RESEARCH ARTICLE

Attractor neural networks with double well synapses

Yu Feng[1], **Nicolas Brunel**[1,2]*

1 Department of Physics, Duke University, Durham, North Carolina, United States of America, **2** Department of Neurobiology, Duke University, Durham, North Carolina, United States of America

* nicolas.brunel@duke.edu

Abstract

It is widely believed that memory storage depends on activity-dependent synaptic modifications. Classical studies of learning and memory in neural networks describe synaptic efficacy either as continuous or discrete. However, recent results suggest an intermediate scenario in which synaptic efficacy can be described by a continuous variable, but whose distribution is peaked around a small set of discrete values. Motivated by these results, we explored a model in which each synapse is described by a continuous variable that evolves in a potential with multiple minima. External inputs to the network can switch synapses from one potential well to another. Our analytical and numerical results show that this model can interpolate between models with discrete synapses which correspond to the deep potential limit, and models in which synapses evolve in a single quadratic potential. We find that the storage capacity of the network with double well synapses exhibits a power law dependence on the network size, rather than the logarithmic dependence observed in models with single well synapses. In addition, synapses with deeper potential wells lead to more robust information storage in the presence of noise. When memories are sparsely encoded, the scaling of the capacity with network size is similar to previously studied network models in the sparse coding limit.

Author summary

A long-lasting question in neuroscience is whether synaptic efficacies should be described as continuous variable or discrete variables. Recent experiments indicate that it is a combination of both: synaptic efficacy changes continuously, but its distribution peaks at several discrete values. In this study, we introduce a synapse model described by a double well potential, and investigate the memory properties of networks of neurons connected with such synapses. Our results show in networks with a bimodal weight distribution, the storage capacity depends on network size as a power law. In addition, we demonstrate that networks with such synapses store information more robustly in the presence of noise, compared to networks with synapses with a single well potential.

1 Introduction

Synapses can change their strength in response to neuronal activity [1–5], and synaptic changes are thought to be critical for the brain to build memories [6]. A popular theoretical framework for studying learning and memory is the attractor neural network scenario [7–9]. In this class of models, memories correspond to the stable fixed points of the neuronal activity dynamics. External stimuli can modify the synaptic connectivity of the network following certain plasticity rules, and these synaptic modifications can create new stable fixed points of neuronal dynamics. An external input is said to be stored in the network if there exists a stable fixed point that is highly correlated with it. An extensively studied question is the storage capacity of such networks [10–12], i.e. the number of stored memories, or the amount of information that can be stored. One key question in theoretical neuroscience is to find biologically plausible learning rules that lead to reasonably large storage capacities.

Multiple synaptic plasticity experiments have suggested that the strengths of individual synapses are modified in a discrete manner. A few experiments have shown that synaptic plasticity is well described by switches between two discrete states (all-or-none), rather than by arbitrary continuous changes in efficacy [13, 14]. In addition, super-resolution imaging experiments have shown that dendritic spines contain a small discrete (1-4) number of nanomodules (i.e., clusters of receptors) [15], consistent with plasticity experiments and indicating that synapses are quantized. This picture of discrete synapses is in contrast with most studies of learning and memory in neuronal network models, in which synapses are taken to be real continuous variables.

Multiple efforts have been made to compute the storage capacity of networks with discrete synapses. Sompolinsky studied a network with a specific binarized Hebbian rule [16], and showed its capacity is close to the capacity of the Hopfield network [17]. Sompolinsky's model assumes synapses can still store all continuous changes in synaptic efficacy during the learning phase, and synapses get clipped after they have learned all patterns. In contrast, Tsodyks [18], and Amit and Fusi [19, 20] introduced online learning rules in which synapses are always discrete. These studies found that discreteness of synapses can cause a large drop in storage capacity [18, 19], unless stored patterns become extremely sparse [20, 21]. Later studies introduced models with hidden states [22–25], but it remains unknown how these synaptic plasticity models perform in attractor neural networks.

While theoretical work has focused on either fully analog or fully discrete synapses, recent experimental results suggest an intermediate scenario in which the synaptic efficacy can be described by a continuous variable, but whose distribution is peaked around a small set of discrete values [26]. Motivated by this data, we introduce and explore a model in which synapses are described by a continuous variable that evolves in a potential with multiple minima ('potential wells', see Fig 1A for an example with two wells). Synapses remain in the same potential well when receiving weak or short-lasting stimuli, but can switch to a new state when receiving strong or long-lasting stimuli. Our analytical and numerical results show this model can connect models with discrete synapses which correspond to the deep potential limit [20], and models in which synapses evolve in a single quadratic potential [27]. We show that the storage capacity of the network with double well synapses (and thus a bimodal weight distribution) has a power law dependence on network size. In contrast, models with a single-well potential show a logarithmic dependence when the parameters of the potential are not allowed to vary with network size N, as shown by previous studies [28]. We also show that the networks with double well synapses store information more robustly in the presence of noise.
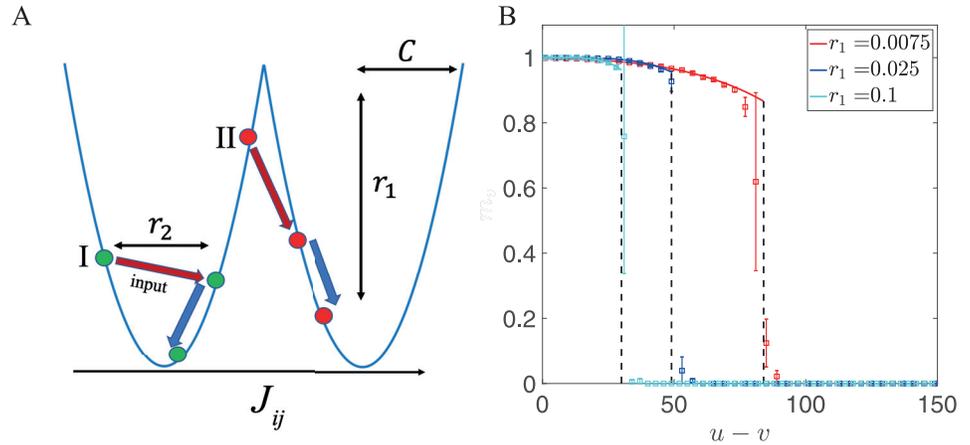
Fig 1. (A) Sketch of the synaptic model. In the presence of external input, a synapse can stay in the same well (I), or jump into another well (II), depending on its current state and the amplitude of the input. C describes the distance between the two wells, while r_1 characterizes its depth. The larger r_1, the faster synapses decay towards the minima of the potential. (B) Overlap between the input presented at time v and its corresponding attractor state, m_v, as a function of time elapsed since the presentation of this input $u - v$, for different values of r_1. The solid lines represent the theoretical prediction and squares represent the simulation results with a network of size $N = 30,000$ (mean and standard deviation computed over ten independent realizations). Dashed lines mark the storage capacity where m_v vanishes. For all lines, the value of C is chosen to optimize storage capacity. Other parameters are: $r_2 = 1, r_3 = 0, \theta = 0, f = 0.5, c = 0.05$.

https://doi.org/10.1371/journal.pcbi.1011354.g001

2 Model

We study a sparsely connected network of N binary neurons whose states $V_i = 0, 1$ ($i = 1, \ldots, N$) are given by a parallel update rule:

$$V_i(t + dt) = \Theta[h_i(t) - \theta] \tag{1}$$

where $dt \ll 1$ is the timescale of neuronal dynamics relative to synaptic dynamics, θ is a threshold, Θ is the Heaviside step function, and $h_i(t)$ is the local field of neuron i:

$$h_i(t) = \frac{1}{N} \sum_{j=1}^{N} c_{ij} J_{ij} V_j(t), \tag{2}$$

where J_{ij} is the strength of the synapse from neuron j to neuron i, and c_{ij}s are i.i.d Bernoulli random variables, with $c_{ij} = 1$ with probability $c \ll 1$, and 0 otherwise for all $i \neq j$, while $c_{ii} = 0$ for all i.

Synapses are assumed to be evolving on a much slower time scale than neuronal dynamics, and synaptic dynamics are supposed to be driven by external inputs that drive the network to specific states. Each synapse evolves according to

$$\frac{dJ_{ij}(t)}{dt} = -r_1 \frac{dU(J_{ij})}{dJ_{ij}} + \sum_{v \leq t} \left[r_2 I(\eta_i^v, \eta_j^v) + r_3 \epsilon_{ij}^v \right] \delta(t - v), \tag{3}$$

where U is a potential function with multiple minima, δ is the Dirac delta function, $\{\eta_i^v\}$ are i.i.d random binary variables that represent the state of the neuron imposed by external inputs at time $v \in \mathbb{Z}$, with $P(\eta_i^v = 1) = f, P(\eta_i^v = 0) = 1 - f$, and ϵ_{ij}^v are i.i.d random Gaussian variables describing the noise at synapse $j \to i$ at time v. Note that the chosen time unit is the interval

between presentations of two successive external inputs. The dynamics of J_{ij} is determined by three terms, the potential U, external inputs I and noise ϵ, of respective strengths r_1, r_2 and r_3.

The first term in the r.h.s. of Eq (3) determines the dynamics of the synapse in the absence of inputs. For simplicity, we use a double well potential U in which each well is given by a quadratic function,

$$U(J) = \begin{cases} (J - C)^2, & J > 0, \\ (J + C)^2, & J < 0. \end{cases} \qquad (4)$$

Thus, in the absence of inputs, each synapse decays to one of the two potential wells—a low efficacy state $J_{ij} = -C$, or a high efficacy state $J_{ij} = C$—exponentially, with a time constant $1/(2r_1)$. Small values of r_1 describe shallow potentials (and hence long decay times towards the minima) while large values of r_1 describe deep potentials (and hence fast decay times towards the minima).

The second term $I(\eta_i^\nu, \eta_j^\nu)$ in the r.h.s. of Eq (3) is activity-dependent. For binary neurons, the function I can take at most four values, one for each combination of pre and post-synaptic activity. We study two scenarios: In the **balanced input** case, the coding level is $f = 0.5$, and we use

$$I(\eta_i^\nu, \eta_j^\nu) = \begin{cases} 1, & \eta_i^\nu = \eta_j^\nu, \\ -1, & \eta_i^\nu \neq \eta_j^\nu. \end{cases} \qquad (5)$$

Note that for $U = 0$ (i.e. the flat potential case), and in the absence of noise, the model leads to a connectivity matrix J that is identical to the Hopfield model [17]. This model however suffers from a 'blackout' catastrophe after a critical number of inputs, all memories suddenly become irretrievable. In the case of a potential with a single well ($C = 0$) the model leads to a connectivity matrix that is identical to a 'palimpsest' model introduced by Mézard et al [27]. This model instead can retrieve the most recently presented memories, but forgets memories some time after they have been presented.

In the **sparse input** case, the coding level is $f \ll 1$. In this case it is no longer a good idea to use a balanced function I, since this choice would lead to the vast majority of synapses being in the high efficacy potential well. In this scenario, we use two different I functions. The first one is inspired by the Tsodyks-Feigelman model [29, 30]

$$I(\eta_i^\nu, \eta_j^\nu) = \begin{cases} (1 - f)(1 - f), & \eta_i^\nu = \eta_j^\nu = 1, \\ -f(1 - f), & \eta_i^\nu \neq \eta_j^\nu, \\ f^2, & \eta_i^\nu = \eta_j^\nu = 0. \end{cases} \qquad (6)$$

In the sparse input case, we also use another I function, motivated by the plasticity rule in [20]:

$$I(\eta_i^\nu, \eta_j^\nu) = \begin{cases} I_1, & \eta_i^\nu = \eta_j^\nu = 1, \\ I_2, & \eta_i^\nu \neq \eta_j^\nu, \\ 0, & \eta_i^\nu = \eta_j^\nu = 0. \end{cases} \qquad (7)$$

where $I_1 > 0$ sets the strength of potentiation when both pre and post-synaptic neurons are active, $I_2 < 0$ sets the strength of depression when one neuron is active while the other is inactive, and no changes occur when both neurons are inactive. Here, we focus on $I_1 = 2C$, $I_2 = -f(1 - f)$. The rationale is that when the distribution of synaptic strengths peaks around $\pm C$, a potentiation strength of $2C$ leads to a transition from a lower to a higher state with high

probability. When the connection is depressed, I_2 takes the same value as in Eq (6), leading to a small transition probability from a higher to a lower state. This scheme mirrors the learning rule described in [20] for binary synapses.

The third term in the r.h.s. of Eq (3) is the noise term caused by fluctuations. These fluctuations could be of multiple origins: Fluctuations of neural activity; or fluctuations in the state of the synapse, due to fluctuations in vesicle release, numbers of activated receptors at the post-synaptic side, and fluctuations in biochemical reactions involved in synaptic efficacy changes.

Once the input patterns $\{\eta_i^v\}$ are given, the learning dynamics (3) are determined by four non-negative parameters: C determines the distance between two minima of U; r_1 how strongly the potential U affects the dynamics (i.e. how fast synapses decay to the minima of the potential); r_2 is the input signal strength and r_3 is the noise strength. In the presence of inputs, a synapse can stay in its current state, or jump into another state depending on its position in the potential well. This dynamics is illustrated in Fig 1A.

With such synaptic plasticity dynamics and appropriate parameters, the network described by Eq (1) has multiple fixed points, that are close to the patterns imposed by the external inputs that were presented most recently. A pattern is said to be stored in memory if a network configuration close to the pattern is a stable fixed point of the dynamics. For this online synaptic plasticity rule Eq (3), we found that the most recently shown patterns $\eta^u, \eta^{u-1}, \ldots, \eta^{u-p}$ can be retrieved as stable fixed-point attractors at a given time $u \in \mathbb{Z}$, while older patterns cannot be retrieved. p is defined as the maximal number of patterns that are still retrievable or, equivalently, the maximal age at which patterns can still be retrieved. This scenario is similar to previously studied attractor neural network models with online learning and decay term in the synaptic updates [27, 31], or networks with binary synapses [18, 32, 33] in which memories are forgotten exponentially.

In order to calculate the storage capacity for the network defined by Eqs (1) and (3), we consider a network at a particular time u when pattern u has just been presented, and investigate whether a pattern shown in the past $v < u$ is still stored in the connectivity matrix. To determine whether there still exists an attractor state close to pattern v, we introduce the overlap between the pattern shown at time v and the network state at time u:

$$m_v = \frac{1}{Nf(1-f)} \sum_{i=1}^{N} (\eta_i^v - f) V_i^u, \tag{8}$$

where V_j^u represent the network state at time u. The overlap m_v measures the retrieval quality of the pattern η^v stored in memory (see e.g. [30, 34]). States with $m_v \sim 1$ mean that the network can correctly retrieve the stored memory. In practice, in numerical simulations, the network is initialized at $V_i = \eta_i^v$, and the dynamics is run until convergence to a fixed point. Simulations reveal that sometimes, the network converges to the most recently shown pattern instead of the pattern v. Thus, we need to consider also the overlap with the most recently shown pattern, m_u. In the balanced input case, we use analytical calculations to compute the mean-field equations for the overlap m_v, and whether solutions to these equations with $m_v > 0$ are stable with respect to perturbations in the direction of the most recently shown pattern [31]. Details of analytical calculations are described in Methods I and II. These calculations rely on approximating distributions of synaptic inputs, conditioned on η_i^v, η_i^u, by Gaussian distributions. In the sparse input case, we used numerical simulations only, since this approximation is not accurate in this case for the values of f and N used in this paper.

The two overlaps m_u and m_v can be obtained by calculating the distribution of the local field h_i at the time $t = u$. In the large N limit (i.e., $cN \gg 1$), the local field defined in Eq (2) follows a Gaussian distribution and its mean and variance are determined by the distribution of

J_{ij}. The time-dependent distribution of synaptic weights whose dynamics obey Eq (3) can be found numerically (see Methods I for details). Eventually, we obtain two self-consistency equations for m_v and m_u:

$$
\begin{aligned}
m_v &= \Phi\left(\frac{\mu^{10}}{\sqrt{2}d}\right) - \Phi\left(\frac{-\mu^{11}}{\sqrt{2}d}\right), \\
m_u &= \Phi\left(\frac{-\mu^{10}}{\sqrt{2}d}\right) - \Phi\left(\frac{-\mu^{11}}{\sqrt{2}d}\right),
\end{aligned}
\tag{9}
$$

where $\Phi(x) = \frac{1}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$, and μ^{11}, μ^{10}, d are functions of m_u, m_v defined in Eqs (32) and (33) (see Methods II for details). In practice, one can fix v and gradually increase u until there is no solution satisfying $m_v \neq 0$, or this solution becomes unstable when a small overlap m_u is present. The storage capacity p is then obtained at the maximum value of $u - v$ where a stable solution $m_v \neq 0$ exists.

3 Results

3.1 Learning and retrieval dynamics for balanced input in the absence of noise

We start by considering the balanced input case ($f = 0.5$), and zero noise ($r_3 = 0$). Without loss of generality, we can fix the value of $r_2 = 1$, since the dynamics are invariant with respect to the transformation

$$
(r_2, C, J_{ij}, r_1, r_3, \theta) \rightarrow (xr_2, xC, xJ_{ij}, r_1/x, xr_3, x\theta).
$$

We also set $\theta = 0$, since in this case the distribution of synaptic weights is symmetric around zero. For a given set of parameters, we use both mean field and numerical simulations to compute the overlap of a retrieved memory as a function of its age, and the storage capacity, defined as the maximum age at which memory can be retrieved. Fig 1B shows the overlap of a memory learned at time v, m_v as a function of age (time $u-v$ elapsed since the pattern was learned), for different potential depths r_1 (other parameter values are indicated in the caption). We can see that the overlap decreases with age until it drops abruptly. The figure shows that analytic results are in good agreement with simulations, using a network with 30,000 neurons. It also shows that increasing r_1 (i.e. the depth of the potential wells) decreases storage capacity. This is consistent with previous observations that clipping synaptic strength reduces capacity [35], and that online plasticity rules with discrete synapses typically have much lower capacity than rules with continuous synapses [18–20].

The storage capacity p thus strongly depends on the depth of potential wells as measured by r_1, but it also depends on the width of the potential wells C relative to the jumps caused by input patterns. To investigate the storage capacity of the network, we begin by examining the single potential case, where the width of the potential wells C is set to zero.

Single well potential. When $C = 0$, the potential defined in Eq (4) becomes a single parabolic function. In this case, each synapse decays to zero exponentially with a time constant $\tau = 1/(2r_1)$ in the absence of inputs, according to the dynamics defined in Eq (3). This scenario is similar to the previously investigated palimpsest model [27, 36] or the pure forgetting model of ref. [28]. The storage capacity p in this case depends on the time constant τ as [28]:

$$
p = \frac{\tau}{2}\log(\frac{\tau_0}{\tau}),
\tag{10}
$$

where $\tau_0 = 2A(f)N$, and where $A(f)$ is a function of the coding level and thus is a constant in the

balanced input case. When $\tau > \tau_0$, no patterns are retrievable, similar to the Hopfield model scenario when the number of patterns exceeds the critical value. When $\tau < \tau_0$, the model can always retrieve recent patterns and becomes a palimpsest model. By allowing the time constant τ in Eq (10) to scale linearly with the network size N, the storage capacity p can have a linear dependence with N, as investigated in previous studies [27, 36]. However, this scaling assumes an unrealistically large synaptic decay time for large networks. In this study, we consider the decay constant τ or potential depth r_1 to be a property of individual synapses, rather than being dependent on network size, and thus $r_1 \sim O(1)$ in the large N limit. In this scenario, the weight dynamics with a single well potential only leads to a storage capacity that increases logarithmically with N, as shown in Fig 2C.

In the following, we will demonstrate that the storage capacity of the palimpsest model with fixed τ can be substantially enhanced by introducing the double well potential in the synaptic dynamics.

Double well potential.   The capacity of the model with double well potential is influenced by the width of the potential wells C, and the model displays distinct behaviors depending on r_1. When r_1 is small, the storage capacity p is optimized for $C = 0$ (single well potential), but decreases only weakly with C, as shown in Fig 2A. In this case, the model is similar to the previously studied pure forgetting model [28], and the storage capacity only increases logarithmically with N, as demonstrated by the solid lines in Fig 2C.

However, when r_1 is larger than a critical value r_1^*, so that the potential significantly influences the dynamics of the weights in between presentations, the capacity is no longer maximized at $C = 0$. Rather, there is an optimal value C^* that maximizes capacity, as demonstrated in Fig 2B. When $C = 0$, the storage capacity increases logarithmically with the network size. However, the storage capacity at the optimal value of C increases much faster than the storage capacity with $C = 0$, as shown in Fig 2B. When C is increased beyond the optimal value, the storage capacity drops abruptly since jumps induced by inputs are too small to cross the boundary between the two wells. The optimal C depends on both r_1 and N. We find that the critical value of r_1 at which the optimal C becomes nonzero is $r_1^* \approx 0.0075$ as shown in Fig 3A. In a broad parameter region, the optimal value C is larger than one, which means that multiple pattern presentations are required to induce the switch between two stable states of synapses. Our results also indicate that smaller r_1 or larger N lead to a higher value of the optimal C when $r_1 > r_1^*$.

For each value of r_1 and N, we focused on the value of C that maximizes capacity. For the region $r_1 > r_1^*$, we found that the optimal storage capacity p with double well potential can be well described by a power law of the network size N

$$p \sim N^a, \tag{11}$$

as indicated in Fig 2D. We show the dependence of the exponent a on r_1 in Fig 3B, together with representative distributions of weights at a few values of r_1. This plot shows three qualitatively different regions:

Shallow potential region. In the flat potential limit (i.e., $r_1 < r_1^*$), p has a logarithmic dependence on N, and thus the exponent $a \sim 0$. The distribution of the weights in this region is unimodal. The model in this limit is equivalent to models with exponentially decaying memories [27, 36].

Intermediate region. When r_1 increases beyond r_1^*, the storage capacity of the network can be significantly enhanced by optimizing the potential width C. In this region, the network exhibits a bimodal weight distribution, and the exponent a is slightly larger than 0.5 (solid line
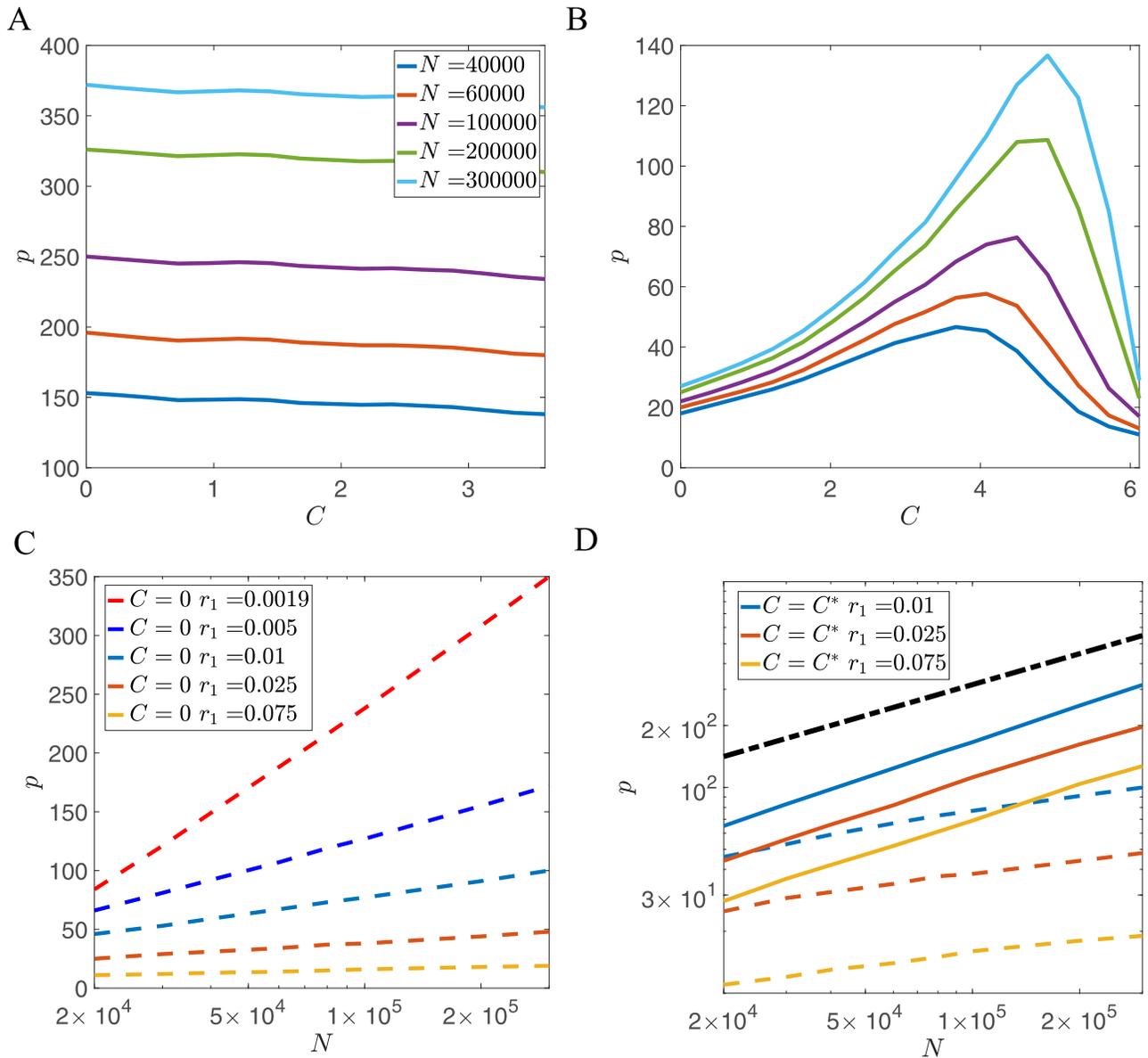
Fig 2. Dependence of storage capacity p on potential width C and network size N, for representative examples of the decay rate r_1. (A) Dependence of p on C for $r_1 = 2 \times 10^{-3} < r_1^*$. In the small r_1 limit, the optimal potential width C^* is zero (i.e., a single well potential), and p decreases weakly with C. The number of stored patterns increases logarithmically with network size, as demonstrated in (C). (B) Dependence of p on C for $r_1 = 0.05 > r_1^*$. The number of stored patterns p reaches its maximum at a nonzero value of C. The optimal storage capacity increases as a power law of N, as demonstrated in (D). (C) Storage capacity of the single-well potential model ($C = 0$) as a function of with network size N. The storage capacity decreases when r_1 increases as indicated by [Eq (10)](#). (D) Storage capacity of the double well potential model ($C = C^*$, full lines) as a function of network size N. When $r_1 > r_1^*$, the optimal storage capacity is much larger than the storage capacity of the single well potential model (dashed lines with the same color). The dashed-dotted black line represents $p = \sqrt{N}$ for reference. Other parameters in this figure are $r_2 = 1$, $r_3 = 0$, $\theta = 0$, $f = 0.5$, $c = 0.05$.

in [Fig 3B](#)), while the network with unimodal weight distribution ($C = 0$) still gives $a = 0$ (dashed line in [Fig 3B](#)).

Deep potential region. When r_1 further increases, the weights are more and more attracted to the minima of the potential well in between pattern presentations, and the model becomes more and more similar to a binary synapse model. In the large r_1 limit, the exponent a tends
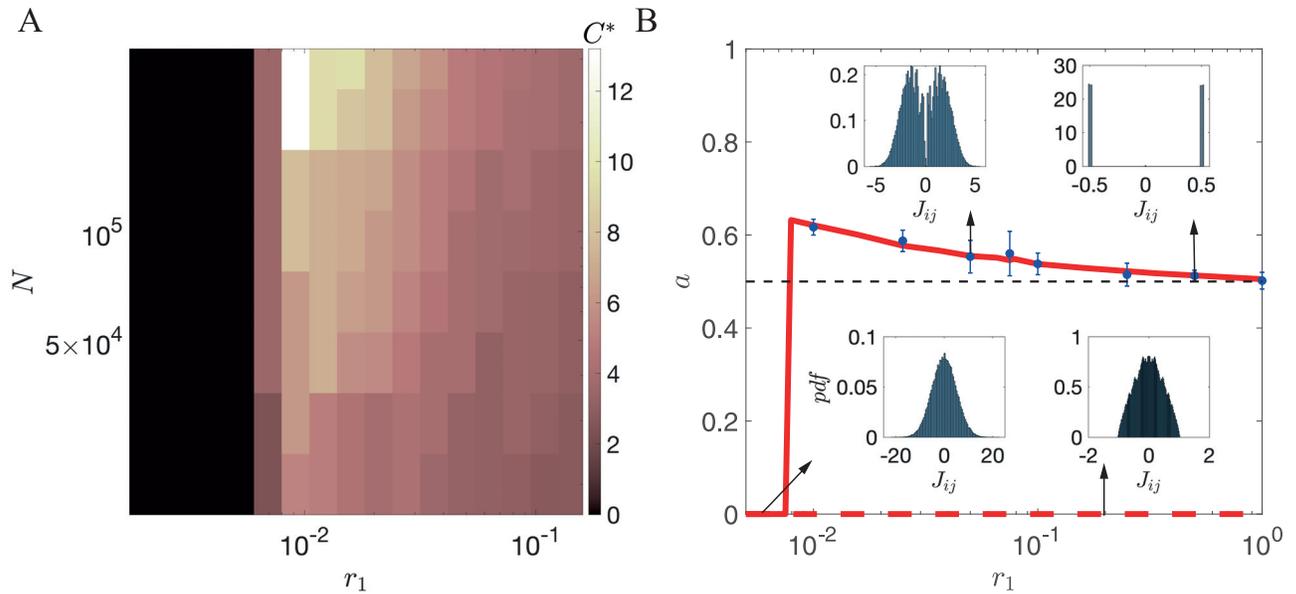
Fig 3. (A) Dependence of optimal potential width C^* on network size N and potential depth r_1. There is a critical value r_1^* that separates two regimes: When $r_1 < r_1^*$, capacity is optimized for a single well potential, $C^* = 0$. When $r_1 > r_1^*$, C^* is non-zero, and increases as N increases and r_1 decreases, indicating more pattern presentations are required to induce the switch between two stable states of synapses. Our analytical and numerical computation suggest that $r_1^* \approx 0.0075$, as indicated by the boundary of the black region. (B) Scaling exponent between p and N as a function of r_1, with $r_3 = 0$. Solid red lines: theoretical prediction for optimal C. Dashed red line: Theoretical prediction for $C = 0$. Blue circles: Simulation results for optimal C (mean and standard deviation computed over ten independent realizations). For $r_1 < r_1^*$, $p \sim \log(N)$ gives $a = 0$. When $r_1 > r_1^*$, the model with $C = C^*$ has a power law dependence on N (red solid line), while the model with $C = 0$ gives $a = 0$ (dashed red line). The exponent a decreases with r_1 and converges to 0.5 in the large r_1 limit. Insets show distributions of synaptic weights for representative examples. The distribution of J_{ij}s goes from unimodal when $C = 0$, to bimodal for the optimal C and $r_1 > r_1^*$ to quasi-discrete when r_1 becomes large.

to 0.5, and the distribution of weights becomes close to the sum of two delta functions $\delta(J + C) + \delta(J - C)$. This scenario is similar to the model with stochastic binary synapses studied in [19, 20].

3.2 Comparison between double well synapses and binary Markovian synapses

We now turn to a comparison between our model and models with binary Markovian synapses. For a fair comparison between these two models, we binarize weights in the double well model,

$$w_{ij}(t) = \text{sign}(J_{ij}(t)). \tag{12}$$

$w_{ij}(t)$ is thus a binarized connectivity matrix whose entries take $+1$ or -1 values, depending on which well the current value of J_{ij} belongs to. The binary Markov model has two states represented by $+1$ and -1, and each weight has a probability of switching from -1 to $+1$ when it is potentiated (i.e., $I(\eta_i^\nu, \eta_j^\nu) = 1$), or from $+1$ to -1 when it is depressed (i.e., $I(\eta_i^\nu, \eta_j^\nu) = -1$). The transition probability between the two states is set to be the same as the transition probability between the wells in the double well model. Note that while the dynamics of J_{ij} in the double well model is Markovian, the dynamics of binarized weights is not, since transitions between states do not depend only on the current presented pattern and the current value of w_{ij}, but also on the current value of J_{ij} which is influenced by patterns shown in the past.

For the comparison between the two models, we use two measures that quantify how much information is stored in such models. In attractor neural networks, stored information is typically measured by the number of stored patterns p, as introduced in the previous section. Another common measure is the signal-to-noise ratio (SNR) [22, 24, 37] defined for a connectivity matrix w at some time t after a particular pattern has been presented as

$$\text{SNR}(t) = \frac{\sum_{ij} c_{ij} w_{ij}^{ideal} \cdot w_{ij}(t) - \sum_{ij} c_{ij} w_{ij}^{ideal} \cdot w_{ij}(\infty)}{\text{std}(\sum_{ij} c_{ij} w_{ij}^{ideal} \cdot w_{ij}(\infty))}. \tag{13}$$

The SNR describes the degree to which the connectivity matrix at time t is correlated with an ideal connectivity matrix, in which all synaptic weights are entirely determined by a particular pattern stored at some time in the past. For instance, for binary inputs and binary synapses, $w_{ij}^{ideal} = I(\eta_i^0, \eta_j^0)$ where $I(\eta_i^0, \eta_j^0)$ is given by Eq (5). In Eq (13) the standard deviation is calculated over all elements in the matrix while $w_{ij}(\infty)$ is the state of the connectivity matrix at $t \to \infty$. w_{ij}^{ideal} is the ideal state of the weight determined by the learning rule Eq (5). At time $t = 0$, w_{ij}^{ideal} is 1 if it experiences a potentiation event and is −1 if it experiences a depression event (i.e., in Eq (5)). Note that SNR ignores the neuronal dynamics and is usually considered an upper bound on how memories stored in synaptic connectivity can be read out using neural activity. In the binarized double well model, the SNR can be calculated using methods described in Methods III. The SNR and the upper bound of SNR for the Markov model are calculated using Eqs (41) and (43) [24].

Fig 4 shows comparisons between the two models. Fig 4A shows that the storage capacity, defined by the number of patterns that can be retrieved as fixed point attractors, is much larger in the model with double well synapses than in the model with binary synapses, with transition probabilities chosen to match the probabilities of switching potential well in the double well model. Fig 4B shows a comparison of the SNR of the double well model with that of a corresponding Markov model. The Markov model and the double well model have the same initial SNR when a pattern is first learned, because transition probabilities are identical in both models. However, the SNR of the Markov model initially decays slowly and then rapidly drops off after $t \sim 10$, while the SNR of the double well model does not decrease monotonically due to the analog depth provided by the double well synapses, which allows the synaptic state to have a richer dependence on neuronal activity history. Consequently, the SNR of the double well model can surpass the upper bound of SNR in the Markov model and decay more slowly than the SNR of the Markov model, as shown in Fig 4B.

To better understand the history-dependent nature of double well synapses, we show the changes in the distribution of potentiated connection $J_{ij}^t(\eta_i^{t=0}, \eta_j^{t=0})$ over time in Fig 4C. After a pattern is learned, the distribution of $J_{ij}^t(\eta_i^{t=0}, \eta_j^{t=0})$ becomes asymmetric, and subsequent uncorrelated patterns can cause a probability flux towards the positive side, as demonstrated in Fig 4C. This happens in particular for the parameters of Fig 4 at $t = 2$, which leads to an increased SNR from $t = 2$ to $t = 3$, as seen in Fig 4B. After a sufficient amount of time, the distribution eventually returns to a symmetric asymptotic distribution, resulting in a decay of the SNR towards zero.

3.3 Robustness of stored information to noise

Neural circuits are required to not only learn new information but also to retain old memories. Recent experiments suggest that a significant fraction of synaptic plasticity is random noise, independent of learning [38, 39]. How can neural networks maintain long-lasting memories in the presence of noise? One long-lasting hypothesis is that discrete-like synapses could
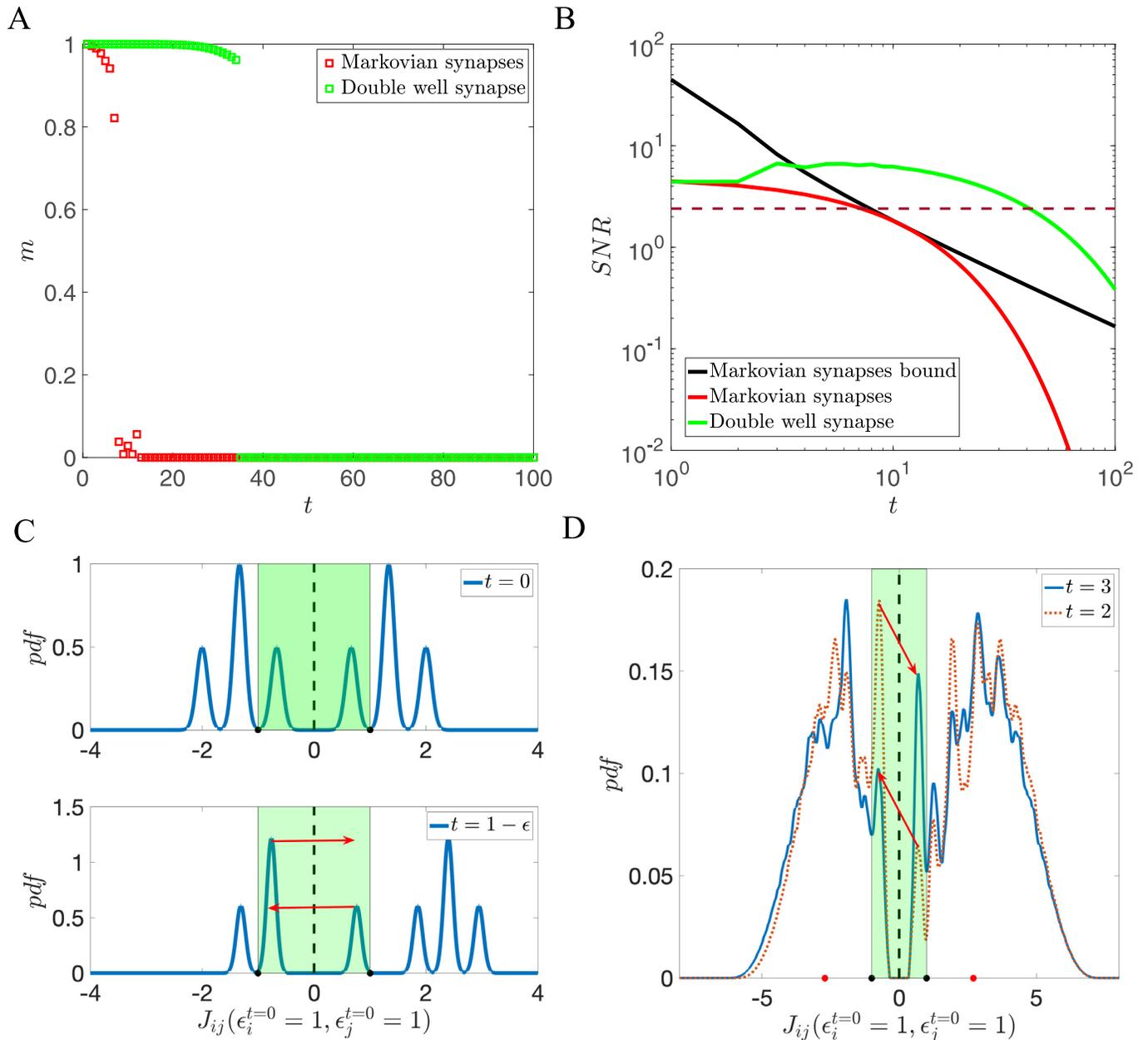
Fig 4. (A) Comparison between the storage capacity of Markovian binary synapses and binarized double well synapses defined in Eq (12). The parameters of the double well model are $r_1 = 0.1$, $r_2 = 1$, $r_3 = 0$, $\theta = 0$, $f = 0.5$, $cN = 2000$. The width of the potential C is chosen to maximize storage capacity. The transition probabilities of binary Markovian model are chosen to match probabilities of switching wells in the double well model. (B) Comparison between the SNRs of Markovian synapses and binarized double well synapses, as a function of time elapsed since the presentation of a given pattern. Green curve: SNR of the double well model with $C = C^*$ calculated using Eq (40). Red line: SNR of the Markov model with the same transition probability. Black line: Upper bound of SNR for two-state Markov models. Dashed line: critical SNR below which the memories are no longer retrievable [28]. The intersection between the SNR curve and this dashed line determines the storage capacity of the network. (C) Cartoon of the temporal evolution of the distribution of synapses that undergo potentiation at $t = 0$, i.e. those for which $\eta_i^{t=0}\eta_j^{t=0} = 1$ just before potentiation (top) and after (bottom). The highlighted region denotes the region around the maximum of the potential where synapses that reside in one of the wells can make a transition to the other well ($r_2 = 1$). Just before potentiation ($t = 0$), the distribution is symmetric. The distribution is then shifted up by r_2, before decaying towards the respective wells. As the distribution becomes asymmetric near 0 (within the range of $0 \pm r_2$), the next presented uncorrelated patterns will cause more probability mass to shift towards the positive side than towards the negative side, as demonstrated in the bottom figure. (D) Distribution of connections that have been potentiated at time $t = 0$, at later times $t = 2$ and $t = 3$. Black dots indicate $\pm r_2 = \pm 1$, while red dots indicate the minima of the potential, $\pm C = \pm 2.7$. As demonstrated in (C), an uncorrelated pattern leads to a larger probability of switching from the low well to the high well than the opposite transition, resulting in an increase in SNR from $t = 2$ to $t = 3$, as seen in (B). At longer times, the distribution eventually returns to a symmetric distribution, leading to a decrease in SNR.

benefit neural networks by increasing their robustness with respect to noise (see e.g. [40]). Here, we can verify this hypothesis in the model with double well synapses by adding noise to the model. This noise is described by the last term in Eq (3), where r_3 quantifies the magnitude of the noise term.

We first investigated how the potential width affects the network's storage capacity in the presence of noise. We set the potential depth to a fixed value of $r_1 = 0.1$ and calculated the storage capacity for every combination of r_3 and C. As illustrated in Fig 2B, there is an optimal value C^* that maximizes the storage capacity of the network without noise. If the potential width increases beyond C^*, the Hebbian learning term defined in Eq (5) is insufficient to induce the transition between different states, resulting in an abrupt decrease in storage capacity. For models with $C < C^*$, the network's storage capacity decreases monotonically with the noise strength r_3, as shown in Fig 5. However, for models with $C > C^*$ (e.g., $C = 4$, $C = 5$), there is a peak in the storage capacity as r_3 increases, indicating that noise facilitates the learning of new memories.

In summary, there are two potential ways in which noise can impact the double-well model. In instances where the potential width C is small, noise interferes with the information stored in the synapses, resulting in a reduction of storage capacity. When C is larger, the Hebbian learning term is insufficient by itself to induce transitions of sufficient numbers of synapses, leading to an inability of the network to learn new memories. In this scenario, introducing noise can help synapses make transitions between wells, therefore increasing the storage capacity of the network. This scenario is comparable to the model presented in [19, 20].

We also analyzed the effect of potential depth on the robustness of stored memories. For each fixed potential depth r_1, we can choose the potential width C that optimizes the storage capacity. With the optimal C, the storage capacity of the network always decreases when the
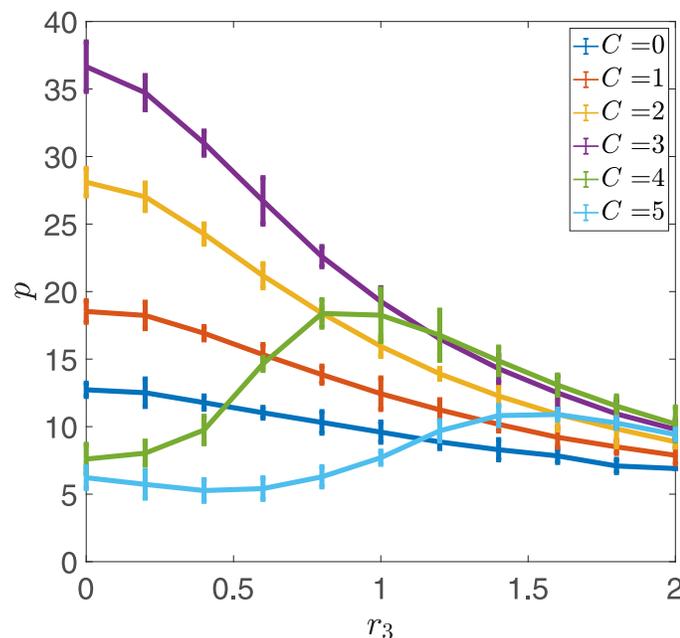
Fig 5. Dependence of Storage Capacity on Noise Strength r_3 for different values of Potential Width C. When C is small (e.g., $C = 0, 1, 2, 3$), the storage capacity p monotonically decreases with the noise strength r_3. However, for larger values of C (e.g., $C = 4, 5$), the storage capacity first increases with noise intensity, reaches a peak and then decreases. Thus, noise facilitates the learning of new memories in such cases. Other parameters are given as $N = 10,000$, $r_1 = 0.1$, $r_2 = 1$, $\theta = 0$, $f = 0.5$, $c = 0.05$.
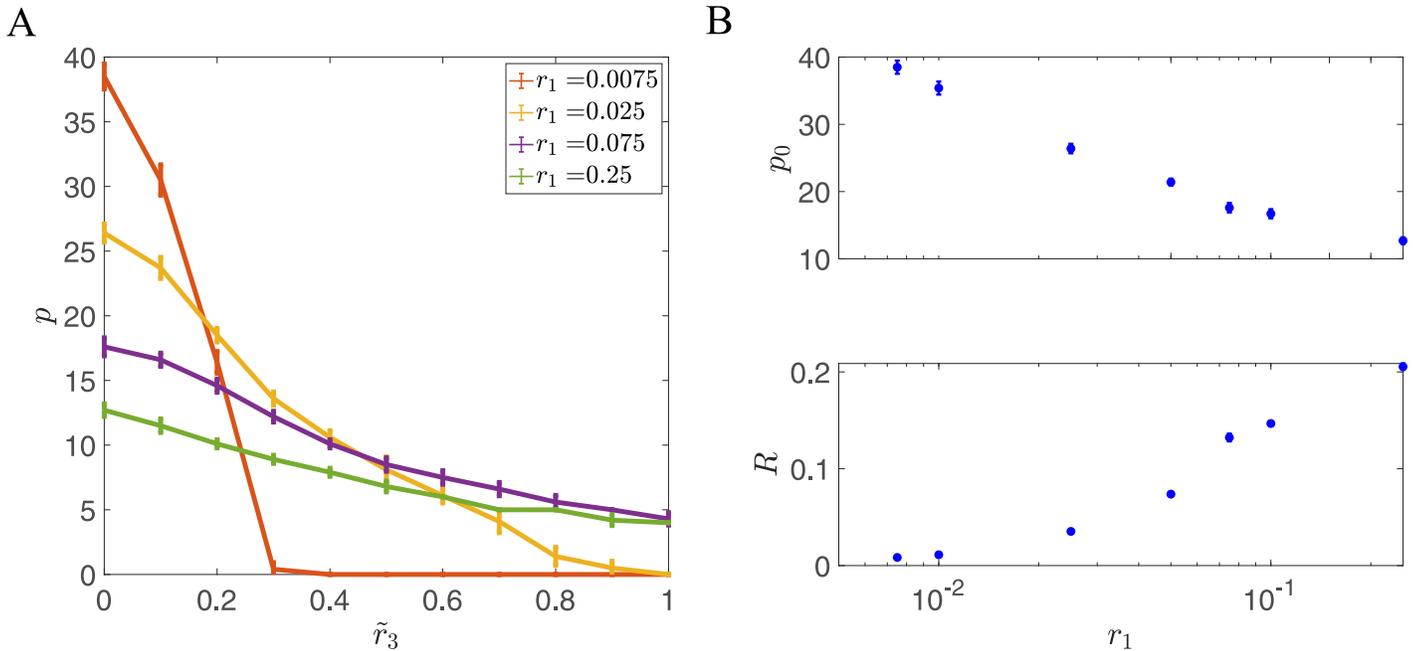
Fig 6. Trade-off between capacity and robustness when varying potential depth r_1. (A) The storage capacity decreases when the noise strength increases strongly depends on r_1. Small values of r_1 lead to large storage capacity, but storage is highly sensitive to noise. Large values of r_1 lead to small capacity, but storage is highly robust. (B) The trade-off between p_0 and R, where p_0 is the storage capacity in the absence of noise ($r_3 = 0$), and R quantifies robustness to noise, Eq (14). With deeper potentials, the storage capacity is smaller but storage is more robust. Here, C is optimized for each value of r_1 and other parameters are given as: $N = 10,000$, $r_2 = 1$, $\theta = 0$, $f = 0.5$, $c = 0.05$.

noise strength increases, as shown in Fig 6. However, the storage capacity decreases more slowly if the potential is deeper, which indicates the model is more robust to input noise. To quantify this effect, we introduced a robustness parameter as

$$R = \left| \frac{dp}{d\tilde{r}_3} \right|^{-1} \tag{14}$$

where $\tilde{r}_3 = r_3/M$ and M is the standard deviation of weight strength defined in Eq (22). $\tilde{r}_3$ measures the perturbation to the weight caused by the input noise. A larger R means the storage capacity decreases more slowly in the presence of input noise.

We explored how the robustness R depends on the depth of the potential r_1. As shown in Fig 6, when r_1 increases, the storage capacity of the network in the absence of noise decreases. However, the robustness R becomes larger, indicating a trade-off between storage capacity and robustness as potential wells become deeper. From a biological perspective, each minimum of the potential can correspond to a cluster of receptors [15]. The parameter r_1 may be related to the rate at which these clusters form. The timescale $\tau = 1/2r_1$ provides synapses with the analog depth to retain the history information of neuronal activity. Our results indicate that the balance between the memory storage capacity and the robustness of stored information would lead to an optimal r_1.

3.4 Learning dynamics in the sparse coding limit

We next turn to a more biologically realistic scenario in which neurons are described by 0,1 variables, and the probability that a neuron is active in a given pattern, f, is $f \ll 0.5$. We study

the storage capacity of the network with double well synapses with the learning rules described in Eqs (6) and (7) in the sparse coding limit $f \sim \ln N/N$ [41]. In this limit, it was shown that networks with binary synapses and online learning have a storage capacity that is comparable to the optimal capacity, unlike with other scalings of f with N where the capacity is sub-optimal [20, 41–43]. We first focus on the noiseless case ($r_3 = 0$). For each value of r_1, we optimize the storage capacity by adjusting the neuronal threshold θ and the potential width C. The relationship between the storage capacity p and the network size N is shown in Fig 7 for different choices of the learning rule and different potential depths r_1.

We start with the learning rule defined in Eq (6), which is similar to the Tsodyks-Feigel'-man learning rule [29, 30]. For this learning rule, we found that the optimal C is zero for all r_1, and the distribution of synaptic strengths is unimodal. As demonstrated in Fig 7, the storage capacity of the network increases logarithmically with the network size N, with a prefactor that is proportional to $\tau = 1/(2r_1)$ as indicated by Eq (10). This case is similar to the model with continuous synapses and exponentially decaying memories, as discussed in [27, 28, 31].

In contrast, the learning rule described in Eq (7), which is similar to the rule introduced by Amit and Fusi [20], exhibits a large storage capacity when the potential is deep, as illustrated in Fig 7B. In the deep potential regime, the weight distribution becomes strongly bimodal, with two peaks centered around $\pm C$. When there is a coincidence of pre and post-synaptic activity, the learning rule in Eq (7) allows the synaptic connection to cross the potential barrier with probability ~ 1. When one neuron is active while the other is silent, synapses switch to the low state only with low probability. This scenario is similar to the model discussed in [20]. As shown in Fig 7D, the storage capacity of this learning rule leads to $p \sim (N/\log(N))^2$ in the sparse coding limit.

Our result demonstrates that the double well model can achieve a much larger storage capacity in the sparse coding scenario than when memories are densely coded, similarly to previously studied models. In the flat potential case, the storage capacity scales as $p \sim \log(N)$, but with a prefactor that increases when coding becomes sparse. However, in the deep potential case, the capacity scales supralinearly with network size, $p \sim (N/\log(N))^2$, in the sparse coding limit, while it is sublinear when coding is dense.

We also investigated the robustness of the network with learning rules defined in Eqs (6) and (7) in the sparse coding case. We explored how the storage capacity changes in the presence of noise, in both shallow and deep potential scenarios. As shown in Fig 8, when the noise strength r_3 increases, the storage capacity of the network with deep potentials decays much more slowly than the storage capacity of the network in the flat potential case. This result is consistent with the result for the balanced input case. Notice that for the given network size $N = 1,500$, the storage capacity in the deep potential case is smaller than the storage capacity in the flat potential case for small r_3, despite the fact that it should increase faster with N. This is because the flat potential case has a much larger decay constant, $\tau = 1/2r_1$, leading to a larger prefactor in the relationship between N and p as indicated in Eq (10).

4 Discussion

A long-standing question in neuroscience is whether individual synaptic plasticity events are better described as small changes on a continuum of possible values that are then stable over long time scales, or as large changes among a small discrete set of states [13, 26, 44]. These two scenarios have been studied essentially independently of one another, in networks with unconstrained analog synapses [17, 27, 30], or in networks with binary synapses [18–21].

In this study, we introduced a model that can interpolate between these two scenarios. Synapses in our model evolve in a double well potential, whose shape is described by its depth r_1
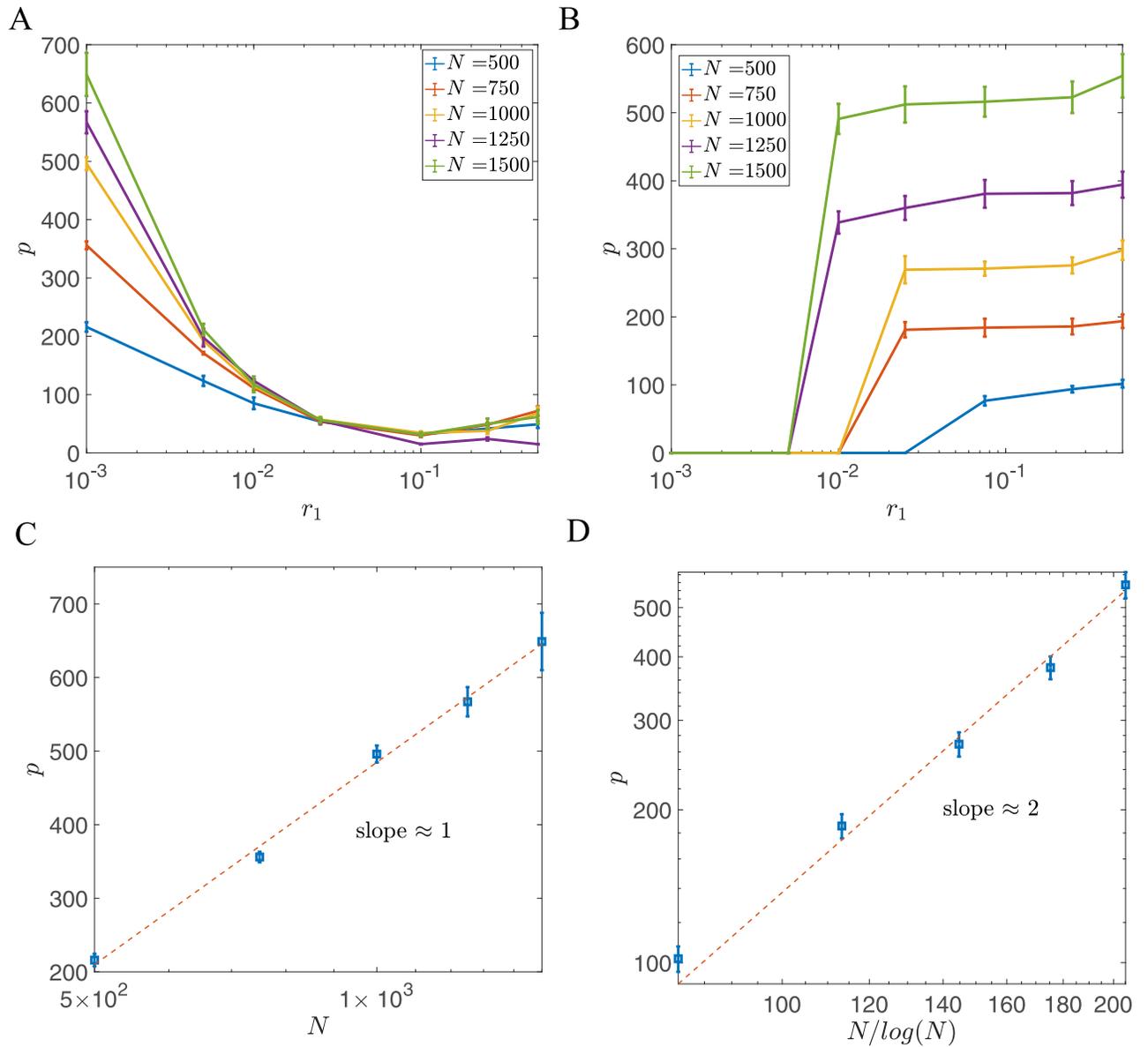
Fig 7. Storage capacity of the model with double well synapses in the sparse coding limit. (A) Storage capacity of the network as a function of r_1 with the learning rule defined in Eq (6) (B) Storage capacity of the network as a function of r_1 with the learning rule defined in Eq (7) (C) p as a function of N in the flat potential case ($r_1 = 10^{-3}$) for the model with learning rule defined in Eq (6). Optimal p is plotted as a function of N in a semi-log plot. Dashed red line: $p = a\log(N/b)$, $a = 396.5$, $b = 294.5$, $R^2 = 0.9964$. (D) p as a function of N in the deep potential case ($r_1 = 0.5$) for the model with the learning rule defined in Eq (7). Dashed red line: $p = a(N/\log(N))^b$, $a = 0.212$, $b = 1.91$, $R^2 = 0.9927$. In all panels, data points are averaged over 10 realizations, and error bars represent the standard deviation. C and θ are chosen to optimize for storage capacity. Other parameter are given as: $r_2 = 1$, $r_3 = 0$, $c = 1$, $f = 4$ $\log(N)/N$.

https://doi.org/10.1371/journal.pcbi.1011354.g007

and width C. When the depth goes to zero, the model becomes equivalent to classic models with continuous synapses, while in the opposite limit of a large depth, the model becomes similar to models with binary synapses. In the absence of inputs, a synapse decays to one of the potential minima exponentially, with a time constant $\tau = 1/2r_1$. External input can cause synapses to jump from one well to the other, leading to a stable change in the absence of noise or additional external inputs leading to subsequent jumps. When the size of the jumps is smaller
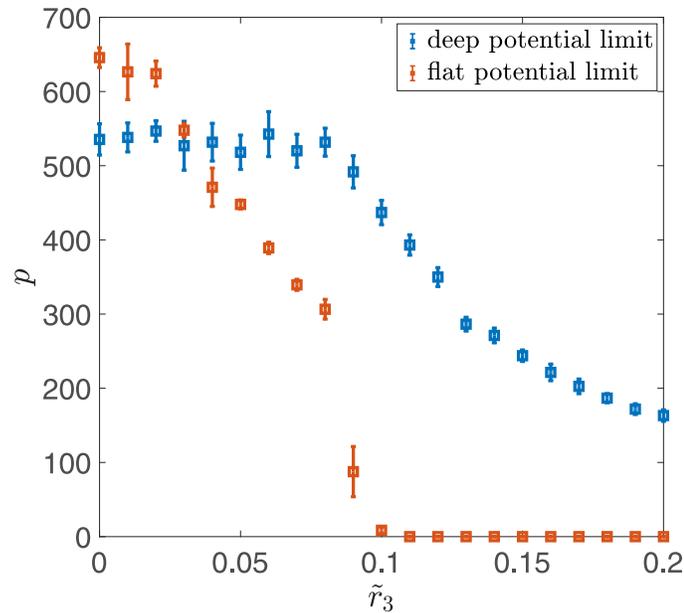
Fig 8. Robustness of the double well model in the sparse coding case. Orange squares represent the storage capacity for the model with the learning rule defined in Eq (6) in the flat potential case ($r_1 = 10^{-3}$). Blue squares represent the storage capacity for the model with the learning rule defined in Eq (7) in the deep potential case ($r_1 = 0.5$). Each point is averaged over 10 independent realizations, and the error bar represents the standard deviation. Other parameters are given as: $r_2 = 1, c = 1, N = 1,500, f = 4\log(N)/N$.

https://doi.org/10.1371/journal.pcbi.1011354.g008

than the distance between the minimum of a well and the barrier that separates the two wells, repetitive stimulation is needed for a synapse to jump from one well to the other. The number of required stimuli is determined by the potential width C. These dynamics are reminiscent of phenomena such as synaptic consolidation or late long-term potentiation. Experimental studies on plasticity in the hippocampus have shown that weak stimulation leads to synaptic changes that decay to their initial value in a relatively short period. However, a strong stimulation results in an enhanced synaptic connection, that persists for the entirety of the recording [45]. This synaptic consolidation process involves multiple complex biochemical dynamics, and our double well potential model could be seen as a minimal implementation of this idea in the framework of attractor neural networks.

There are multiple lines of evidence that suggest that synapses have a small set of stable states. Synaptic plasticity has been shown to be implemented by the addition of unitary synaptic 'nanomodules' to dendritic spines, the loci of excitatory synapses onto pyramidal cells, the main excitatory neuronal type in the cortex and hippocampus [15]. These nanomodules contain spatially localized clusters of pre-and postsynaptic proteins that are aligned with each other. These clusters form due to a reversible diffusion-trapping process of receptors and scaffold proteins on the membrane [46]. Our model can be seen as a minimal model for this scenario, where the two minima of our potential well could correspond to spine configurations with one and two clusters of receptors, which represent close to 90% of the observed spines in [15]. The depth of the potential could be related to the rate at which these clusters form, which depends on various factors, such as the fluidity of the synaptic membrane and the affinity of the receptors for scaffold proteins. The potential width C determines the synaptic plasticity gap, describing the number of stimuli required to induce the formation of new nanomodules. The idea of discrete stable states is also consistent with electron microscopy data showing

distributions of spine sizes in the cortex are bimodal [26], and models of synaptic plasticity relying on biochemical interaction networks of the post-synaptic density [47–49]. Our model predicts that measurements of the efficacy of single synapses on long time scales, in conditions in which learning occurs, should reveal a bimodal distribution.

Our results show that a network in which synapses have two potential wells, and consequently there is a bimodal distribution of synaptic strengths, has a larger storage capacity than a model in which synapses evolve in a single well, and therefore the distribution of weights is unimodal, provided $r_1 > r_1^*$ or equivalently $\tau < \tau^* \equiv 1/(2r_1^*)$. Furthermore, our results show that C^*, the optimal potential width, is greater than one whenever r_1 is larger than its critical value, as shown in Fig 3A. The optimal value of C increases with both network size N and the time constant τ. For large networks, this implies that many repetitive stimuli are needed to induce transitions between potential wells. In this regime, the storage capacity of the network increases as a power law of the network size, instead of a logarithmic relationship. The exponent is close to 0.5 for densely encoded memories, but it becomes equal to 2 (with logarithmic corrections) with sparsely encoded memories when the number of active neurons per memory is logarithmic in network size. These exponents are similar to networks with binary synapses [20, 21]. Our results also suggest that synaptic noise can facilitate the learning of new memories, in situations in which the Hebbian learning term is insufficient to induce transitions between potential wells. Overall, this suggests that a network with such synapses can store memories robustly without requiring high precision in synaptic dynamics.

Our model captures the process of synaptic consolidation in an abstract fashion, as the decay of the synaptic state towards one out of two potential minima. The decay is characterized by a single time constant, τ, which results in a storage capacity that depends sublinearly on the network size in the balanced input case. Experimental and theoretical work has revealed that synaptic plasticity contains several phases across multiple time scales [50–54]. It will be interesting to generalize the model we have proposed to include dynamics with multiple timescales, which would account for various biochemical processes during synaptic consolidation. Such an extension may increase the storage capacity of the network, leading to an almost linear dependence on network size, as suggested by previous research [37]. The phenomenon of synaptic consolidation can be well described by the theory of synaptic tagging and capture [50, 52, 55]. According to this theory, stimulation can result in the tagging of synapses. When weak stimulation is applied, the tag gradually decays. Strong stimulation instead can tag more synapses, initiating the production of plasticity-related proteins that lead to the consolidation of the tagged synapses. Although our double well synapse model has some similarities with the theory of synaptic tagging and capture, it does not account for a number of biological details. For instance, our model only considered independent synapses, while in reality, different synapses on the same dendritic tree can share tagging, since plasticity-related proteins are synthesized in the soma of postsynaptic neurons, which are available to all presynaptic pathways. Moreover, experiments suggest that synapses have varying stimulation thresholds for discrete changes in their efficacy [13], while we assumed homogeneous synapses. Future research will be necessary to include these biological details or investigate these phenomena with more realistic neuron and network models. Another extension of our model would be to increase the number of wells of the potential. We also note that other forms of memory consolidation are likely to play an important role. A recent study demonstrated that adding a pattern reactivation mechanism in attractor neural networks with synapses evolving in a single well potential, mimicking systems memory consolidation [56], can also produce power law scaling storage capacity and lead to lifelong learning scenarios [28]. It will be interesting to incorporate synaptic consolidation and pattern reactivation together in an attractor neural network to explore how double well synapses affect memory function in the context of lifelong learning.

5 Methods

I. The probability density function of synaptic weights

The goal of this section is to derive an equation for $g(J, t)$, the probability density of synaptic strength J at time t, and distributions of synaptic strengths conditioned by inputs experienced in the past. We focus on the noiseless case (i.e., $r_3 = 0$) and the balanced input case with input function defined in Eq (5). From Eq (3), one can derive a master equation for $g(J, t)$:

$$g(J, t^+) = \frac{1}{2} \sum_{I=\pm 1} g(J - r_2 I, t^-), \quad t \in \mathbb{Z}, \tag{15}$$

$$\frac{\partial g(J, t)}{\partial t} = r_1 \frac{\partial}{\partial J} \left(\frac{\partial U}{\partial J} g(J, t) \right), \quad t \notin \mathbb{Z}, \tag{16}$$

where $t^{-,+}$ denote times right before or after a new pattern is presented. Eq (15) describes the transitions from states $g(J + r_2, t^-)$ and $g(J - r_2, t^-)$ to $g(J, t^+)$ when an input is presented. Eq (16) describes the temporal evolution of the distribution g in between presentations, due to the potential term. With the piece-wise parabolic potential defined by Eq (4), the solution of Eq (16) is given as:

$$g(J, t + \Delta t) = e^{2r_1 \Delta t} g(J_0, t^+), \quad \text{for all } t \in \mathbb{Z} \text{ and } 0 < \Delta t < 1, \tag{17}$$

where

$$J_0 = \begin{cases} (J - C)e^{2r_1 \Delta t} + C, & J > C(1 - e^{-2r_1 \Delta t}), \\ (J + C)e^{2r_1 \Delta t} - C, & J < -C(1 - e^{-2r_1 \Delta t}), \end{cases} \tag{18}$$

Note that $g(J, t + \Delta t) = 0$ when $-C(1 - e^{-2r_1 \Delta t}) < J < C(1 - e^{-2r_1 \Delta t})$. After a sufficiently long time, the distribution of synaptic weights should only depend on the time elapsed since the last input presentation, and should therefore be periodic with period 1:

$$g(J, t + 1) = g(J, t). \tag{19}$$

In addition, the probability density function should satisfy the usual boundary conditions,

$$\lim_{J \to \pm \infty} g(J, t) = 0, \tag{20}$$

and the normalization condition

$$\int g(J, t) dJ = 1.$$

With Eqs (15) and (20), one can solve the asymptotic distribution of the weights right before a new pattern is presented, $g_\infty(J) \equiv g(J, t^-)$ for sufficiently large $t \in \mathbb{Z} \to \infty$. To obtain the asymptotic distribution of weights, we discretize J to take $N_d = 4000$ regularly spaced values in the range of $[J_{min}, J_{max}]$. The value of J_{max} (J_{min}) is large (small) enough so that, values outside the range are never visited. In practice, we set $J_{min} = -20$, $J_{max} = 20$. The initial values of the distribution are taken to be random uniform i.i.d., such that the initial distribution is normalized to 1. We then use a discretized version of equations (16,18,19) to update the distribution until convergence.

The mean and root mean square (RMS) of asymptotic distribution $g_\infty(J)$ are given as

$$M = \int g_\infty(J) J dJ \tag{21}$$

$$\Omega = \left(\int g_\infty(J) J^2 dJ \right)^{\frac{1}{2}} \tag{22}$$

We next turn to the distribution of synaptic strengths, conditioned on plasticity events at time $t = v$ and $t = u$. We use the η^u to denote the most recently presented pattern and η^v to denote the previously stored pattern where $v < u$. Distributions of synaptic strengths at time t, conditioned on a plasticity event at $t = v$, are denoted by $g_{\sigma^v}(J, t)$, where $\sigma^v = I(v) = \pm 1$ denotes potentiation or depression. Just after the presentation of the pattern, these distributions are given by:

$$g_{\sigma^v}(J, v^+) = g_\infty(J - \sigma^v r_2), \tag{23}$$

The distributions then evolve according to Eqs (15) and (16), replacing $g(J, t)$ by $g_{\sigma^v}(J, t)$. At any given time, one can calculate the mean $M_{\sigma^v}(t)$ and RMS $\Omega_{\sigma^v}(t)$ of the weight distributions $g_{\sigma^v}(J, t)$:

$$M_{\sigma^v}(t) = \int g_{\sigma^v}(J, t) J dJ \tag{24}$$

$$\Omega_{\sigma^v}(t) = \left(\int g_{\sigma^v}(J, t) J^2 dJ \right)^{\frac{1}{2}} \tag{25}$$

Distributions of synaptic strengths at the time $t = u$, conditioned on plasticity events at time $t = v$ and $t = u$ are denoted by $g_{\sigma^v \sigma^u}$, where $\sigma^v = I(v) = \pm 1$ and $\sigma^u = I(u) = \pm 1$. The distribution of $g_{\sigma^v \sigma^u}$ is given by

$$g_{\sigma^v \sigma^u}(J, u^+) = g_{\sigma^v}(J - \sigma^u r_2, u^-) \tag{26}$$

Its mean and RMS are

$$M_{\sigma^v \sigma^u}(t) = \int g_{\sigma^v \sigma^u}(J, t) J dJ \tag{27}$$

$$\Omega_{\sigma^v \sigma^u}(t) = \left(\int g_{\sigma^v \sigma^u}(J, t) J^2 dJ \right)^{\frac{1}{2}} \tag{28}$$

II. Storage capacity of the double well potential model

We now turn to the question of whether at time u, the network is still able to retrieve a pattern η^v that was stored at a previous time $v < u$. Here, we focus on the standard coding ('balanced input') scenario $f = 0.5$, and $\theta = 0$. Simulations show that sometimes, when initialized at a pattern stored in the past, the network mistakenly retrieves the most recently shown pattern η^u

instead [31]. To calculate the storage capacity, we thus need to define two overlaps:

$$m_v = \frac{4}{N}\sum_{i=1}^{N}(\eta_i^v - \frac{1}{2})V_i^u,$$

$$m_u = \frac{4}{N}\sum_{i=1}^{N}(\eta_i^u - \frac{1}{2})V_i^u, \tag{29}$$

where V_j^u represents the fixed point of Eq (1) with initial condition $V_j = \eta_j^v$, $u > v \in \mathbb{Z}$ and η^u represent the most recently shown pattern.

When the network has overlaps m_v and m_u with patterns shown at times u and v, the probability that a neuron is in a given state V_i, conditioned on η_i^u, η_i^v can be written as:

$$P(s_v, s_u) \equiv P(V_i^u = 1 | \eta_i^v = s_v, \eta_i^u = s_u) = \left(s_v m_v + \frac{1 - m_v}{2}\right)\left(s_u m_u + \frac{1 - m_u}{2}\right), \tag{30}$$

where $s_v, s_u \in \{0, 1\}$. In Eq (30), the first(second) parenthesis on the r.h.s. gives the probability that $V_i^u = 1$ when $\eta_i^u = s_u$ ($\eta_i^v = s_v$) in a network state with overlap m_u (m_v) with this pattern.

To proceed, we need to compute the distribution of local fields conditioned on the state of the neuron in patterns shown at times u and v. The local fields are given by

$$h_i = \frac{1}{N}\sum_{j=1}^{N}c_{ij}J_{ij}V_j^u, \tag{31}$$

where on the right hand side, the distribution of J_{ij} are given by $g_{\sigma^v\sigma^u}$ where $\sigma^{u,v} = (2\eta_i^{u,v} - 1)(2\eta_j^{u,v} - 1)$, and the distribution of V_j^u are given by Eq (30). For large cN, ($cN \gg 1$), we can use the central limit theorem and approximate the local field distributions by a normal distribution $\mathcal{N}(\mu^{s_v,s_u}, d^{s_v,s_u})$, where μ^{s_v,s_u} (d^{s_v,s_u}) is the mean(standard deviation) of the local field h^{s_v,s_u}, conditioned on $\eta_i^v = s_v, \eta_i^u = s_u$. These two moments are given by

$$\mu^{s_v,s_u} = c\sum_{s,s'=0,1}M_{(2s_v-1)(2s-1),(2s_u-1)(2s'-1)}(t = u^+)P(s, s') \tag{32}$$

$$d^{s_v s_u} = \sqrt{\frac{cf}{N}}\Omega_{\sigma^v}(t = u), \tag{33}$$

where $s_v, s_u, s, s' \in \{0, 1\}$. Here we used the fact that $\Omega_{\sigma^v}(t = u) = \Omega_{\sigma^v\sigma^u}(t = u)$ since the pattern u only changes the average of the local field at $t = u$.

The overlaps m_v, m_u can be written as:

$$
\begin{aligned}
m_v &= \frac{4}{N}\sum_{i=1}^{N}(\eta_i^v - \frac{1}{2})\Theta(h_i) \\
&= \frac{1}{2}(P(h^{11} > 0) + P(h^{10} > 0) - P(h^{01} > 0) - P(h^{00} > 0)), \\
m_u &= \frac{4}{N}\sum_{i=1}^{N}(\eta_i^u - \frac{1}{2})\Theta(h_i - 0) \\
&= \frac{1}{2}(P(h^{11} > 0) + P(h^{01} > 0) - P(h^{10} > 0) - P(h^{00} > 0)).
\end{aligned}
\tag{34}
$$

In the balanced input case, $M_{\sigma^v}(t=u) = -M_{-\sigma^v}(t=u)$, $\Omega_{\sigma^v}(t=u) = \Omega_{-\sigma^v}(t=u)$ and thus

$$
\begin{aligned}
h^{00} &= -h^{11}, \\
h^{01} &= -h^{10}, \\
d &\equiv d^{s^v s^u}.
\end{aligned}
\tag{35}
$$

The Eq (34) can be simplified to:

$$
\begin{aligned}
m_v &= P(h^{11} > 0) - P(h^{10} < 0), \\
m_u &= P(h^{11} > 0) - P(h^{10} > 0),
\end{aligned}
\tag{36}
$$

Using the notation defined in Eqs (32) and (33), we obtain self-consistency equations for m_v and m_u:

$$
\begin{aligned}
m_v &= \Phi\left(\frac{\mu^{10}}{\sqrt{2}d}\right) - \Phi\left(\frac{-\mu^{11}}{\sqrt{2}d}\right), \\
m_u &= \Phi\left(\frac{-\mu^{10}}{\sqrt{2}d}\right) - \Phi\left(\frac{-\mu^{11}}{\sqrt{2}d}\right),
\end{aligned}
\tag{37}
$$

where $\Phi(x) = \frac{1}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$. Eq (37) can be solved numerically by updating Eqs (30), (32) and (33) using m_v, m_u. The pattern η_v is retrievable if there exists a solution $m_v \sim 1$, $m_u = 0$, that is stable to small perturbations $m_u = \epsilon$. In order to solve this mean-field equation, we convert Eq (37) to a map, in which the values of $m_{u,v}$ at a given time step are given by the r.h.s. of Eq (37) where $\mu^{11,10}$ and d are computed using the values of $m_{u,v}$ at the previous time step. To check the stability of the state with $m_v > 0$, $m_u = 0$, we take as initial conditions m_v as 1 and m_u as ϵ. We then proceed to update the values of m_v and m_u iteratively until convergence is reached. In practice, we used $\epsilon = 5 \times 10^{-2}$.

III. Signal-to-Noise Ratio

We focus on the calculation of the signal-to-noise ratio in the balanced input case. We first derive the SNR for the binarized double well model. In this case, w_{ij} has the equal probability of being $+1$ and -1, thus we have:

$$
\begin{aligned}
&\langle c_{ij} w_{ij}^{ideal} \cdot w_{ij}(\infty)\rangle = 0 \\
&\text{std}(\sum_{ij} c_{ij} w_{ij}^{ideal} \cdot w_{ij}(\infty)) = \sqrt{c}N.
\end{aligned}
\tag{38}
$$

Therefore, the SNR can be simplified as follows:

$$
\begin{aligned}
SNR(t) &= \frac{\sum_{ij} c_{ij} w_{ij}^{ideal} \cdot w_{ij}(t)}{\sqrt{c}N} \\
&= \sqrt{c}N_{ij}^{ideal} \cdot w_{ij}(t)\rangle.
\end{aligned}
\tag{39}
$$

$w_{ij}(t)$ is drawn from the distribution of $g_{\sigma^v=1}(J,t)$ and the distribution of $g_{\sigma^v=-1}(J,t)$ with the same probability in the balanced input case. The average $w_{ij}^{ideal} \cdot w_{ij}(t)$ can be calculated by:

$$
\begin{aligned}
SNR(t) &= \frac{\sqrt{c}}{2}N\left(\int g_{\sigma^v=1}(J,t)\text{sign}(J)dJ - \int g_{\sigma^v=-1}(J,t)\text{sign}(J)dJ\right) \\
&= \sqrt{c}N\int g_{\sigma^v=1}(J,t)\text{sign}(J)dJ,
\end{aligned}
\tag{40}
$$

where $g_{\sigma^\nu}(J, t)$ are defined through Eqs (15), (18) and (23), and we used the identity $g_{\sigma^\nu}(J, t) = g_{-\sigma^\nu}(-J, t)$ in the balanced input case.

The SNR for Markovian binary synapses SNR_b is given in [57]:

$$SNR_b = \frac{\sqrt{c}N}{2}\mathbf{p}^\infty(\mathbf{M}^{\mathbf{pot}} - \mathbf{M}^{\mathbf{dep}})\mathbf{e}^{\mathbf{tW}}\mathbf{w}. \tag{41}$$

where

$$\mathbf{W} = \frac{1}{2}M^{pot} + \frac{1}{2}M^{dep} - \mathbf{I} \tag{42}$$

where $\mathbf{I}$ is a 2×2 identity matrix, $\mathbf{w} = [-1, 1]$ is the efficacy of the synapse, $\mathbf{p}^\infty = [0.5, 0.5]$ is the stationary distribution of synapses. $\mathbf{M}^{\mathbf{pot}}(\mathbf{M}^{\mathbf{dep}})$ is the 2×2 transition matrix when the synapse is potentiated(depressed). Their value is determined by the transition probability between two states.

SNR_b depends on $\mathbf{M}^{\mathbf{pot}}$ and $\mathbf{M}^{\mathbf{dep}}$, and the upper bound SNR_b^* is given as [57]:

$$
\begin{aligned}
SNR_b^* &= \sqrt{c}Ne^{-t}, t \leq 1 \\
SNR_b^* &= \frac{\sqrt{c}N}{et}, t > 1
\end{aligned}
\tag{43}
$$

Author Contributions

Conceptualization: Yu Feng, Nicolas Brunel.

Formal analysis: Yu Feng, Nicolas Brunel.

Investigation: Yu Feng.

Methodology: Yu Feng, Nicolas Brunel.

Project administration: Nicolas Brunel.

Software: Yu Feng.

Supervision: Nicolas Brunel.

Validation: Yu Feng, Nicolas Brunel.

Visualization: Yu Feng.

Writing – original draft: Yu Feng.

Writing – review & editing: Yu Feng, Nicolas Brunel.

References

1. Bliss TVP, Lomo T. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. J Physiol London. 1973; 232:331–56. https://doi.org/10.1113/jphysiol.1973.sp010273 PMID: 4727084

2. Dudek SM, Bear MF. Bidirectional long-term modification of synaptic effectiveness in the adult and immature hippocampus. J Neurosci. 1993; 13:2910–8. https://doi.org/10.1523/JNEUROSCI.13-07-02910.1993 PMID: 8331379

3. Markram H, Lubke J, Frotscher M, Sakmann B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science. 1997; 275:213–5. https://doi.org/10.1126/science.275.5297.213 PMID: 8985014

4. Sjöström PJ, Turrigiano GG, Nelson SB. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. Neuron. 2001; 32:1149–64. https://doi.org/10.1016/S0896-6273(01)00542-6 PMID: 11754844

5. Bittner KC, Milstein AD, Grienberger C, Romani S, Magee JC. Behavioral time scale synaptic plasticity underlies CA1 place fields. Science. 2017; 357:1033–6. https://doi.org/10.1126/science.aan3846 PMID: 28883072

6. Hayashi-Takagi A, Yagishita S, Nakamura M, Shirai F, Wu YI, Loshbaugh AL, et al. Labelling and optical erasure of synaptic memory traces in the motor cortex. Nature. 2015; 525:333–8. https://doi.org/10.1038/nature15257 PMID: 26352471

7. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA. 1982; 79:2554–8. https://doi.org/10.1073/pnas.79.8.2554 PMID: 6953413

8. Amit DJ. Modeling brain function. Cambridge University Press; 1989.

9. Khona M, Fiete IR. Attractor and integrator networks in the brain. Nat Rev Neurosci. 2022; 23:744–66. https://doi.org/10.1038/s41583-022-00642-0 PMID: 36329249

10. Amit DJ, Gutfreund H, Sompolinsky H. Storing infinite numbers of patterns in a spin-glass model of neural networks. Phys Rev Lett. 1985; 55:1530–1. https://doi.org/10.1103/PhysRevLett.55.1530 PMID: 10031847

11. Amit DJ, Gutfreund H, Sompolinsky H. Information storage in neural networks with low levels of activity. Phys Rev A. 1987; 35:2293–303. https://doi.org/10.1103/PhysRevA.35.2293 PMID: 9898407

12. Gardner EJ. The phase space of interactions in neural network models. J Phys A: Math Gen. 1988; 21:257–70. https://doi.org/10.1088/0305-4470/21/1/030

13. Petersen CC, Malenka RC, Nicoll RA, Hopfield JJ. All-or-none potentiation at CA3-CA1 synapses. Proceedings of the National Academy of Sciences. 1998; 95(8):4732–7. https://doi.org/10.1073/pnas.95.8.4732 PMID: 9539807

14. O'Connor DH, Wittenberg GM, Wang SSH. Graded bidirectional synaptic plasticity is composed of switch-like unitary events. Proceedings of the National Academy of Sciences. 2005; 102(27):9679–84. https://doi.org/10.1073/pnas.0502332102 PMID: 15983385

15. Hruska M, Henderson N, Le Marchand SJ, Jafri H, Dalva MB. Synaptic nanomodules underlie the organization and plasticity of spine synapses. Nature neuroscience. 2018; 21(5):671–82. https://doi.org/10.1038/s41593-018-0138-9 PMID: 29686261

16. Sompolinsky H. The theory of neural networks: The Hebb rule and beyond. In: Heidelberg colloquium on glassy dynamics. Springer; 1987. p. 485–527.

17. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the national academy of sciences. 1982; 79(8):2554–8. https://doi.org/10.1073/pnas.79.8.2554 PMID: 6953413

18. Tsodyks M. Associative Memory in Neural Networks with Binary Synapses. Mod Phys Lett B. 1990; 4:713. https://doi.org/10.1142/S0217984990000325

19. Amit DJ, Fusi S. Constraints on learning in dynamic synapses. Network: Computation in Neural Systems. 1992; 3(4):443–64. https://doi.org/10.1088/0954-898X_3_4_008

20. Amit DJ, Fusi S. Learning in neural networks with material synapses. Neural Computation. 1994; 6 (5):957–82. https://doi.org/10.1162/neco.1994.6.5.957

21. Dubreuil AM, Amit Y, Brunel N. Memory capacity of networks with stochastic binary synapses. PLoS computational biology. 2014; 10(8):e1003727. https://doi.org/10.1371/journal.pcbi.1003727 PMID: 25101662

22. Fusi S, Drew PJ, Abbott LF. Cascade models of synaptically stored memories. Neuron. 2005; 45 (4):599–611. https://doi.org/10.1016/j.neuron.2005.02.001 PMID: 15721245

23. Fusi S, Abbott L. Limits on the memory storage capacity of bounded synapses. Nature neuroscience. 2007; 10(4):485–93. https://doi.org/10.1038/nn1859 PMID: 17351638

24. Lahiri S, Ganguli S. A memory frontier for complex synapses. Advances in neural information processing systems. 2013; 26.

25. Benna MK, Fusi S. Computational principles of synaptic memory consolidation. Nature neuroscience. 2016; 19(12):1697–706. https://doi.org/10.1038/nn.4401 PMID: 27694992

26. Dorkenwald S, Turner NL, Macrina T, Lee K, Lu R, Wu J, et al. Binary and analog variation of synapses between cortical pyramidal neurons. Elife. 2022; 11. https://doi.org/10.7554/eLife.76120 PMID: 36382887

27. Mézard M, Nadal J, Toulouse G. Solvable models of working memories. Journal de physique. 1986; 47 (9):1457–62. https://doi.org/10.1051/jphys:019860047090145700

28. Shaham N, Chandra J, Kreiman G, Sompolinsky H. Stochastic consolidation of lifelong memory. Scientific Reports. 2022; 12(1):13107. https://doi.org/10.1038/s41598-022-16407-9 PMID: 35907920

29. Tsodyks MV, Feigel'man MV. The enhanced storage capacity in neural networks with low activity level. EPL (Europhysics Letters). 1988; 6(2):101. https://doi.org/10.1209/0295-5075/6/2/002

30. Tsodyks M. Associative memory in asymmetric diluted network with low level of activity. EPL (Europhysics Letters). 1988; 7(3):203. https://doi.org/10.1209/0295-5075/7/3/003

31. Pereira U, Aljadeff J, Brunel N. Forgetting leads to chaos in attractor networks. Phys Rev X. 2022; 13:011009.

32. Amit DJ, Fusi S. Constraints on learning in dynamics synapses. Network. 1992; 3:443–64. https://doi.org/10.1088/0954-898X_3_4_008

33. Amit DJ, Fusi S. Dynamic learning in neural networks with material synapses. Neural Computation. 1994; 6:957–82. https://doi.org/10.1162/neco.1994.6.5.957

34. Amit DJ, Gutfreund H, Sompolinsky H. Statistical mechanics of neural networks near saturation. Annals of physics. 1987; 173(1):30–67. https://doi.org/10.1016/0003-4916(87)90092-3

35. Sompolinsky H. Neural networks with nonlinear synapses and a static noise. Physical Review A. 1986; 34(3):2571. https://doi.org/10.1103/PhysRevA.34.2571

36. Derrida B, Nadal JP. Learning and forgetting on asymmetric diluted neural networks. J Stat Phys. 1987; 49:993–1009. https://doi.org/10.1007/BF01017556

37. Benna MK, Fusi S. Computational principles of synaptic memory consolidation. Nat Neurosci. 2016; 19:1697–706. https://doi.org/10.1038/nn.4401 PMID: 27694992

38. Yasumatsu N, Matsuzaki M, Miyazaki T, Noguchi J, Kasai H. Principles of long-term dynamics of dendritic spines. Journal of Neuroscience. 2008; 28(50):13592–608. https://doi.org/10.1523/JNEUROSCI.0603-08.2008 PMID: 19074033

39. Holtmaat AJ, Trachtenberg JT, Wilbrecht L, Shepherd GM, Zhang X, Knott GW, et al. Transient and persistent dendritic spines in the neocortex in vivo. Neuron. 2005; 45(2):279–91. https://doi.org/10.1016/j.neuron.2005.01.003 PMID: 15664179

40. Petersen CC, Malenka RC, Nicoll RA, Hopfield JJ. All-or-none potentiation at CA3-CA1 synapses. ProcNatlAcadSciUSA. 1998; 95:4732–7. https://doi.org/10.1073/pnas.95.8.4732 PMID: 9539807

41. Willshaw D, Buneman OP, Longuet-Higgins H. Non-holographic associative memory. Nature. 1969; 222:960–2. https://doi.org/10.1038/222960a0 PMID: 5789326

42. Nadal JP. Associative memory: on the (puzzling) sparse coding limit. J Phys A: Math Gen. 1991; 24:1093–101. https://doi.org/10.1088/0305-4470/24/5/023

43. Dubreuil AM, Amit Y, Brunel N. Memory capacity of networks with stochastic binary synapses. PLoS Comput Biol. 2014; 10:e1003727. https://doi.org/10.1371/journal.pcbi.1003727 PMID: 25101662

44. Montgomery JM, Madison DV. Discrete synaptic states define a major mechanism of synapse plasticity. Trends Neurosci. 2004; 27:744–50. https://doi.org/10.1016/j.tins.2004.10.006 PMID: 15541515

45. Frey U, Morris RG. Synaptic tagging and long-term potentiation. Nature. 1997; 385(6616):533–6. https://doi.org/10.1038/385533a0 PMID: 9020359

46. Choquet D. Linking nanoscale dynamics of AMPA receptor organization to plasticity of excitatory synapses and learning. Journal of Neuroscience. 2018; 38(44):9318–29. https://doi.org/10.1523/JNEUROSCI.2119-18.2018 PMID: 30381423

47. Zhabotinsky AM. Bistability in the Ca^{2+}/calmodulin-dependent protein kinase-phosphatase system. Biophys J. 2000; 79:2211–21. https://doi.org/10.1016/S0006-3495(00)76469-1 PMID: 11053103

48. Miller P, Zhabotinsky AM, Lisman JE, Wang XJ. The stability of a stochastic CaMKII switch: dependence on the number of enzyme molecules and protein turnover. PLoS Biol. 2005; 3:e107. https://doi.org/10.1371/journal.pbio.0030107 PMID: 15819604

49. Graupner M, Brunel N. STDP in a bistable synapse model based on CaMKII and associated signaling pathways. PLoS Comput Biol. 2007; 3:e221. https://doi.org/10.1371/journal.pcbi.0030221 PMID: 18052535

50. Clopath C, Ziegler L, Vasilaki E, Büsing L, Gerstner W. Tag-trigger-consolidation: a model of early and late long-term-potentiation and depression. PLoS computational biology. 2008; 4(12):e1000248. https://doi.org/10.1371/journal.pcbi.1000248 PMID: 19112486

51. Ziegler L, Zenke F, Kastner DB, Gerstner W. Synaptic consolidation: from synapses to behavioral modeling. Journal of Neuroscience. 2015; 35(3):1319–34. https://doi.org/10.1523/JNEUROSCI.3989-14.2015 PMID: 25609644

52. Barrett AB, Billings GO, Morris RG, Van Rossum MC. State based model of long-term potentiation and synaptic tagging and capture. PLoS computational biology. 2009; 5(1):e1000259. https://doi.org/10.1371/journal.pcbi.1000259 PMID: 19148264

53. Gastaldi C, Muscinelli S, Gerstner W. Optimal stimulation protocol in a bistable synaptic consolidation model. Frontiers in computational neuroscience. 2019; 13:78. https://doi.org/10.3389/fncom.2019.00078 PMID: 31798436

54. Yasuda R, Hayashi Y, Hell JW. CaMKII: a central molecular organizer of synaptic plasticity, learning and memory. Nat Rev Neurosci. 2022; 23:666–82. https://doi.org/10.1038/s41583-022-00624-2 PMID: 36056211

55. Redondo RL, Morris RG. Making memories last: the synaptic tagging and capture hypothesis. Nature Reviews Neuroscience. 2011; 12(1):17–30. https://doi.org/10.1038/nrn2963 PMID: 21170072

56. Hofer SB. Structural traces of past experience in the cerebral cortex. Journal of molecular medicine. 2010; 88:235–9. https://doi.org/10.1007/s00109-009-0560-2 PMID: 19904481

57. Lahiri S, Ganguli S. A memory frontier for complex synapses. In: Burges CJC, Bottou L, Welling M, Ghahramani Z, Weinberger KQ, editors. Advances in Neural Information Processing Systems. vol. 26. Curran Associates, Inc.; 2013. p. 1034–42.