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

Jiacheng Xu,[1*] Daniel L Cox,[1#] Steven J Luck,[2#]

[1]Department of Physics and Astronomy,
University of California Davis, Davis CA 95616 USA
[2]Center for Mind & Brain and Department of Psychology,
University of California Davis, Davis CA 95616 USA

[*]To whom correspondence should be addressed; E-mail: jiacheng.xu@dartmouth.edu
[#]Senior authors, listed alphabetically

Working memory is a core feature of cognition that enables items to be maintained and manipulated over short durations of time. Stored information can be binary, such as the presence or absence of an object, or multivalued, such as the graded intensity or location of a feature. Remarkably, there are currently no computational models of working memory that can robustly maintain both the graded intensity and spatial location of a stored item. Here, we show how this limitation can be overcome if neurons contain multiple bistable dendritic compartments. We first illustrate the core mechanism in a simple network architecture for both a spiking model with conductance-based dendrites and a rate-based model that permits analytic understanding. We then implement this mechanism within a "ring model" architecture for spatial working memory, in which neurons are arranged in a one-dimensional line or ring such that

the spatial location of an item is encoded by the set of neurons that are active. In contrast to classic ring models, which encode the binary presence and continuous location of an item, we show that the multi-dendrite-neuron model can robustly (without fine tuning and with minimal drift in response to noise) encode both the amplitude and the location of an item in working memory. This work provides a solution to the problem of encoding graded information in spatial working memory and demonstrates how dendritic computation can increase the capacity and robustness of working memory.

Significance Statement

Working memory refers to the ability to actively hold information. This ability provides a fundamental basis for cognitive functions like decision-making and recall. We focus on the modeling of working memory with two different dimensions, such as a hue with its intensity. Conventionally, two dimensions are represented by local activity and activity amplitude. However, previous models fail unless requiring unrealistic parameter fine-tuning. Here, we incorporate dendritic bistability, where each dendrite acts as a separate memory unit, into the classical model. The model can store both the location and amplitude of an input without parameter fine-tuning. This work highlights the power of dendritic computation.

Introduction

"Working memory" refers to a mechanism that provides a temporary storage buffer for cognitive tasks such as search, reasoning, and decision-making.

Working memory must often precisely store analog quantities such as the location, amplitude, or color of an object. How such analog quantities are stored robustly in working memory has been a popular field for mechanistic modeling. Classic models of graded working memory are based on line attractor representations that are defined by having a continuous set of stably maintainable neural activity states, where each state represents the value of a distinct analog feature. Such models typically fall into two classes. In one class, which we term "location-coding" [1,2], the encoded quantity is represented by a fixed amplitude pattern of neural activity that can shift between different neural locations, with the location of the pattern representing the analog stimulus [3]. This class of models has commonly been used to describe the encoding of variables such as the spatial location of an object [4], the orientation of an animal's head in space [5], or the location (hue) of a color on a color wheel. The canonical model of this class is the 'ring' model in which neurons are arranged in a one-dimensional ring, with each neuron preferentially encoding a particular location in the stimulus space. This model has found direct experimental support in the microscopic connectivity of the fly head direction system [6] and has been widely used to explain many cognitive tasks [7]. The second class of models, which we term "amplitude-encoding" [1, 8, 9], encodes memorized quantities in the amplitude of a largely fixed spatial pattern of neural activity. This class of models has been used to represent the memory of amplitude- or intensity-like variables such as luminance [10] or vibration frequency [11]. Such models have direct anatomical support in the connectome of the oculomotor integrator of the eye movement system [12] and are the canonical neural implementation of accumulation-to-evidence based decision-making [13, 14].

In actual organisms, working memory can simultaneously encode both the location and amplitude of an item [15]. Surprisingly, however, few models exist of such joint amplitude-location encoding, and the few existing models exhibit a lack of robustness to maintain analog amplitudes of neural firing. This lack of robustness reflects fundamental limitations of classic models of location and amplitude encoding. Classic location-coding models like the ring model exhibit only a single all-or-none "bump" of activity and do not represent the amplitude at a given location (Fig. 1). Classic amplitude-coding models achieve stable representations of analog stimulus amplitudes only when the synaptic connectivity parameters are fine-tuned to precisely offset intrinsic neuronal decay processes at every network activity level. Previous models of joint amplitude-location encoding have inherited these shortcomings and either require fine-tuning of parameters to store analog amplitudes [16, 17, 18, 19] or trade-off discreteness in representation against robustness [20].

A separate set of amplitude-coding memory models has succeeded in overcoming the fine-tuning problem by incorporating bistable memory components. In these models, analog memory values are generated through summing up the contributions of many bistable neurons [8] or dendritic compartments [21, 22]. These models inherit the robustness of the bistable processes upon which they are built, while also being capable of storing a graded representation of amplitude limited only by the number of bistable components. However, in location-coding memory models, bistable components have only been used to increase memory stability [23], but not to jointly memorize amplitude information. Biophysically, bistability can be generated by positive feedback among small clusters of neurons [8] or, at a cellular level, through positive-feedback among the ionic currents that underlie somatic [23,24,25,26,27,28,29] or local dendritic [30,31] plateau potentials.

Here, we build upon the bistable models of amplitude-coding to provide, to our knowledge, the first model that jointly and robustly encodes location and amplitude information. We first

demonstrate the core amplitude-storage biophysical mechanism through construction of a spiking, multi-dendrite ”autapse” model consisting of a single neuron connected to itself [32]. We then simplify this biophysical model to a rate model that provides analytic insight and use this simplification as the basis for a multi-neuron ring model that can represent both the analog amplitude and analog location of a stimulus. The model exhibits robustness to noise in its encoding of both memory amplitude and location, thus demonstrating how bistable neuronal processes can enhance the performance of working memory.

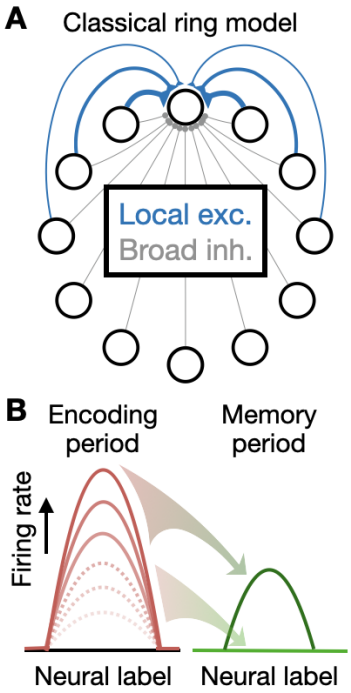

Figure 1: Schematic overview of inability of classic ring models to memorize graded amplitudes. **A**, Classical ring model. Neurons are arranged in a one-dimensional ring with local excitation between nearby neurons (blue) and broad inhibition across all neurons (gray). **B**, Inability to memorize graded amplitudes. In classical ring models, only two stable memory states exist: a fixed amplitude pattern of activity (dark green line) that occurs when inputs during the encoding period have high firing rates (solid red lines), or an untuned baseline level of activity (light green line) that occurs when inputs during the encoding period have low firing rates (dotted red lines). Graded values of input amplitudes are not memorized.

Materials and Methods

We present three models based on multiple bistable dendritic compartments. First, to demonstrate the essential mechanism by which dendritic bistability can lead to robust maintenance of graded-amplitude representations, we consider a spiking 'autapse' consisting of a single neuron with excitatory synapses onto its own conductance-based bistable dendrites. Second, we simplify the spiking autapse to a rate-based model that permits analytical study. Third, we construct a 'ring network' based on the rate-based autapse to show how a network can simultaneously and robustly maintain both the location and the amplitude of a stimulus.

Spiking autapse model with multiple conductance-based, bistable dendrites

The spiking autapse model consists of an integrate-and-fire soma with $N_d = 10$ conductance-based dendrites. The somatic voltage V_s is governed by

$$C\frac{dV_s}{dt} = -I_L - I_d - I_e - I_{ton} \tag{1}$$

$$= -\bar{g}_L(V_s - E_L) - \sum_{i}^{N_d} \bar{g}_{ds}(V_s - V_{d,i}) - \bar{g}_e s_e(V_s - E_{AMPA}) - I_{ton}, \tag{2}$$

where I_L is the leak current, I_d is the summed current received from all dendrites, I_e is the excitatory current from external input through AMPA receptors, and I_{ton} is the tonic background current. We chose parameter values similar to those of a previous model [31]. The membrane capacitance is $C = 10$ nF/mm^2. The maximal conductances are $\bar{g}_L = 0.5$ μS/mm^2 and $\bar{g}_e = 5.2$ μS/mm^2. The conductance mediating the flow of currents from dendrite to soma is $\bar{g}_{ds} = 0.007$ μS/mm^2. The equilibrium potentials are $E_L = -80$ mV and $E_{AMPA} = 0$ mV. The tonic current is $I_{ton} = -17.2$ nA/mm^2. The synaptic activation variable s_e obeys 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, after which it resets the somatic voltage to $V_{reset} = -80$ mV.

The voltage dynamics of the i-th dendritic compartment, $V_{d,i}$, were adapted from [31], with modified maximum conductance values and the inward rectifier potassium (KIR) conductance simplified to not include a synaptically driven component:

$$C\frac{dV_{d,i}}{dt} = -I_{L,i} - I_{KIR,i} - I_{s,i} - I_{AMPA,i} - I_{NMDA,i}, \tag{3}$$

where the first two terms give the intrinsic dendritic leak current $I_{L,i}$ and inward-rectifying potassium current $I_{KIR,i}$, the third term $I_{s,i}$ gives the current arriving from the soma, and the final two terms give the AMPA and NMDA components of the synaptic currents driven recurrently by somatic spiking. These currents are described by the following equations:

$$I_{L,i} = \bar{g}_L(V_{d,i} - E_L), \tag{4}$$

$$I_{KIR,i} = \bar{g}_{KIR}\frac{V_{d,i} - E_{KIR}}{1 + e^{0.1(V_{d,i} - E_{KIR} + 10)}}, \tag{5}$$

$$I_{s,i} = \varkappa\bar{g}_{ds}(V_{d,i} - V_s), \tag{6}$$

$$I_{AMPA,i} = \bar{g}_{AMPA}s_{AMPA}(V_{d,i} - E_{AMPA}), \tag{7}$$

$$I_{NMDA,i} = \bar{g}_{NMDA,i}s_{NMDA}\frac{V_{d,i} - E_{NMDA}}{1 + 0.15e^{-0.08V_{d,i}}}, \tag{8}$$

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, $\bar{g}_{AMPA} = 0.38$ μS/mm^2 and $\bar{g}_{KIR} = 8.6$ μS/mm^2. The maximum conductance for the NMDA receptors is $\bar{g}_{NMDA,i} = \frac{48}{(i+4.5)^{0.54}}$ μS/mm^2.

The synaptic activation s_α obeys:

$$\frac{ds_\alpha}{dt} = -s_\alpha/\tau_\alpha, \text{ at all times} \tag{9}$$

$$\Delta s_\alpha = 0.5(1 - s_\alpha), \text{ each time a spike is received,} \tag{10}$$

where α denotes AMPA (s_e and s_{AMPA}) or NMDA (s_{NMDA}) receptors with a decay time constant $\tau_{AMPA} = 2$ ms or $\tau_{NMDA} = 100$ ms, respectively.

Each dendrite has a bistable input-output relation as shown in Fig. 2B. To derive these intrinsic bistable relations for each dendrite, for this panel, we removed all of the self-connections

shown in Fig. 2A. Next, a periodic train of spikes, which steps up and down across a wide range of frequencies, was delivered to a dendrite while its dendritic voltage was recorded. During this process, the dendritic voltage may also be affected by somatic voltage changes through the electrical coupling between the somatic and dendritic compartments. To account properly for the fact that the somatic voltage is driven jointly by all dendrites, the same periodic train is delivered to other dendrites simultaneously so as to generate a somatic spike train resembling that of the periodic inputs. For the simulation, V_s and $V_{d,i}$ were initialized to –70 mV and the periodic train of input spikes delivered to each dendrite started at a rate of 0 Hz. The spike train lasted for a period of 1400 ms, and the averaged dendritic voltages over the last 1000 ms were recorded for all dendrites simultaneously. Next, the frequency of the periodic input train was increased by 1 Hz, and the recording repeated. Once the frequency reached 70 Hz, it was decreased by 1 Hz at each recording until returning to 0 Hz.

To demonstrate how the network maintains multiple levels of activity, in Fig. 2C, we implemented the following procedure: V_s and $V_{d,i}$ were initialized to –70 mV and the network dynamics were allowed to stabilize for a period of 1200 ms before providing external inputs. External input consisted of steps on and off of periodic spikes for 400 ms, followed by an 800 ms memory period with no external input. The frequency of external input started at 20 Hz and increased by 5 Hz for each step until reaching 65 Hz. For the parameter values given above, this procedure activated one additional dendrite with each step increase in frequency, as shown in Fig. 2C.

Rate-based autapse model

Simplified autapse dynamics

To facilitate analytic understanding, we simplified the spiking autapse to a rate-based autapse. To do so, we assumed each dendrite has a bistable steady-state input-output relation as a

function of the firing rate r, as shown in Fig. 3A:

$$D_i(r) = \begin{cases} \beta\Theta(w_i r - T_u) & \text{Activation from down- to up-state} \\ \beta\Theta(w_i r - T_d) & \text{Deactivation from up- to down-state,} \end{cases} \tag{11}$$

where β is the dendritic contribution to firing rate for an up-state dendrite, $\Theta(x)$ indicates the Heaviside (unit step) function, w_i is the weight of the connection onto the ith dendrite, and T_u and T_d are constant up and down thresholds. Dendrites are ordered such that wi is a monotonically decreasing function of the dendrite index i. A larger synaptic weight w_i, which corresponds to a larger value of $\bar{g}_{NMDA,i}$ in the spiking model, effectively lowers the up and down thresholds. We can then rewrite $D_i(r)$ in terms of threshold firing rates for activating or deactivating a dendrite,

$$D_i(r) = \begin{cases} \beta\Theta(r - r^{th}_{u,i}) & \text{Activation from down- to up-state} \\ \beta\Theta(r - r^{th}_{d,i}) & \text{Deactivation from up- to down-state,} \end{cases} \tag{12}$$

where the effective up and down thresholds are:

$$r^{th}_{u,i} = \frac{T_u}{w_i} \text{ and } r^{th}_{d,i} = \frac{T_d}{w_i}, \tag{13}$$

For analytic simplicity, we assumed the somatic firing rate dynamics are given by an exponential approach to a steady state given by the sum of the dendritic input contributions and the external input to be remembered I_e:

$$\tau\frac{dr}{dt} = -r + \sum_{i=1}^{N_d} D_i(r) + I_e \tag{14}$$

where N_d is the number of dendrites. τ is chosen to capture, in a simple manner, the slowest timescale of neural activity, which is set by the NMDA receptor kinetics in the spiking model. Explicit inclusion of additional synaptic or dendritic dynamics does not qualitatively change the primary results of this work.

Determination of the shape of the multistable band

The total dendritic output to the soma $\sum_{i=1}^{N_d} D_i(r)$ graphically defines a band of multi-stable memory storage. Here, we provide the methods by which the edges of this band and their continuum approximation (Fig. 3) are defined. Following Eq. 14, the right side of the band, which consists of the set of up thresholds of sequentially activated dendrites is given by

$$B_u(r) = \sum_{i=1}^{N_d} \beta\Theta(r - r_{u,i}^{th}) \tag{15}$$

For a large number of dendrites, the weight function w_i and the edge of the band can be approximated by a continuous set of values. Replacing the discrete index i by a continuum index $x \in [0, N_d]$ and assuming w_i changes gradually, we approximate the discrete set of w_i by a continuous monotonically decaying function w(x). Then we have

$$r_u^{th}(x) = \frac{T_u}{w(x)} \tag{16}$$

and B_u becomes:

$$B_u = \int_0^{N_d} \beta\Theta(r - r_u^{th}(x))dx. \tag{17}$$

To derive an analytical expression for the right side of the band, we evaluate the above integral. Since $r_u^{th}(x)$ is a monotonically increasing function of x, the integrand is zero for $r < r_u^{th}(0)$ and a constant value of βN_d for all values of x such that $r > r_u^{th}(N_d)$. For intermediate firing rate values, using Eq. 16, the integrand contributes a constant value of beta to the integral for all x values such that $r > T_u/w(x)$ so that:

$$B_u(r) = \begin{cases} 0, & \text{for } r < r_u^{th}(0) \\ \beta w^{-1}(\frac{T_u}{r}), & \text{for } r_u^{th}(0) \leq r \leq r_u^{th}(N_d) \\ \beta N_d, & \text{for } r > r_u^{th}(N_d) \end{cases} \tag{18}$$

Similar reasoning applies in deriving $B_d(r)$.

Determination of the memory amplitude as a function of the stimulus strength

To obtain the continuum limit relationship between the steady-state memory activity amplitude r_m and the stimulus strength I_e (Fig. 3F), we proceeded as follows. During the encoding period, activity is non-decreasing, so that dendrites are only activated but not deactivated. Thus, during this period, the total dendritic activity $\sum_{i=1}^{N_d} D_i(r)$ in Eq. 14 simplifies to $B_u(r)$, so that the steady-state value of activity that is approached during the encoding period is determined from the relation:

$$B_u(r) = r - I_e. \tag{19}$$

This represents the intersection of the band edge $B_u(r)$ (Fig. 3C-E, solid line) and the line $y = r - I_e$ (Fig. 3C-E, dashed line) and is indicated by a red dot in Figs. 3C-E. We assume that the encoding period is sufficiently long to reach this steady state and that all dendrites activated during the encoding period remain on during the memory period, which is the case if the firing rate does not decrease past the down-threshold line B_d. Then, the steady-state firing rate r_m maintained during the memory period (Fig. 3C-E, green dot) is related to the steady-state firing rate r during the encoding period by $r_m = r - I_e$, where the presence and subsequent removal of I_e are represented by the rightward and leftward pointing arrows in Fig. 3C-E. Substituting this expression into Eq. 19 gives:

$$B_u(r_m + I_e) = r_m. \tag{20}$$

For a given form of the weight function w(x), and using Eq. 18, Eq. 20 can be solved to give the function relating the external input I_e and the steady-state memory activity r_m. As an example, we solve for a linear function (up to thresholding and saturation) here. From Eq. 20, a linear function requires $B_u(r)$ to be linear in the non-thresholded, non-saturated regions, which corresponds to the same amount of input increase always activating the same number of extra dendrites (Fig. 3C). From Eq. 91 (and its discrete-valued equivalent from Eq. 13, this can be

achieved by a weight function of the form $w_i = \frac{a}{i+b}$ in the discrete case or $w(x) = \frac{a}{x+b}$ in the continuous case, where a and b are constants. Solving for $w^{-1}(x)$ and substituting into Eq. 18 gives the linear equation for the (continuum) band:

$$B_u(r) = \frac{a\beta}{T_u} r - b\beta, \text{for } \frac{T_u b}{a} \leq r \leq \frac{T_u(N_d + b)}{a}. \tag{21}$$

Based on Eq. 20, the memory amplitude as a function of external input is given by the linear relationship:

$$r_m = \frac{a\beta(I_e - I_{e,th})}{T_u - a\beta}, \tag{22}$$

where $I_{e,th} = \frac{T_u b}{a}$ is the minimal threshold value of external input. We note that such a linear function only exists when the slope of $B_u(r)$ in Eq. 21 is less than unity, $\frac{a\beta}{T_u} < 1$, so that it can intersect with the shifted decay line $y = r - I_e$ (Fig. 3C, right dashed line).

In Fig. 3, all plots used parameters $\beta = 1$, $T_u = 9$, and $T_d = 2$. For the steady-state behavior focused on here, $\frac{dr}{dt} = 0$, the value of τ is irrelevant. For Fig. 3B, the dendritic weights are given by $w_i = \frac{4}{i+1}$, with i from 1 to $N_d = 10$ for the discrete case, and the continuum approximation is identical to that in Fig. 3C.

Rate-based model with ring architecture

The ring architecture network underlying Figures 4-7 and illustrated in Fig. 4A was constructed as follows. The network consists of N neurons, with each neuron having $N_d = N$ dendrites.The recurrent excitatory weights $w_{ij}(d)$ were chosen to be symmetric, with a falloff in strength as a function of the distance d between neurons along the periodically arranged ring. Each excitatory connection projects to a separate bistable dendrite on the postsynaptic neuron. In addition, there is uniform recurrent inhibition with summed strength proportional to the total firing rate across the network, received by the soma of each neuron. The firing rate dynamics of

each neuron is given by:

$$\tau \frac{dr_i}{dt} = -r_i + \sum_{j=1}^{N} D_{ij}(r_j) - w_{inh} \sum_{j=1}^{N} r_j + I_{e,i}, \tag{23}$$

where $D_{ij}(r_j)$ is the bistable dendritic input-output relation, w_{inh} is the weight for the global inhibition, and $I_{e,i}$ is the external input. The dendritic relation $D_{ij}(r_j)$ for the contribution to the firing rate of neuron i from the dendrite receiving input from neuron j, is defined analogously to $D_i(r)$ in the rate-based autapse model:

$$D_{ij}(r_j) = \begin{cases} \beta\Theta(w_{ij}r_j - T_u) & \text{Activation from down- to up-state} \\ \beta\Theta(w_{ij}r_j - T_d) & \text{Deactivation from up- to down-state.} \end{cases} \tag{24}$$

Unlike the autapse network, it is difficult to provide a simple, analytic characterization of the system dynamics described by Eq. 23. This is because, although each individual neuron's behavior for a given location of the memory-storing bump can be qualitatively conceptualized by a multistable band defined by the sum of its dendritic inputs, each dendrite in the ring model receives a different input r_j. Therefore, the shape of such a multistable band does not just depend on the local excitatory weights w_{ij}, but also on a specific pattern of firing rates r_j. Because r_j changes during the encoding period, and depends on the amplitude and shape of $I_{e,i}$, the exact shape of the multistable band is not uniquely defined. However, as shown in the following section, an approximate multistable band (Fig. 4F) and relationship between input amplitude and memory amplitude (Fig. 4G) can be derived analytically.

Calculation of how the shape of the weight function determines the mapping between the external input and memory amplitude in the ring network

Here, analogous to the case of the autapse, we show how (1) the amplitude of the external input can be approximately related to the memory amplitude for a given weight function w_{ij}, where the memory amplitude is defined as the firing rate of the peak firing neuron in the bump of activity, and (2) the form of the weight function required to get a linear mapping between external input amplitude and memory amplitude.

During the encoding period, we assume for simplicity that neurons start from zero firing rate and only increase their firing rates due to the external stimulus. In this case, the term $D_{ij}(x)$ can be replaced in Eq. 23 by a step function, of height β, corresponding to the condition for activating a dendrite from its down to its up state:

$$\tau \frac{dr_i}{dt} = -r_i + \beta \sum_{j=1}^{N} \Theta[w_{ij} r_j - T_u] - w_{inh} \sum_{j=1}^{N} r_j + I_{e,i}. \tag{25}$$

In the continuum limit, the stabilized activity $r_e(x)$ during the encoding period obeys:

$$r_e(x) = \beta \int_0^N \Theta[w(x-z) r_e(z) - T_u] dz - w_{inh} \int_0^N r_e(z) dz + I_e(x), \tag{26}$$

where $w(x-z)$ is the excitatory weight as a function of the shortest distance between x and z along the ring. A dendrite is activated if $w(x-z)r(z) > T_u$. Thus, the first integral gives the total number of activated dendrites. In Fig. 4E, we illustrate schematically how this integral can be graphically solved for the maximally firing neuron of a bump of activity (postsynaptic neuron x in the above equation) in response to a local symmetric input $I_e(x)$. The total dendritic input ($w(x-z)r(z)$, black) is plotted as a function of the presynaptic neuron index z. Activated dendrites are those (labeled in green) that exceed the threshold value T_u (plotted as a horizontal dashed line).

To derive an approximate multi-stable band and an approximate relationship between the external input amplitude and bump height, we make two assumptions. First, we assume T_d is low enough so that no dendrites flip from their up- to down-state following the offset of the external stimulus (i.e. during the memory period). Therefore, the stabilized activity $r_m(x)$ is

$$r_m(x) = \beta \int_0^N \Theta[w(x-z) r_e(z) - T_u] dz - w_{inh} \int_0^N r_m(z) dz. \tag{27}$$

Note that the integral in Eq. 27 is identical to the first integral in Eq. 26, indicating that the total dendritic contribution is unchanged. Second, we assume that $I_e(x) = I_A R(x)$, $r_e(x) = E_A R(x)$,

and $r_m(x) = M_A R(x)$, where $R(x)$ is the shared shape profile across three quantities and is defined to be symmetric, unimodal (bump-like) with a peak value of 1. I_A, E_A, and M_A then give the amplitudes of the corresponding quantities.

Deriving the approximate shape of the multistable band

Given these assumptions, we first derive the approximate shape of the multistable band of recurrent network feedback analogous to that done for the autapse case (Eq. 18). We focus our analysis on the neuron at the peak of the bump of activity, which we label without loss of generality as $x = 0$. The total dendritic activity for neuron $x = 0$, denoted as $D_{tot}(E_A)$, in Eq. 26 or Eq. 27 follows:

$$D_{tot}(E_A) = \beta \int_{-\frac{N}{2}}^{\frac{N}{2}} \Theta[w(z)R(z)E_A - T_u]dz \tag{28}$$

$$= 2\beta \int_{0}^{\frac{N}{2}} \Theta[w(z)R(z)E_A - T_u]dz, \tag{29}$$

where in the second line we have used the symmetry of the weight matrix and the neural activity profile. Performing the integral analogously to the calculation for the autapse case then gives:

$$D_{tot}(E_A) \begin{cases} 0, & \text{for } E_A < \frac{T_u}{w(0)} \\ 2\beta(w \circ R)^{-1}(\frac{T_u}{E_A}), & \text{for } \frac{T_u}{w(0)} < E_A < \frac{T_u}{w(\frac{N}{2})R(\frac{N}{2})} \\ \beta N, & \text{for } E_A > \frac{T_u}{w(\frac{N}{2})R(\frac{N}{2})}, \end{cases} \tag{30}$$

where the inverse function of the composite function from $w(z)$ and $R(z)$, $(w \circ R)^{-1}(x)$, is needed.

During the memory period, from Eq. 27, the peak of the activity bump is given by:

$$(1 + w_{inh}\varkappa)M_A = D_{tot}(E_A), \tag{31}$$

where $\varkappa = \int_0^N R(x)dx$ is a constant for a given shape based on a known $R(z)$, and we have used the symmetry of the weight matrix and the input and neural activity profiles.The stable point

is given by the intersection of $D_{tot}(E_A)$ and the effective decay line, which has the additional effect of recurrent inhibition, $y = (1 + w_{ihn}\varkappa)M_A$.

We note that analogous reasoning can be applied to determine the downward band limit B_d in response to steady inhibitory input.

Deriving the weight required to provide a given relationship between the external input I_A and memory activity M_A

We next derive the symmetric weight $w(|z|)$ required to generate a given relationship $M_A = f(I_A)$ between the external input and the height of the activity bump during the memory period. To do this, we first observe that the firing rate M_A of the neuron at the peak of the bump of activity during the memory period is determined by the total number of activated dendrites on this neuron. Dendrites are activated when the presynaptic input that arrives on them during the encoding period, $w(z)R(z)E_A$, exceeds the threshold T_u. The last presynaptic neurons to activate their postsynaptic dendrite thus are defined by the equation:

$$w(|z_{max}|)R(|z_{max}|)E_A = T_u, \tag{32}$$

where $|z_{max}|$ gives the absolute value of the continuum index of such neurons. Next we want to rewrite this equation for $w(|z_{max}|)$ in terms of M_A. To do so, we make two observations. First, from Eq. 26 and Eq. 31, the steady-state amplitude E_A of the bump at the end of the encoding period is related to M_A and I_A as follows:

$$E_A = D_{tot}(E_A) - w_{inh}\varkappa E_A + I_A \tag{33}$$

$$(1 + w_{inh}\varkappa)E_A = (1 + w_{inh}\varkappa)M_A + I_A \tag{34}$$

$$E_A = M_A + \frac{I_A}{1 + w_{inh}\varkappa}. \tag{35}$$

Second, from Eq. 31, the total dendritic activity $2\beta|z_{max}|$ during the memory period, M_A, is given by

$$2\beta|z_{max}| = (1 + w_{inh}\varkappa)M_A. \tag{36}$$

From Eq. 32, and using that $I_A = f^{-1}(M_A)$ for non-zero values of M_A, we thus have:

$$w(|z_{max}|)R(|z_{max}|)[M_A + \frac{f^{-1}(M_A)}{1 + w_{inh}\varkappa}] = T_u \tag{37}$$

$$w(|z_{max}|)R(|z_{max}|)[2\beta|z_{max}|+f^{-1}(M_A)] = (1 + w_{inh}\varkappa)T_u. \tag{38}$$

Since the above equation is a requirement on w(x) for any distance from the peak of the bump $x = |z_{max}|$ that could occur with different amplitudes of input I_A, we finally have the general relation:

$$w(x) = \frac{T_u(1 + w_{inh}\varkappa)}{R(|x|)[2\beta|x|+f^{-1}(\frac{2\beta|x|}{1+w_{inh}\varkappa})]} \tag{39}$$

For the particular case of a linear relationship $M_A = aI_A + b$ between the external input and memory period bump amplitudes, we have:

$$I_A = f^{-1}(M_A) = \frac{M_A - b}{a}, \tag{40}$$

which gives:

$$w(x) = \frac{aT_u(1 + w_{inh}\varkappa)}{R(x)(2a\beta x + \frac{2\beta x}{1+w_{inh}\varkappa} - b)}. \tag{41}$$

We note that, for a flat input shape, $R(x) =$ constant, this reduces to the formula found for the rate-based autapse case. In addition, note that, although $w(x)R(x)$ decreases monotonically for monotonically decreasing input profiles, w(x) from the above formula may not be monotonically decreasing in $|x|$ or may even blow up if R(x) approaches zero at the edge of the bump. To offset this, and ensure that w(x) is monotonically decreasing as $|x|$ increases, in all simulations, we manually set w(x) to zero once it stopped monotonically decreasing.

Simulation details in the noise-free case

The default setting for the ring network model simulations uses $N = 360$ neurons, with $w_{inh} = \frac{1}{360}$, $\beta = 1$, $T_u = 9$, $T_d = 2$ and $\tau = 50$ ms, unless specified otherwise. For the model of Fig. 4B-D, the local excitatory weights were defined by a power law function $w_{ij} = \frac{18}{(d_{ij}+2)^2}$

and the external input to each neuron, $I_{e,i}$, was a Gaussian function, $I_A e^{-\frac{(i-180)^2}{2\times 10^2}}$. The input amplitude values simulated are $I_A = 8, 16, 24$, or 32.

With all dendrites originally in the down-state, $I_{e,i}$ was applied for 1000 ms (the 'encoding period'). The input was then turned off, and the subsequent 'memory period' activity was simulated for another 1000 ms. The duration is sufficient for the network activity to reach equilibrium.

The approximate multistable band of network feedback shown in Fig. 4B corresponds to the neuron $i = 180$ at the center of the bump of activity that determines the height of the activity bump. The B_u and B_d curves that define the edges of the band were obtained by simulation as follows.

- Tracing the up-threshold B_u. During the encoding period, we plot the total dendritic activity $\beta \sum_{j=1}^{N} \Theta(w_{ij} r_j - T_u)$ in Eq. 25 against the firing rate r_i, with time as the parametric variable. This numerically generates the up-threshold line B_u. The input amplitude used in generating this plot was $I_A = 80$.

- Tracing the down-threshold B_d. The firing rate is first allowed to reach stable memory activity after undergoing a delay period. We then apply a uniform inhibition, $I_{e,i} = -10$ for 1000 ms. This is large enough to shut off the memory activity.

Finally, to get the decay line as in the autapse case, we connected the stable point during the memory period with the origin, Fig. 4B, dashed line. Different from the autapse case, this line has a slope larger than 1, as the recurrent inhibition in the ring model effectively makes activity decay more. The overlap between the dashed line and the multistable band approximately locates stable points during the memory period.

To test the sensitivity of the shape of the multistable band to the specific form and pattern of excitatory and inhibitory inputs to the network, in Fig. 5 we re-generated the band of Fig.

4B under different parameter settings. In Fig. 5A-C, we changed the input during the encoding period. In Fig. 5A, we used a lower input amplitude $I_A = 40$. In Fig. 5B, we applied two successive inputs centered at the same location. First, a smaller input with $I_A = 20$ was applied, leading the subsequent memory activity to stabilize at the location of the red dot. Then a larger input with $I_A = 80$ was applied, followed by a second memory period. In Fig. 5C, we used a local cosine input $80\cos(\frac{i-180}{12})$, i from 162 to 198. In Fig. 5D-F, we changed the inhibition applied for generating B_d. In Fig. 5D, we used a smaller uniform inhibition $I_{e,i} = -6$. In Fig. 5E, we applied two successive uniform inhibitions, separated by a memory period. The first one had $I_A = -4$, leading the memory activity to drop from the black dot to the red dot. The second one had $I_A = -10$, erasing the memory. In Fig. 5F, we changed the uniform inhibition to a Gaussian one, $I_{e,i} = -10e^{-\frac{(i-180)^2}{2\times 10^2}}$.

In Fig. 4F, G, we checked the accuracy of our analytical calculation for the weight function required to produce a linear memory amplitude versus input amplitude relationship. From Eq. 41 above, we calculated the weight function required to produce the linear relation $I_A = aM_A + b$, with $a = 0.5$ and $b = -0.6$. We used an input $I_e(x) = I_A R(x)$, with $R(x) = e^{-\frac{(x-180)^2}{2\times 10^2}}$, corresponding to a value $\varkappa \approx 10\sqrt{2\pi}$. To better compare the analytical result obtained under continuum limit with discrete simulations, Eq. 25, we ran simulations also in the continuum limit, with neuronal index x finely discretized into $dx = 0.2$. We also extended the duration of the encoding and memory periods to 2000 ms to ensure activity equilibrated. In Fig. 4G, I_A takes integer values from 0 to 35, with the inset gives some exemplar memory activities for I_A from 5 to 20 in steps of 5. The desired multistable band in Fig. 4F, shaded red, is obtained by plugging in relevant values into Eq. 31 and Eq. 30.

Simulation details under noise

For all simulations with noise, spatially and temporally independent Gaussian noise of mean 0 and standard deviation σ was applied to the soma of each neuron at each time step of duration $dt = 1$ ms throughout the entire session. All other model parameters are identical to Fig. 4B-D, unless specified otherwise.

In Fig. 6B, $\sigma = 2$, 10, or 15, and the external input amplitude I_A = 10, 30, or 60. 10 trials were conducted for each condition. For each trial, the memory amplitude was estimated by averaging the firing rates of the 10 highest firing neurons and plotted for a memory period of $10,000$ ms.

In Fig. 7A, B, we simulated diffusion of the bump location under noise for different bistable ranges, starting from the step function case with $T_u = T_d$ and increasing the bistable range $T_u = T_d$. However, there is a complication. As T_u increases, not just the range is increased, but the memory activity gets smaller, which potentially makes it more susceptible to noise. Vice versa for the decrease of T_d. Therefore, to approximately keep the memory activity the same, we changed the protocol. Initially, we set $T_u = T_d = 3$ and noise zero. For an encoding period of 1000 ms, we applied an external input amplitude $I_A = 15$. The subsequent first memory period takes 1000 ms. Then, we increased the bistable range by increasing T_u and decreasing T_d. This procedure does not make up-state dendrites drop nor make down-state activate, keeping the memory activity unchanged. Next, for the second memory period of 1000 ms, we set the standard deviation of noise to be 10. Following this protocol, before the noise was applied, memories across for different sets of bistable ranges are identical. In Fig. 7A, during the second memory period, the blue trace takes $T_u = T_d = 3$ unchanged, the orange trace has $T_u = 3.6$ and $T_d = 2.4$, and the green trace has $T_u = 4.2$ and $T_d = 1.8$. For each trace, we record the variances of memory locations, which were estimated by the center of mass and averaged over 400 trials, throughout the second delay period. Every trial maintains a stable local memory

throughout the time. In Fig. 7B, the final location variance is plotted against the bistable range, with $T_u = 3 + 0.2s$ and $T_d = 3 - 0.2s$ for an integer s from 0 to 10.

In Fig. 7C, we simulated diffusion of the bump location under noise for different maintained amplitudes. We used the default protocol as in Fig. 6, instead of the one used in Fig. 7A, B. We set $T_u = 4$, $T_d = 1$, and $\sigma = 15$. In Fig. 7C, left, we record the memory amplitude for 400 trials with $I_A = 8$ (light gray), 15 (dark gray), or 30 (black). The corresponding averaged variances of the bump locations are shown in Fig. 7C, right.

Implementations of numerical simulation

For all simulations, we integrated the differential equations using the forward 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 Figures 4B, 4F, 5, and 6A, where 0.01 ms was used for better approximation of the simulated bistable band. Code was written in python version 3.9.16 and is available upon request.

Results

Memorizing amplitudes in autapse network with bistable dendrites

To construct a model for memorizing both amplitude and location of a stimulus, we deal with three questions progressively: 1. How persistent activity is maintained; 2. How the maintained activity shows graded amplitudes; 3. How this mechanism can be extended to memorize both amplitude and location of a stimulus. The simplest model to maintain persistent activity might be a single neuron connected back to itself, an autapse [32]. In the absence of an external stimulus, the autapse can be self-sustained by feeding back excitation to itself, with a stable amplitude where feedback excitation equals activity decay. By properly tuning the strength of feedback excitation through synaptic weight, amplitude can be stable with multiple values. However, classically, synaptic weight has to be fine-tuned such that a slight deviation breaks the

system down completely. We show how our model can avoid this problem in this section. Later, we show how this idea is generalized to memorize both amplitude and location of a stimulus with a ring architecture, where the self feedback excitation is replaced by feedback excitation from nearby neurons.

In this section, we focus on the autapse model. First, we introduce a spiking autapse to demonstrate the key mechanism for the memorization of input amplitudes. Next, we simplify to a rate-based autapse with a detailed explanation of the mechanism. We find that autapse robustly memorizes memory in synaptic driven activity over a wide range of parameter values with no need for fine-tuning, and that synaptic weights determine memory performance.

A spiking autapse with multiple bistable dendrites

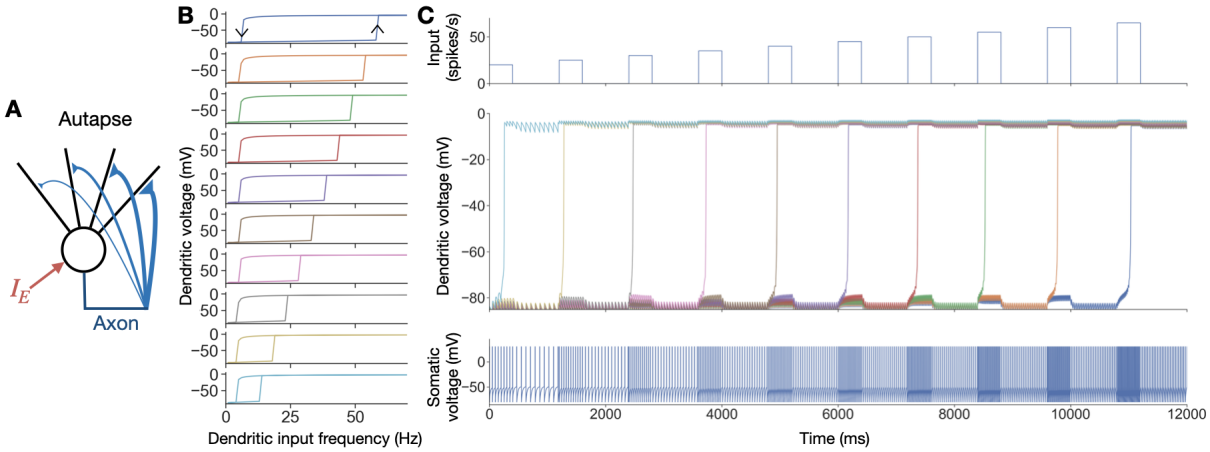

Figure 2: Maintenance of graded firing rates in the spiking autapse model. **A**, Model architecture. The model consists of an integrate-and-fire soma that projects to ten conductance-based dendritic compartments with bistability based on the voltage-dependent properties of the NMDA and KIR channels (see Methods). Each compartment receives input with a different weight (the maximal NMDA receptor conductance, $\bar{g}_{\mathrm{NMDA,i}}$), so that different dendritic compartments recruit sequentially with increased amplitude of input. External input (red) goes into the soma. **B**, Hysteretic input-output relations of the ten dendrites. To characterize the bistable input-output relations of each dendrite, the synaptic inputs from the soma were removed and replaced by periodically firing, externally applied synaptic inputs that stepped across a range of different frequencies (x-axis). The shown voltages represent averages over 1000 ms of recording. Arrows in the top panel indicate the directions of the hysteretic response. Top to bottom panels show the 10 dendrites, ordered by increasing NMDA conductance $\bar{g}_{\mathrm{NMDA,i}}$ (or equivalently, decreasing up threshold). **C**, Memorization of different amplitudes. For the model in **A**, an external input spike train to be encoded by the network is briefly applied, followed by a memory period with no input. This is repeated 10 times for successively increasing external input frequencies (top). Dendrites are activated during the encoding periods, starting from the one with the largest $\bar{g}_{\mathrm{NMDA,i}}$, and self-sustain their activity during the memory period (middle). Somatic activity, which represents the memorized stimulus during the memory period, reflects the number of activated dendrites (bottom).

To memorize graded amplitudes, a spiking autapse with multiple bistable dendrites are constructed. Each bistable dendrite is like a small binary memory unit such that graded amplitude is achieved by progressively recruiting dendrites. As illustrated in Fig. 2A, an integrate-and-fire soma is self-connected. Each self-connection projects to a separate dendrite. Each dendrite follows conductance-based dynamics modified from a previous model [31]. As shown in Fig. 2B, this makes each dendrite follow a bistable dendritic input-output relation. The dendritic voltage stays around –80 mV (down-state) as the dendritic input frequency increases from 0 Hz. Once the dendritic input frequency exceeds a certain (up) threshold, the dendritic voltage activates to around 0 mV (up-state) due to the sudden opening of NMDA receptors. This NMDA input can then keep the voltage stably up even for smaller external input, so that For deactivation from up-state to down-state occurs at a lower, dendritic input frequency should reduce below another (down) threshold. In total, there are 10 such dendrites, identical except for different NMDA conductance values, $\bar{g}_{\mathrm{NMDA,i}}$. A higher $\bar{g}_{\mathrm{NMDA,i}}$ gives lower up and down thresholds, which is plotted at the bottom of Fig. 2B. Fig. 2C shows how amplitude memorization is achieved by recruiting more dendrites. Initially an external input with a frequency of 20 Hz is applied to the soma. This causes the dendrite with the largest NMDA conductance to be activated to a voltage plateau (cyan line, middle). The activation leads to sustained memory activity during the memory period where the external input is turned off. A larger external input frequency can successively activate other dendrites with smaller NMDA conductances, which results in larger memory activity. More simulation details can be found in Materials and Methods.

A rate-based autapse: the core mechanism for memorizing amplitudes

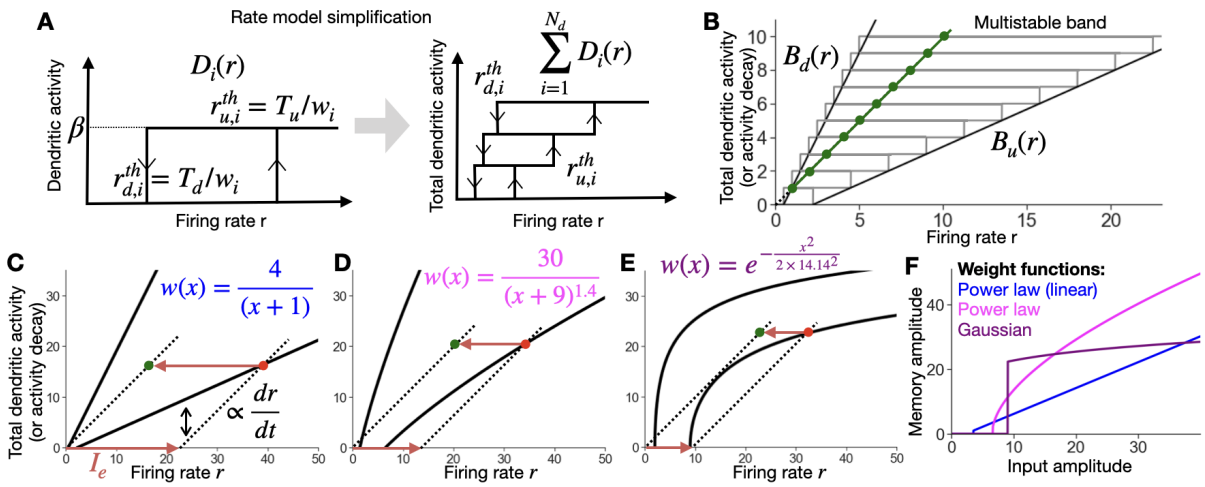

Figure 3: Rate-based autapse model for storage of graded amplitude inputs. **A, B,** Rate based model captures the key behavior of the spiking model of Fig. 2B. **A**, Left, bistable input-output relation $D_i(r)$ of a dendrite i. The input is the somatic firing rate, which feeds back to the dendrite through the self (autaptic) connection. $D_i(r)$ is hysteretic with jumps at up- and down-thresholds, whose values are inversely related to the synaptic weight w_i. Right, summed dendritic output to the soma from all dendrites (3 shown). Each rectangle represents the contribution from one dendrite. Dendrites with smaller w_i values have smaller thresholds and are plotted at the bottom. **B**, Multistable band in the discrete or continuous case. Multiple dendrites form a band with two right and left edges B_u, B_d consisting of up and down thresholds, either in discrete steps (gray) or the continuum limit (black). The dashed line represents the decay of firing rate activity (magnitude of term -r in Eq. 14. Intersections of the line and multistable band (the green line in the continuous case or green dots in the discrete case) represent points at which dendritic feedback offsets firing rate decay to lead to persistent firing (i.e., memory) states in the absence of external input. **C-E** Different weight functions lead to different shapes of the multistable band and different responses to external input. **C** Linear band derived from a power law decay weight function. During the encoding period, an external input I_e shifts the dashed line (now representing the combined effect of neuronal decay and external input) rightwards, with the activity stabilizing at the red dot. The rate at which the firing rate increases is proportional to the difference between the dashed line and rightward band edge B_u. In the memory period following the offset of the external input I_e (red leftward arrow), the memory activity stabilizes at the green dot. The power law decay with power $p = 1$ leads to the linear band shown in blue (Materials and Methods). **D**, Same as **C**, but for a power law weight decay with $p = 1.4$. This case leads to a multistable band with slightly curved B_u and B_d. **E**, Same as **C**, but with an example Gaussian weight function. For this case, when the external input barely exceeds the smallest up-threshold, the memory activity abruptly jumps to a high level, amplifying the external input. **F**, Weight function determines the relationship between input amplitude and the amplitude of stored memory activity. Traces show the mappings between input and memory amplitude for each weight function used in **C-E**. The blue trace represents linear encoding of inputs in memory (Materials and Methods).

For analytical tractability, we simplified the spiking autapse to a rate-based autapse. The dynamics of the firing rate r follows:

$$\tau \frac{dr}{dt} = -r + \sum_{i=1}^{N_d} D_i(r) + I_e, \tag{42}$$

where N_d is the number of dendrites, and I_e is the external input to the soma. This equation is reproduced from Eq. 14 of Materials and Methods. Fig. 3A shows each dendritic input-output relation $D_i(r)$. The total dendritic activity $\sum_{i=1}^{N_d} D_i(r)$ effectively captures the spiking model

results in Fig. 2B, with more activatable dendrites with larger weights at the bottom and piling up. See Materials and Methods for more details.

How does such an autapse maintain finely discretized values in memory? As shown in Fig. 3B, $\sum_{i=1}^{N_d} D_i(r)$ forms two stair-like lines B_u and B_d along its left and right ends, each consisting of a list of $r_{u,i}^{th}$ or $r_{d,i}^{th}$ values. The equilibrium memory states exist if the total dendritic activity equals the neural decay term, $\sum_{i=1}^{N_d} D_i(r) = r$ (green dots). In the continuum limit, B_u and B_d become smooth lines and the discretized stable values become a memory line, represented by the intersection of the multistable band with the dashed line in Fig. 3B. Notably, the existence of the memory line does not require fine-tuning of parameters, as it is easy to find the overlap between the multistable band and the unity slope line.

Different memory activities are achieved by recruiting more dendrites to the up-state. Starting with all dendrites in the down-state, an external input I_e moves the dashed line rightward by I_e, as shown in Fig. 3C. As long as I_e is large enough to cause dendritic activation, the activity stabilizes at the red dot during the encoding period. Turning off the external input reduces the firing rate, moving the stable point horizontally leftward by I_e. The new stable point (green) is still within the band, indicating that it gives a stable memory. 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, moving the green dot higher, more dendrites get recruited, which results in a larger memory amplitude. Further details are presented in Materials and Methods.

For simplicity, we assume that once a dendrite is activated, its up-state can be maintained without forgetting. That is, the stable point is always within the band. In this way, we do not need to worry about the B_d curve. However, as we discuss below, in the presence of noise the shape of B_d matters.

How the weight function affects memory performance in the rate-based autapse

Each external input value maps to a memory amplitude. This mapping is determined by the synaptic weight function. As shown in Fig. 3C, for an increase in I_e, the amount of additionally activated dendrites is determined by the B_u slope at the green dot. In general, the shape of B_u is determined by the synaptic weight function. In particular, with $w(x) = a/(z + b)$ for positive constants a and b, a linear mapping between input and memory amplitude is possible. An example of this case is plotted in Fig. 3, with the band shape in Fig. 3C and the mapping between input and memory amplitude in Fig. 3F, with the blue line.

More generally, the existence of stable memory states does not need fine-tuning in the weight functions. As shown in Fig. 3C-E, memory states of amplitudes exist for different weight functions.

For a different w(x), the mapping can be a gradually saturating curve (Fig. 3F, magenta curve). This gives a relatively higher resolution in memory amplitude for smaller input values, which 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 by changing w(x). See more details in Materials and Methods.

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 multistable band shapes as shown in Fig. 3E. Once the small signal activates the dendrite with the largest weight, the feedback excitation can activate these dendrites, leading to a big jump in activity, as shown by the purple line in Fig. 3F. By adjusting the weight function, the model can control the trade-off between amplifying a small signal [33] or faithfully representing the small signal.

Amplitude increasing during the encoding period

During the encoding period, the speed at which the firing rate increases is proportional to the vertical difference between line B_u and $r - I_e$. It is based on Eq. 42 and illustrated in Fig. 3C by the double arrow. If the right dashed line in Fig. 3C is parallel to B_u, the system will never stabilize, and the activity will keep increasing linearly over time, making it a perfect integrator for evidence accumulation [34, 35, 36]. Under this condition, the performance is similar to that of other integrators based on cellular or dendritic bistability [8, 21, 22], despite a different network connectivity.

However, such an integration during the encoding period is rarely observed in working memory tasks [37, 38], except for some transient integration within 100 ms [39]. Therefore, for this paper, we focus on the stabilized dynamics for memory formation.

Working memory in the rate-based ring model with bistable dendrites

Here we implemented dendritic bistability into the classical ring network to memorize both the location and amplitude of a local input. We take the firing rate of the peak neuron, which receives the largest local input value, as the memory amplitude. As shown in Fig. 4A, our model has $N = 360$ neurons arranged in a periodic ring, with rotationally invariant, local excitatory weights and uniform recurrent inhibition, as similar to classic ring models [2, 3, 4]. The firing rate r_i, as reproduced from Eq. 23, follows:

$$\tau \frac{dr_i}{dt} = -r_i + \sum_{j=1}^{N} D_{ij}(r_j) - w_{inh} \sum_{j=1}^{N} r_j + I_{e,i}, \tag{43}$$

where $D_{ij}(r_j)$ is the bistable dendritic input-output relation, w_{inh} is the strength factor for global inhibition, and $I_{e,i}$ is the external input. Similar to the rate-based autapse model, each dendrite is bistable with a relation $D_{ij}(r_j)$ as examples shown in Fig. 4A, right. See more details in Materials and Methods.

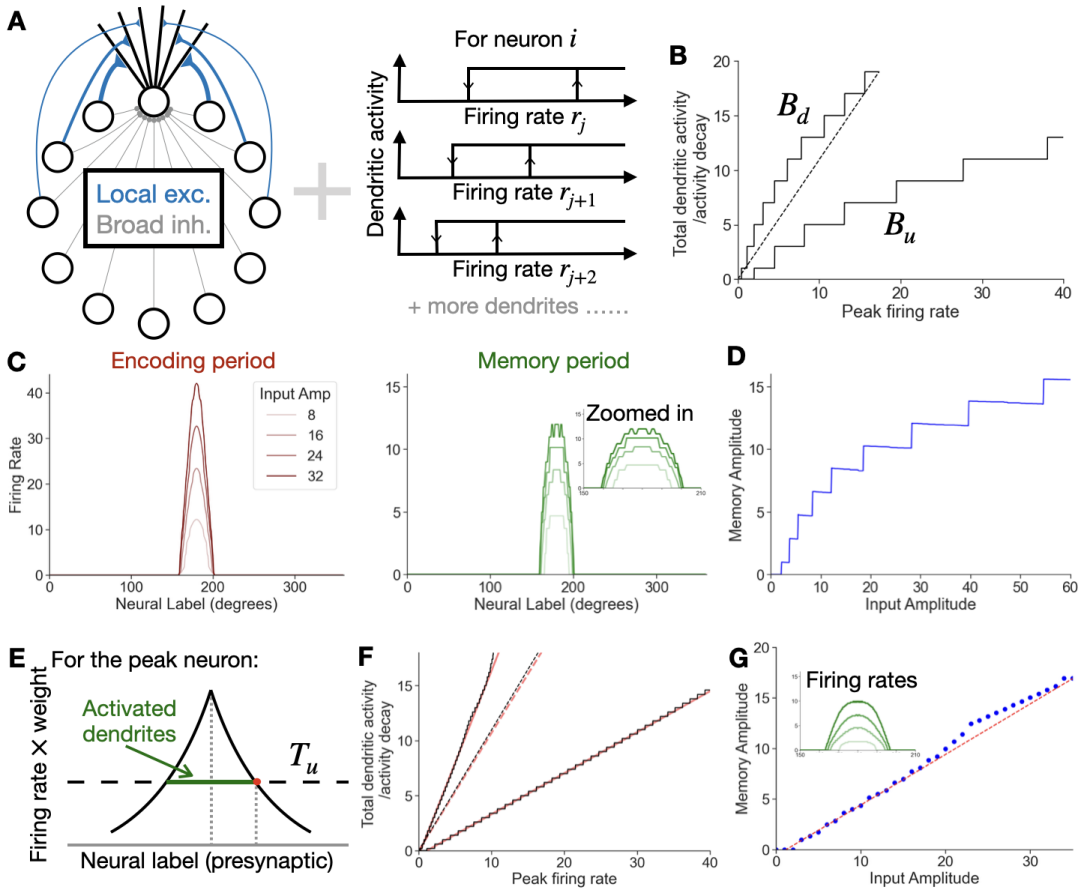

Figure 4: Simultaneous storage of input location and amplitude in the ring model with dendritic bistability. **A**, Network architecture. Left, each neuron receives local excitatory input through connections onto separate dendrites and broad (uniform) inhibitory input through connections onto the soma. Right, each dendrite is bistable. Dendrites receiving inputs through larger synaptic weights (bottom) have lower thresholds for presynaptic input to flip the dendrite to its up state. Note that, unlike the autapse case, the presynaptic input to each dendrite is not the same. **B**, Approximate multistable band. Similar to the rate-based autapse case, the model forms a multistable band defined by the activations of its multiple bistable dendrites. The shown band is simulated for the neuron at the peak of the activity bump, but approximately represents the multistability condition and thresholds for changes in activity for the entire network. Note the decay line now is slightly steeper than 45 degrees, in contrast to the autapse case, which accounts for recurrent inhibition in the ring model. See Materials and Methods for details. **C**, Storage of input amplitude at a given input location. Left, Input amplitudes presented during the encoding period for an input with location centered at 180 degrees. Right, Stable pattern of persistent activity during the memory period. Each input amplitude is faithfully encoded during the memory period. Inset shows a zoomed-in version of the neuronal activities during the memory period. **D**, Mapping between input and memory amplitude. Instead of just four amplitudes in **C**, we record memory amplitude as a function of input amplitude. A jump in memory amplitude happens every time a new dendrite is activated to the up-state. The stair shaped memory amplitude has a slight decrease in each step due to recurrent inhibition as explained in text. **E**, Schematic illustrating how different presynaptic inputs contribute to a postsynaptic neuron's level of activity during the encoding period. The effective input a postsynaptic neuron receives at each dendrite is given by the product of the presynaptic neuron's firing rate and synaptic weight onto the dendrite. If this product exceeds the up-threshold value T_u (dashed line), the dendrite is activated. The total number of activated dendrites of a neuron equals the length of the green segment. The example shown is for the peak neuron at the center of a local input pattern. For other neurons, the effective input curve (black line) will not be symmetric. **F**, Multistable bands for the network chose a weight function calculated for a linear relation between input and memory amplitude. Black curves show the simulated results in a near-continuum ring model. Shaded red curves show the targeted band shape to obtain a linear relation. **G**, Memory performance approaching a linear relation. The simulated mapping (blue dots) is close to the desired linear mapping (red dashed line), using the same calculated weight as in **F**. The inset shows a zoomed-in version of the memory activity. See Materials and Methods for details.

Similar to the rate-based autapse case, this network memorizes amplitude by activating den-

drites. The main difference from the rate-based autapse case is that each dendrite depends upon a different presynaptic firing rate for activation, instead of self-excitation. Despite this difference, it is still possible to numerically generate a multistable band for the peak neuron, which receives the largest local input value, as shown in Fig. 4B (Materials and Methods). Extra recurrent inhibition in the ring model makes the decay line effectively steeper than 45 degrees (dashed line), compared to the autapse case. This indicates that the network model is qualitatively analogous to the rate-based autapse model for dynamics, and, in particular, a line of attractors for amplitudes can be achieved without fine-tuning. Starting with all dendrites in the down-state, different input amplitudes can be maintained by different memory amplitudes, as shown in Fig. 4C.

The mapping between input and memory amplitude is determined by the weight function similar to the autapse case. An example is shown in Fig. 4D, where memory amplitude gradually saturates as input amplitude increases. Note that it is not monotonic due to discreteness in dendrites. It is possible that an input amplitude increase is not enough to make a memory amplitude increase, but inducing a wider memory. A wider memory gives more recurrent inhibition, which makes its amplitude drop.

Mechanistically, dendritic activation is illustrated in Fig. 4E in the continuum limit. For a neuron, the effective input it receives through its dendrites is given by the product of firing rate and weight function (black curve). A dendrite is activated if its effective input exceeds a fixed up-threshold, T_u. The total number of activated dendrites are given by the length of the green segment. Fig. 4E shows an illustration for the peak neuron, which receives the largest input value.

In the continuum limit, a flexible mapping can be achieved approximately by a corresponding weight function. Fig. 4F, G shows such an example for a linear relation, with a weight function following Eq. 41. Fig. 4F shows the simulated multistable band (black), which is

similar to the desired band (shaded red). Blue dots in Fig. 4G show the linear mapping between input and memory amplitude. This agrees well with the theoretical prediction, represented by the red dashed line, except in low input amplitudes, where its induced activity is too small to be sustained for memory. See more details in Materials and Methods

There exists a trade-off between the amplitude and the minimal width of a local memory. As shown in insets of Fig. 4C, G, 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 further requires nearby neurons having large firing rates to support these dendrites, inevitably making the bump wider. While we focus on the memorization of a unimodal, Gaussian input, in the case of holding multiple items in memory, a wider bump is more likely to interfere with other memory bumps.

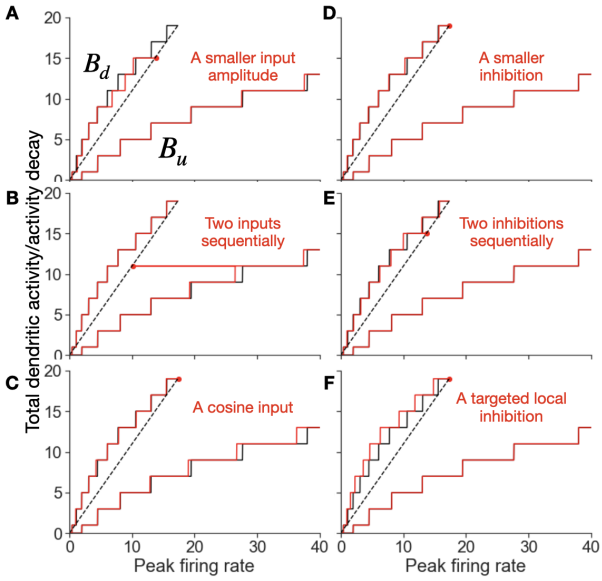

Figure 5: Similar multistable bands under different conditions. The multistable band used in Fig. 4B is an approximation. To demonstrate this approximation represents a typical multistable band in general, we show multistable bands in various conditions in red lines. The black lines are reproduced from Fig. 4B for comparison, which is generated with a local Guassian input with an amplitude 80, followed by a constant uniform inhibition of -10 in the memory period. **A-C**, Input changes. **A**, Making the input amplitude smaller, 40. **B**, Applying input in two steps. First, a smaller input amplitude is applied, resulting in a small memory amplitude (red dot). Then, a larger input amplitude is applied, resulting in a large memory amplitude (black dot). **C**, Making the input a cosine form instead of a Gaussian. **D-F**, Inhibition changes. **D**, Making the uniform inhibitory smaller. **E**, Applying inhibition in two steps. First, a smaller uniform inhibition is applied, reducing the memory amplitude from the black dot to the red dot. Then, a larger uniform inhibition is applied. **F**, Making the inhibitory connectivity profile localGaussian instead of uniform. See Materials and Methods.

Note that the multistable band generated in Fig. 4B is an approximation, such that the actual network dynamics, or dendritic activation, do not exactly follow the band in general. The actual dendritic activation of the peak neuron depends on the product of local weight function and firing rate pattern. That is to say, if the firing rate pattern changes, the dendritic activation will also change, leading to a different multistable band shape. The change in firing rate pattern can be caused by simply applying an input with a smaller amplitude, which leads to different recurrent activity. Another example is to replace a Gaussian input to a cosine one of the same amplitude and similar width. In Fig. 5, we generated multistable bands under six different conditions in red, with the reference reproduced in black. The overall band shape is not changing much for each condition. For conditions which give different stable points (Fig. 5B, E, red dots) during the memory period, those points align well with the dashed line, with only small deviations. This indicates that Fig. 4B is a good representation of the general multistable band for describing dendritic activation similar to the rate-based autapse model. See Material and Methods for more details.

Memory robustness under somatic noise

In general, our model is robust under the presence of somatic noise. This robustness is caused by several factors. A classical one is recurrent inhibition. Noise can trigger more activity, the increase of which can cause recurrent excitation. This gives a positive feedback loop that can blow up activity. However, recurrent inhibition introduces negative feedback, by providing more inhibition when activity goes larger. This counteracts the excess excitation and restores stability.

On the other hand, dendritic bistability adds an additional stability mechanism to memory activity. When the noise level is low, a noisy firing rate a dendrite receives does not alter dendritic state/output. That is, the noise can not drive the down-state to the up-state or vice

versa. Therefore, noise does not induce any recurrent changes in dendrites. When the noise level is high, an excitatory perturbation may activate a dendrite to the up-state and induce extra excitation. Does it lead to a positive feedback as the classical case above? Not necessarily. The induced extra excitation into a dendrite may not change dendritic state/output, if the excitation is insufficient to activate that dendrite, or if that dendrite is already in the up-state. In either case, there is zero positive feedback.

Classically, recurrent inhibition is required to maintain a stable local memory [3, 40]. However, due to the presence of dendritic bistability in our model, a stable local memory can be maintained even without any recurrent inhibition. Such simulations (not shown) reproduces all results in this paper qualitatively.

In the following, we present simulations to explore the robustness of memory amplitude and location against somatic noise.

Robustness of memory amplitude under noise

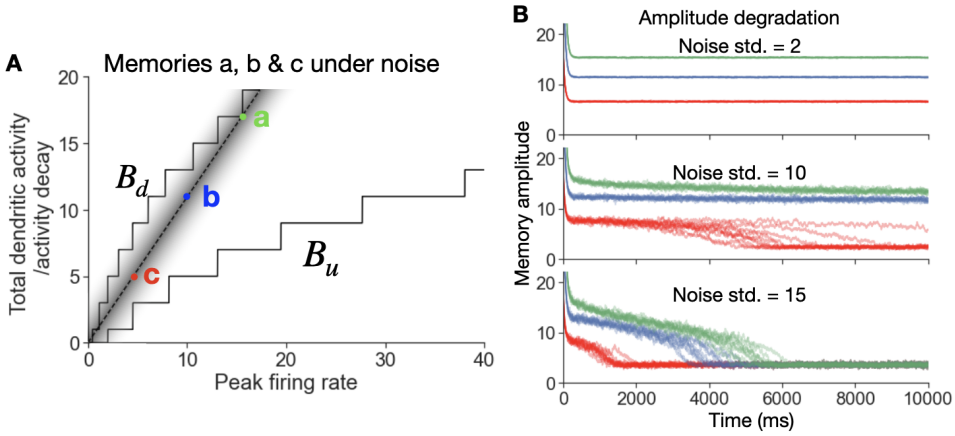

Figure 6: Robustness of memory amplitude under noise and amplitude drift. **A**, Robustness of memory amplitude. Three illustrative memory amplitudes (a, b, c) are represented by three stable points in the multistable band. Noise can drive stable points left or right horizontally (gray bar). An amplitude is stable if noise does not drive the stable point across the band (point 'b'), indicating no change in any dendritic state. A memory will degrade when noise drives the stable point across B_d (point 'a' or 'c'), causing some up-state dendrites to fall back to the down-state. This degradation depends on the noise level and geometry of the multistable band. **B**, Examples of amplitude drift. The idea in **A** is further demonstrated with simulation results with three input amplitudes. Under small noise, all three memories remain stable with different amplitudes, with ten trials for each case (top). Under moderate noise, the memories with a higher amplitude (green) go lower, while the memories with a lower amplitude (red) get erased (middle). Under large noise, all three memories are erased (bottom).

For memory amplitude which sits in a multistable band as shown in Fig. 4B, it can be stable under noise, as long as noise fluctuation is limited within the band. Otherwise, amplitude degradation happens depending on the geometry of the multistable band. As illustrated in Fig. 6A, three memories stabilized in points a, b, and c within a multistable band. Under a moderate level of noise, the point 'b' can move back and forth horizontally (gray shaded) without crossing the multistable band. Therefore, the fluctuation does not change any dendritic state and has zero averaged effect. In contrast, the same noise is enough to drive the point 'a' across the line B_d, deactivating some dendrites and leading to a reduction of memory amplitude. With a slightly lower amplitude of 'a', the distance between the stable point and B_d is wider, which can prevent further reduction. We define such restabilization of memory at a lower amplitude as the 'reduction mode'. On the other hand, when noise deactivates point 'c' by crossing B_d, a lower 'c' makes the distance between the stable point and B_d narrower. It makes 'c' more susceptible to deactivation and leads to further amplitude drop till it is completely erased. We define this way of degradation as 'erasure mode'. Note that, for the shown example, it is less likely that the noise drives the stable point across B_u, which requires the noise fluctuation to be greater than the input amplitude applied during the encoding period.

Simulations show the performance of memory amplitudes under somatic noise. In Fig. 6B, memory traces are plotted for three different input amplitudes with 10 trials each. When the standard deviation of the noise is too weak to drive any stable point across the multistable band, amplitudes are stable (top). A larger noise causes the higher memories, green lines, reducing to a lower level and the lower memories, red lines, is erased (middle). The performance is as illustrated in Fig. 6A. Further increasing the noise erases all memories (bottom). Overall, memory degradation depends on the shape of the multistable band, the location of the memory stable point, and the noise level. See the Discussion for possible interpretations of these modes. See Materials and Methods for more simulation detail.

Robustness of memory location under noise

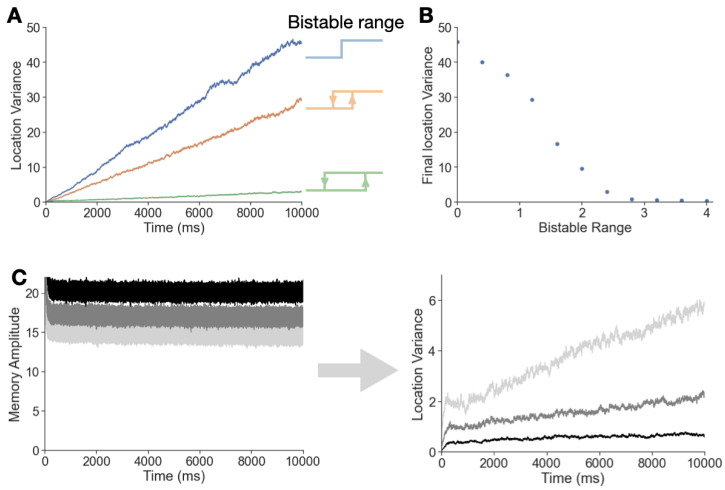

Figure 7: Robustness of memory location under noise and location diffusion. **A**, Dendritic bistability reduces location diffusion. The memorized location may diffuse under noise. We record the variance of memory location over time for 400 trials. If each dendrite is a step function, $T_u = T_d$, the diffusion is severe with a large location variance over time (blue). If the dendritic bistable range, $T_u - T_d$, is larger, the diffusion is reduced (orange or green). **B**, The reduction of location diffusion. Location variance at the last time step of **A** is recorded for each bistable range, from 0 to 4. As bistable range increases, the location diffusion decreases and eventually becomes negligible, indicating noise can hardly change any dendritic state. **C**, Location diffusion for different amplitudes. Left, three memory amplitudes are maintained, with each having 400 trials plotted. This shows that no amplitude change happens during the diffusion of memory location. Right, for a larger memory amplitude, location is more stable. See Materials and Methods for simulation details.

Under noise, memory location can undergo diffusion. This makes the memorized location less and less accurate as the time goes. First, we show how this process is affected by bistable range $T_u - T_d$. The blue curve in Fig. 7A shows the increase of location variance in time across 400 trials if each dendrite follows a step input-output relation, $T_u = T_d$. As the bistable range increases from zero by increasing T_u and reducing T_d, the diffusion of approximately the same memory activity is reduced (orange and green curves). Interestingly, the variance of location still grows approximately linearly with time, indicating a diffusion process as classical models [19, 41]. Fig. 7B plots the variance of final location for each bistable range. When the range is small, noise is large enough to ignore the gap and still leads to large diffusion. When the range gets larger, it is harder for noise to flip on or off a dendritic state, thus memory activity is more stable. A similar stability has been demonstrated previously with a model having bistable

neurons [23].

On the other hand, how location diffusion is affected by different memory amplitudes? We compared the location diffusion for three memory amplitudes, with each having 400 trials. Fig. 7C, left, records memory amplitudes under noise for each trial. It shows that different amplitudes are still maintained despite noise. Fig. 7C, right, shows that the location variance for a larger amplitude is smaller. It is because a memory with a higher amplitude has overall larger activity, making the effect of noise relatively smaller and exerting more recurrent inhibition to reduce the diffusion. See the Discussion for a potential relation of this observation to the precision of the input. See Materials and Methods for more simulation details

Discussion

Memorizing input amplitudes in a given location is a challenging problem for neural circuit models. Here we propose a network model based on dendritic bistability. As long as there are enough bistable dendrites in each neuron, the model can approximately maintain a continuous amplitude by recruiting different dendrites. The dendrites form a multistable band requiring no fine-tuning in parameters to achieve a stable memory. The amplitude resolution is proportional to the total number of dendrites. The weight function controls the shape of the multistable band and thus the memory performance. Furthermore, the memory is robust under small noise and shows different amplitude degradation modes for larger noise.

While in this paper we focus on unimodal input, it is possible to extend to an input with multiple local peaks. The same dynamic equation, Eq. 43, still works in this case. Two peaks may have two wide widths, recall the trade-off between amplitude and width, such that if they are nearby, their widths overlap and lead to interference. On the other hand, if the network is no longer a ring structure but consists of a pool of separate autapses as in Fig. 3, each autapse can memorize a separate input value, which can together memorize a graded activity pattern.

Biological plausibility and alternative implementations

Dendritic bistability is the building block of our model. One way to realize it is by widely distributed NMDA receptors [30]. A subsequent work [31] further demonstrated that with the help of the KIR current, dendritic bistability can be achieved for a wide range of cellular parameters. Consistent with this, NMDA receptors have been demonstrated to be critical to working memory function [42, 43, 44]. However, it is still unclear whether this is simply due to their slow kinetics, which makes memory activity more stable [45], or it is also related to their ability to induce dendritic bistability. Physiological recordings have shown that NMDA receptors in dendrites can cause prolonged voltage plateaus lasting hundreds of milliseconds [46,47,48], but more direct evidence is needed to determine whether dendritic bistability can sustain itself for seconds in working memory tasks.

In addition to dynamics based on NMDA receptors, bistability can be achieved in alternative ways. They include cellular bistability/multistability mechanisms [23, 24, 25, 26, 27, 28, 29] or network-level bistability [8]. Cellular bistability [49], or multistability [50], have also been observed experimentally. It is possible that these mechanisms can be used to achieve dendritic (or synaptic, cellular, see below) bistability to realize this model. But more work is needed.

On the other hand, each dendrite is treated as a separate memory unit. This simplifies a lot of intracellular interactions existing in a neuron such as the back-propagation of action potentials from the soma to each dendrite. It is possible that such back-propagation is too weak to affect dendritic voltage [51]. Even if the back-propagation is strong, a previous work has demonstrated that each dendrite can still maintain a high level of independence [52].

In addition, the total number of bistable dendrites for each neuron limits the amplitude resolution for working memory. As suggested byprevious experiments [53, 54, 55], in simulations, each neuron recruits no more than 20 dendrites. However, it is possible to make this value much higher by alternative implementations of this model. One way is to make bistable units smaller.

For example, instead of dendrites, neurons may use spines as basic bistable units [56, 57]. Another way is to make bistable units larger. For example, instead of dendrites project to a soma as the current case, the implementation could be neurons, which receive recurrent inputs, project to a common neuron, which provides a recurrent output. Such a group of neurons may exist in a minicolumn, while multiple such groups can form a ring network like Fig. 4. In this way, those neurons receiving recurrent inputs need cellular bistability.

Sensitivity to external input and robustness under noise

Our model involves a trade-off between sensitivity and robustness. A wide bistable range in dendrites makes memory robust under noise. However, this comes at the cost of reduced sensitivity to external input, as a small input may not be able to overcome the thresholds, nor affect the memory. This insensitivity may not be a problem if the small actual stimulus increase induces a large input jump in I_e, which is enough to overcome the thresholds. However, at least in experiments studying the maintenance of head direction coding in Drosophila [58], such an input jump is not observed and a small head velocity input can change the maintained head direction memory.

One potential way to mitigate the trade-off may be making the dendritic input-output relation switch between bistable and sigmoid/linear. For a longer time scale, this might be done by changing the maximal conductances of different channels [31]. For a shorter time scale, this might be done by changing dendritic disinhibition [59], particularly through GABAb/KIR current as demonstrated previously [31].

Interpretations for memory amplitude and corresponding behaviors

Memory amplitude can be interpreted in two common ways: as the input intensity at a location or as the precision of the location. For the former, there are two quantities maintained

simultaneously, the location and intensity. Technically, two quantities should be independent. However, the model maintains them in a correlated way. That is, for a high intensity memory, its location is better memorized with less location diffusion through time, as shown in Fig. 7C. In addition, if intensity is inevitably degrading during the memory period, how should it behave? It is unlikely that intensity reduces from a high value to a lower one, because it means an erroneous lower intensity is maintained, causing a waste in metabolic energy. Thus, it is reasonable to expect a high intensity being erased directly, which aligns with the 'erasure mode' mentioned above. This further predicts that a higher intensity should be more robust against degrading due to a wider bistable range, as is the case of memory 'c' in Fig. 6A shows.

Bayesian models of sensory coding [19, 60, 61] have alternatively suggested that the amplitude of neural firing can represent the precision of a location, with higher firing rates indicating higher precision. However, the explicit connection is still controversial. When the maintained precision degrades during the memory period, two possibilities exist. First, the memory may still keep a fixed amplitude but diffuse around the original location, as suggested by some experiments [62, 63, 64]. Consistent with this idea, Fig. 7C shows how a higher amplitude leads to higher precision and less diffusion, without much amplitude reduction. Second, the memory may not diffuse much but its amplitude gradually reduces [65]. In this case, the "reduction mode", which allows gradual amplitude reduction, would be suitable to represent precision. This then predicts that higher precision is more susceptible to noise and more likely to reduce to lower precision.

ACKNOWLEDGMENTS

This work was supported by the Air Force Office of Scientific Research (AFOSR) under award number FA9550-22-1-0532.

References

[1] Mark S Goldman, Albert Compte, and Xiao-Jing Wang. Neural integrator models. *Encyclopedia of neuroscience*, pages 165–178, 2009.

[2] Christos Constantinidis and Xiao-Jing Wang. A neural circuit basis for spatial working memory. *The Neuroscientist*, 10(6):553–565, 2004.

[3] Shun-ichi Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological cybernetics*, 27(2):77–87, 1977.

[4] Albert Compte, Nicolas Brunel, Patricia S Goldman-Rakic, and Xiao-Jing Wang. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral cortex*, 10(9):910–923, 2000.

[5] Rani Ben-Yishai, R Lev Bar-Or, and Haim Sompolinsky. Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences*, 92(9):3844–3848, 1995.

[6] Sung Soo Kim, Hervé Rouault, Shaul Druckmann, and Vivek Jayaraman. Ring attractor dynamics in the drosophila central brain. *Science*, 356(6340):849–853, 2017.

[7] Xiao-Jing Wang. 50 years of mnemonic persistent activity: quo vadis? *Trends in Neurosciences*, 44(11):888–902, 2021.

[8] Alexei A Koulakov, Sridhar Raghavachari, Adam Kepecs, and John E Lisman. Model for a robust neural integrator. *Nature neuroscience*, 5(8):775–782, 2002.

[9] Carlos D Brody, Ranulfo Romo, and Adam Kepecs. Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. *Current opinion in neurobiology*, 13(2):204–211, 2003.

[10] Christos Constantinidis, Matthew N Franowicz, and Patricia S Goldman-Rakic. The sensory nature of mnemonic representation in the primate prefrontal cortex. *Nature neuroscience*, 4(3):311–316, 2001.

[11] Carlos D Brody, Adrián Hernández, Antonio Zainos, and Ranulfo Romo. Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cerebral cortex*, 13(11):1196–1207, 2003.

[12] Ashwin Vishwanathan, Alex Sood, Jingpeng Wu, Alexandro D Ramirez, Runzhe Yang, Nico Kemnitz, Dodam Ih, Nicholas Turner, Kisuk Lee, Ignacio Tartavull, et al. Predicting modular functions and neural coding of behavior from a synaptic wiring diagram. *Nature Neuroscience*, pages 1–12, 2024.

[13] Philip L Smith and Roger Ratcliff. Psychology and neurobiology of simple decisions. *Trends in neurosciences*, 27(3):161–168, 2004.

[14] Joshua I Gold and Michael N Shadlen. The neural basis of decision making. *Annu. Rev. Neurosci.*, 30(1):535–574, 2007.

[15] Steven J Luck and Edward K Vogel. Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in cognitive sciences*, 17(8):391–400, 2013.

[16] Sukbin Lim and Mark S Goldman. Balanced cortical microcircuitry for spatial working memory based on corrective feedback control. *Journal of Neuroscience*, 34(20):6790–6806, 2014.

[17] Sam Carroll, Krešimir Josić, and Zachary P Kilpatrick. Encoding certainty in bump attractors. *Journal of computational neuroscience*, 37(1):29–48, 2014.

[18] Weronika Wojtak, Stephen Coombes, Daniele Avitabile, Estela Bicho, and Wolfram Erlhagen. A dynamic neural field model of continuous input integration. *Biological cybernetics*, 115(5):451–471, 2021.

[19] Anna Kutschireiter, Melanie A Basnak, Rachel I Wilson, and Jan Drugowitsch. Bayesian inference in ring attractor networks. *Proceedings of the National Academy of Sciences*, 120(9):e2210622120, 2023.

[20] Kathleen P Champion, Olivia Gozel, Benjamin S Lankow, G Bard Ermentrout, and Mark S Goldman. An oscillatory mechanism for multi-level storage in short-term memory. *Communications biology*, 6(1):829, 2023.

[21] Mark S Goldman, Joseph H Levine, Guy Major, David W Tank, and H Sebastian Seung. Robust persistent neural activity in a model integrator with multiple hysteretic dendrites per neuron. *Cerebral cortex*, 13(11):1185–1195, 2003.

[22] Maxim Nikitchenko and Alexei Koulakov. Neural integrator: A sandpile model. *Neural computation*, 20(10):2379–2417, 2008.

[23] Marcelo Camperi and Xiao-Jing Wang. A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability. *Journal of computational neuroscience*, 5(4):383–405, 1998.

[24] Eve Marder, LF Abbott, Gina G Turrigiano, Zheng Liu, and Jorge Golowasch. Memory from the dynamics of intrinsic membrane currents. *Proceedings of the national academy of sciences*, 93(24):13481–13486, 1996.

[25] Eugene M Izhikevich. Simple model of spiking neurons. *IEEE Transactions on neural networks*, 14(6):1569–1572, 2003.

[26] Erik Fransén, Babak Tahvildari, Alexei V Egorov, Michael E Hasselmo, and Angel A Alonso. Mechanism of graded persistent cellular activity of entorhinal cortex layer v neurons. *Neuron*, 49(5):735–746, 2006.

[27] Michael E Hasselmo and Chantal E Stern. Mechanisms underlying working memory for novel information. *Trends in cognitive sciences*, 10(11):487–493, 2006.

[28] Guy Major and David Tank. Persistent neural activity: prevalence and mechanisms. *Current opinion in neurobiology*, 14(6):675–684, 2004.

[29] Joel Zylberberg and Ben W Strowbridge. Mechanisms of persistent activity in cortical circuits: possible neural substrates for working memory. *Annual review of neuroscience*, 40:603–627, 2017.

[30] John E Lisman, Jean-Marc Fellous, and Xiao-Jing Wang. A role for nmda-receptor channels in working memory. *Nature neuroscience*, 1(4):273–275, 1998.

[31] Honi Sanders, Michiel Berends, Guy Major, Mark S Goldman, and John E Lisman. Nmda and gabab (kir) conductances: the "perfect couple" for bistability. *Journal of Neuroscience*, 33(2):424–429, 2013.

[32] H Sebastian Seung, Daniel D Lee, Ben Y Reis, and David W Tank. The autapse: a simple illustration of short-term analog memory storage by tuned synaptic feedback. *Journal of computational neuroscience*, 9:171–185, 2000.

[33] Rani Ben-Yishai, R Lev Bar-Or, and Haim Sompolinsky. Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences*, 92(9):3844–3848, 1995.

[34] Emre Aksay, Georgi Gamkrelidze, H Sebastian Seung, Robert Baker, and David W Tank. In vivo intracellular recording and perturbation of persistent activity in a neural integrator. *Nature neuroscience*, 4(2):184–193, 2001.

[35] Emre Aksay, Guy Major, Mark S Goldman, Robert Baker, H Sebastian Seung, and David W Tank. History dependence of rate covariation between neurons during persistent activity in an oculomotor integrator. *Cerebral Cortex*, 13(11):1173–1184, 2003.

[36] Kayvon Daie, Mark S Goldman, and Emre RF Aksay. Spatial patterns of persistent neural activity vary with the behavioral context of short-term memory. *Neuron*, 85(4):847–860, 2015.

[37] Steven J Luck and Edward K Vogel. The capacity of visual working memory for features and conjunctions. *Nature*, 390(6657):279–281, 1997.

[38] Weiwei Zhang and Steven J Luck. Discrete fixed-resolution representations in visual working memory. *Nature*, 453(7192):233–235, 2008.

[39] Paul M Bays, Nikos Gorgoraptis, Natalie Wee, Louise Marshall, and Masud Husain. Temporal dynamics of encoding, storage, and reallocation of visual working memory. *Journal of vision*, 11(10):6–6, 2011.

[40] Bard Ermentrout. Neural networks as spatio-temporal pattern-forming systems. *Reports on progress in physics*, 61(4):353, 1998.

[41] Yoram Burak and Ila R Fiete. Fundamental limits on persistent activity in networks of noisy neurons. *Proceedings of the National Academy of Sciences*, 109(43):17645–17650, 2012.

[42] Min Wang, Yang Yang, Ching-Jung Wang, Nao J Gamo, Lu E Jin, James A Mazer, John H Morrison, Xiao-Jing Wang, and Amy FT Arnsten. Nmda receptors subserve persistent neuronal firing during working memory in dorsolateral prefrontal cortex. *Neuron*, 77(4):736–749, 2013.

[43] Bram van Vugt, Timo van Kerkoerle, Devavrat Vartak, and Pieter R Roelfsema. The contribution of ampa and nmda receptors to persistent firing in the dorsolateral prefrontal cortex in working memory. *Journal of Neuroscience*, 40(12):2458–2470, 2020.

[44] Shengtao Yang, Hyojung Seo, Min Wang, and Amy FT Arnsten. Nmdar neurotransmission needed for persistent neuronal firing: potential roles in mental disorders. *Frontiers in Psychiatry*, 12:337, 2021.

[45] Xiao-Jing Wang. Synaptic basis of cortical persistent activity: the importance of nmda receptors to working memory. *Journal of Neuroscience*, 19(21):9587–9603, 1999.

[46] Jackie Schiller, Guy Major, Helmut J Koester, and Yitzhak Schiller. Nmda spikes in basal dendrites of cortical pyramidal neurons. *Nature*, 404(6775):285–289, 2000.

[47] Dong-Sheng Wei, Yan-Ai Mei, Ashish Bagal, Joseph PY Kao, Scott M Thompson, and Cha-Min Tang. Compartmentalized and binary behavior of terminal dendrites in hippocampal pyramidal neurons. *Science*, 293(5538):2272–2275, 2001.

[48] Guy Major, Alon Polsky, Winfried Denk, Jackie Schiller, and David W Tank. Spatiotemporally graded nmda spike/plateau potentials in basal dendrites of neocortical pyramidal neurons. *Journal of neurophysiology*, 99(5):2584–2601, 2008.

[49] RH Lee and CJ Heckman. Bistability in spinal motoneurons in vivo: systematic variations in rhythmic firing patterns. *Journal of neurophysiology*, 80(2):572–582, 1998.

[50] Alexei V Egorov, Bassam N Hamam, Erik Fransén, Michael E Hasselmo, and Angel A Alonso. Graded persistent activity in entorhinal cortex neurons. *Nature*, 420(6912):173–178, 2002.

[51] Nace L Golding, Nathan P Staff, and Nelson Spruston. Dendritic spikes as a mechanism for cooperative long-term potentiation. *Nature*, 418(6895):326–331, 2002.

[52] Bardia F Behabadi and Bartlett W Mel. Mechanisms underlying subunit independence in pyramidal neuron dendrites. *Proceedings of the National Academy of Sciences*, 111(1):498–503, 2014.

[53] Attila Losonczy and Jeffrey C Magee. Integrative properties of radial oblique dendrites in hippocampal ca1 pyramidal neurons. *Neuron*, 50(2):291–307, 2006.

[54] Alon Polsky, Bartlett W Mel, and Jackie Schiller. Computational subunits in thin dendrites of pyramidal cells. *Nature neuroscience*, 7(6):621–627, 2004.

[55] Greg J Stuart and Nelson Spruston. Dendritic integration: 60 years of progress. *Nature neuroscience*, 18(12):1713–1721, 2015.

[56] Victor Hugo Cornejo, Netanel Ofer, and Rafael Yuste. Voltage compartmentalization in dendritic spines in vivo. *Science*, 375(6576):82–86, 2022.

[57] Roberto Araya. Dendritic morphology and function. *Neuroscience in the 21st Century: From Basic to Clinical*, pages 571–606, 2022.

[58] Angel M Pastor, Rosa R De la Cruz, and Robert Baker. Eye position and eye velocity integrators reside in separate brainstem nuclei. *Proceedings of the National Academy of Sciences*, 91(2):807–811, 1994.

[59] Panayiota Poirazi and Athanasia Papoutsi. Illuminating dendritic function with computational models. *Nature reviews neuroscience*, 21(6):303–321, 2020.

[60] H Sebastian Seung and Haim Sompolinsky. Simple models for reading neuronal population codes. *Proceedings of the national academy of sciences*, 90(22):10749–10753, 1993.

[61] Robert Wilson and Leif Finkel. A neural implementation of the kalman filter. *Advances in neural information processing systems*, 22, 2009.

[62] Sebastian Schneegans and Paul M Bays. Drift in neural population activity causes working memory to deteriorate over time. *Journal of Neuroscience*, 38(21):4859–4869, 2018.

[63] Klaus Wimmer, Duane Q Nykamp, Christos Constantinidis, and Albert Compte. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature neuroscience*, 17(3):431–439, 2014.

[64] Phui Cheng Lim, Emily J Ward, Timothy J Vickery, and Matthew R Johnson. Not-so-working memory: Drift in functional magnetic resonance imaging pattern representations during maintenance predicts errors in a visual working memory task. *Journal of Cognitive Neuroscience*, 31(10):1520–1534, 2019.

[65] Frans W Cornelissen and Mark W Greenlee. Visual memory for random block patterns defined by luminance and color contrast. *Vision research*, 40(3):287–299, 2000.