 2$ represents the ratio of the area of the soma to that of each dendrite, $E_{NMDA} = 0$ mV, $E_{KIR} = –90$ mV, $g_{AMPA} = 0.38$ μS/mm^2 and $g_{KIR} = 8.6$ μS/mm^2. The maximum NMDA conductance $g_{NMDA,i} = 48/((i + 4.5)^{0.54})$ μS/mm^2 was chosen such that each dendrite i can be successively activated, starting from the one with the smallest value, as shown in Fig. 2.2C.

The synaptic activations s_{α} obey the dynamics

$$\frac{\mathrm{ds}_{\alpha}}{\mathrm{dt}} = -\mathrm{s}_{\alpha}/\tau_{\alpha}, \tag{2.9}$$

$$\Delta \mathrm{s}_{\alpha} = 0.5(1-\mathrm{s}_{\alpha}), \text{ each time a spike is received,} \tag{2.10}$$

where α denotes AMPA (s_{e} and $\mathrm{s}_{\mathrm{AMPA}}$) or NMDA ($\mathrm{s}_{\mathrm{NMDA}}$) type synapses with decay time constants $\tau_{\mathrm{AMPA}} = 2$ ms or $\tau_{\mathrm{NMDA}} = 100$ ms, respectively.

To trace out the bistable hysteretic relation for each dendrite in Fig. 2.2B, we replaced the recurrent synaptic input to each dendrite by a manually applied excitatory spike train and tested the response of each dendrite to different fixed levels of presynaptic input. Starting with 0 Hz, each excitatory spike train lasted for a period of 1400 ms, and the averaged dendritic voltages over the last 1000 ms were recorded. Next, the frequency was increased by 1 Hz, and the recording repeated. Once the frequency reached 70 Hz, it was decreased by 1 Hz each step until returning to 0 Hz.

To illustrate how larger external inputs successively recruit higher-threshold dendrites, enabling the neuron to encode graded input amplitudes, in Fig. 2.2C, we implemented the following procedure: V_{s} and $\mathrm{V}_{\mathrm{d,i}}$ were initialized to –70 mV and the system stabilized after a period of 1200 ms before providing external inputs to the soma. Each external input was modeled as a periodic burst of spikes for 400 ms, followed by an 800 ms delay period with no external input. This cycle of applying external input, followed by a delay period when the input is off, was repeated for different external input amplitudes, starting at 20 Hz and increasing by 5 Hz for each subsequent input.

2.2.2. Rate-based autapse model

The rate-based autapse model was constructed to phenomenologically capture the key features of the conductance-based model in an analytically tractable manner. Each dendrite exhibits a bistable response to synaptic input, and the somatic firing rate r is driven by the sum of dendritic contributions and external input:

$$\tau \frac{\mathrm{dr}}{\mathrm{dt}} = -\mathrm{r} + \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{D_i(r)} + \mathrm{I_e}, \tag{2.11}$$

where $\mathrm{D_i(r)}$ is the bistable dendritic input-output relation, $\mathrm{I_e}$ is the external input to the soma, and $\mathrm{N_D}$ is the number of dendrites. τ is chosen to capture the slowest timescale of neural activity in this simple single-equation rate model, which is set by the NMDA receptor kinetics in the spiking model. For the current model, we focus primarily on the ability of the model to maintain multiple stable states ($\mathrm{dr/dt} = 0$), in which case the value of τ is irrelevant.

$\mathrm{D_i(r)}$ is illustrated in Fig. 2.3A and given mathematically by:

$$\mathrm{D_i(r)} = \begin{cases} \beta\Theta(\mathrm{w_i r} - \mathrm{T_u^0}) & \text{Activation from down- to up-state} \\ \beta\Theta(\mathrm{w_i r} - \mathrm{T_d^0}) & \text{Deactivation from up- to down-state,} \end{cases} \tag{2.12}$$

where β is the increment in firing rate due to a dendrite flipping from its down- to its up-state. $\Theta(\mathrm{x})$ indicates the Heaviside (unit step) function, and $\mathrm{T_u^0}$ and $\mathrm{T_d^0}$ are constant up and down thresholds with respect to the current entering dendrite i, $\mathrm{w_i r}$. This expression is equivalent to:

$$\mathrm{D_i(r)} = \begin{cases} \beta\Theta(\mathrm{r} - \mathrm{T_{u,i}}) & \text{Activation from down- to up-state} \\ \beta\Theta(\mathrm{r} - \mathrm{T_{d,i}}) & \text{Deactivation from up- to down-state,} \end{cases} \tag{2.13}$$

where the effective up and down thresholds with respect to the presynaptic firing rate r into dendrite i are $\mathrm{T_{u,i}} = \frac{\mathrm{T_u^0}}{\mathrm{w_i}}$ and $\mathrm{T_{d,i}} = \frac{\mathrm{T_d^0}}{\mathrm{w_i}}$.

We derive the graphical band of dendritic feedback $\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{D_i(r)}$ in Eq. 2.11, shown in Figures 2.3A and 2.3B, as follows. The total dendritic activity $\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{D_i(r)}$ can be plotted as a relation of the firing rate r by stacking the corresponding hysteretic rectangles on top of each other, forming a band. We denote the stair-like edges of the band, which are defined by the set of individual dendrite up and down thresholds $\mathrm{T_{u,i}}$ and $\mathrm{T_{d,i}}$, as $\mathrm{B_u(r)}$ and $\mathrm{B_d(r)}$, respectively. The coordinates, with the firing rate r along the horizontal axis and the total dendritic activity along the vertical axis, of the upper left corner of each step in $\mathrm{B_u}$ and $\mathrm{B_d}$ are:

$$[\frac{\mathrm{T_u^0}}{\mathrm{w_i}}, \beta \mathrm{i}] \text{ and } [\frac{\mathrm{T_d^0}}{\mathrm{w_i}}, \beta \mathrm{i}]. \tag{2.14}$$

In this manner, the shape of the band reflects the specific set of weights $\mathrm{w_i}$.

Stable memory states in the absence of external input ($\mathrm{I_e} = 0$) occur when the total dendritic feedback $\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{D_i(r)}$ term in Eq. 2.11 balances the intrinsic neuronal decay term –r during the delay period, i.e. when $\mathrm{r} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \mathrm{D_i(r)}$. Graphically, this occurs when the band in Figure 3B intersects the diagonal, 45° line that represents the amplitude of the decay term –r.

For ease of analysis, if $\mathrm{w_i}$ changes gradually in the range $\mathrm{N} \geq \mathrm{i} \geq 0$, we can approximate the discrete set of $\mathrm{w_i}$ by a continuous function w(x) so that the coordinates of the points in the continuum plot are given by:

$$[\frac{\mathrm{T_u^0}}{\mathrm{w(x)}}, \beta \mathrm{x}], \tag{2.15}$$

An analytic formula for the continuum approximation of the edges of the band can be derived as follows by approximating the sum of dendritic activations by an integral:

$$\mathrm{B_u} = \sum_{\mathrm{i}=1}^{\mathrm{N}} \beta \Theta(\mathrm{r} - \frac{\mathrm{T_u^0}}{\mathrm{w_i}}) \rightarrow \mathrm{B_u} \approx \int_0^{\mathrm{N}} \beta \Theta(\mathrm{r} - \frac{\mathrm{T_u^0}}{\mathrm{w(x)}}) \mathrm{dx}. \tag{2.16}$$

The Heaviside function in the integral equals 1 for $\mathrm{w(x)} \geq \mathrm{T_u^0/r}$ and 0 otherwise. For a weight function w(x) that decreases monotonically with x, if a dendrite $\mathrm{x^*}$ is activated, all dendrites with labels $\mathrm{x < x^*}$ are also activated. The analytical formula for the upper band edge is:

$$\mathrm{B_u(r) = \beta w^{-1}(\frac{T_u^0}{r}), for\ r \leq T_u^0/w(N)} \tag{2.17}$$

and similar reasoning applies to deriving the lower edge of the band.

From the above expression, we can derive the minimum value of a steady-state stimulus $\mathrm{I_e}$ that can be encoded in memory. From Eq. 2.11, the steady-state value of the firing rate before any dendrites turn on is given by $\mathrm{r = I_e}$. The first dendrite turns on when this value reaches the x-intercept of the band $\mathrm{B_u(r) = 0}$. Thus, the minimum value of stimulus that can be encoded is given by $\mathrm{B_u(I_e) = 0}$.

From Eq. 2.11, during the encoding period, the steady-state equation is given by:

$$\mathrm{B_u(r_m + I_e) = r_m}, \tag{2.18}$$

where $\mathrm{r_m}$ is the memory amplitude and we assume the firing rate does not decrease past the down-threshold line $\mathrm{B_d(r)}$ during the memory period. This equation defines the location of the green dots in Figures 2.3C and 2.3D, and we note that such a stable fixed point can only exist at values for which the band $\mathrm{B_u(r)}$ has a slope less than unity, $\mathrm{\frac{a\beta}{T_u^0} < 1}$, so that it can intersect the shifted decay line $\mathrm{y = r - I_e}$ (Figures 2.3C and 2.3D, rightmost dashed lines).

During the delay period, the firing rate decreases due to the removal of $\mathrm{I_e}$, but the firing rate is maintained (Figures 2.3C and 2.3D, tip of top red arrow).

As an example, we next derive the conditions for achieving a linear relation between the strength of external input $\mathrm{I_{ext}}$ and the memory amplitude. Equal spacings are required between $\mathrm{T_{u,i}}$ of successively recruited dendrites. This can be achieved by a weight function of the form $w_i = \frac{a}{i+b}$ in the discrete case, where a and b are constants. In the continuous case, the analogous equation is $w(x) = \frac{a}{x+b}$, giving a straight line for B_u as shown in Fig. 2.3C. Solving for $w^{-1}(x)$ and substituting into Eq. 2.17 gives:

$$B_u(r) = \frac{a\beta}{T_u^0} r - b\beta, \tag{2.19}$$

which has an x-intercept $B_{u,th} = \frac{T_u^0 b}{a}$ that defines the minimal steady-state external input required to activate any dendrites. Based on Eq. 2.18, we can solve for the memory amplitude for $I_e > B_{u,th}$:

$$r_m = \frac{a\beta(I_e - B_{u,th})}{T_u^0 - a\beta}. \tag{2.20}$$

Fig. 2.3 shows plots for the autapse model with $\beta = 1$ Hz, $T_u^0 = 9$ Hz and $T_d^0 = 2$ Hz. For the discrete case in Fig. 2.3B, dendritic weights follow $w_i = \frac{30}{(i+9)^{1.4}}$, with i from 1 to $N_D = 10$. For the continuous limit, the B_u and B_d lines are plotted in Figures 2.3B-D with different weight functions. In Figures 2.3B and 2.3C, the weights obeyed power law decays $w(x) = \frac{30}{(x+9)^{1.4}}$ (Fig. 2.3B) and $w(x) = 4/(x+1)$ (Fig. 2.3C). In Fig. 2.3D, the weights obeyed a (half) Gaussian $w(x) = e^{-\frac{x^2}{2\times 14.14^2}}$. Fig. 2.3E summarizes the relation between external input and memory amplitude for all three weight functions.

2.2.3. Rate-based model with ring architecture

2.2.3.1. Basic dynamics with the ring architecture

Similar to classic ring models [5, 23, 26], our ring network has N neurons arranged along a one-dimensional ring, with symmetric and rotationally invariant excitatory weights w_{ij}. The weight value decays as the neuronal distance $|i-j|$, with a value limited to $[0, N/2]$ for periodicity. Formally in the model, each of the N=360 neurons in the network excitatory connection projects to a separate dendrite on itself and every other neuron in the network.

Thus, the number of neurons formally equals the number of dendrites of a total of N dendrites. However, in practice, for synaptic weight functions that are spatially local, there is effectively only input to a much smaller number of dendrites on each neuron. The firing rate r_i of the neuron i follows the dynamics:

$$\tau\frac{dr_i}{dt} = -r_i + \sum_{j=1}^{N} D_{ij}(r_j) + I_{e,i}, \tag{2.21}$$

where $D_{ij}(r_j)$ is the dendritic bistable input-output relation of the jth dendrite, receiving input from neuron j, $I_{e,i}$ is the external input. $D_{ij}(r_j)$ is defined analogously to $D_i(r)$ in the autapse model:

$$D_{ij}(r_j) = \begin{cases} \beta\Theta(w_{ij}r_j - T_u^0) & \text{Activation from down- to up-state} \\ \beta\Theta(w_{ij}r_j - T_d^0) & \text{Deactivation from up- to down-state.} \end{cases} \tag{2.22}$$

Note a change in labeling. In the autapse case, each dendrite is specified by i. However, in this context, each dendrite is specified by its presynaptic neural label j and postsynaptic neural label i. Here, as in the autapse model above, τ is chosen to capture the slowest timescale of neural activity in a simple single-equation rate model of the neuron. Explicit inclusion of synaptic or dendritic dynamics would not change the steady-state dynamics during the memory period that are the primary focus of this work, but would change the detailed dynamics of approach or decay from such steady states due to external inputs or mistuning.

For a local $I_{e,i}$, all dendrites of a neuron i can still be conceived of as forming a bistable band consisting of individual effective thresholds, analogous to the autapse case. However, because each dendrite receives input r_j from a different neuron, the pattern of dendritic activations onto neuron i now depends not only on r_i but also on the specific, non-unique set of firing rates of the other neurons in the network. Thus, any plots of such a bistable band as a relation of the neuron's firing rate need to be understood as either specific to the exact manner in which the other neurons were activated or, if bands generated for different network patterns of firing are not too different, as an approximation, as shown in Fig. 2.4A, left. In the analytic calculations below, we find that we can approximate certain features of the network response, like the peak of the bump of firing as a function of the external stimulus strength, by making the following approximation:

$$\mathrm{D_{ij}(r_j)} = \begin{cases} \beta\Theta(\mathrm{w_{ij}R_jr_A} - \mathrm{T_u^0}) & \text{Activation from down- to up-state} \\ \beta\Theta(\mathrm{w_{ij}R_jr_A} - \mathrm{T_d^0}) & \text{Deactivation from up- to down-state,} \end{cases} \tag{2.23}$$

where we approximate that $\mathrm{r_j}$ can be decomposed into the product of a fixed shape $\mathrm{R_j}$, whose peak rate is 1, and a changing scalar $\mathrm{r_A}$, which increases as the overall activity rises. This approximated decomposition is effective for analysis, even though, in real simulations, a larger $\mathrm{r_A}$ can change the overall shape $\mathrm{R_j}$. Next, we have:

$$\mathrm{D_{ij}(r_j)} = \begin{cases} \beta\Theta(\mathrm{r_A} - \mathrm{T_{u,ij}}) & \text{Activation from down- to up-state} \\ \beta\Theta(\mathrm{r_A} - \mathrm{T_{d,ij}}) & \text{Deactivation from up- to down-state,} \end{cases} \tag{2.24}$$

where the effective thresholds are $\mathrm{T_{u,ij}} = \frac{\mathrm{T_u^0}}{\mathrm{w_{ij}R_j}}$ and $\mathrm{T_{d,ij}} = \frac{\mathrm{T_d^0}}{\mathrm{w_{ij}R_j}}$. Working under this assumption, in this paper we focus on the bistable band of the peak neuron, for which the bump amplitude $\mathrm{r_A}$ is the same as the peak firing rate $\mathrm{r_p}$.

2.2.3.2. How weight determines the relation between external input and memory amplitude

Similar to the autapse case, the synaptic weight function $\mathrm{w_{ij}}$ approximately determines the relation between external input and memory amplitude. Below, we analytically show how to derive a weight function that approximately gives a linear mapping between the strength of the external input stimulus and the height of the memory pattern of activity (as defined by the firing rate of the neuron at the peak of the bump of activity).

For simplicity, we assume that T_d^0 is low enough to prevent any dendrites from flipping from their up to their down states during the delay period. Therefore, we can focus on the dynamics during the encoding period, where $D_{ij}(x)$ is effectively a step function:

$$\tau \frac{dr_i}{dt} = -r_i + \beta \sum_{j=1}^{N} \Theta[w_{ij} r_j - T_u^0] + I_{e,i}. \tag{2.25}$$

Next, we proceed to the continuous limit as was done in the autapse case. If we choose $N = 360$, neural label x and z can be also viewed as degrees in the ring:

$$\tau \frac{dr(x)}{dt} = -r(x) + \beta \int_0^N \Theta[w(x-z)r(z) - T_u^0] dz + I_e(x). \tag{2.26}$$

where the x-z term indicates the distance, with a value limited to [-N/2, N/2] for periodicity and with the weight profile function w(x-z) a function of the absolute value of this modulo difference. $w(x-z)r(z)$ gives the effective input current received in dendrite z of neuron x. If this input exceeds T_u^0, $\Theta(x)$ gives 1, with the corresponding dendrite activated. The integration counts the total number of activated dendrites, satisfying $w(x-z)r(z) > T_u^0$. Multiplying by the factor β then gives the total dendritic contribution to the steady-state neural firing. Since this contribution is also the total firing following the offset of the external input as no forgetting happens, we will refer to this below as the memory activity M(x).

If we assume that the memory activity pattern M(x) approximately follows the pattern of the external input $I_e(x)$, with both having the fixed symmetric profile R(x) in the continuum limit with its maximal value of 1, then we can define $I_e(x) = I_p \times R(x)$, $M(x) = M_p \times R(x)$, and $r(x) = (I_p + M_p) \times R(x)$, where I_p is the peak rate of input and M_p is the peak rate of memory.

Without loss of generality, we set $x = 0$ as the location of the peak of R(x). Since we are primarily concerned with the magnitude of the peak firing rate for a fixed shape, we focus on $x = 0$. With these considerations, the stable solution based on Eq. 2.26 can be rewritten as:

$$\mathrm{M_p} = \beta \int_0^{\mathrm{N}} \mathrm{H}[(\mathrm{I_p} + \mathrm{M_p})\mathrm{w(z)R(z)} - \mathrm{T_u^0}]\mathrm{dz}. \tag{2.27}$$

To get a linear input-memory relation $\mathrm{M_p} = \mathrm{aI_p} + \mathrm{b}$, we need to solve:

$$\mathrm{aI_p} + \mathrm{b} = \beta \int_0^{\mathrm{N}} \Theta[(\mathrm{I_p} + \mathrm{aI_p} + \mathrm{b})\mathrm{w(z)R(z)} - \mathrm{T_u^0}]\mathrm{dz}. \tag{2.28}$$

This integral equation can be solved graphically, as shown in Fig. 2.4E. The product of two symmetric functions w(z) and R(z), both centered at $\mathrm{x} = 0$, results in the dendritic input, another local concentric symmetric function w(z)R(z), as illustratively plotted. The threshold line is $\mathrm{T_u^0}/((1{+}\mathrm{a})\mathrm{I_p}{+}\mathrm{b})$. For any z where the dendritic input exceeds this threshold, it activates a dendrite. The total number of activated dendrites is represented by the length of the green segment, which equals $\frac{\mathrm{aI+b}}{\beta}$. Altogether, we can express the coordinate of the red dot in Fig. 2.4E along the composite function w(z)R(z) as:

$$(\frac{\mathrm{aI} + \mathrm{b}}{2\beta}, \frac{\mathrm{T_u^0}}{(1 + \mathrm{a})\mathrm{I} + \mathrm{b}}). \tag{2.29}$$

We can reparametrize it as:

$$\mathrm{R(x)w(x)} = \frac{\mathrm{T_u^0}}{\frac{(1+\mathrm{a})(2\beta|\mathrm{x}|-\mathrm{b})}{\mathrm{a}} + \mathrm{b}}. \tag{2.30}$$

Finally, the desired weight function is:

$$\mathrm{w(x)} = \frac{\mathrm{T_u^0}}{\frac{(1+\mathrm{a})(2\beta|\mathrm{x}|-\mathrm{b})}{\mathrm{a}} + \mathrm{b}} \frac{1}{\mathrm{R(x)}}. \tag{2.31}$$

Thus, the weight function for producing a linear output is a shifted power-law term as in the autapse case, multiplied by $\frac{1}{\mathrm{R(x)}}$ to compensate for the effect of input shape. For a perfectly flat output, this reduces to the formula found for the autapse case. For maintaining a finite width bump of activity, R(x) from the above formula would go to 0, leading to an infinite weight w(x). Furthermore, even before this blow up, w(x) in the above formula becomes non-monotonic due to the rapidly increasing values of R(x) near the edges of the bump. Therefore, in our simulations, we enforce a monotonically decreasing weight function by truncating (setting to 0) the values of w(x) at the point that w(x) stops monotonically decreasing. As shown in Figures 2.4F and 2.4G, this still results in a highly linear memory activity versus input amplitude relationship.

2.2.3.3. Simulation details in the noise-free case

Figures 2.4-2.7 show the simulations of the ring under different conditions. $\mathrm{I_{e,i}}$ followed a (truncated) Gaussian function, $\mathrm{Ae^{-\frac{|i-j|^2}{2\times 10^2}}}$, where A is the amplitude. $\mathrm{|i-j|}$ refers to the distance between neuron i and neuron j, which is the peak of the Gaussian function is located at, $\mathrm{j = 180}$. The distance is bound between 0 to N/2 due to the periodicity of the ring model network. $\mathrm{I_{e,i}}$ was applied for a duration of 1000 ms (the 'encoding period'), which is sufficient for the network activity to reach equilibrium. The input was then turned off, and the subsequent 'memory period' activity was simulated for a duration of 1000 ms for Fig. 2.4. A shifted power law weight function $\mathrm{w_{ij} = \frac{15}{(|i-j|+2)^2}}$ was used. Other parameters remained unchanged from Fig. 2.3, $\beta = 1$ Hz, $\mathrm{T^0_u} = 9$ Hz, $\mathrm{T^0_d} = 2$ Hz, and we set $\tau = 50$ ms. For Fig. 2.4, the input amplitude A = 12, 20, 28, or 36.

In Figures 2.4B and 2.5A, to give intuition for the relationship between the ring model and the simpler autapse model, we numerically generated a bistable band for a particular neuron for a particular strength and profile of external excitatory and inhibitory input. For the up band, for a given neuron i in the presence of a stimulus, we can plot the total dendritic response, $\sum_{\mathrm{j}}^{\mathrm{N}} \mathrm{D_{ij}}$ against the neuron's firing rate $\mathrm{r_i}$ with time parametric in the plot, i.e. the resulting band edge $\mathrm{B_u}$ indicates the trajectory of the firing rate over time. In Fig. 2.4B, this was done for the neuron with peak firing rate and for initial condition $\mathrm{r_i(t = 0) = 0}$, and for applying a constant external stimulus input A = 100. We note that the resulting band shape depends upon the stimulus strength and profile because the flipping on of dendrites depends upon the firing rates of the neurons afferent to them – while these other neurons' firing rates approximately go up and down together with the firing rate of neuron i that is plotted along the x-axis, the exact relationship depends upon the exact shape of the bump of network activity, which in turn depends upon the stimulus.

To identify the down threshold band $B_d(r)$, we first shut off the input, allowing the firing rate to stabilize. We then apply a uniform inhibition (-15 Hz in Fig. 2.4B) large enough to shut off the memory-storing activity. The resulting $B_d(r)$ is a trace in time of the memory decay from the initial value until zero firing rate. Similar to the discussion in the paragraph above, while activity of the different neurons in the network approximately goes up and down together, the exact form of $B_d(r)$ depends on the exact shape of the bump of activity at any time during the shutting off of the memory, and this shape depends both on the initial stable memory activity profile before inhibition was applied and also on the shape and strength of the inhibitory input used to turn off the memory activity.

In Fig. 2.4F, G, we checked the performance for a linear input memory relation. The linear relation $I_p = aM_p + b$ has $a = 0.5$ and $b = –1$. We further used a fixed Gaussian shape, $R(x) = e^{-\frac{x^2}{2\times10^2}}$. Together, from Eq. 2.31, this yields a weight function $w(x) = \frac{9}{6|x|+2}e^{\frac{x^2}{2\times10^2}}$. In the simulation, we used this new weight function, but truncated after $|x| > 9$ to ensure $w(x)$ is a local decay function, and with x discretized in increments of size $dx = 0.05$. With input amplitudes taking integer values from 6 to 15, the memory changes agree with the desired linear relation, as shown in the red dashed line in Fig. 2.4G.

2.2.3.4. Simulation details with noise

For all simulations with noise, spatially and temporally independent Gaussian noise with a zero mean was applied to each soma with a time step of 1ms, with standard deviations as specified below. All other model parameters are identical to Fig. 2.4, unless otherwise specified.

We show the drifting of memory amplitude in the presence of noise in Fig. 2.5B for a memory period of duration 9000 ms. The standard deviation of noise is 2, 10, or 15, and the external input amplitude A = 8, 26, or 80. For each condition, 5 trials were conducted, and the peak firing rate was estimated by averaging the firing rates of the 10 highest firing neurons.

In Fig. 2.6, we compare the bistable model to its classical limit. A recurrent inhibition term $-\frac{\alpha}{N}\sum_{j=1}^{N} r_j$ was added to the right hand side of Eq. 2.21. To compensate for the additional inhibition, we adjusted T_d^0 to 1 Hz. The standard deviation of noise equaled 10, and A = 12. For each α value, from 0 to 2, with a step size of 0.4, 10 trials were simulated. We estimated the memory width by averaging across trials the width of the bump of activity at half of its peak rate, calculated at the end of the delay period. The classical limit was taken by setting T_u^0 also to be 1 Hz. For Fig. 2.6C, a smaller distractor input centered at neuron 100 was applied, with A = 8 and $\alpha = 1$.

In Fig. 2.7, we study location diffusion under noise. The memory period is set to 9000ms. The same recurrent inhibitory term as Fig. 2.6A were used with $\alpha = 1$. We adjusted T_d^0 to 1 Hz. In Fig. 2.7A, T_u^0 is 1, 1.2, or 1.4 Hz. Memory location was estimated by the center of mass. Traces in Fig. 2.7B, with increasing darkness, represent location diffusion with A = 12, 15, or 18. Each trace in Fig. 2.7 was obtained by averaging over 400 trials, with no spontaneous decay of activity happening throughout the delay period.

2.2.4. Numerical simulation of differential equations

For all simulations, we integrated the differential equations using Euler's method with a time step of 0.01 ms for the spiking model and 1 ms for the rate-based ring model, except for Fig. 2.4B, where 0.01 ms was used for better approximation. Code was written in python version 3.9.16.

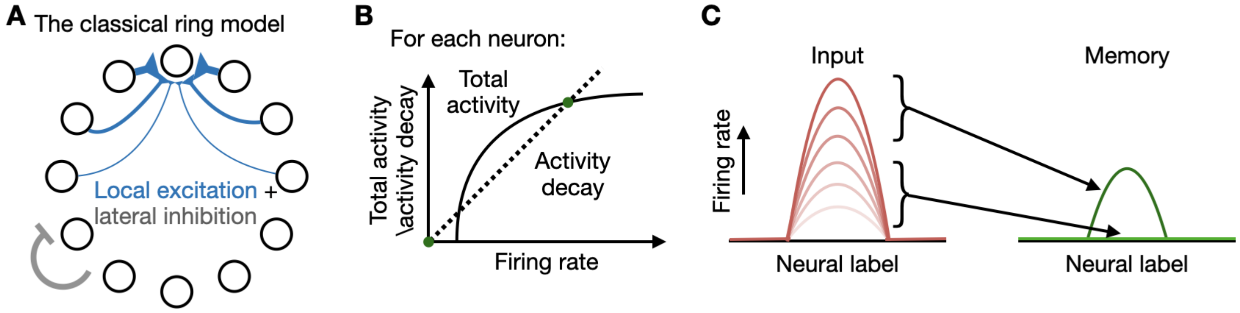

Figure 2.1. A classical ring model and its limit. A, The classical ring model. Neurons are arranged in a one-dimensional ring with local excitation between close neurons (blue) and board inhibition between distant neurons (gray). B, Excitatory and inhibitory contribution in each neuron. As neural firing rate (x-axis) increases, more excitatory activity and activity decay are induced (y-axis). The stable activity is given by the intersection (green) where the total activity (gray) equals the activity decay (dashed). C, Incapability of amplitude encoding. Under a local input with different firing rate amplitudes, only two stable memory states exist, either a fixed amplitude or baseline activity.

2.3. Results

2.3.1. Amplitude encoding in autapse with bistable dendrites

While we primarily focus on working memory of amplitude at a given location, we start by studying amplitude encoding alone with an autapse model. The autapse is a self-connected neuron with simple discrete line attractor dynamics. Its performance is qualitatively identical to the network model mentioned later which also includes location memory. In this section, we first introduce a spiking autapse to demonstrate amplitude encoding. This is followed by an explanation of the underlying mechanism using a rate model simplification. We further demonstrate how different synaptic weight functions affect memory performance.

2.3.1.1. A spiking autapse with multiple bistable dendrites

We built a spiking autapse to demonstrate amplitude encoding in a biologically plausible way. Here, each connection projects to a separate dendrite, as shown in Fig. 2.2A. All 10 dendrites are identical except for different NMDA conductance values, g_{NMDA}. Each dendrite shows

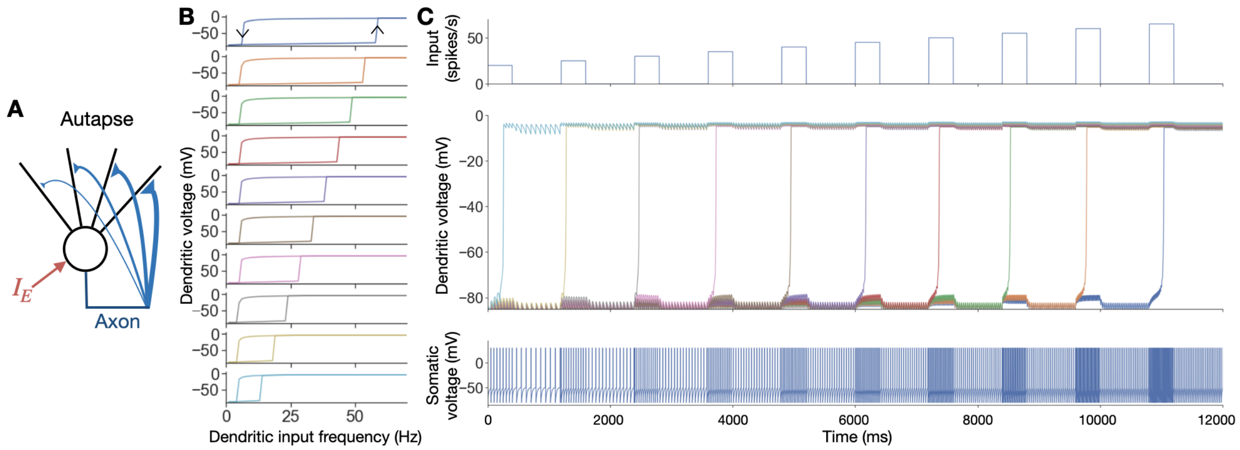

Figure 2.2. The spiking autapse model for amplitude encoding. A, The autapse model with dendritic self-connections. The autapse implements conductance-based dynamics with an integrate-and-fire soma. The external input (red) is directly into the soma. The recurrent input is directed into 10 dendrites, with different g_{NMDA} values, see Materials and Methods for more details. B, The hysteretic input-output relations of ten dendrites. The input frequency in x-axis is periodic impulses, and the dendritic voltage in y-axis is averaged over 1000 ms. The arrows (top) show the directions for hysteresis. Dendrites with smaller g_{NMDA} values are plotted at the bottom. C, Amplitude encoding by activating dendrites. As input frequency increases in each temporal burst (top), dendrites are activated by order starting from the one with the largest g_{NMDA} (middle). The dendrite colors match those in B. During the delay period, the activated dendrites are self-sustained, with frequency increasing in proportion to the number of activated dendrites (bottom).

a bistable dendritic input-output relation, with conductance-based dynamics modified from a previous model [82] (See Materials and Methods). As shown in Fig. 2.2B, starting with a voltage around –80 mV (up-state) or 0 mV (down-state), once the synaptic activation exceeds a up (down) threshold, the dendrite voltage activates (drops) to an up/down-state due to the sudden opening (closing) of NMDA receptors. Moreover, a higher g_{NMDA} gives lower thresholds, as plotted at the bottom.

Amplitude encoding is possible with the help of bistable dendrites. In Fig. 2.2C, we apply multiple inputs, each with a duration of 400 ms, followed by a delay period of 800 ms when the input is off. The input starts with a frequency of 20 Hz, causing the dendrite with the largest NMDA conductance to be activated to a voltage plateau (blue line, middle). The activation leads to sustained memory activity during the delay period. The stepwise increase in input frequency by 5 Hz activates successively more dendrites, resulting in larger memory

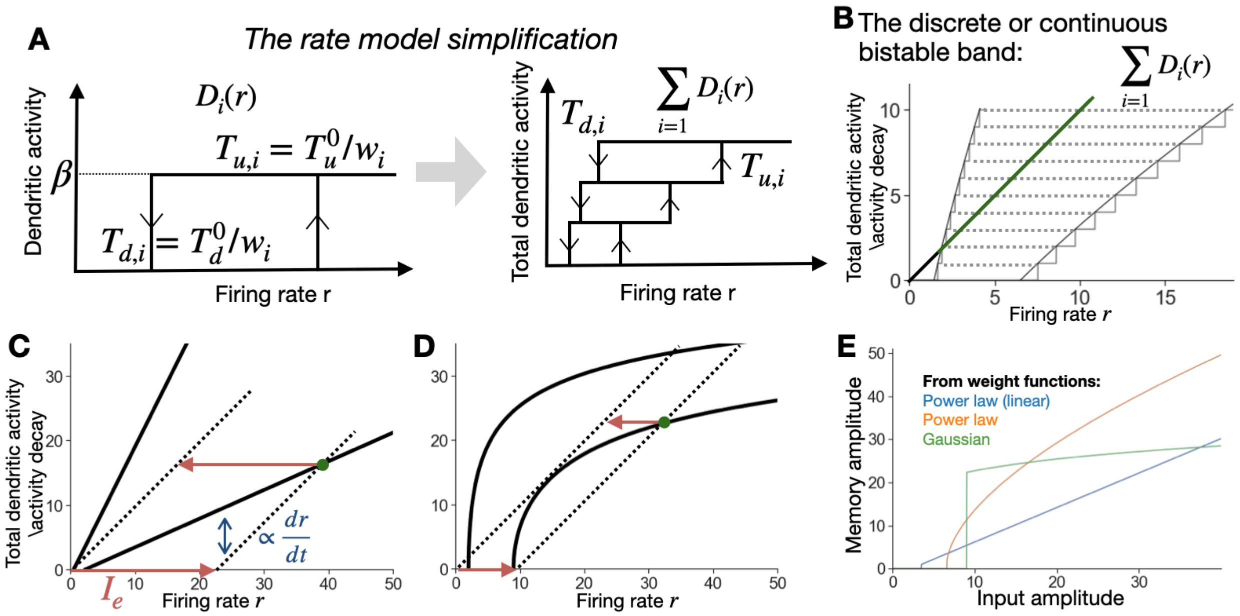

Figure 2.3. [The rate-based autapse model for amplitude encoding. A, Rate model to capture key behavior in the spiking model in Fig. 2.2B. Left, the input-output relation $D_i(r)$ of a dendrite i. The y-axis represents dendritic activity contributed to the somatic firing rate. $D_i(r)$ is flat with jumps at up and down thresholds, whose values are inversely related to synaptic weight value w_i. Right, multiple $D_i(r)$ with a common x-axis. Dendrites have smaller g_{NMDA} values at the bottom. The y-axis is changed to the total dendritic activity of all dendrites. B, A bistable band in the discrete or continuous case. Multiple dendrites form a band with two lines B_u, B_d for up and down thresholds, either in discrete jumps or smooth approximations. The green line represents the memory attractor, where the total dendritic activity, the gray lines, equals the activity decay, the diagonal line. The curve in the continuous limit agrees with the discrete case through step corners. C, How to encode memory. An external input I_e raises the stable point from baseline to the green dot. The rate at which the firing rate changes during the encoding period is proportional to the difference indicated by the blue double arrow. Its removal moves the stable point horizontally leftward to a point within the line attractor. D, Same as C, but with a different weight function. It leads to a different bistable band, where memory is relatively high even when the external input is barely encodable. In other words, the memory amplifies the external input. E, The mappings between input and memory amplitude under different weight functions.

activity. The process demonstrates how dendrites can function as separated memory units, enabling amplitude encoding through dendrite-wise recruitment.

2.3.1.2. The core mechanism for amplitude encoding

For ease of analysis, we simplified the spiking autapse to a rate model, as shown in Fig. 2.3. Some approximations are made: 1. we assume there are N independent dendrites; 2. for dendrite i, the dendritic input-output relation $D_i(r)$ has flat up and down-states, as shown in Fig. 2.3A, left. We treat the contribution of each up-state dendritic voltage to the firing rate as a constant β, while a down-state causes no firing; 3. the conductance dependence of thresholds is simplified to be $T_{u,i} = T^0_u/w_i$ and $T_{d,i} = T^0_d/w_i$, where g_{NMDA} in the spiking model is simply captured in the rate model by the synaptic weight w_i, and T^0_u, T^0_d are constants. Because all $D_i(r)$ relations share the same horizontal axis, firing rate r, they can be stacked to reflect the summed effect of all dendrites, $\sum_{i=1}^{N} D_i(r)$. This is plotted in Fig. 2.3A, right, with the vertical axis showing the total dendritic activity of all up-state dendrites. We will use this rate model simplification for the rest of the paper.

How does a neuron with multiple bistable dendrites maintain a finely discretized set of values in memory? The summed effect of dendrites, $\sum_{i=1}^{N} D_i(r)$, forms two lines B_u, B_d, consisting of effective up and down threshold values. The region between them creates a bistable band. In the continuous limit, as shown in Fig. 2.3B, the B_u and B_d lines become two lines. The equilibrium memory states exist if the total dendritic activity equals the firing rate, $\sum_{i=1}^{N} D_i(r) = r$ (See Materials and Methods). This results in a finely discretized set of stably maintained values, represented by the intersection of the bistable band with the green unity slope line in Fig. 2.3B. Notably, the existence of the near-continuum of stable states does not require fine-tuning of parameters due to the bistable band.

Amplitude encoding is achieved by recruiting dendrites to the up-state. Starting with all dendrites in the down-state, an external input I_e sets the new stable condition to be $\sum_{i=1}^{N} D_i(r) = r - I_e$, moving the dashed line rightward, as shown in Fig. 2.3C. Once I_e exceeds the minimum up threshold, it causes dendritic activation, leading to recurrent excitation, and the system stabilizes at the green dot. The removal of I_e reduces dendritic input, moving the stable point horizontally leftward to the line attractor. The formed memory has an amplitude proportional to the number of up-state dendrites, multiplied by a factor of β. If larger I_e is applied, more dendrites get recruited, moving the green dot higher and resulting in a larger memory.

During the encoding period, the speed at which the firing rate changes is proportional to the vertical difference between line $\mathrm{B_u}$ and $\mathrm{r - I_e}$, as indicated by the double arrow in Fig. 2.3C (see Materials and Methods, Eq. 2.11). For this paper, we focus on maintaining a stable memory of the amplitude of an input; see previous works [47,56,74] on temporal integration of an input in models with a similar implementation based on bistability.

2.3.1.3. How the weight function affects memory performance

The mapping of input and memory amplitude is determined by the synaptic weight function. As shown in Fig. 2.3C, any increase in $\mathrm{I_e}$ moves the stable point upward along line $\mathrm{B_u}$, recruiting more dendrites for memory. The shape of line $\mathrm{B_u}$ (or $\mathrm{B_d}$) is controlled by the synaptic weight function. For example, a linear mapping between input and memory amplitude is possible, with $\mathrm{w(x) = a/(z+b)}$ for positive constants a and b (see Materials and Methods). An example is shown by the blue line in Fig. 2.3E. On the other hand, for different weight profiles, the mapping can exhibit a saturating shape (Fig. 2.3E, orange curve). Such a saturating shape implies that a larger portion of the firing rate range of the neuron is devoted to smaller input values and could be useful if a task demands more sensitivity to low-amplitude stimuli. Generally, depending on the task demand, the corresponding optimal weight function may be formed through plasticity.

In addition, it may be desirable for a system to be able to have a small threshold for inputs to generate large, very stable memory activity. This is possible if there are many dendrites with weights similar to the largest weight, leading to bistable band shapes such as that shown in Fig. 2.3D. Once the small signal activates the dendrite with the largest weight, the recurrent excitation can activate these dendrites, leading to a big jump in activity, as shown by the green line in Fig. 2.3E. By adjusting the weight function, the model can control the trade-off between amplifying a small signal [13] or faithfully encoding the small signal.

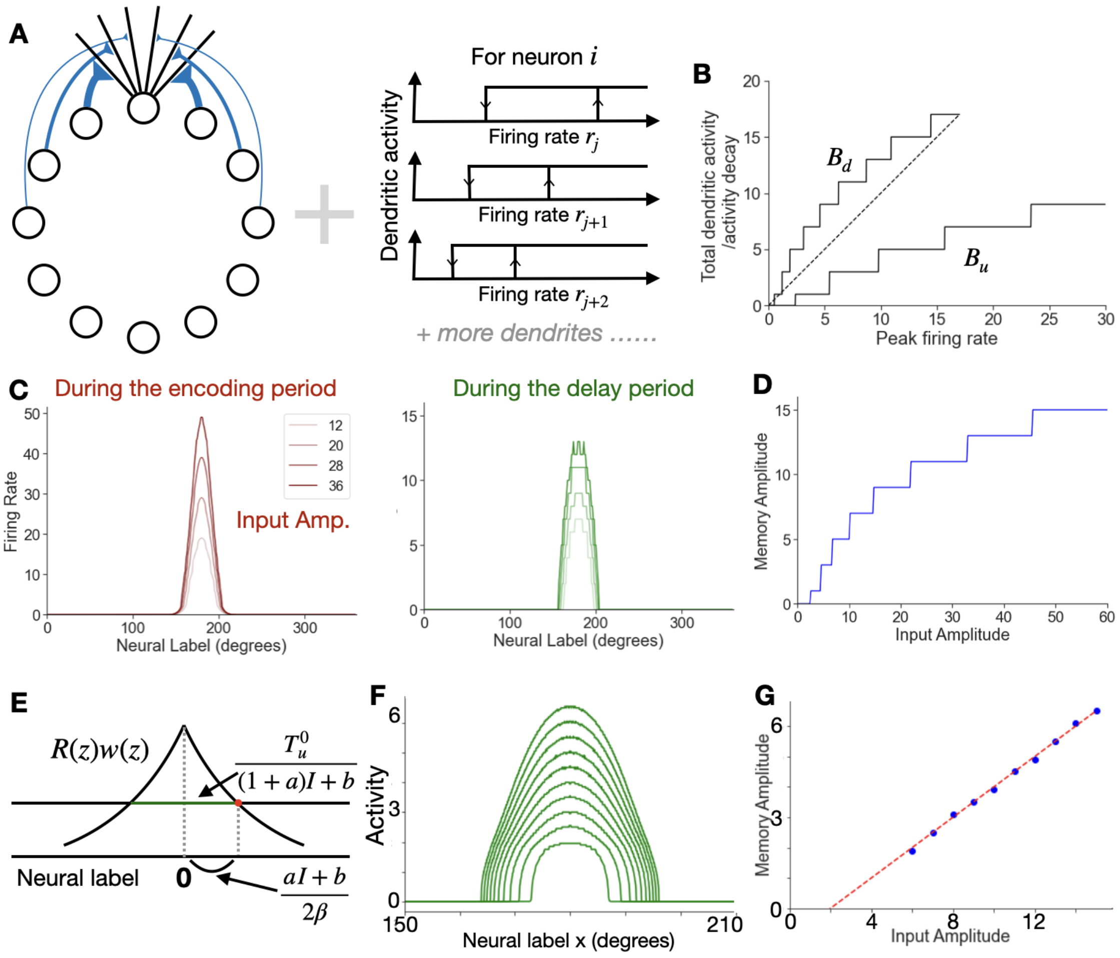

Figure 2.4. Amplitude encoding in the ring model with dendritic bistability. A, Network architecture. Left, each neuron receives recurrent inputs through separate dendrites. The connectivity is excitatory and stronger for nearby neurons. Right, each dendrite is bistable with thresholds smaller for larger connectivity strength. B, Approximated bistability band of the peak firing rate neuron, acquired from simulations with time parametric (See Materials and Methods). C, Amplitude encoding. The input is Gaussian. Each input amplitude induces a different memory amplitude of the same shadeness. D, The mapping between input and memory amplitudes in C. E, Comparing effective up threshold and effective inputs. It is an illustrative step of the underlying network dynamics during the encoding period. See Materials and Methods. F, Memory activities for under a particular weight function. As the input amplitude increases linearly, the resulting memory activity is plotted. The simulation is done in the continuous limit, with a weight function calculated to approximate a linear mapping between input and memory. G, Matching a desired linear mapping. Blue dots: the simulated mapping between input and memory amplitudes for F. The red dashed line: the desired linear mapping.

2.3.2. Working memory in the ring model with bistable dendrites

Here we implemented dendritic bistability into the classical ring network conventionally used to store location information. Our model has N = 360 neurons, as shown in Fig. 2.4A, with two main modifications. First, instead of point neurons, each neuron has N dendrites. Each recurrent connection from neuron j goes into a separated dendrite of neuron i with weight w_{ij}. Second, each dendrite is bistable (See Materials and Methods).

This network encodes amplitude in a way similar to the autapse. The main difference is that instead of self-excitation, this network relies on mutual excitation to maintain the memory. Unlike the autapse, each dendrite of a given neuron depends upon a different presynaptic neuron for its activation. As a result, the bistable band of activity defined by the activations of each dendrite of a neuron as that neuron's firing rate increases and decreases now depends explicitly on the pattern of activity generated in other neurons. Nevertheless, for a given amplitude and shape of network activity, one can numerically plot such a bistable band for a given neuron, providing a qualitative analog to the analytical insights afforded by the autapse model (Materials and Methods). This is demonstrated for the peak firing neuron for a Gaussian-shaped external input in Fig. 2.4C. Under different input amplitudes, the network can maintain different memory amplitudes, as shown in Fig. 2.4C.

The input-memory amplitude relations are determined by the weight function similar to the autapse case. An example is shown in Fig. 2.4D. In general, it is possible to derive an approximate weight function to achieve any desired relation. An example is shown in the continuum limit in Fig. 2.4F, G, where a linear relation is achieved for the calculated weight function, see Materials and Methods. This relation agrees well with the theoretical prediction, represented by the red dashed line in Fig. 2.4G.

There exists a trade-off between the amplitude and the minimal width of a local memory. As shown in Figures 2.4D and 2.4F, the memory bump is wider for a larger amplitude. This is because, to achieve a larger amplitude, more dendrites need to be maintained in the up-state for the peak neuron. This requires these dendrites to receive strong enough recurrent input, which necessitates strong firing of more surrounding neurons, making the bump wider. In the case of holding multiple items in memory, a wider bump is more likely

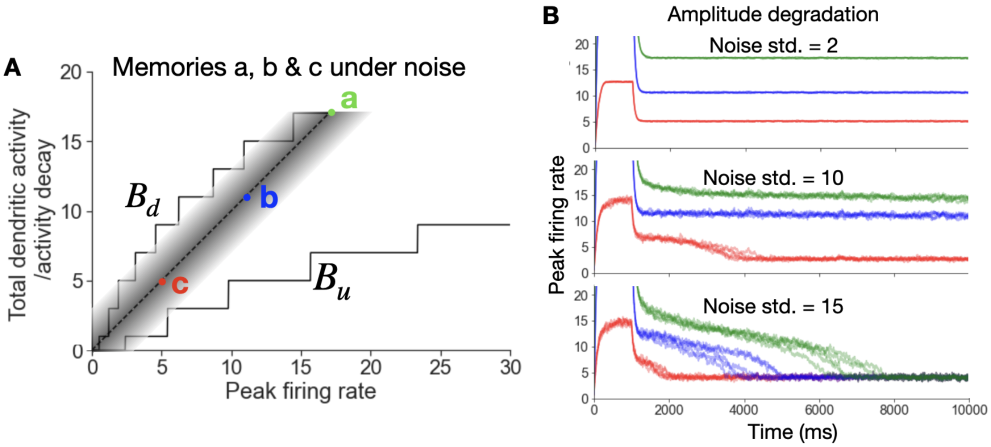

Figure 2.5. Drift of memory amplitudes under noise. A, How noise affects existing memory amplitudes. Three illustrative stable points (a, b, c) are plotted to represent encoded memory amplitudes. Noise will drive them left or right along the flat up- or down-state unless the stable point crosses the thresholds. B, How memory amplitude drifts over time. The y-axis represents the estimated peaking firing rate. Memories remain stable under small noise (top). Under moderate noise (middle), the high-amplitude memories (green) decay, while the low-amplitude memories drop to the baseline activity. Sufficiently large enough noise erases all memories (bottom). 5 trials are shown under each condition. See Materials and Methods.

to interfere with other memory bumps and cause errors in maintaining the items. Indeed, the memory network here can maintain quite general multimodal shapes, although we focus on the storage of unimodal, Gaussian inputs in this paper.

2.3.3. Memory robustness under perturbations

2.3.3.1. Amplitude drift

Noise may degrade memorized amplitude in different ways, depending on the local geometry of the memory activity. This idea is illustrated in Fig. 2.5A, with three exemplar memories stabilized in points a, b, and c. For memory ‘b’, a moderate level of noise drives its stable point back and forth horizontally (gray shaded) without crossing line B_d. Therefore, the fluctuation has no effect on average, although this robustness comes with a trade-off in sensitivity (See Discussion). In contrast, the same noise is enough to drive the memory ‘a’ out of the bistable band, deactivating some dendrites and leading to memory amplitude decay. With a slightly lower amplitude of ‘a’, the distance between the stable point and line $\mathrm{B_d}$ is wider, which can prevent further decay. We define such restabilization of memory at a lower amplitude as the ‘decay mode’. When noise drives memory ‘c’ across line $\mathrm{B_d}$, the reduction in amplitude leads to more and more dendritic deactivation, as the distance goes narrower. We define the reduction of amplitude to the baseline as the ‘drop mode’. Note that, for the shown example, it is less likely that the noise drives the stable point rightwards across line $\mathrm{B_u}$, leading to a higher amplitude, as this requires the noise magnitude to be greater than the input amplitude.

In Fig. 2.5B, we plot the memory traces in the presence of independent Gaussian noise applied to each soma for three amplitudes with 5 trials each (Materials and Methods). When the standard deviation of the noise input applied at each 1 ms time step is 5, it is too small to drive the stable point across the bistable band, and all amplitudes are maintained (top). A larger noise, std = 10, causes the memories with a higher amplitude, green lines, to slowly decay to a lower level (middle). In contrast, the memories with a lower amplitude, red lines, drop to the baseline level activity. Further increasing the noise makes all memories drop to the baseline activity (bottom). The details of how memory degradation occurs depend on the shape of the bistable band, the input magnitude and the noise level. See the Discussion for possible interpretations of these modes.

2.3.3.2. Local memory and the effect of recurrent inhibition

Different from the classical ring model, the ring model with dendritic bistability does not require recurrent inhibition to stabilize a localized memory. Classically, a ring model [5] needs recurrent inhibition to maintain a localized bump of neural activity [37]. Our model with dendritic bistability provides another solution. As long as the noise in dendritic input does not change the dendritic state, it causes no further change to the activity and does not destabilize the memory. In Fig. 2.6A, we compare the bistable model with its classical

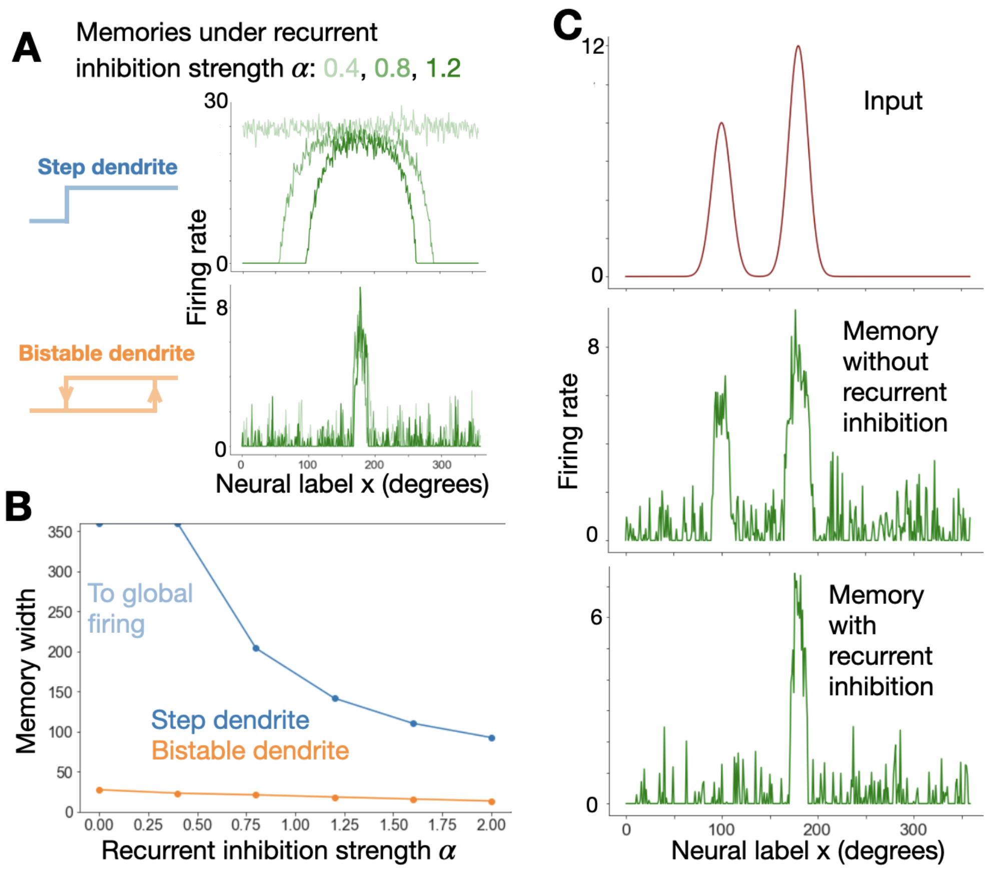

Figure 2.6. The effect of recurrent inhibition. A, Under different inhibition strength values. Three memories are plotted, each with a different value of recurrent inhibition strength value. For a step function dendrite, the existence of local memory requires sufficient recurrent inhibition. But for a bistable dendrite, a local memory always exists. B, Memory width under changing inhibition strength. Memory width is firstly estimated at the half value of the maximum memory activity and then averaged over 10 trials. See Materials and Methods. C, Recurrent inhibition reduces distraction. For a central, strong input with an additional smaller perturbation aside, recurrent inhibition is desired to suppress irrelevant information.

limit, where each dendrite follows a step function. For the step function, a local memory is possible only if the recurrent inhibition is strong enough. In contrast, for the illustrated set of parameters, the bistable dendrite model is much less sensitive to the recurrent inhibition values. Fig. 2.6B further shows the averaged width for a range of recurrent inhibition values. However, recurrent inhibition can still be helpful even in the bistable model. Adding recurrent inhibition can suppress irrelevant, distractor signals. An example of this is shown in Fig. 2.6C, where the larger input peak is the real signal and the smaller peak is from the distractor input.

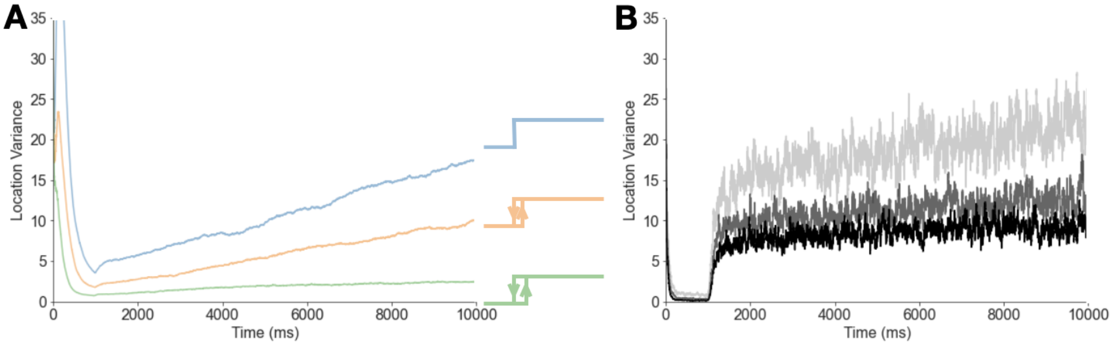

Figure 2.7. Diffusion of memory locations under noise. The input is turned off after 1000 ms. The location variance is obtained by averaging over 400 trials. A, Dendritic bistability reduces location diffusion. Each line is generated with a different dendritic bistable range. The reduction is significant even when the range of bistability is very small. Note that the bistable range drawn is for illustration, not drawn to scale relative to ones in Fig. 2.6. B, Larger input amplitude reduces location diffusion. A larger input amplitude (darker trace) results in a smaller location diffusion over time.

2.3.3.3. Location diffusion

Memory location undergoes diffusion under noise. Compared with the diffusion in the classical limit in Fig. 2.7A (blue), the diffusion is drastically reduced when dendrites become bistable (orange and green). Note that even for a bistable band width as small as 0.4 (green), the diffusion is almost negligible. Similar reduction has also been previously modeled with bistable neurons [18]. Interestingly, the variance of location still grows approximately linearly with time, exhibiting a diffusion process as proposed by previous analytical models [17,57]. Mechanistically, bistability reduces drift because changing the pattern of firing rates requires turning on or off dendrites, which requires having sufficiently large noise to cross dendritic thresholds. In addition, a higher memory amplitude leads to smaller diffusion, as shown in Fig. 2.7B. This occurs because the memory with larger activity exerts more recurrent inhibition to suppress the diffusion. See the Discussion for a potential relation of this observation to the precision of encoding stimuli.

2.4. Discussion

Amplitude encoding in a given location is a challenging problem for neural circuit models. Here we propose a network model based on dendritic bistability. By recruiting different dendrites, the model is able to encode amplitude in a biologically plausible way without fine-tuning parameters. The bistable band formed by the weight function controls the input-memory mapping, to satisfy different memory requirements. Furthermore, we have shown memory robustness under noise, both in amplitude and location, is provided by dendritic bistability.

It is possible to extend from the current focus on a single local memory to a multimodal memory; the same dynamic equation, Eq. 2.21, still works if the input is multimodal with different amplitudes. Interference can be caused by the widths of local memories, where the trade-off between amplitude and width mentioned above can play a role. On the other hand, if the network consists of a pool of separate autapses, a graded input pattern with multiple values can be encoded with pixel precision by mapping them to these autapses.

2.4.1. Biological plausibility and alternative implementations

Dendritic bistability is the building block of our model. The utility of this concept was first modeled by Lisman, et al. in a conductance-based model based on widely distributed NMDA receptors [61]. A subsequent work [82] demonstrated that, with the help of the KIR current, dendritic bistability can exist for a wide range of cellular parameters. NMDA receptors have been demonstrated to be critical to working memory function [90, 91, 100]. However, it is unclear whether this is simply due to their slow kinetics or is more fundamentally related to their potential ability to mediate bistable dendritic activity. Physiological recordings have shown that NMDA receptors in dendrites can cause prolonged voltage plateaus lasting hundreds of milliseconds [68,83,95], but more direct evidence is needed to determine whether dendritic bistability can sustain itself for seconds in working memory tasks.

In our model, each dendrite functions as a separate memory unit, and the total number of dendrites per neuron limits the maximum resolution for amplitude encoding. Long-term plasticity and neuron morphology may help to increase this number. Experiments suggest that each dendrite can function as a separate unit [63, 79, 88], possibly even at the level of individual spines [6, 27]. In simulations, we set the synaptic weight to decay rapidly so that only about 20 dendrites are recruitable for each neuron, while other dendrites have effectively zero synaptic weight.

In addition to dendritic bistability induced by NMDA-receptors, there are alternative ways to achieve bistability. They include cellular bistability/multistability mechanisms [69, 103] or network-level bistability [56]. Such mechanisms may offer alternative implementations for our model. For example, dendrites of a currently modeled neuron can be mapped to separate neurons which receive recurrent inputs. These neurons then project to a common neuron, which is corresponding to the soma and provides recurrent output to other neurons. Such an organization with receiving neurons converging to an output neuron may exist in minicolumn. In this alternative implementation, those neurons mapped by dendrites need cellular bistability, while cellular bistability [58], or multistability [34], have been observed in the brain. In this alternative implementation, layer 4 neurons need cellular bistability. In this way, there could be a much larger number of separated bistable units compared to bistable dendrites in a single neuron, potentially increasing amplitude resolution.

Moreover, the memory is essentially stored as activated dendrites, which function similarly to potentiated synapses. Interestingly, a small number of studies have shown that short-term potentiation can depend on presynaptic input and NMDA receptors [35, 76].

2.4.2. Sensitivity to external input and robustness under noise

Our model involves a trade-off between sensitivity and robustness. A wide bistable range in dendrites is good for robustness under noise. However, this comes at the cost of reduced sensitivity to small external stimulus change, as it becomes more difficult to overcome the thresholds. This insensitivity may not be a problem if the small stimulus increase actually induces a large input jump as can occur for a band shape such as that shown in Fig. 2.3D. However, in general, if a system needs to encode a smaller input than that required to cross the threshold for activating dendrites, this can be a problem. For example, experiments studying the encoding of head direction coding in Drosophila even small head velocity inputs can move the bump of activity that encodes head direction. At least in some systems, the trade-off between sensitivity and robustness may be mitigated by allowing the dendritic input-output relation to switch between bistable and sigmoid. This can be realized by effectively adjusting synaptic weights or, on a shorter time scale, by modifying dendritic current [82]. A potential mechanism for the latter could be dendritic disinhibition [99] which could provide a fast, active switch when the task requirements shift between sensitivity and robustness.

2.4.3. Interpretations for memory amplitude and experimental connections

Memory amplitude can be interpreted in two common ways: as the input intensity at a location or as the precision of the location. Input intensity could be the strength of a stimulus in tasks where the stimulus amplitude needs to be encoded, regardless of the amplitude over which it is presented. Alternatively, the input intensity could represent an accumulation of the external input, as occurs in accumulation of evidence tasks [3,4,30]. In this case, a band with edges parallel to the unity line would enable inputs beyond a threshold strength to be mathematically integrated, although inputs smaller than this threshold would be ignored [47, 56, 74]. Bayesian models of sensory coding have alternatively suggested that the amplitude of neuronal firing can encode the precision of a stimulus, with higher firing rates indicating higher precision. In this case, the ability to robustly sustain multiple firing rates could be useful for stably remembering the precision of stimuli, for example for use in cue combination tasks that require combining stimuli according to their uncertainties [57,97]. Altogether, this diverse set of applications suggests a need for the nervous system to be able to robustly store activity in a manner that encodes not only categorical or positional information, but also graded amplitudes. While many models of working memory focus primarily on mechanisms governed by network connectivity, with very simple single-neuron response properties, this work demonstrates how the nonlinear capabilities of active dendrites may provide a cellular mechanism that complements and makes robust network interactions.

CHAPTER 3

Encoding a novel pattern in working memory via dendritic bistability

Working memory can hold a variety of information and is crucial for understanding cognition. Commonly, people study the encoding of familiar and simple information, such as hue, orientation, or head direction. However, real world information can be novel and complex. In this study, we investigate how to encode a novel pattern with graded values. We propose an unstructured network model in which each neuron has multiple bistable dendrites. The dendritic dynamics effectively give fast Hebbian plasticity and enable robust encoding of a novel pattern under various perturbations. Through analytical study, we characterize the network dynamics of memory encoding and retention. We further identify conditions for perfect memory, and characterize different error patterns due to neuron interaction and saturation. Additionally, our model exhibits resistance to strong inhibitory perturbations. Our model provides an example of how dendritic computation can solve a hard working memory task.

3.1. Introduction

Working memory refers to a diverse range of information that can be temporarily held in mind. The content can include hues, locations, words, odors, music excerpts, or images. Working memory serves as the foundation for many advanced cognitive processes, such as decision-making, thinking, and recall. Despite its importance, the underlying mechanisms of working memory are not well understood.

The majority of working memory models focus on familiar and simple information. For example, binary information, such as a decision out of two choices, can be represented by the activity of one of two neural groups [42,86,93]; a scalar, such as the shade of a stimulus or the muscle strength for controlling eye position, can be represented by the amplitude of memory activity in a non-periodic line attractor [20,46,59,84]; and a periodic scalar, such as a hue on a color wheel or a specific orientation among 360 degrees, can be represented by local persistent activity in a periodic, type of line attractor known in one-dimensional as a ring attractor [5,23,26].

In these models, an attractor in activity state space is required to attract memory activity into a certain stable and stereotypical form. However, the formation of such an attractor is typically assumed to require a prolonged training to adjust network connectivity, demonstrating a lack of flexibility in encoding a novel memory without a pre-existing attractor.

Novel encoding, which refers to the ability to hold a working memory without any prior training, is crucial for human cognition. In everyday life, we often encounter novel information and can make flexible decisions based on it. Such information may include a new word, a new odor, a new mathematical rule and so on. On the other hand, a prolonged training signal may not always be available during attractor formation, especially when dealing with a fleeting novel input. Therefore, novel encoding is necessary for holding the input and providing the necessary prolonged signal [50].

Despite its importance, there are few efforts to model novel encoding and they primarily focus on encoding simple information. They are generally divided into two approaches: no training at all or fast training. The first approach often relies on cellular mechanisms [40, 49, 69, 103]. When memory is maintained by intrinsic mechanisms within a single neuron, it may bypass the training happening between synaptic connections between neurons. On the other hand, although training is usually slow, the second approach assumes fast Hebbian plasticity, enabling one-shot training to encode a novel input [38,81]. Experimental evidence exists for fast Hebbian plasticity [36,76,80], but it is not yet strong. Additionally, for other forms of fast plasticity, such as short-term plasticity or behavior time-scale plasticity [14], the connection to novel encoding remains unclear.

The memorized information for the novel encoding models mentioned above is unrealistically simple. For models based on cellular mechanisms, only a binary or graded value can be held in a neuron. For models using fast Hebbian plasticity, memory is represented as a binary local pattern, neglecting any graded-value information [38,81].

In this paper, we model novel encoding of a pattern with graded firing rates across neurons. This novel pattern can represent flexible content, whether explicit or implicit. For example, it could be a newly defined word, where each neuron maps to a word with a known definition, and its firing rate reflects the relevance of that word to the new one. Alternatively, it may represent an abstract, encrypted form of information, such as a new odor that elicits a distributed firing pattern.

We propose a network model for encoding a novel pattern. The model consists of a homogeneous network where each neuron possesses multiple bistable dendrites. These dendrites can be achieved by NMDA receptor dynamics [61,82] and effectively exhibit a form of fast-Hebbian plasticity, enabling novel pattern encoding. The structure of the paper is as follows: First, we discuss the basic performance and show the robustness of novel encoding under various perturbations. Next, we explain the detailed network dynamics during the encoding and memory periods. Then we analytically derive the input function for a perfect memory and identify various error patterns. Additionally, we show memory resilience to a strong, localized external inhibitory input that silences a small portion of the network. Finally, we discuss biological plausibility, different model interpretations, and comparisons with two other pattern encoding approaches.

3.2. Results

3.2.1. Network structure and basic performance

To demonstrate the principle of novel encoding in a network of neurons with multiple bistable dendritic compartments, we first constructed an idealized network with $\mathrm{N} = 2500$ neurons, all-to-all connected with weights equal to 1 (Fig. 3.1A). Each neuron is identical and has $\mathrm{N} = 2500$ identical dendrites (Fig. 3.1B). Each dendrite receives a recurrent dendritic input (Fig. 3.1B). All dendritic outputs are summed in the soma along with the external input to determine the firing rate. The core structure lies in each dendrite, which has a bistable input-output relation with up and down-thresholds $\mathrm{T_u}$, $\mathrm{T_d}$ (Fig. 3.1C). Each dendritic up-state contributes β to the somatic firing. Importantly, the *somatic effect* of a neuron x can backwardly lower the effective up-threshold $\mathrm{T_u^e}$ of all its dendrites:

$$\mathrm{T_u^e(x)} = \mathrm{T_u} - \alpha \mathrm{f(x)}, \tag{3.1}$$

where $\mathrm{f(x)}$ is the firing rate and α is the strength factor. We also require $\mathrm{T_u^e(x)} > \mathrm{T_d}$. This bistable dendrite, with its a box-like bistable range and the somatic effect, qualitatively captures the performance of previous conductance-based models [61, 82] based on NMDA receptors and the back-propagation of action potentials, which is also reproduced in the spiking model discussed below.

These bistable dendrites function as basic associative memory units. During the encoding period, a dendritic up-state can be activated by the combination of a high presynaptic to

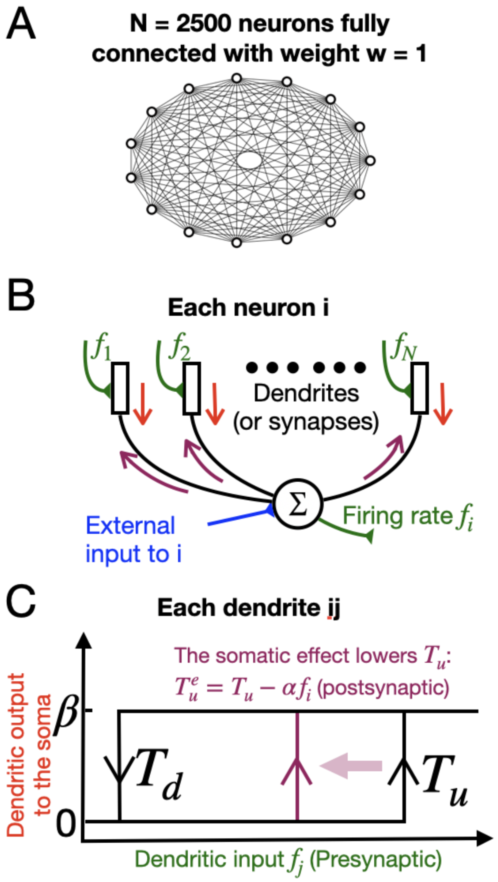
A
N = 2500 neurons fully connected with weight w = 1
B
Each neuron i
f_1
f_2
f_N
Dendrites (or synapses)
Σ
External input to i
Firing rate f_i
C
Each dendrite ij
The somatic effect lowers T_u:
$T_u^e = T_u - \alpha f_i$ (postsynaptic)
Dendritic output to the soma
β
0
T_d
T_u
Dendritic input f_j (Presynaptic)

Figure 3.1. Network structure. A, Unstructured neural network. A fully connected network with N=2500 neurons and uniform weight w=1. B, Multicompartment neuron. Each neuron has N separate computational units, which are assumed to be dendrites. See Discussion for more details about biological plausibility. Each dendrite receives recurrent input from other neurons (with self-connection) and outputs to the soma. The firing rate equals the summation of all dendritic outputs and the external input in the soma. C, Bistable dendrite input-output relationship. The dendritic input and output follow a bistable relation with up and down thresholds T_u and T_d. In addition, the somatic firing f_i can effectively lower T_u. The effective up threshold is modeled by $T_u^e = T_u - \alpha f_i$, which qualitatively captures the effect from the voltage backpropagation from the soma to all dendrites of a neuron. See Fig. 3.8 for a spiking neuronal network version.

the dendrite and a high postsynaptic firing rate, which lowers T_u^e. This effectively acts as fast Hebbian plasticity with binary values, but is based on activity instead of actual weight change. During the memory period, as long as the dendritic input remains higher than T_d, a dendritic up-state is maintained, and the memory activity of a neuron is proportional to the total number of its up-state dendrites.

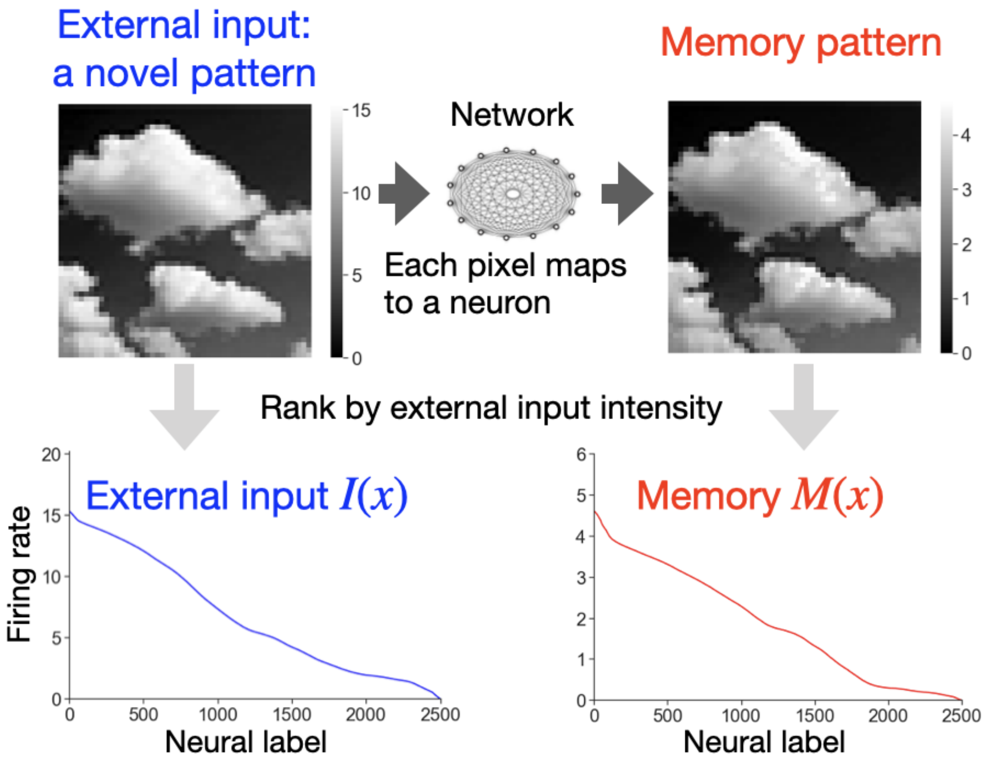

Figure 3.2. Example of novel pattern encoding. Top panels: Performance in two-dimensional representation. Left, a graded cloud picture (50x50) is chosen as the input. Each pixel maps to a neuron in the network, with its intensity equal to the neural firing rate. Right, the resulting memory maintains a similar graded pattern despite some small distortion. Lower panels: One-dimensional representation of the firing rates of the neurons, ordered according to the input firing rate, which equals their input intensity. Left, external input to each neuron. Right, memory activity of each neuron following the offset of the external input.

An example of the performance of our model is shown in Fig. 3.2, top. A graded cloud picture with 50×50 pixels is chosen as a novel pattern. Each pixel maps to a separate neuron within the network. During the encoding period, the external input strength received by each neuron equals the corresponding pixel intensity. During the memory period, after the external input is turned off, a stable activity pattern maintains graded values similar to the original input despite some distortion. See Appendix A for simulation details.

For ease of analysis, we convert both the external input and the resulting memory into a one-dimensional coordinate. Because the network is unstructured, each neuron is anatomically equivalent. Without loss of generality, we can relabel the neurons based on the external input intensity in descending order, resulting in a non-increasing intensity values $\mathrm{I_i}$, which is approximated to be a continuous function I(x), as shown in Fig. 3.2, bottom left. In addition, as explained in more details below, because a larger input induces a larger memory response, the memory is also a non-increasing M(x) under the same neural label.

Before we proceed, there are some clarifications: 1. We chose a large value of N for ease of analysis. While N = 2500 separate dendrites may be too large for a neuron, and a novel pattern with N = 2500 entries may also be too hard for human memory, a large N value allows the performance of this discrete network simulation to approach its continuous limit, simplifying analysis. See discussion for equivalent model implementations at the synaptic or cellular level, which allow for more computational units. We do discuss a spiking model later in the paper with fewer neurons. 2. The cloud picture is just an example to visualize a novel pattern. Our model is not directly related to how a cloud picture is actually encoded in the brain, which involves multilayer information processing from the photoreceptors of the eye to visual cortex. 3. The minimum external input is set to zero for simplicity. The following results still hold if it is larger than zero. 4. We focus on the memorization of relative intensities within an input pattern. The averaged firing rate difference between input and memory is ignored. We assume that the difference, which is captured by an overall scaling factor, can be encoded by a separate parametric working memory system [16,46,56].

3.2.2. Memory robustness against perturbations

A major feature of our model besides encoding a novel pattern flexibly is that the encoded memory is robust against various perturbations. We first show the robustness against a small somatic noise level during the memory period in Fig. 3.3A. The noisy memory can be decomposed into two parts: a fluctuating part, where the noise drives the memory up or down (shaded band), and a stable part (dark red line) contributed by up-state dendrites. Notably, the stable part aligns well with the noise-free memory shown in Fig. 3.2 (dotted red line), whose activity is entirely due to up-state dendrites. The reason for this robustness

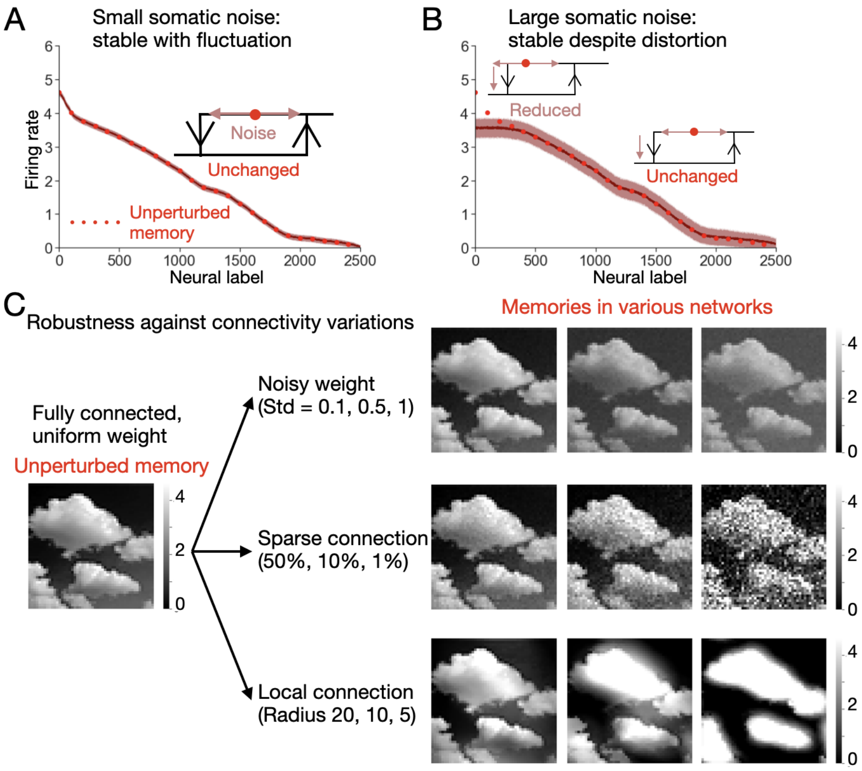

Figure 3.3. Memory performance with somatic noise or connectivity variations. A, Performance in the presence of noise injected to the somatic compartment. Gaussian random noise with zero mean and standard deviation 0.1 is applied to each soma independently. The averaged memory (dark red line) is computed over 500 ms, with a band representing its standard deviation. The dotted line is the memory replotted from Fig. 3.2 for comparison. Inset: A schematic plot illustrating how noise can drive a stable point in a bistable dendrite back and forth along its original stable point. See more details in the text. B, Same as A but with a noise standard deviation of 0.3. Noise not only leads to fluctuations, but also causes the high firing rate part of the memory to drop. Inset: Similar to A, but now some dendrites flip to the down-state due to noise. C, Performance under connectivity variations. Three possibilities are implemented. Left: The unperturbed memory replotted from Fig. 3.2. Top Right: Memories resulting from noisy weight values from a Gaussian distribution of connection strengths with mean 1 and standard deviation 0.1, 0.5, or 1, respectively. Middle Right: Memories resulting from uniform weight sparse connections with a connection probability of 50%, 10%, or 1%. Bottom Right: Memories resulting from a connectivity pattern in which connections only exist between local neurons, within a radius of 20, 10, or 5, in a space defined by two-dimensional pixels. The memory encoding is surprisingly robust for all but the most extreme connectivity modifications. The same graded scale bar is used to help direct comparison. See simulation details in appendix A.

against noise lies in dendritic bistability. As shown in Fig. 3.3A (inset), noise changes the dendritic input, shifting the stable point (red) back and forth. However, small noise is insufficient to drive the stable point across $\mathrm{T_u}$ or $\mathrm{T_d}$, thus causing no change in the state or the resulting dendritic output.

On the other hand, larger somatic noise can trigger a positive feedback loop. Unlike the small noise case, larger noise may cause a dendritic input to cross T_u, leading to larger dendritic output which recurrently increases dendritic inputs. This can create a positive feedback loop, where excitation leads to more recurrent excitation, it causes runaway excitation that potentially can destabilize the system. However, it is not always the case in our model. If a dendrite is already saturated in the up-state, the increased dendritic input does not change the dendritic output. Similarly, if a dendrite is in the down-state, the increased dendritic input may not be strong enough to activate the dendrite to the up-state. Therefore, the threat of a positive feedback loop can be mitigated by zero feedback. Similarly for the case related to T_d. An example is shown in Fig. 3.3B, where noise drives some dendrites in high-firing neurons from the up-state to the down-state, while low-firing neurons remain almost unaffected by both the noise and the recurrent changes.

Somatic noise has other effects. In either the small or large noise case shown in Figures 3.3A or 3.3B, the averaged memory in low-firing neurons is slightly larger than the unperturbed memory. This occurs because firing rates are non-negative, and the noise can not drive them below zero, resulting in a net excitatory effect. In addition, if noise is applied during the encoding period, it is encoded along with the novel input pattern, as the network makes no distinction between input arising from the novel pattern or the noise.

We further show the memory robustness against variations in connectivity. For visual clarity, memories are presented as two-dimensional pixels with the same graded scale bar as in Fig. 3.2B. Three common variations are considered in Fig 3C: random weight values in a uniformly connected network (top), sparse connectivity (middle), and local connectivity, where connections only exist between neurons that are close enough in two-dimensional space (bottom). In each case, the network can robustly encode a novel pattern similar to the unperturbed case (left), unless the variations are extreme. For example, the memory is clearly degraded when the standard deviation of weights approaches its mean value 1 (top, rightmost); when the connection probability drops to 1% (middle, rightmost); or when the connectivity exists only within a narrow radius of 5, in a space of 50 × 50 pixels. See Appendix A for implementation details.

3.2.3. Detailed network dynamics during the encoding and the memory periods

How can a graded pattern be encoded? In this section, we provide a detailed breakdown of the network dynamics during the encoding and memory periods. Because of the particular importance of the encoding dynamics, we start by developing some intuition for this dynamics. Fig. 3.4A shows neurons in the network, where a higher-firing neuron has a lower $\mathrm{T_u^e}$ value. All dendrites of a particular neuron $*$, as shown in Fig. 3.4B, receive graded dendritic inputs from recurrent connections and share the same $\mathrm{T_u^e}$ determined by the firing rate of neuron $*$. A dendrite is activated if the input it receives exceeds $\mathrm{T_u^e}$. As shown in Fig. 3.4C, the total dendritic output is given by counting all activated dendrites and multiplying by the up-state contribution factor β. The soma then integrates this total dendritic output with the external input it receives to get an updated firing rate. Through this procedure, each neuron receives identical but graded dendritic inputs due to the uniform connectivity. What differentiates the activation in each neuron is the somatic effect. A neuron with a higher firing rate has a lower $\mathrm{T_u^e}$ for all its dendrites, allowing at least the same amount of, or possibly more, dendrites to activate. As a result, as shown in Fig. 3.2, the memory activity proportional to the number of activated dendrites follows a non-increasing function if ranked by the input intensity.

Next, we formally describe the encoding dynamics in the continuous, noise-free limit in which the continuum identity of neurons is labeled by x. During the encoding period, firing rates are non-decreasing such that only the $\mathrm{T_u}$ branch of the dendritic bistability gets involved.

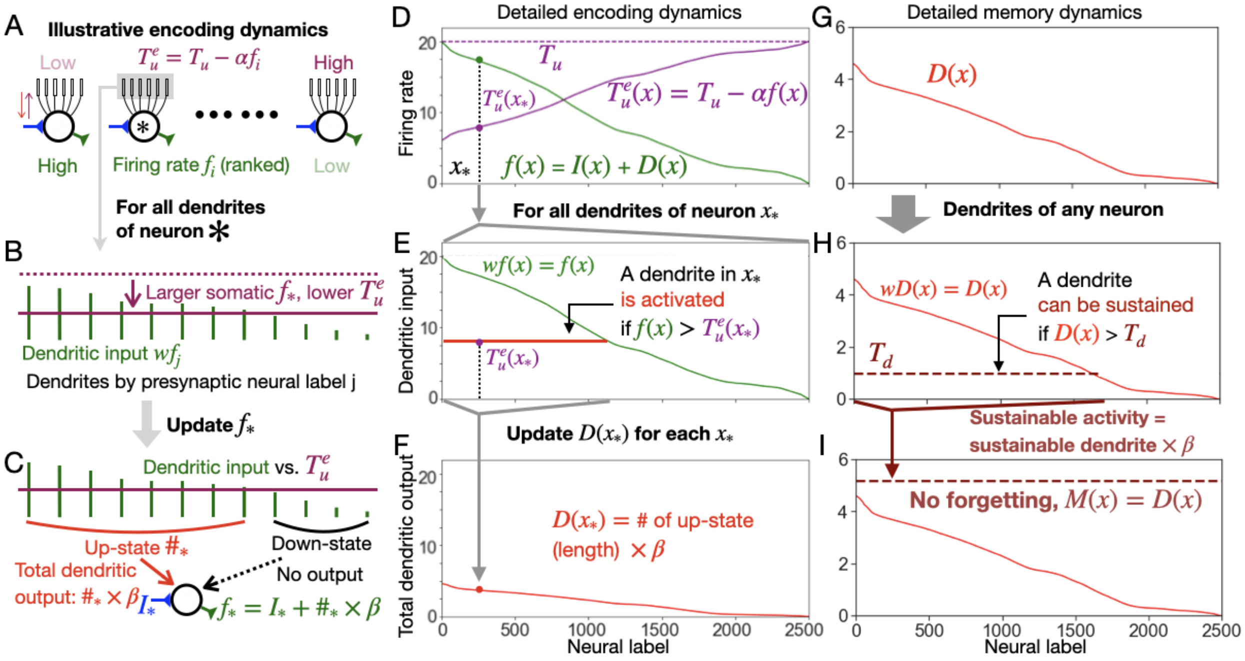

Figure 3.4. Detailed dynamics during the encoding and memory periods. A-C, Illustrative plots of encoding dynamics. A, A list of network neurons ranked by external input intensity, each of which has multiple bistable dendrites. A neuron with a high firing rate has a low $\mathrm{T_u^e}$ due to the somatic effect. B, All dendrites of a given neuron. Dendrites are labeled by their presynaptic neurons. The somatic effect uniformly reduces $\mathrm{T_u^e}$ (purple bar) of all its dendrites. Dendrites also receive recurrent inputs of graded values (green sticks) through uniform weight $\mathrm{w} = 1$. C, How firing rate is determined. If a dendritic input exceeds $\mathrm{T_u^e}$, the dendrite is activated. Each activation contributes β to the somatic firing. The soma sums an external input (blue) with all up-state dendrites, with each contributing β to the somatic firing. Down-state dendrites have no effect. D-F, Detailed encoding dynamics, drawn to illustrate Eq. 3.3. D, The green curve indicates the firing rate $\mathrm{f(x)}$ at a particular time step. $\mathrm{f(x)}$ is the summation of external input $\mathrm{I(x)}$ and the total dendritic output $\mathrm{D(x)}$. The purple curve indicates the effective up-threshold $\mathrm{T_u^e(x)}$, which is inversely related to $\mathrm{f(x)}$. The plot is done at the equilibrium limit, while the dynamics shown are generally true. E, For a particular neuron $\mathrm{x_*}$, all its dendritic inputs, whose values are simply the uniform weight times the firing rates, are plotted. Dendritic inputs are ranked by the same input intensity order as in D. All dendrites in $\mathrm{x_*}$ shares a common $\mathrm{T_u^e(x_*)}$. The portion of dendrites activated is shown by the red segment. F, $\mathrm{D(x_*)}$ is updated to a new value, which is proportional to the total number of activated dendrites, by a factor of β. This new $\mathrm{D(x_*)}$ is fed back to A, iteratively in time, until the system stabilizes. See more details in the text in Results. G-I, detailed memory dynamics, drawn to illustrate Eq. 3.4. G, After $\mathrm{I(x)}$ is turned off, $\mathrm{D(x)}$ remains in the system. Note that the y-axis is rescaled from F. H. The dashed dark red segment shows the portion of dendrites with inputs above $\mathrm{T_d}$. I, The dashed dark red segment shows the maximally sustainable activity. If the total dendritic output $\mathrm{D(x)}$ falls below this line, it means existing dendritic activity is sustainable such that $\mathrm{D(x)}$ becomes $\mathrm{M(x)}$. See more details in the text in Results.

In Fig. 3.4D, the firing rate, $\mathrm{f(x) = I(x) + D(x)}$, sums the external input $\mathrm{I(x)}$, and the total summed output of all dendrites $\mathrm{D(x)}$. For each neuron x, its $\mathrm{T_u^e}$ is lowered by the somatic effect: $\mathrm{T_u^e(x) = T_u - \alpha f(x)}$. Zooming in on all dendrites of a neuron $\mathrm{x_*}$ in Fig. 3.4E, the synaptic inputs to the dendrites of a given neuron is given by $\mathrm{wf(x) = f(x)}$ (due to the uniform weight $\mathrm{w} = 1$), where each dendrite is specified by its presynaptic neural label. If a dendritic input exceeds the effective up-threshold, $\mathrm{f(x)} > \mathrm{T_u^e(x_*)}$, a dendrite is activated. Because of the non-increasing nature of f(x), the total number of up-state dendrites equals the length of the solid red line:

$$\mathrm{f^{-1}(T_u^e(x_*)) = f^{-1}(T_u - \alpha f(x_*))}, \tag{3.2}$$

where we assume the existence of the inverse function $\mathrm{f^{-1}(x)}$. Finally, we have new D(x): $\mathrm{D(x) = f(x) - I(x)}$, as shown in Fig. 3.4F. This is fed back to Fig 4D iteratively until the system is stabilized. It gives a dynamical equation as:

$$\tau \frac{\mathrm{df(x)}}{\mathrm{dt}} = \beta \mathrm{f^{-1}(T_u - \alpha f(x)) + I(x) - f(x)}, \tag{3.3}$$

with a time constant τ, capturing the slowest timescale of neural dynamics. See Appendix A for the discrete version of dynamics with dendritic dynamics.

After dendrites are activated, we turn off I(x), reducing the steady-state value of f(x) to D(x) and check if the activated dendrites can be recurrently sustained. Similar to the encoding case, we zoom in on all dendrites of a neuron $\mathrm{x_*}$, as shown in Fig. 3.4H. An activated dendrite can still be sustained if its dendritic input exceeds $\mathrm{T_d}$, $\mathrm{wD(x)} > \mathrm{T_d}$. Technically, we need to check this inequality for every dendrite. However, given the non-increasing nature of D(x), the total number of sustainable dendrites equals the length of dashed dark red line, $\mathrm{D^{-1}(T_d)}$. If the length exceeds the number of actually activated up-state dendrites for each neuron, $\mathrm{D^{-1}(T_d)} > \mathrm{D(x)}/\beta$, all activated up-state dendrites are preserved with the memory $\mathrm{M(x) = D(x)}$. Equivalently, it happens when the maximum sustainable activity (indicated by the dashed dark red line) exceeds the total dendritic activity (solid red line) as shown in Fig. 3.4I.

Next we examine the conditions under which forgetting of the novel encoded memory can happen . If $\mathrm{T_d}$ is increased, $\mathrm{D^{-1}(T_d)}$ decreases, and the inequality, $\mathrm{D^{-1}(T_d)} > \mathrm{D(x)}/\beta$, may not hold for some dendrites. As the number of actually activated up-state dendrites can never exceed the total number of sustainable dendrites, forgetting happens, described by the dynamics:

$$\tau\frac{\mathrm{df(x)}}{\mathrm{dt}} = \beta\mathrm{Min}[\mathrm{D}^{-1}(\mathrm{T_d}), \mathrm{D(x)}/\beta] - \mathrm{f(x)}, \tag{3.4}$$

where the memory M(x) is given by the stabilized D(x). Notably, the forgetting may not lead to instability, and a graded pattern can still be maintained in this case. The intuition is provided in the text describing Fig. 3.3B.

The above treatment of dynamics assumes that the network is noise-free, and I(x) (or the induced f(x), D(x), or M(x)), is approximately a non-increasing function. In this way, we can view the total number of up-state dendrites as the length of the dashed red line in Fig. 3.4E. In the presence of noise, the dendritic input to the leftmost neuron may be accidentally below $\mathrm{T_u^e}$, making the total number of up-state dendrites slightly smaller than the length of the dashed red line.

3.2.4. Perfect memory and functional separation of two neural groups

Given the dynamics in Equations 3 and 4, we can solve for when a perfect memory arises, i.e., a memory proportional to the input function by a scaling factor so there is no distortion. This requires D(x) ∝ I(x) ∝ f(x) at the end of the encoding period, and no forgetting during the memory period, M(x) = D(x). With detailed algebra in Appendix B, we arrive at a solution for the perfect memory:

$$\mathrm{I(x)} = \mathrm{T_u} - \beta\mathrm{A} - (\frac{\mathrm{T_u}\alpha}{\mathrm{A}} - \alpha\beta)\mathrm{x}, \tag{3.5}$$

$$\mathrm{T_u} \leq (1+\alpha)(\mathrm{T_u} - \beta\mathrm{A}), \tag{3.6}$$

$$\mathrm{T_d} \leq (1-\alpha)\beta\mathrm{A}, \tag{3.7}$$

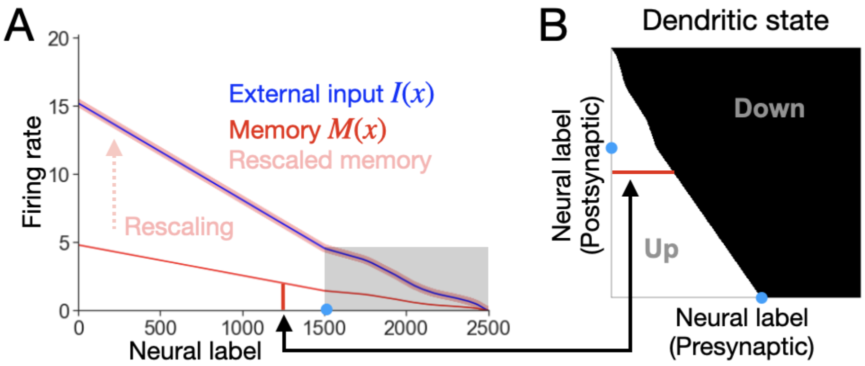

Figure 3.5. Memory that perfectly matches a partially linear input up to a scaling factor. A, A particular form of external input gives a perfect memory. The external input I(x) (blue) follows a partially linear form, separated by the blue dot. For neural labels smaller than the blue dot, the input is linear with a fixed slope determined by the exact location of the blue dot. For neural labels larger than the blue dot, the input is arbitrary within the gray box. The resulting memory (red) is proportional to I(x), as it perfectly overlaps with I(x) after rescaling (thick shaded red line). See more details in the text. B, The dendritic state of all dendrites under the memory. Each pixel represents a dendrite, with white indicating an up-state and dark indicating a down-state. The memory activity is proportional to the total number of up-state dendrites, as summed along the y-axis (double arrow line), by a factor of β. A memory forms an effective network, with each neuron having some up-state dendrites (in-degree) and activating up-state dendrites in other neurons (out-degree). The location of the blue dot in A separates the network into two groups: the mutual-correlated group in high firing neurons, where they have both in-degree and out-degree and the mutual-separated group in low firing neurons, where they have only in-degree.

where the non-negativity of $\mathrm{T_u}$, I(x) requires $\mathrm{T_u} < \beta \mathrm{A}$, $1 \geq \alpha > 0$, and A takes an arbitrary value between 0 to 2500.

To achieve a perfect memory, the external input I(x) needs to be partially tuned. For x between 0 and A, I(x) needs to be linear; for x between A and N, I(x) can be any pattern. An exemplar I(x) is given in Fig. 3.5A alongside the resulting memory M(x). The manually chosen point A, indicating the horizontal position where the linear part ends, is marked by the blue dot. M(x) perfectly matches I(x) after rescaling, as indicated by the shaded red line. This holds true for arbitrary I(x) values within the gray box. See Appendix A for more simulation details.

To understand why this partially linear form of I(x) works, we take a digression to examine all dendritic states of a given memory. As visualized in Fig. 3.5B, each pixel represents a dendrite, specified by its pre and postsynaptic neuron labels. More up-state dendrites (white) are found in the lower left corner, where both pre and postsynaptic neurons have high firing rates. Neurons form an effective network through dendritic states in such a way that each neuron has some up-state dendrites (in-degree), whose summation is proportional to its firing rate, and activates up-state dendrites in other neurons (out-degree). Importantly, we define two neural groups: the *mutual-correlated group*, which consists of neurons with both in-degree and out-degree connections, and the *mutual-separated group*, which consists of neurons lacking out-degree connections.

We can gain some intuition as to why this partially linear I(x) works. In the mutual-correlated group, neurons are highly interactive. To prevent unwanted interaction-induced error, the input to this region must be well-tuned to follow a particular linear pattern. In contrast, in the mutual-separated group, each neuron has no out-degree and thus does not affect others. Therefore, it can fire at any rate, without any interaction-induced error.

3.2.5. Different errors formed during the encoding period

If I(x) does not follow the special partially linear form in Eq. 3.5, a graded memory can still be encoded, but distortion can occur even in the noise-free case. In this section, we examine four types of errors that happen during the encoding period, with the assumption that no forgetting happens in the later memory period.

3.2.5.1. In-degree error

Ideally, a local increase in the external input should induce a memory increase at the same location, whose magnitude is proportional to the input increase. If this is true, an input pattern with arbitrary intensities can be memorized perfectly, as it just consists of an aggregation of many small input increases.

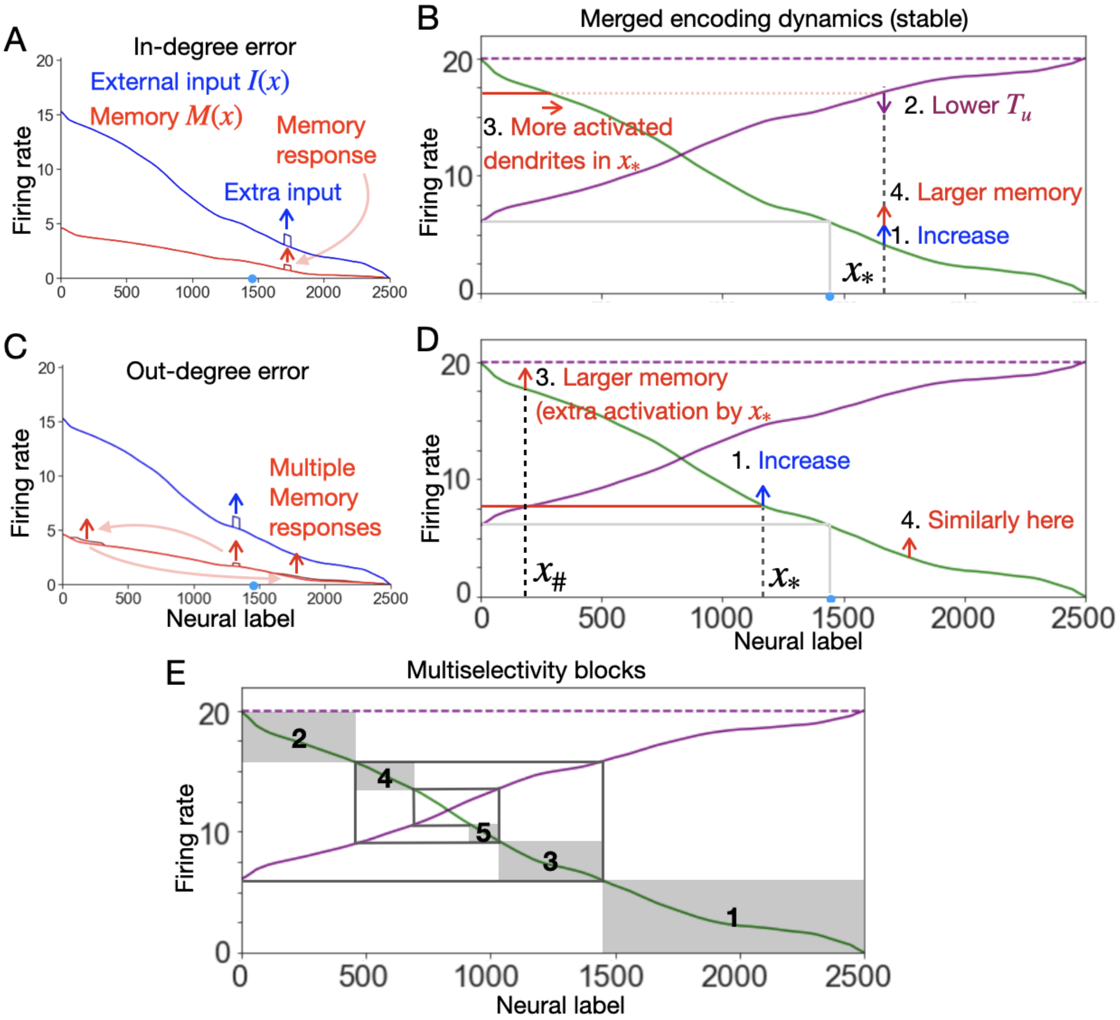

Figure 3.6. Distortion of the memory activity in response to localized perturbations of neural firing rates. A, How memory responds to a local extra input to neurons in the mutual-separated group. The memory response is local. The blue dot separates the mutual-correlated group (left) from the mutual-separated group (right). B, The detailed encoding dynamics of A. This plot merges dynamics from Figures 3.4D-F into a single plot with the same legends. The gray lines indicate how the blue dot is determined mechanistically. See more details in text. C, How memory responds to a local extra input in the mutual-correlated group. Memory changes in three locations, as indicated by the red arrows. D, The detailed encoding dynamics of C. E, An illustration of how interactions lead to multiselectivity. Neurons are separated into different regions, with region 1 being just the mutual-separated group. These groups are determined by the spiraling gray line starting from the point of the smallest $\mathrm{T}_\mathrm{u}^\mathrm{e}$ value. See more details in the text.

However, this does not always hold for this network. In general, the increased magnitude is determined by a nonlocal effect in addition to the input increase, and the induced memory can be in multiple locations. Let's first look at the first case, where "in-degree error" happens. The local memory response upon a local input increase is shown in Fig. 3.6A. Fig. 3.6B illustrates the encoding dynamics if the input increase is infinitesimal, where we merged all three panels in Fig. 3.4D-F into a single panel for simplicity. As in Fig. 3.6B, the memory in neuron x_* is proportional to the length of the red segment. An increase in input to neuron x_* lowers $\mathrm{T}^{\mathrm{e}}_{\mathrm{u}}(\mathrm{x}_*)$, activating more dendrites/in-degree and increasing $\mathrm{M}(\mathrm{x}_*)$. The extent of this increase depends on the slope of $\mathrm{f}(\mathrm{x})$ at the location of the red arrow. In other words, memory increase locally in x_*, which is proportional to its in-degree, depends on a nonlocal slope, which thus causes in-degree error.

Fig. 3.6B also shows mechanistically how neurons are separated into mutual-correlated and mutual-separated groups during the encoding period. The blue dot indicates the separation neuron, which has just 0 out-degree: its firing rate is just below the effective up-threshold of the highest firing neuron, $\mathrm{T}^{\mathrm{e}}_{\mathrm{u}}(0)$, such that it can not activate any dendrite, as illustrated by the solid gray line.

As x_* belongs to the mutual-separated group, the in-degree error remains localized to that neuron. No further interaction happens as a neuron in this group does not influence others. However, if x_* belongs to the mutual-correlated group, additional out-degree error exists due to the interaction, as explained next.

3.2.5.2. Out-degree error

Additional out-degree error exists if the input increase is in the mutual-correlated group. An example is shown in Fig. 3.6C where it causes memory changes in three locations. The mechanistic explanation is illustrated in Fig. 3.6D through an infinitesimal input increase. For a neuron $\mathrm{x}_\#$ indicated by the vertical dashed line, it receives recurrent dendritic inputs. Some dendritic inputs are above $\mathrm{T}^{\mathrm{e}}_{\mathrm{u}}(\mathrm{x}_\#)$, and some are below. There exists a dendrite whose dendritic input is right below $\mathrm{T}^{\mathrm{e}}_{\mathrm{u}}(\mathrm{x}_\#)$. We denote the presynaptic neuron of that dendrite as x_*. An increase in the firing rate of x_*, $\mathrm{f}(\mathrm{x}_*)$, can activate that dendrite in $\mathrm{x}_\#$, thus inducing a nonlocal memory increase. Similarly, the increase in $\mathrm{x}_{\#}$ can trigger a nonlocal memory increase in another neuron. It sequentially induces nonlocal memory increases until it propagates to a neuron in the mutual-separated region. That neuron has no out-degree, thus not affecting others further. For all three locations illustrated, each still has the in-degree error, in parallel with the nonlocal out-degree error.

The in-degree and out-degree errors define a particular form of multiselectivity. In Fig. 3.6E, with a horizontal line starting from the y-axis with the lowest $\mathrm{T}_{\mathrm{u}}^{\mathrm{e}}$, the merged plot for the encoding dynamics can be separated into different regions, gray boxes, as guided by the spiraling gray line. Each gray box includes a portion of neurons in the network depending on their firing rates. In Fig. 3.6D we have shown how an increase in $\mathrm{f}(\mathrm{x}_*)$, within neurons in region 3, gives three memory responses, in region 3, 2, 1 respectively. It shows a level of multiselectivity, such that an increase in $\mathrm{f}(\mathrm{x}_*)$ can induce the firing of other neurons, while x_* itself is susceptible to firing changes in neurons from a larger region label. In general, a neuron in a region R can induce R changes in region R, R-1, . . . , 1, while can also be susceptible to changes in region R+1, R+2. . . . For a neuron with a smaller region label, it is less influential to other regions but more susceptible under small incremental perturbations. For neurons with a larger region label, they are more influential but less susceptible in the sense that they can only be influenced by higher numbered regions. Therefore, multiselective neurons exhibit a level of anticorrelation in influential and susceptible levels, and the multiselectivity level depends on firing rate of each neuron. Such properties can potentially be tested in the brain through perturbation experiments.

3.2.5.3. Encoding failure or saturation

We can consider another source of error based upon the peak of input intensity, which is denoted as I(0). In Fig. 3.2, we scale the input pattern to have its maximal value I(0) equal to 15.33 such that the example shows no encoding failure nor saturation. However, if we scale the input pattern down, from $\mathrm{I}(0) = 15.33$ to $\mathrm{I}(0) = 13$, while keeping the overall shape unchanged, the high firing portion of input could be encoded, while the rest gives encoding failure with a lack of dendritic activation, as shown in Fig. 3.9A. If the scale is reduced further, I(0) can be too small such that no dendrite is activated, leading to a complete encoding failure. Therefore, as shown in Appendix C, we require $\mathrm{T_u} \leq \mathrm{I}(0)(1+\alpha)$ such that at least one dendrite can be activated.

On the other hand, if we scale the input pattern used in Fig. 3.2 up with a too large I(0), memory may saturate. The saturation can happen in two ways. One is out-degree saturation, where a high firing neuron activates all dendrites it projects to. It causes all neurons to have at least one activated dendrite, leading to a constant background firing, as shown in Fig. 3.9D with $\mathrm{I}(0) = 16$. The other case is in-degree saturation, where a high firing neuron has such a low $\mathrm{T_u^e}$ that it can no longer activate more dendrites, making its firing rate saturate during the memory period, as shown in Fig. 3.9G with $\mathrm{I}(0) = 19$. For a more comprehensive treatment of encoding failure or saturation, see Appendix C.

3.2.5.4. Instability and runaway error

As mentioned in the text of Fig. 3.3B, the interaction between neurons can form a positive feedback loop which may destabilize the system, causing runaway error. It is particularly true during the encoding period, where an increase in $\mathrm{D}(\mathrm{x}_*)$ lowers $\mathrm{T_u^e}(\mathrm{x}_*)$, which can cause larger $\mathrm{D}(\mathrm{x}_*)$ recurrently. To keep $\mathrm{D}(\mathrm{x}_*)$ stable, mathematically, we want the recurrent increase at each iteration smaller than the previous increase. The stability is guaranteed if the slope of I(x) in the mutual-correlated group satisfies:

$$\alpha\beta < |\mathrm{I}'(\mathrm{x})|. \tag{3.8}$$

A detailed derivation can be found in Appendix D.

The stability requires $\mathrm{I}'(\mathrm{x})$ not be too flat. If $\mathrm{I}'(\mathrm{x})$ is too flat such that some neurons receive almost identical input intensities, certain dendrites for a given neuron may receive similar dendritic inputs from them. A slight activation in that neuron leads to a lower $\mathrm{T}_{\mathrm{u}}^{\mathrm{e}}$ such that it may activate all these dendrites, leading to a blow up. Failure to satisfy the inequality leads to runaway error, which includes a discrete jump or a flat line in memory activity, but a partially graded memory is still possible. See more details in Appendix D.

3.2.6. Memory survives strong local inhibition

How does an existing memory get affected if a subset of neurons in the memory are silenced? From a physiological perspective, such inhibition ideally could be performed by optogenetic silencing of a portion of a population. Therefore, it is interesting to see how a stored novel memory reacts to applied strong local inhibition.

After strong local inhibition is applied to neurons with memory activity, neurons at the location of inhibition are silenced (light red line, Fig. 3.7A). This further induces a recurrent effect: the out-degrees of inhibited neurons will drop to down state due to the silence, which thus reduces the firing rates of not-inhibited neurons. This is also represented in dendrite states (Fig. 3.7B). Due to the potential breakdown of the positive feedback loop, this activity reduction may not lead to further forgetting.

Notably, up-state dendrites (in-degree) still exist in inhibited neurons, supported by dendritic inputs from other neurons, even with zero somatic firing. Upon the removal of inhibition, the inhibited neurons recover most of their firing rates due to these up-state dendrites (Fig. 3.7A, solid red line).

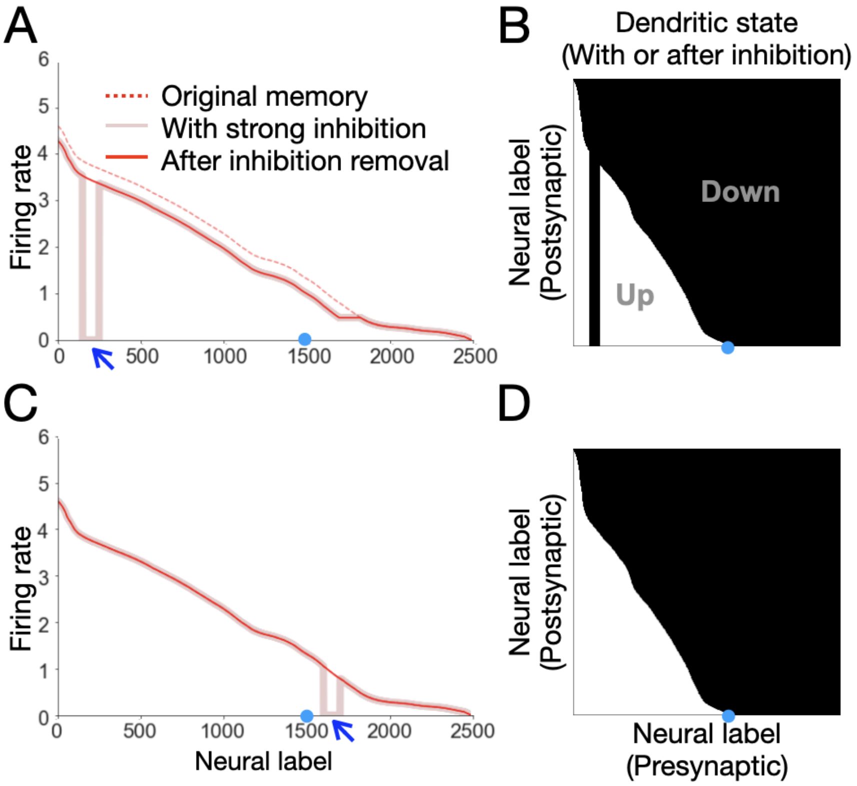

Figure 3.7. Memory performance with strong local inhibition. A, Memory changes under a strong local inhibition. The dashed red line represents the original memory. The shaded red line shows the memory after strong local inhibition is applied. The inhibition kills local activity and reduces activity in other regions. The solid red line represents the memory after the inhibition is removed, where local activity reactivates. The blue dot separates the mutual-correlated group of neurons (left) from the mutual-separated group of neurons (right). Inhibition is applied to the mutual-correlated group of neurons. B, The dendritic state plot for the memory in A, during the inhibition or removal period. The inhibited neurons cannot support the activation of dendrites in other neurons (no out-degree). C, D, similar to A, B, but with inhibition applied to the mutual-separated group of neurons. The inhibition does not reduce activity in other regions, and the memory is perfectly recovered after the inhibition is removed.

Interestingly, if local inhibition is applied to neurons in the mutual-separated group, due to a lack of out-degrees for those neurons, inhibition does not reduce any activity except locally (Figures 3.7C and 3.7D). And its removal can fully recover all suppressed activities.

3.2.7. A spiking neuronal network for novel pattern encoding

In this section, we present a spiking network model which can perform novel encoding similar to the rate model. We built a network of 40 randomly connected neurons, where the

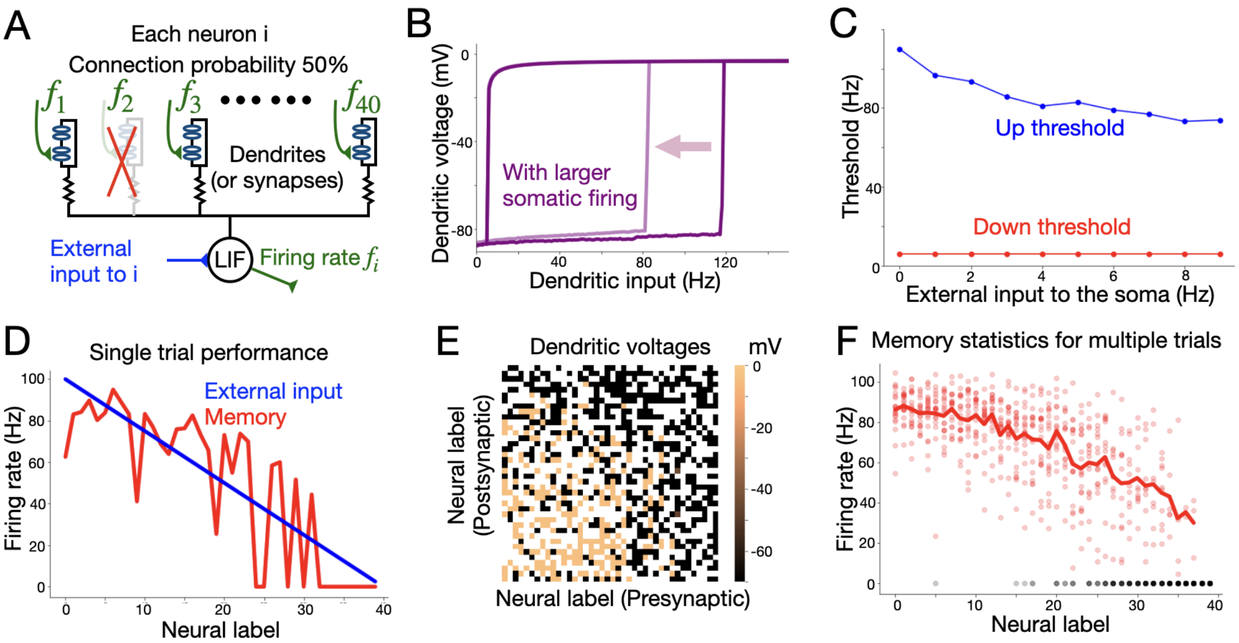

Figure 3.8. Spiking model performance. A, Network structure. We construct a network with 40 neurons. The connection probability between any two neurons is 50%. Each existing connection goes into a separate dendrite (or synapse) with conductance-based dynamics. The soma operates as a leaky integrate-and-fire model. See Appendix F for more details. B, Dendritic bistable voltage. A dendrite shows bistability in its voltage under changing dendrite input frequency. A higher somatic firing can lower its up threshold. C, Changing dendritic threshold. Threshold values are recorded as the soma fires more. Up threshold drops (blue), while the down threshold remains mostly unchanged (red). D, Memory in firing rate. For a single trial, a linear input is delivered (blue), and the remaining memory (red) is similar to the input but noisy. E, Memory in dendritic voltages. For the memory in D, the heat map shows voltages of all dendrites, with white pixels indicating the absence of a recurrent connection. F, Memory statistics. We ran 20 trials, each time with a random connectivity and input spiking train and recorded all memories. The red dots represent non-zero memory activity of neurons across trials. The gray dots represent zero memory activity of neurons across trials, with lower firing neurons having more gray dots overlapped. The solid red line represents the average of all non-zero memory activities.

probability for a recurrent connection is 50%. Each recurrent connection projects to a separate dendrite, such that, on average, a neuron has 20 conductance-based dendrites. Each dendrite is equipped with NMDA receptor based dynamics [61, 82], which gives dendritic bistability, as shown in Fig. 3.8B. See Appendix F for more details on the model dynamics.

Larger somatic firing significantly lowers the up-threshold of a given neuron, as demonstrated in Fig. 3.8B and C for different values of somatic firing. Physiologically, it occurs due to the following steps: 1. The back-propagation of somatic voltage reaches a dendrite; 2. The depolarization removes the Mg^{2+} block from the NMDA channels, facilitating the opening of NMDA receptors in these dendrites; 3. The increased current from NMDA receptors provides further depolarization which leads to more current influx until saturation.

In contrast, the down-threshold behaves differently. Because NMDA receptors easily saturate even when the dendritic input is low, the down threshold is low. This may be beneficial as it makes memory forgetting harder. In addition, the down-threshold is insensitive to the somatic effect, as the simulation shows. This is because most NMDA receptors are already unblocked upon dendritic activation. The voltage back-propagation does not significantly affect the Mg^{2+} block.

This network model is capable of novel pattern encoding. Fig. 3.8D depicts the performance under an exemplar linear input in a single trial. The memory is noisy but still encodes graded values. Fig. 3.8E shows the corresponding dendritic voltages during the memory period, which are indeed higher for those dendrites with both their pre and postsynaptic firing rate high, indicating the associative nature of the dendritic voltage. Fig. 3.8F shows all memory activities (dots) across 20 trials. We can see that for neurons with lower input intensities, encoding failure is more likely, with moregray dots overlapped at zero activity, and the averaged activity (red line) for these non-zero memory activities (red dots) is also lower, indicating the encoding of graded values.

3.3. Discussion

In this paper, we model a challenging working memory task involving the encoding of a novel pattern. The model is based on a uniformly connected network where each neuron possesses multiple bistable dendrites. Each dendrite follows a box-like dendritic relation with the somatic effect, effectively making a fast-Hebbian plasticity. The network is highly interactive but a stable, graded novel encoding can be achieved without requiring finely tuned model parameters. Notably, the failure of dendritic bistability—either box-like dendritic relation or somatic effects—leads to failure in performing the task, as detailed in Appendix E.

3.3.1. Biological plausibility and alternative implementations

Dendritic bistability is central to our model, and it may be achieved in several ways. In the spiking model, we choose dynamics based on wide-spreading NMDA receptors, as done in previous models [61, 82]. However, other multistability mechanisms [69, 103] may achieve the same functional network.

A major candidate for the somatic effect is back-propagated action potentials(bAPs) [39, 87, 88]. But this is not the only possibility, especially if an action potential is missing, or the propagation is weak [44]. Other mechanisms which can similarly provide the required depolarization include: i) A subthreshold voltage in the soma or an input current roughly colocalized with bistable dendrites. Similar ideas have been tested in [48] for long term potentiation. ii) If bAPs are very strong, [12] shows that dendrites can still maintain high independence.

Instead of a different input intensity to each neuron, it may be helpful to provide the same input to a small cluster of neurons. This potentially provides more bistable units and smooths noise. For example, each neuron in our model might correspond to a cluster of neurons in a network that is wired with a connectivity like that of our bistable-dendrites-based model.

3.3.2. Two interpretations of the model and implications/predictions

3.3.2.1. Single-system interpretation

If treating the entire network as a whole, as is the default assumption, our model shows multiselectivity. This multiselectivity is essentially due to the highly recurrent, functional connection established by up-state dendrites. It causes distortion to the memory, with different performance in two groups as shown in Fig. 3.6, and exhibits an interesting anticorrelation between its susceptibility and influence. Meanwhile, this recurrent connection can also be beneficial, providing extra robustness against strong local inhibition. It is because the local inhibition can not erase all dendritic activity of the receiving neuron, as these dendrites are nonlocally maintained through other neurons. Interestingly, dendritic bistability in our model is functionally similar to synaptic plasticity, where information is stored in dendrites and thus escapes from strong somatic inhibition.

3.3.2.2. Dual-system interpretation

The multiselectivity behaves differently in two groups of neurons. For those in the mutual-correlated group, a perturbative local increase in firing rate during the encoding period leads to more activity increase in other neurons, spreading out the perturbation. For those in the mutual-separated group, a perturbative local increase only changes activity locally, although the changing magnitude depends on nonlocal neurons. In addition, for the perfect memory case as shown in Fig. 3.5, the input separates into a linear part and an arbitrary part. These performances motivate us to interpret the network as a dual-system.

Suppose that the input is from, instead of a single external source, two sources simultaneously. One is an auxiliary internal input from a certain brain region, providing fixed high intensities to form the mutual-correlated group. Another is an arbitrary external input, providing relatively low intensities to form the mutual-separated group. The final memory consists of one part due to the auxiliary input and another part due to the external input, where information is actually encoded. The major benefit for this dual-system interpretation is that there is generally no out-degree error for the external input. If the auxiliary input is further trained to be linear in a form of Eq. 3.5, an arbitrary external input can be memorized perfectly.

The dual-system interpretation may be realistic. The fixed internal input may represent a general control of the current task setting, and the external input can represent a more explicit and flexible task-relevant content. The task content part can be encoded only if the task setting part is active to provide it dendritic inputs.

Appendix

A: Simulation details of the rate network model

The default input pattern first appearing in Fig. 3.2 was generated based on a cloud picture ($\mathrm{CCSN_v2/Cu/Cu-N083.jpg}$) in [101]. The original picture was compressed to a resolution of 50×50, converted into gray scales and ranked by its intensity in the descending order. To get rid of intensity degeneracy, which exists as intensity is an integer value out from 0 to 255, the ranking was smoothed by using the 'nearest' mode of ndimage.convolve in Python. This gets rid of degeneracy by giving a small distortion to each intensity value. Further, the minimal intensity was subtracted from all intensities. Finally, the input pattern is scaled to have a maximum value of 15.33. Eventually it gives discrete input values $\mathrm{I_i}$ with i from 1 to N=2500, which are approximated to be continuous, I(x), for the analysis in the main text.

A rate network model is used in Figures 3.1-3.7. For the default setting introduced in Fig. 3.1, parameters are defined as $N = 2500$, $T_u = 20$, $T_d = 1$, $\beta = 0.0032$, $\alpha = 0.7$, $\tau = 50$ ms and $w = 1$. For memory encoding, the network starts with all dendrites in the down-state, an input is applied for 1000 ms during the encoding period, and it is turned off for another 1000 ms during the following memory period. Equilibrium is reached in each period. The model dynamics are governed by:

$$\tau \frac{df_i}{dt} = -f_i + \sum_{j=1}^{N} B(f_i, f_j) + I_i, \tag{3.9}$$

where $B(f_i, f_j)$ is the dendritic bistable relation described in Fig. 3.1C, which is mathematically expressed as:

$$B(f_i, f_j) = \begin{cases} \beta\Theta(f_j - T_u + \alpha f_i) & \text{Activation from down- to up-state} \\ \beta\Theta(f_j - T_d) & \text{Deactivation from up- to down-state,} \end{cases} \tag{3.10}$$

where $\Theta(x)$ is a Heaviside step function, and the somatic effect is bound from below $T_u - \alpha f_i > T_d$. Note that T_d is not affected by the somatic firing, which qualitatively captures the conductance-based dynamics of a bistable dendrite. See Fig. 3.8B for more details. Simulations are done using Euler's method with $dt = 1$ ms in Python [Version: 3.9.16].

In Figures 3.3A and 3.3B, we implement Gaussian independent noise throughout the memory period. The noise has a zero mean and a standard deviation of 0.1 or 0.3, applied with a time step of 1 ms. It adds to the firing rate directly without the effect of τ. The average (dark line) and standard deviation (light red shadiness) of the resulting noisy memory were calculated over a period of 500 ms, after 1000 ms into the memory period. The noise-free memory in Fig. 3.2 is reproduced as a dotted red line for comparison.

In Fig. 3.3C, we implement three variations to the original connectivity. For the top panels, weight values were drawn from a Gaussian distribution with mean 1 and standard deviation

0.1, 0.5 or 1 respectively, with a minimal value set to 0. For the middle panels, the connection probability p was changed to 50%, 10% or 1% respectively. As a dendrite exists only if a connection exists, this change reduces the total number of dendrites for each neuron. To balance this reduction, β was normalized correspondingly $\beta \leftarrow \beta/p$. For the bottom panels, the connectivity exists only locally, within the two-dimensional space defined by horizontal and vertical pixel labels with periodic boundary conditions. If two neurons are located within a radius of 20, 10 or 5, the weight value is one, otherwise zero. To balance the reduction of dendrites, β was normalized correspondingly $\beta \leftarrow \beta N/N_{loc}$, where N_{loc} is the total number of dendrites in each neuron. For all panels in Fig. 3.3C, we use the same graded scale bar for ease of direct comparison. However, due to changes in connectivity, the actual firing rates of some neurons may exceed the maximum scale shown. This procedure does not change qualitative results.

In Fig. 3.5, the external input for a perfect memory has a linear part $I(x) = 14.4 - 0.00576x$, for x between 0 and A=1500. For the arbitrary part, we chose another picture, $CCSN_v2/Cu/Cu-N004.jpg$ in [101], followed the same steps as mentioned above and took the 1000 lowest intensity pixels. This portion was rescaled so that its maximum value matches the minimal value of the linear part in neuron 1500.

In Fig. 3.6, the extra input was given to neurons 1700 to 1750 (Fig. 3.6A) or 1300 to 1350 (Fig. 3.6C) during the encoding period with a magnitude of 1 adding to the existing I(x).

In Fig. 3.7, an inhibitory signal with an intensity of 20 was given to neurons 150 to 250 (Fig. 3.7A) or 1600 to 1700 (Fig. 3.7C) at the end of the memory period. It lasted for 1000 ms, and another 1000 ms was simulated after its removal.

In Figures 3.9A, 3.9D, and 3.9G, the maximum external input intensities are $I(0) = 13, 16,$ and 19 respectively.

In Fig. 3.10, the derivative was estimated through 4 nearby neurons from a discrete function $\mathrm{f(x)}$: $\mathrm{f'(x)} \approx \frac{-\mathrm{f(x+2\Delta x)+8f(x+\Delta x)-8f(x-\Delta x)+f(x-2\Delta x)}}{12\Delta \mathrm{x}}$, with step size $\Delta\mathrm{x} = 5$. In Fig. 3.10C, we chose the same picture as used in Fig. 3.5, with a maximum external input intensity $\mathrm{I(0)} = 14.17$. In Fig. 3.10E, with the default input pattern, the input pattern is rescaled such that $\mathrm{I(0)} = 15.33$ becomes $\mathrm{I(0)} = 15.8$ and all intensities after neuron 1200 were decreased to 80% of their original values, so that a discrete jump was created in the input pattern.

In Fig. 3.11, we set $\alpha = 0$ and the maximum external input intensities are $\mathrm{I(0)} = 22$, 30, 38 or 46.

B. Solving for a perfect memory

As discussed in the text for Fig. 3.5, for a perfect memory, the stable state during the encoding period needs to satisfy:

$$\beta \mathrm{f}^{-1}(\mathrm{T_u} - \alpha \mathrm{f(x)}) = \mathrm{D(x)} = \frac{\beta \mathrm{A}}{\mathrm{T_u}}\mathrm{f(x)}, \tag{3.11}$$

where $\frac{\beta \mathrm{A}}{\mathrm{T_u}} < 1$ is the proportionality ratio, which is smaller than 1 as firing rate $\mathrm{f(x)}$ is always larger than the total dendritic activity $\mathrm{D(x)}$, with an arbitrary neural label A. The choice of this particular form will be clear below. It simplifies to:

$$\mathrm{T_u} = \alpha \mathrm{f(x)} + \mathrm{f}(\frac{\mathrm{A}}{\mathrm{T_u}}\mathrm{f(x)}). \tag{3.12}$$

Recall that $\mathrm{f(x)}$ is a non-increasing function that decays to 0, $\mathrm{f(N)} = 0$. Substituting it into Eq. 3.12 gives: $\mathrm{f(0)} = \mathrm{T_u}$.

On the other hand, taking the derivative of Eq. 3.12 gives:

$$0 = \alpha \mathrm{f'(x)} + \mathrm{f'}(\frac{\mathrm{A}}{\mathrm{T_u}}\mathrm{f(x)})\mathrm{f'(x)}\frac{\mathrm{A}}{\mathrm{T_u}}. \tag{3.13}$$

For non trivial solution, $f'(x) \neq 0$, We then have:

$$f'(\frac{A}{T_u} f(x)) = -\frac{T_u \alpha}{A}. \tag{3.14}$$

Note that $f(x)$ can take values from $f(N) = 0$ to $f(0) = T_u$. This means that $\frac{A}{T_u} f(x)$ takes values from 0 to A. The above equation requires a constant slope $-\frac{T_u \alpha}{A}$ for $f(x)$ with x ranging from 0 to A. Note that this value A just maps to the location of the blue dot as shown in Fig. 3.5.

Altogether, we have the solution for Eq. 3.11:

$$f(x) = T_u - \frac{T_u \alpha}{A} x, (1 \geq \alpha > 0), \tag{3.15}$$

$$I(x) = T_u - \beta A - (\frac{T_u \alpha}{A} - \alpha\beta) x, \tag{3.16}$$

$$D(x) = \beta A - \alpha\beta x, \tag{3.17}$$

where x ranges from 0 to A. We require $1 \geq \alpha > 0$ such that $f(x)$ is non negative. To ensure $I(x)$ can activate a dendrite in the absence of any dendritic activity, as discussed in Appendix C, we further require: $T_u \leq (1 + \alpha)(T_u - \beta A)$

On the other hand, to ensure no forgetting happens, the maximum sustainable line should be above $D(x)$ as shown in Fig. 3.4I, but now with $D(x)$ defined in Eq. 3.17. We require:

$$T_d \leq (1 - \alpha)\beta A. \tag{3.18}$$

We can check if T_u^e hits its lower bound, T_d, as well. Because the ratio is $\frac{\beta A}{T_u} < 1$, we have $(1 - \alpha)T_u > (1 - \alpha)\beta A$. With the above equation, we have $T_u^e = (1 - \alpha)T_u > (1 - \alpha)\beta A \geq T_d$. Therefore, T_u^e is always above its lower bound, T_d, for the parameter range considered here and does not saturate.

Altogether, we get the condition for a perfect memory as shown in the main text.

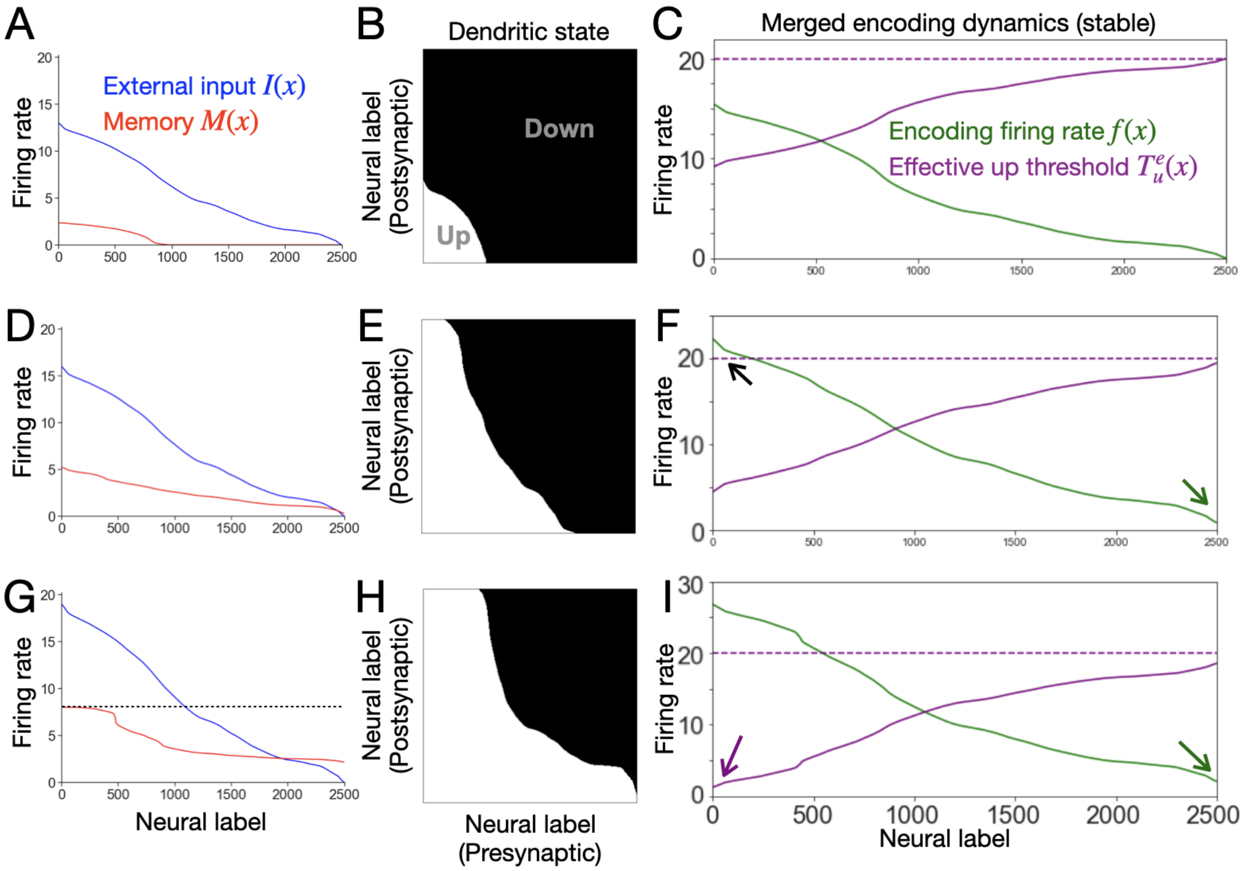

Figure 3.9. The scaling of I(x) and encoding failure or saturation error. A-C, A memory with encoding failure. A, An input with a small scaling results in a memory that has graded values in neurons receiving high input intensities, while neurons receiving low input intensities remain silent. B. All dendrite states of the memory, with up-state dendrites in white and down-state dendrites in dark. C, The merged encoding dynamics. See more details in the text. D-F, A memory with out-degree saturation. Similar to A-C, but with a higher scaled input pattern that leads to constant background activity. F. The maximum firing neurons exceed T_u (black arrow) making them capable of activating all dendrites it projects to. It gives a constant firing in all neurons (dark green arrow). G-I, A memory with in-degree saturation. Similar to D-F, but with an even higher scaled input, which leads to further saturation. G. the memory reaches a maximum firing rate (dashed horizontal line). I. The effective up-threshold of high firing-neurons (purple arrow) is lower than the smallest firing rate (dark green arrow) making all dendrites in high firing neurons activate.

C. Memory with encoding failure or saturation

In this section, we take $0 < \alpha \leq 1$ for analysis, but the idea is qualitatively similar for $\alpha > 1$ case.

A minimal requirement is that input can at least activate one dendrite so that memory can be nonzero. The most activatable dendrite is the one that projects from neuron 0 back to itself, which has the largest pre and postsynaptic firing rate $\mathrm{I}(0)$ in the absence of any dendritic activity. The activation of this dendrite requires $\mathrm{T_u} \leq \mathrm{I}(0)(1+\alpha)$.

Further, during the encoding period, for any neuron receives nonzero input, we want nonzero memory. It requires that a neuron receiving an infinitesimal input to have at least one dendrite activated. The most activatable dendrite in such a neuron is the one that receives dendritic input from neuron 0, with maximum rate $\mathrm{f}(0)$. Therefore we need $\mathrm{f}(0) \geq \mathrm{T_u}$. The input in Fig. 3.2 is tuned with $\mathrm{I}(0) = 15.33$ to make $\mathrm{f}(0) = \mathrm{T_u}$ such that no encoding failure happens. If the input is scaled lower, $\mathrm{I}(0) = 13$, the memory loses some information of the region with low input intensity (Fig. 3.9A), as the firing rate is not enough to activate some dendrites to the up-state.

A large input may saturate the memory. Fig. 3.9D shows an example with $\mathrm{I}(0) = 16$ for the out-degree saturation. The high firing neurons exceed $\mathrm{T_u}$, such that out-degree saturation happens, with all dendrites they project to activate. These up-state dendrites, distributed among all neurons, provide a constant background activity (Fig. 3.9F). This out-degree saturation is benign, as a constant background term still allows a graded memory. In contrast, in-degree saturation happens when a high firing neuron can no longer activate more dendrites as input goes larger. It may happen either because its $\mathrm{T_u^e}$ hits the minimal value, $\mathrm{T_u^e} = \mathrm{T_d}$, or the dendritic input from the lowest firing neuron N exceeds $\mathrm{T_u^e}$ such that it has already activated all its dendrites. An example of the latter case is provided with $\mathrm{I}(0) = 19$, as shown in Figures 3.9G and 3.9I.

Why do we focus on the range $0 < \alpha \leq 1$? We want the memory to have no encoding failure, $\mathrm{f}(0) \geq \mathrm{T_u}$. However, if we choose $\alpha > 1$, $\mathrm{T_u^e}(0)$ reaches its minimum value $\mathrm{T_d}$, as $\mathrm{T_u} - \alpha \mathrm{f}(0) < \mathrm{T_d}$. This leads to in-degree saturation. Choosing $0 < \alpha \leq 1$ makes it possible to have no encoding failure nor saturation.

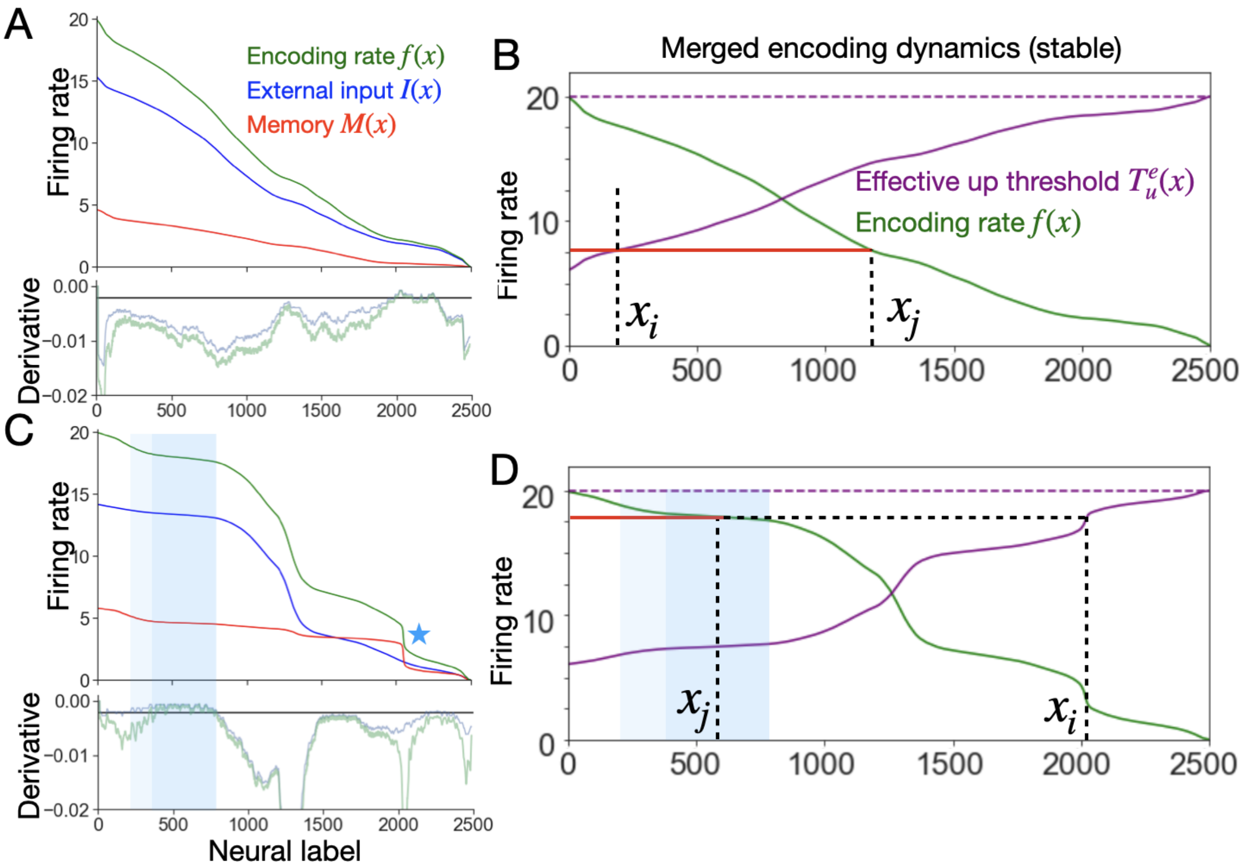

Figure 3.10. The slope of I(x) and runaway error. A, B, A memory without runaway error. A, Top: Results reproduced from Fig. 3.2. Bottom: The derivatives of f(x) (green) and I(x) (blue). A horizontal line indicates the value of −αβ. B. The encoding dynamics merged into a single plot. This plot merges dynamics in Figures 3.4D-F in a single plot with the same legends. See the text for a detailed explanation. C, D, Similar to A, B, but with a different input pattern. The light shaded region includes neurons whose slope of I(x) (negative) is less negative than −αβ. The dark shaded region includes neurons whose f(x) exceeds −αβ. The star indicates a local discrete jump in the memory, even though I(x) has no jump.

D. How input slope affects the formation of memory and runaway error

How does input slope affect encoding dynamics? As mentioned above in Fig. 3.6B, there exists a positive feedback loop. It is reproduced in Fig. 3.10B with more details. A local infinitesimal input increase $\delta I(x_i)$, or equivalently $\delta f(x_i)$, lowers $T_u^e(x_i)$ by $\alpha \delta f(x_i)$. This expands the total number of up-state dendrites, which equals the neural label x_j in value, by $-\frac{\alpha}{f'(x_j)}\delta f(x_i)$. Therefore, the induced change is given by:

$$\delta D(x_i) \leftarrow -\frac{\alpha\beta}{f'(x_j)}\delta f(x_i), \tag{3.19}$$

where $\delta D(x_i)$ feeds back into $\delta f(x_i)$. To prevent the blowing up of recurrent activity and to make $D(x)$ sensitive to a small increase, we limit the gain factor:

$$-\frac{\alpha\beta}{f'(x_j)} < 1. \tag{3.20}$$

As $f'(x_j) = I'(x_j) + D'(x_j)$, we have $\alpha\beta < |I'(x_j)| + |D'(x_j)|$. Because $I(x)$ is what can be manipulated externally, we get the sufficient condition to prevent runaway error:

$$\alpha\beta < |I'(x_j)|. \tag{3.21}$$

This condition applies to neurons in the entire mutual-correlated group, where any x_i has a corresponding x_j. An example, where the memory has no blow-up and is sensitive, is shown in Fig. 3.10A, where $I'(x)$ (bottom) is always below $-\alpha\beta$, consistent with the inequality.

Here, we only consider how a local increase to neuron i directly causes x_i to gain more up-state dendrites. However, it may cause more activity in other neurons, which feeds back to activate x_i more. This indirect, recurrent path can be ignored because an increase in x_i maximally activates one dendrite in other neurons. In the continuous limit, the contribution of a single dendrite is negligible such that we ignore such activation and further recurrent interaction.

Failing to satisfy Eq. 3.21 may bring extra errors to the memory. An example is shown in Fig. 3.10C, where $\alpha\beta > |I'(x)|$ in the light shaded region. The violation of the sufficient condition is fine for the lighter shaded region, as the exact requirement, Eq. 3.20 is still satisfied (Fig. 3.10C, bottom). However, for the darker shaded region, neurons have similar rates such that the dendrites they project to in neuron x_i are similarly prone to be activated

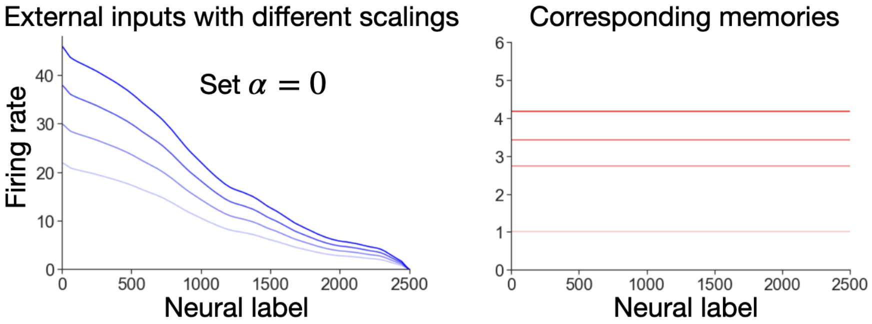

Figure 3.11. Memory performance when the somatic effect is absent. If there is no somatic effect, $\alpha = 0$, a graded novel pattern leads to a uniform memory, with no graded values. However, different input scalings result in different uniform values.

(Fig. 3.10D). That is, a small increase $\delta f(x_i)$ can lead to the activation of all these dendrites, resulting in a discontinuity in $D(x_i)$.

E. The necessity of both dendritic bistability and somatic effect for novel encoding in this model

To achieve novel encoding in this network model, dendritic bistability is necessary, with both the box-like dendritic input-output relation and the somatic effect. To demonstrate the necessity further, we consider two potential changes here. First, instead of a bistable relation, now a step function is applied, with a single threshold T_0. We assume the somatic effect still exists. The stable firing rate during the memory period is:

$$\beta M^{-1}(T_0 - \alpha M(x)) = M(x). \tag{3.22}$$

Similar to the calculation for Eq. 3.16, we can solve this as $M(x) = T_0 - \alpha\beta x$, with x from 0 to T_0/β. However, this solution is unstable, as $M'(x) = \alpha\beta$ does not satisfy the stability inequality Eq. 3.20.

Secondly, with no somatic effect, $\alpha = 0$, all neurons are now functionally identical. Therefore, the memory gives a uniform line, as shown in Fig. 3.11. The overall scaling difference is still memorized.

F. Simulation details about the spiking network

We constructed a single neuron model with $n = 20$ dendrites to show the dendritic bistability based on dynamics of NMDA receptors. This model is based on our above autapse model with multiple dendrites and other dendritic bistability models with more detailed gating kinetics [61, 82]. Euler's method is used for the simulation, with step size 0.25 ms, which was performed in python [Version: 3.9.16].

The somatic voltage V_s follows dynamics:

$$C\frac{dV_s}{dt} = -I_L - I_d - I_{ext} - I_{ton} - I_{s,noise}, \tag{3.23}$$

$$= -g_L(V_s - E_L) - \sum_{i}^{n} g_{ds}(V_s - V_{d,i}) \tag{3.24}$$

$$-g_{ext}s_{ext}(V_s - E_{ext}) - I_{ton} - I_{s,noise}, \tag{3.25}$$

where $C = 10$ nF/mm^2, $g_L = 1$ µS/mm^2, $E_L = -80$ mV, $g_{ds} = 0.01$ µS/mm^2, $g_{ext} = 30$ µS/mm^2, $E_{ext} = 0$ mV and $I_{ton} = -34$ nA/mm^2. $I_{s,noise}$ is independent Gaussian noise with mean 0 and a standard deviation of 1 applied at every time step. The synaptic activation s_{ext} was driven by an external spike train with dynamics specified below. Once V_s crossed the firing threshold –50 mV, a spike was triggered and the somatic voltage was set to 30 mV for 3 ms before resetting to –55 mV for another 3 ms.

The dendritic voltage $V_{d,i}$ follows dynamics:

$$C\frac{dV_{d,i}}{dt} = -I_{L,i} - I_{s,i} - I_{NMDA,i} - I_{KIR,i} - I_{d,noise,i} \tag{3.26}$$

and each term on the right hand side is:

$$\mathrm{I_{L,i}} = \mathrm{g_L}(\mathrm{V_{d,i}} - \mathrm{E_L}), \tag{3.27}$$

$$\mathrm{I_{s,i}} = \varkappa \mathrm{g_{ds}}(\mathrm{V_{d,i}} - \mathrm{V_s}), \tag{3.28}$$

$$\mathrm{I_{NMDA,i}} = \mathrm{g_{NMDA,i}} \mathrm{s_{NMDA,i}} \frac{\mathrm{V_{d,i}} - \mathrm{E_{NMDA}}}{1 + 0.15\mathrm{e}^{-0.08\mathrm{V_{d,i}}}}, \tag{3.29}$$

$$\mathrm{I_{KIR,i}} = \mathrm{g_{KIR,i}} \frac{\mathrm{V_{d,i}} - \mathrm{E_{KIR}}}{1 + \mathrm{e}^{0.1(\mathrm{V_{d,i}} - \mathrm{E_{KIR}} + 10)}}, \tag{3.30}$$

where $\mathrm{E_{NMDA}} = 0$ mV, $\mathrm{E_{KIR}} = -90$ mV and the area ratio of the soma to a dendrite $\varkappa = 37$. $\mathrm{I_{d,noise,i}}$ was independent Gaussian noise with mean 0 and standard deviation 1. $\mathrm{g_{NMDA,i}}$ was generated from a Gaussian random variable with mean 34.5 $\mu\mathrm{S/mm^2}$ and standard deviation 1. Similarly, $\mathrm{g_{KIR,i}}$ with mean 41 $\mu\mathrm{S/mm^2}$ and standard deviation 1.

The synaptic activation s_α, with $\alpha = \mathrm{ext}$ or NMDA, i, follows dynamics:

$$\frac{\mathrm{ds}_\alpha}{\mathrm{dt}} = -\mathrm{s}_\alpha/\tau_\alpha, \tag{3.31}$$

$$\Delta \mathrm{s}_\alpha = 0.5(1 - \mathrm{s}_\alpha), \text{ each time a spike is received,} \tag{3.32}$$

where $\tau_{\mathrm{ext}} = 2$ ms and $\tau_{\mathrm{NMDA,i}} = 100$ ms respectively. $\mathrm{s_{ext}}$ receives an external spike train, and $\mathrm{s_{NMDA,i}}$ receives the recurrent spike train generated by neuron i.

Such a single neuron model shows dendritic bistability. To generate the solid line in Fig. 3.8B, a Poisson spike train with a frequency of 40 Hz is delivered to the soma, and a fixed period spike train with changing frequencies is delivered to all 20 dendrites. The latter is to mimic recurrent inputs from other neurons, which has a relatively fixed period due to the integrate-and-fire soma. The dendritic input frequency starts from 0 Hz and increases by 1 Hz every 1400 ms until it reaches 150 Hz. Afterwards, it decreases by 1 Hz every 1400 ms until it goes back to 0 Hz. The dendritic voltage shown is averaged over the last 500 ms before the frequency change happens. The light purple line in Fig. 3.8B is generated with another somatic input of 210 Hz. For Fig. 3.8C, we ran 10 trials with the somatic input ranges from 30 Hz to 210 Hz and recorded the averaged threshold values across all trials and all 20 dendrites. Thresholds are estimated by the largest slope.

We extended the neural model to a homogeneous network of N = 40 neurons. The probability a connection exists from one neuron to another is p = 0.5, which makes n a variable with mean 20. Voltages are stabilized at about -80mV initially. The input to the soma of each neuron is a Poisson spike train with a linear frequency from 100 to 0 Hz (Fig. 3.8D). The encoding period lasts for 1200 ms, the memory period lasts for 800 ms. The dendritic voltage and firing rate are averaged over the last 300 ms (Fig. 3.8E). We further ran 20 trials to get the statistical performance (Fig. 3.8F).

CHAPTER 4

Summary

In this thesis, I address two questions related to working memory. The first is how to encode a familiar local input with its amplitude, and the second is how to encode a novel pattern with graded intensity. I purpose two neural circuit models where dendritic bistability is the core unit. Both models can maintain a robust memory against noise. Despite their similarity, the two models differ structurally. The first model requires pretrained local weights and a box-like dendritic relation, whereas the second model does not need any local weight structure but needs a box-like dendritic relation with the somatic effect. The two models could work together for a partially novel and partially familiar input, such as a red apple with its own unique surface dents and spots. My work demonstrates the strength of dendritic computation and calls for more direct experimental evidence to verify these mechanisms.

Bibliography

[1] E. J. Adams, A. T. Nguyen, and N. Cowan, *Theories of working memory: Differences in definition, degree of modularity, role of attention, and purpose*, Language, speech, and hearing services in schools, 49 (2018), pp. 340–355.

[2] Z. Ajabi, A. T. Keinath, X.-X. Wei, and M. P. Brandon, *Population dynamics of head-direction neurons during drift and reorientation*, Nature, 615 (2023), pp. 892–899.

[3] E. Aksay, G. Gamkrelidze, H. S. Seung, R. Baker, and D. W. Tank, *In vivo intracellular recording and perturbation of persistent activity in a neural integrator*, Nature neuroscience, 4 (2001), pp. 184–193.

[4] E. Aksay, G. Major, M. S. Goldman, R. Baker, H. S. Seung, and D. W. Tank, *History dependence of rate covariation between neurons during persistent activity in an oculomotor integrator*, Cerebral Cortex, 13 (2003), pp. 1173–1184.

[5] S.-i. Amari, *Dynamics of pattern formation in lateral-inhibition type neural fields*, Biological cybernetics, 27 (1977), pp. 77–87.

[6] R. Araya, *Dendritic morphology and function*, Neuroscience in the 21st Century: From Basic to Clinical, (2022), pp. 571–606.

[7] A. Baddeley, *The episodic buffer: a new component of working memory?*, Trends in cognitive sciences, 4 (2000), pp. 417–423.

[8] ——, *Working memory: Theories, models, and controversies*, Annual review of psychology, 63 (2012), pp. 1–29.

[9] O. Barak, D. Sussillo, R. Romo, M. Tsodyks, and L. Abbott, *From fixed points to chaos: three models of delayed discrimination*, Progress in neurobiology, 103 (2013), pp. 214–222.

[10] O. Barak and M. Tsodyks, *Working models of working memory*, Current opinion in neurobiology, 25 (2014), pp. 20–24.

[11] O. Barak, M. Tsodyks, and R. Romo, *Neuronal population coding of parametric working memory*, Journal of Neuroscience, 30 (2010), pp. 9424–9430.

[12] B. F. Behabadi and B. W. Mel, *Mechanisms underlying subunit independence in pyramidal neuron dendrites*, Proceedings of the National Academy of Sciences, 111 (2014), pp. 498–503.

[13] R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky, *Theory of orientation tuning in visual cortex.*, Proceedings of the National Academy of Sciences, 92 (1995), pp. 3844–3848.

[14] K. C. Bittner, A. D. Milstein, C. Grienberger, S. Romani, and J. C. Magee, *Behavioral time scale synaptic plasticity underlies ca1 place fields*, Science, 357 (2017), pp. 1033–1036.

[15] C. D. Brody, A. Hernández, A. Zainos, and R. Romo, *Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex*, Cerebral cortex, 13 (2003), pp. 1196–1207.

[16] C. D. Brody, R. Romo, and A. Kepecs, *Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations*, Current opinion in neurobiology, 13 (2003), pp. 204–211.

[17] Y. Burak and I. R. Fiete, *Fundamental limits on persistent activity in networks of noisy neurons*, Proceedings of the National Academy of Sciences, 109 (2012), pp. 17645–17650.

[18] M. Camperi and X.-J. Wang, *A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability*, Journal of computational neuroscience, 5 (1998), pp. 383–405.

[19] S. Carroll, K. Josić, and Z. P. Kilpatrick, *Encoding certainty in bump attractors*, Journal of computational neuroscience, 37 (2014), pp. 29–48.

[20] K. P. Champion, O. Gozel, B. S. Lankow, G. B. Ermentrout, and M. S. Goldman, *An oscillatory mechanism for multi-level storage in short-term memory*, Communications biology, 6 (2023), p. 829.

[21] T. B. Christophel, P. C. Klink, B. Spitzer, P. R. Roelfsema, and J.-D. Haynes, *The distributed nature of working memory*, Trends in cognitive sciences, 21 (2017), pp. 111–124.

[22] A. Compte, *Computational and in vitro studies of persistent activity: edging towards cellular and synaptic mechanisms of working memory*, Neuroscience, 139 (2006), pp. 135–151.

[23] A. Compte, N. Brunel, P. S. Goldman-Rakic, and X.-J. Wang, *Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model*, Cerebral cortex, 10 (2000), pp. 910–923.

[24] C. Constantinidis, M. N. Franowicz, and P. S. Goldman-Rakic, *The sensory nature of mnemonic representation in the primate prefrontal cortex*, Nature neuroscience, 4 (2001), pp. 311–316.

[25] C. Constantinidis, S. Funahashi, D. Lee, J. D. Murray, X.-L. Qi, M. Wang, and A. F. Arnsten, *Persistent spiking activity underlies working memory*, Journal of neuroscience, 38 (2018), pp. 7020–7028.

[26] C. Constantinidis and X.-J. Wang, *A neural circuit basis for spatial working memory*, The Neuroscientist, 10 (2004), pp. 553–565.

[27] V. H. Cornejo, N. Ofer, and R. Yuste, *Voltage compartmentalization in dendritic spines in vivo*, Science, 375 (2022), pp. 82–86.

[28] N. Cowan, *An embedded-processes model of working memory*, Models of working memory: Mechanisms of active maintenance and executive control, 20 (1999), pp. 1013–1019.

[29] C. E. Curtis and M. D'Esposito, *Persistent activity in the prefrontal cortex during working memory*, Trends in cognitive sciences, 7 (2003), pp. 415–423.

[30] K. Daie, M. S. Goldman, and E. R. Aksay, *Spatial patterns of persistent neural activity vary with the behavioral context of short-term memory*, Neuron, 85 (2015), pp. 847–860.

[31] M. D'Esposito and B. R. Postle, *The cognitive neuroscience of working memory*, Annual review of psychology, 66 (2015), pp. 115–142.

[32] S. Druckmann and D. B. Chklovskii, *Neuronal circuits underlying persistent representations despite time varying activity*, Current Biology, 22 (2012), pp. 2095–2103.

[33] D. Durstewitz, J. K. Seamans, and T. J. Sejnowski, *Neurocomputational models of working memory*, Nature neuroscience, 3 (2000), pp. 1184–1191.

[34] A. V. Egorov, B. N. Hamam, E. Fransén, M. E. Hasselmo, and A. A. Alonso, *Graded persistent activity in entorhinal cortex neurons*, Nature, 420 (2002), pp. 173–178.

[35] M. A. Erickson, L. A. Maramara, and J. Lisman, *A single brief burst induces glur1-dependent associative short-term potentiation: a potential mechanism for short-term memory*, Journal of cognitive neuroscience, 22 (2010), pp. 2530–2540.

[36] ——, *A single brief burst induces glur1-dependent associative short-term potentiation: a potential mechanism for short-term memory*, Journal of cognitive neuroscience, 22 (2010), pp. 2530–2540.

[37] B. Ermentrout, *Neural networks as spatio-temporal pattern-forming systems*, Reports on progress in physics, 61 (1998), p. 353.

[38] F. Fiebig and A. Lansner, *A spiking working memory model based on hebbian short-term potentiation*, Journal of Neuroscience, 37 (2017), pp. 83–96.

[39] V. Francioni and M. T. Harnett, *Rethinking single neuron electrical compartmentalization: dendritic contributions to network computation in vivo*, Neuroscience, 489 (2022), pp. 185–199.

[40] E. Fransén, B. Tahvildari, A. V. Egorov, M. E. Hasselmo, and A. A. Alonso, *Mechanism of graded persistent cellular activity of entorhinal cortex layer v neurons*, Neuron, 49 (2006), pp. 735–746.

[41] K. Fukuda, E. Vogel, U. Mayr, and E. Awh, *Quantity, not quality: The relationship between fluid intelligence and working memory capacity*, Psychonomic bulletin & review, 17 (2010), pp. 673–679.

[42] J. I. Gold and M. N. Shadlen, *The neural basis of decision making*, Annu. Rev. Neurosci., 30 (2007), pp. 535–574.

[43] J. M. Gold, C. M. Wilk, R. P. McMahon, R. W. Buchanan, and S. J. Luck, *Working memory for visual features and conjunctions in schizophrenia.*, Journal of abnormal psychology, 112 (2003), p. 61.

[44] N. L. Golding, N. P. Staff, and N. Spruston, *Dendritic spikes as a mechanism for cooperative long-term potentiation*, Nature, 418 (2002), pp. 326–331.

[45] M. S. Goldman, *Memory without feedback in a neural network*, Neuron, 61 (2009), pp. 621–634.

[46] M. S. Goldman, A. Compte, and X.-J. Wang, *Neural integrator models*, Encyclopedia of neuroscience, (2009), pp. 165–178.

[47] M. S. Goldman, J. H. Levine, G. Major, D. W. Tank, and H. S. Seung, *Robust persistent neural activity in a model integrator with multiple hysteretic dendrites per neuron*, Cerebral cortex, 13 (2003), pp. 1185–1195.

[48] J. Hardie and N. Spruston, *Synaptic depolarization is more effective than back-propagating action potentials during induction of associative long-term potentiation in hippocampal pyramidal neurons*, Journal of Neuroscience, 29 (2009), pp. 3233–3241.

[49] M. E. Hasselmo and C. E. Stern, *Mechanisms underlying working memory for novel information*, Trends in cognitive sciences, 10 (2006), pp. 487–493.

[50] D. O. Hebb, *The organization of behavior: A neuropsychological theory*, Psychology press, 2005.

[51] S. Hedayati, R. E. O'Donnell, and B. Wyble, *A model of working memory for latent representations*, Nature Human Behaviour, 6 (2022), pp. 709–719.

[52] A. Johnson, K. Seeland, and A. D. Redish, *Reconstruction of the postsubiculum head direction signal from neural ensembles*, Hippocampus, 15 (2005), pp. 86–96.

[53] M. K. Johnson, R. P. McMahon, B. M. Robinson, A. N. Harvey, B. Hahn, C. J. Leonard, S. J. Luck, and J. M. Gold, *The relationship between working memory capacity and broad measures of cognitive ability in healthy adults and people with schizophrenia.*, Neuropsychology, 27 (2013), p. 220.

[54] M. Khona and I. R. Fiete, *Attractor and integrator networks in the brain*, Nature Reviews Neuroscience, 23 (2022), pp. 744–766.

[55] ——, *Attractor and integrator networks in the brain*, Nature Reviews Neuroscience, (2022), pp. 1–23.

[56] A. A. Koulakov, S. Raghavachari, A. Kepecs, and J. E. Lisman, *Model for a robust neural integrator*, Nature neuroscience, 5 (2002), pp. 775–782.

[57] A. Kutschireiter, M. A. Basnak, R. I. Wilson, and J. Drugowitsch, *Bayesian inference in ring attractor networks*, Proceedings of the National Academy of Sciences, 120 (2023), p. e2210622120.

[58] R. Lee and C. Heckman, *Bistability in spinal motoneurons in vivo: systematic variations in rhythmic firing patterns*, Journal of neurophysiology, 80 (1998), pp. 572–582.

[59] S. Lim and M. S. Goldman, *Balanced cortical microcircuitry for maintaining information in working memory*, Nature neuroscience, 16 (2013), pp. 1306–1314.

[60] ——, *Balanced cortical microcircuitry for spatial working memory based on corrective feedback control*, Journal of Neuroscience, 34 (2014), pp. 6790–6806.

[61] J. E. Lisman, J.-M. Fellous, and X.-J. Wang, *A role for nmda-receptor channels in working memory*, Nature neuroscience, 1 (1998), pp. 273–275.

[62] M. London and M. Häusser, *Dendritic computation*, Annu. Rev. Neurosci., 28 (2005), pp. 503–532.

[63] A. Losonczy and J. C. Magee, *Integrative properties of radial oblique dendrites in hippocampal ca1 pyramidal neurons*, Neuron, 50 (2006), pp. 291–307.

[64] S. J. Luck and E. K. Vogel, *Visual working memory capacity: from psychophysics and neurobiology to individual differences*, Trends in cognitive sciences, 17 (2013), pp. 391–400.

[65] M. Lundqvist, P. Herman, and E. K. Miller, *Working memory: delay activity, yes! persistent activity? maybe not*, Journal of neuroscience, 38 (2018), pp. 7013–7019.

[66] Z. F. Mainen and T. J. Sejnowski, *Influence of dendritic structure on firing pattern in model neocortical neurons*, Nature, 382 (1996), pp. 363–366.

[67] G. Major, M. E. Larkum, and J. Schiller, *Active properties of neocortical pyramidal neuron dendrites*, Annual review of neuroscience, 36 (2013), pp. 1–24.

[68] G. Major, A. Polsky, W. Denk, J. Schiller, and D. W. Tank, *Spatiotemporally graded nmda spike/plateau potentials in basal dendrites of neocortical pyramidal neurons*, Journal of neurophysiology, 99 (2008), pp. 2584–2601.

[69] G. Major and D. Tank, *Persistent neural activity: prevalence and mechanisms*, Current opinion in neurobiology, 14 (2004), pp. 675–684.

[70] J. McAfoose and B. Baune, *Exploring visual–spatial working memory: A critical review of concepts and models*, Neuropsychology review, 19 (2009), pp. 130–142.

[71] E. M. Meyers, *Dynamic population coding and its relationship to working memory*, Journal of neurophysiology, 120 (2018), pp. 2260–2268.

[72] E. M. Meyers, D. J. Freedman, G. Kreiman, E. K. Miller, and T. Poggio, *Dynamic population coding of category information in inferior temporal and prefrontal cortex*, Journal of neurophysiology, 100 (2008), pp. 1407–1419.

[73] G. Mongillo, O. Barak, and M. Tsodyks, *Synaptic theory of working memory*, Science, 319 (2008), pp. 1543–1546.

[74] M. Nikitchenko and A. Koulakov, *Neural integrator: A sandpile model*, Neural computation, 20 (2008), pp. 2379–2417.

[75] A. Papoutsi, K. Sidiropoulou, and P. Poirazi, *Dendritic nonlinearities reduce network size requirements and mediate on and off states of persistent activity in a pfc microcircuit model*, PLoS computational biology, 10 (2014), p. e1003764.

[76] P. Park, A. Volianskis, T. M. Sanderson, Z. A. Bortolotto, D. E. Jane, M. Zhuo, B.-K. Kaang, and G. L. Collingridge, *Nmda receptor-dependent long-term potentiation comprises a family of temporally overlapping forms of synaptic plasticity that are induced by different patterns of stimulation*, Philosophical Transactions of the Royal Society B: Biological Sciences, 369 (2014), p. 20130131.

[77] P. Poirazi, T. Brannon, and B. W. Mel, *Pyramidal neuron as two-layer neural network*, Neuron, 37 (2003), pp. 989–999.

[78] P. Poirazi and A. Papoutsi, *Illuminating dendritic function with computational models*, Nature reviews neuroscience, 21 (2020), pp. 303–321.

[79] A. Polsky, B. W. Mel, and J. Schiller, *Computational subunits in thin dendrites of pyramidal cells*, Nature neuroscience, 7 (2004), pp. 621–627.

[80] S. Remy and N. Spruston, *Dendritic spikes induce single-burst long-term potentiation*, Proceedings of the National Academy of Sciences, 104 (2007), pp. 17192–17197.

[81] A. Sandberg, J. Tegnér, and A. Lansner, *A working memory model based on fast hebbian learning*, Network: Computation in Neural Systems, 14 (2003), p. 789.

[82] H. Sanders, M. Berends, G. Major, M. S. Goldman, and J. E. Lisman, *Nmda and gabab (kir) conductances: the "perfect couple" for bistability*, Journal of Neuroscience, 33 (2013), pp. 424–429.

[83] J. Schiller, G. Major, H. J. Koester, and Y. Schiller, *Nmda spikes in basal dendrites of cortical pyramidal neurons*, Nature, 404 (2000), pp. 285–289.

[84] H. S. Seung, D. D. Lee, B. Y. Reis, and D. W. Tank, *Stability of the memory of eye position in a recurrent network of conductance-based model neurons*, Neuron, 26 (2000), pp. 259–271.

[85] H. S. Seung and H. Sompolinsky, *Simple models for reading neuronal population codes.*, Proceedings of the national academy of sciences, 90 (1993), pp. 10749–10753.

[86] P. L. Smith and D. Vickers, *The accumulator model of two-choice discrimination*, Journal of Mathematical Psychology, 32 (1988), pp. 135–168.

[87] N. Spruston, *Pyramidal neurons: dendritic structure and synaptic integration*, Nature Reviews Neuroscience, 9 (2008), pp. 206–221.

[88] G. J. Stuart and N. Spruston, *Dendritic integration: 60 years of progress*, Nature neuroscience, 18 (2015), pp. 1713–1721.

[89] B. Szatmáry and E. M. Izhikevich, *Spike-timing theory of working memory*, PLoS computational biology, 6 (2010), p. e1000879.

[90] B. van Vugt, T. van Kerkoerle, D. Vartak, and P. R. Roelfsema, *The contribution of ampa and nmda receptors to persistent firing in the dorsolateral prefrontal cortex in working memory*, Journal of Neuroscience, 40 (2020), pp. 2458–2470.

[91] M. Wang, Y. Yang, C.-J. Wang, N. J. Gamo, L. E. Jin, J. A. Mazer, J. H. Morrison, X.-J. Wang, and A. F. Arnsten, *Nmda receptors subserve persistent neuronal firing during working memory in dorsolateral prefrontal cortex*, Neuron, 77 (2013), pp. 736–749.

[92] X.-J. Wang, *Synaptic basis of cortical persistent activity: the importance of nmda receptors to working memory*, Journal of Neuroscience, 19 (1999), pp. 9587–9603.

[93] ——, *Probabilistic decision making by slow reverberation in cortical circuits*, Neuron, 36 (2002), pp. 955–968.

[94] ——, *50 years of mnemonic persistent activity: quo vadis?*, Trends in Neurosciences, 44 (2021), pp. 888–902.

[95] D.-S. Wei, Y.-A. Mei, A. Bagal, J. P. Kao, S. M. Thompson, and C.-M. Tang, *Compartmentalized and binary behavior of terminal dendrites in hippocampal pyramidal neurons*, Science, 293 (2001), pp. 2272–2275.

[96] P. A. Williams, P. Larimer, Y. Gao, and B. W. Strowbridge, *Semilunar granule cells: glutamatergic neurons in the rat dentate gyrus with axon collaterals in the inner molecular layer*, Journal of Neuroscience, 27 (2007), pp. 13756–13761.

[97] R. Wilson and L. Finkel, *A neural implementation of the kalman filter*, Advances in neural information processing systems, 22 (2009).

[98] W. Wojtak, S. Coombes, D. Avitabile, E. Bicho, and W. Erlhagen, *A dynamic neural field model of continuous input integration*, Biological cybernetics, 115 (2021), pp. 451–471.

[99] Y. G. M. J. W. XJ, *A dendritic disinhibitory circuit mechanism for pathway-specific gating nat*, Commun, 7 (2016), p. 14.

[100] S. Yang, H. Seo, M. Wang, and A. F. Arnsten, *Nmdar neurotransmission needed for persistent neuronal firing: potential roles in mental disorders*, Frontiers in Psychiatry, 12 (2021), p. 337.

[101] J. Zhang, P. Liu, F. Zhang, and Q. Song, *Cloudnet: Ground-based cloud classification with deep convolutional neural network*, Geophysical Research Letters, 45 (2018), pp. 8665–8672.

[102] H. D. Zimmer, *Visual and spatial working memory: from boxes to networks*, Neuroscience & Biobehavioral Reviews, 32 (2008), pp. 1373–1395.

[103] J. Zylberberg and B. W. Strowbridge, *Mechanisms of persistent activity in cortical circuits: possible neural substrates for working memory*, Annual review of neuroscience, 40 (2017), pp. 603–627.