Reviewing the Role of Excitatory and Inhibitory Connectivity

Albert Albesa, Dina Chawla, Emily Hird, Anna Hoyle

05/03/2021

Abstract

In recent years, experimental data has demonstrated significant volatility of excitatory connections in the absence of learning. Such data challenges the theory that excitatory connections are the foundation of learning and long-term maintenance of memories. We hereby replicate the main results obtained in the simulations done in (Mongillo et al., 2018), which try to shed some light on the aforementioned discoveries. The results from the original, together with ours, simulate a scenario in which neural activity is primarily determined by inhibitory connectivity. This is despite the fact that excitatory neurons and synapses are in dominant proportion. Furthermore, it is shown that inhibitory connectivity is significantly more essential and effective at storing memories, compared to excitatory. Finally, we model how excitatory connectivity can indeed shape neural activity when rewiring targets a specific fraction of the population, in accordance to the significant role that excitatory plasticity has been traditionally given in learning processes.

Introduction

Neurons receive vast amounts of information, most of which is transmitted via input currents, with the majority of these inputs occuring at dendritic spines (Tønnesen & Nägerl, 2016). Dendritic spines are small protrusions on the dendritic shaft (figure 1) where most neocortical excitatory synapses reside (Mongillo et al., 2017), thus making spines an ideal proxy for excitatory synaptic existence. Spines are dynamic structures and are considered fundamental to information processing, underlying behaviours such as learning and memory (Yu & Lu, 2012). Over the years, scientific rigor in understanding memory has identified neuronal and synaptic changes that underlie learning and memory. Studies have shown that memory has an association with short-term increase in density and size of dendritic spines (Moczulska et al., 2013; Bencsik et al., 2019). Memory is an asset to humans and non-human animals that supports a wide range of skills and behaviours that are necessary for survival (Friedman et al., 2018). New memories are formed through excitatory connectivity in the hippocampus, amygdala and frontolimbic cortex, whereby, new circuits form within seconds of an event occurring (Kennedy, 2016). The initial attempt to understand and locate neural changes used forms of procedural memory, such as classical conditioning (Davis et al., 1994) and sensitisation (Bailey & Chen, 1988). Such studies have enabled scientists to understand memory through a reductionist approach as each investigation explored neuron connectivity in the different parts of the cortex. Such initial investigations were able to pinpoint specific sites in a neural circuit that were specific to memory storage.

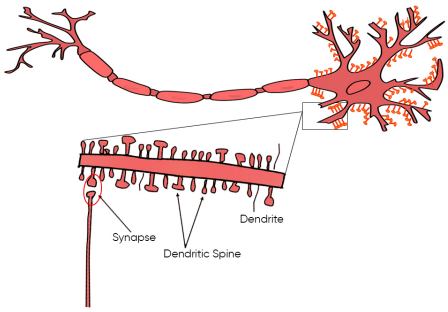

Figure 1: Illustration of dendritic spines (bottom) and their location in a typical neuron (top).

An intriguing characteristic of spines is that they are known to break and reform across different areas of the cortex, with this turnover varying by brain area (Acker et al., 2019). This is often termed spine volatility, and is a discovery that has led to the investigation of how these neural processes persist whilst spines are constantly changing. Many studies have found evidence of spine volatility, with the majority observing spines across imaging sessions to investigate how long they *survive* for (Loewenstein et al., 2015). Investigating the survival rates of spines has proven useful, with research finding that the longer a spine survives, the more likely it is for it to survive and be present in the next imaging session. Spines that survive for a number of sessions are termed *stable*, and those that do not are termed *transient*. Being able to separate the stable spines from the transient has allowed for a number of different findings, including the suggestion that the stability of synaptic

connections allows for the long-term storage of memories (Takeuchi et al., 2014). With spine volatility underlying a number of different processes, as aforementioned, it is no surprise that their dysfunction is linked to neural diseases such as Alzheimer's disease (Dorostkar et al., 2015).

Investigating synaptic activity is important in understanding neuronal connectivity of memory formation, specifically by inspecting changes of connectivity during classical conditioning. Classical fear conditioning and fear extinction has been used by a variety of studies to investigate long-term memory formation and associative learning (Pavlov, 2010). Though, repeated exposure to conditioned stimulus (CS) can diminish manifestation of conditioned response (CR), known as extinction.
Fear extinction creates a new 'safe' association between CS and unconditioned stimulus (UCS) and it has been found that both fear conditioning and extinction regulate neural population in the amygdala, hippocampus and frontolimbic cortex (Milad & Quirk, 2002). Studies have found fear extinction increases the rate of spine formation, whereas, fear conditioning increases the rate of spine elimination (Lai et al., 2012). Furthermore, tone-shock association (CS-UCS) is also found to be correlated with short-term increase in spine formation, which leaves long-lasting structural changes of neural connectivity in the cortex (Moczulska et al., 2013). Such research directly indicates that there is an association between synaptic changes and memories. Additionally, spine formation is dynamic and overtime leads to decrease in spine density with increasing age, which may be due to an increase in elimination of existing spines. Studies have further demonstrated that elimination of spines results in elimination of corresponding memories (Zuo et al., 2005).

This plethora of behavioural studies was the substrate for further investigation of how associations can be imprinted in the brain. With the hope of providing a mechanistic explanation for this, a lot of research, especially that influenced by Donald Hebb's theories of learning (Hebb, 1949), has worked to present different models of synaptic plasticity. These models follow diverse approaches: for example, some (such as the BCM rule, Bienenstock et al. (1982)) describe synaptic dynamics in terms of the firing rate of pre and postsynaptic neurons, while others (such as STDP, Markram et al. (1997)) take into account the actual firing times of each cell. Nevertheless, what all these dynamics have in common is a clear inspiration in the famous mnemonics *cells that fire together wire together*, so that cells that have time coincident increased levels of activity see their recurrent connections strengthened.

Excitatory connectivity, and in particular NMDA channels, have traditionally been regarded as the cornerstone of learning because of the biophysically plausibility *coincidence detector* that they carry with them. NMDA synaptic currents are voltage dependent; the lower the membrane potential of the post-synpatic cell the more it attracts Mg^{2+} ions, which in turn block NMDA channels and prevent current from flowing. For this reason, NMDA currents are only significant when both the pre-synaptic neuron has fired (and thus transmitted an action potential) and the post-synpatic neuron is in a prolonged depolarized state (is in an active state).

Research investigating the volatility of spines and their association with the loss and formation of synapses has mainly used spines as a proxy for excitatory synapses, not inhibitory synapses (Chen et al., 2012). This is because inhibitory neurons do not tend to possess spines. Consequently, this means that there are fewer studies on the role of inhibitory neurons in neural processes such as learning and memory. Research suggests that inhibitory neurons work to rebalance the network after an increase in excitatory neurons during learning. This act of inhibitory neurons re-balancing the network is thought to leave a long-lasting trace on the network, allowing memories to be stored (Froemke et al., 2007). Not only are inhibitory neurons thought to modulate excitability in the network, they are also thought to modulate glutamatergic synaptic release (Pérez-Garci et al., 2013). This points out a gap in the literature, leaving suitable motivation for (Mongillo et al., 2018)'s study.

The present work aims to reproduce some of the results presented in Mongillo et al. (2018). In the original paper, a sparse network is used as a model for evaluating the effects of highly dynamic connections. This is combined with theoretical results obtained by means of mean-field analysis of the network (Bailey & Chen, 1988) and, altogether, presents arguments in favour of the resistance of a network activity to spine volatility. They propose a balanced network, where inhibitory and excitatory inputs are approximately cancelled out (Van Vreeswijk & Sompolinsky, 1996). In this context, fluctuations in the net recurrent input can be considered uncorrelated and the network operates in an asynchronous regime. Thus, the probability of firing depends on the probability of a neuron to be positively unbalanced for a sufficiently prolonged time.
Our work focuses on the simulations and reproduces (i) the network statistics, (ii) both activity and memory retrieval resistance to excitatory rewiring, (iii) sensitivity to an increase in excitatory connections in a subpopulation of neurons.

Methods

Processing of the Dataset

The motivation of the work in (Mongillo et al., 2018) comes from the experimental observation that excitatory connectivity is highly dynamic, which had been presented in the authors' previous publications (Loewenstein et al., 2015, 2011). The dataset from which results were obtained is publicly available in http://bio.huji.ac.il/ yonatanlab/spines/.
Each row is associated to a particular spine and session, and includes different identifiers (cell, dendrite and spine ID) and quantitative information. To associate each spine an efficacy, Principal Component Analysis (PCA) was performed on the pixel map belonging to the spine, which results in two eigenvalues λ_1 (larger), λ_2 (smaller). From these, the strength of the spine (also called shape parameter, S) can be computed by means of the following equation:

$$S = \frac{\lambda_1 - \lambda_2}{\lambda_1 + \lambda_2} \tag{1}$$

The particular efficacy assigned to each synapse was calculated by computing the parameter

$$g = \frac{\langle W_{EE} \rangle}{\langle S \rangle} \tag{2}$$

and for every spine i with shape S_i

$$W(i) = gS_i \tag{3}$$

where $W(i)$ is the assigned efficacy in the connectivity matrix.

Network Model

Dynamics

We follow the network proposal in (Mongillo et al., 2018) to perform the different experiments in the paper. The network consists in Leaky Integrate and Fire (LIF) units, distributed among 10000 inhibitory neurons and a 30000 neurons excitatory population. The dynamics of neuron i inside population a ($a \in \{I, E\}$, $i = 1, 2, ..., N_a$) are described by:

$$\frac{dv_a^i(t)}{dt} = -\frac{v_a^i(t) - H_a^{ext}}{\tau_m} + h_a^i(t) \tag{4}$$

where v_a^i is the membrane potential (volts), τ_m is the membrane time constant (seconds), H_a^{ext} is the external drive (homogeneous among neurons of the same population, in volts) and h_a^i is the afferent input a neuron receives due to synaptic input (in volt · second). It should be noted that every neuron can be uniquely identified with a tuple (a, i), which means *neuron i inside population a*. A neuron *fires at time* t_0 when the condition $\{v_a^i \geq v_\theta\}$ is met (v_θ is called the threshold potential), and a neuron firing is followed by the following operations: $\{t_{a,k}^i\} => \{t_{a,k}^i\} \cup t_0$, $v_a^i(t_0) => v_{res}$. The first operation means that the time of firing is added to the set of firing times of that particular neuron, and the second can be interpreted as the voltage being *reset* to its resting potential (v_{res}).
The synaptic input a neuron receives as a function of time can be expressed in terms of discrete events by the use of Dirac deltas:

$$\begin{aligned} h_a^i(t) = &\sum_{j=1}^{N_E} c_{aE}^{ij} W_{aE}^{ij} \sum_k \delta(t - t_{E,k}^j) \\ &- \sum_{j=1}^{N_I} c_{aI}^{ij} W_{aI}^{ij} \sum_k \delta(t - t_{I,k}^j) \end{aligned} \tag{5}$$

where W_{ab}^{ij} is the synaptic efficacy of the connection corresponding to presynaptic neuron (b, j) and postsynaptic neuron (a, i) and c_{ab}^{ij} is a binary variable that takes value 1 if neuron (b, j) is connected to neuron (a, i) and value 0 otherwise. Upon integration, because the primitive of $\delta(t-t_0)$ is $H(t-t_0)$, a Heaviside function centered at t_0, equation (5) together with (4), can be interpreted as every neuron (a, i) following an exponential decay towards H_a^{ext} with discrete jumps of $+W_{aE}^{ij}$ volts every time neuron (E, j) fires and of -W_{aI}^{ij} every time neuron (I, j) reaches v_θ.

Table 1: LIF model parameters

variable	value	description
single-cell parameters		
v_θ	34 mV	threshold potential
v_{res}	24.75 mV	resting potential
τ_m	10 ms	membrane time constant
τ_{ref}	1 ms	absolute refractory period
network parameters		
N_E	32000	number of E neurons
N_I	8000	number of I neurons
c_{EE}	0.2	probability of E → E connection
c_{EI}	0.3	probability of E → I connection
c_{EI}	0.24	probability of I → E connection
c_{II}	0.4	probability of I → I connection
$\langle W_{EE} \rangle$	0.37 mV	E → E efficacy 1st moment
$\langle W_{IE} \rangle$	0.66 mV	E → I efficacy 1st moment
$\langle W_{EI} \rangle$	0.44 mV	I → E efficacy 1st moment
$\langle W_{II} \rangle$	0.54 mV	I → I efficacy 1st moment
$\langle W_{EE}^2 \rangle$	0.37 mV	E → E efficacy 2nd moment
$\langle W_{IE}^2 \rangle$	0.66 mV	E → I efficacy 2nd moment
$\langle W_{EI}^2 \rangle$	0.44 mV	I → E efficacy 2nd moment
$\langle W_{II}^2 \rangle$	0.54 mV	I → I efficacy 2nd moment
$H_E^{(ext)}$	77.6 mV	E external input
$H_I^{(ext)}$	77.6 mV	I external input

Connectivity

Two models of connectivity define a *random network* (RN) and a *structured network* (SN), respectively. Each network connectivity has a biological interpretation in terms of the question being studied in the original paper: *what is the effect of excitatory connectivity volatility?* As such, the RN tries to represent the experimentally observed variability in synaptic efficacies during different imaging sessions of the same tissue, while the SN has a set of weights chosen to have certain *memories* imprinted, so that one can later rewire the different populations and measure its effect the memories being recalled.

The Random Network The RN defines a probability of connection for every ordered pair (a, b) of populations (which can be done by specifying $\langle c_{ab}^{ij} \rangle = p_{ab}$, probability of a neuron in population b being connected to a neuron in population a). After that, each connection is assigned a synaptic efficacy sampled from a log-normal distribution. The distribution is chosen to be log-normal because so is that of the efficacies measured experimentally. The mean and standard deviation of the log-normal distribution is chosen to have a specific first and second momentum in the final set of weights, which have been chosen by means of mean-field analysis of the network. To do this, for a given pair $(\langle W_{ab} \rangle, \langle W_{ab}^2 \rangle)$ of first and second momentum (respectively) of the final set of weights $\{W_{ab}^{ij}\}$ from population a to population b, once can proceed as follows:

$$\mu_f = \langle W_{ab} \rangle \ , \ \sigma_f^2 = \langle W_{ab} \rangle^2 - \langle W_{ab}^2 \rangle \tag{6}$$

$$\mu_{LN} = \ln\left(\frac{\mu_f^2}{\sqrt{\mu_f^2 + \sigma^2}} \right) \ , \ \sigma_{LN}^2 = \ln\left(1 + \frac{\sigma_f^2}{\mu_f^2} \right) \tag{7}$$

and use these values to define a random variable

$$W = \exp(\mu_{LN} + \sigma Z) \tag{8}$$

which, provided Z is distributed normally, will follow a log-normal distribution with final mean μ_f and variance σ_f^2.

To mimic the effect of the excitatory connections variability across sessions, we generate the connectivity matrix corresponding to the (E, E) pair with a slightly different procedure: (i) generate a binary (connected or not) matrix with probability $p_{EE} = 0.51$, (ii) assign each synapse a spine from the experimental dataset, (iii) for each session 1 to 6, assign each synapse the calculated efficacy for the spine at the session (see *Processing of the Dataset*). This generates 6 connectivity matrices that capture the variability of synaptic efficacies across sessions (1 matrix per session). It should be noted that each spine is uniquely identified by the tuple (cell_id, spine_id), giving a total of 3688 spines in the dataset. Because some of these spines did not exist in the first sessions, or had disappeared at the end of the experiment, the effective number of spines per session is of the order of 1200 spines. This results in some efficacies corresponding to (session_id, cell_id, spine_id) having a value of 0 (the spine did not exist at the moment), which in turn has the effect of having an effective connection probability of $\tilde{p}_{EE} \approx 0.2$ at each session.

The Structured Network As introduced, the SN mimics the process of Long-Term-Memory (LTM) in biological neural networks and is used to model how memory can be affected by spine volatility. To do so, we start by storing $P = 2000$ patterns in the connectivity of the network. Every pattern μ is defined as a binary mapping from μ to $\xi_a^i(\mu)$ that has value 1 if neuron (i, a) participates in the pattern (should be active upon its recalling) and 0 otherwise. Because we are interested in generating a set of weights that both stores the patterns and follows the desired log-normal distribution, we need to perform a set of operations besides the Hebbian and anti-Hebbian standard rules: (i) Compute the Hebbian terms for each connection:[1]

$$z_{ab}^{ij} = \frac{\sqrt{2}}{f(1-f)\sqrt{P}} \sum_{\mu=1}^{P} \epsilon_a^i(\mu, 1)\epsilon_b^j(\mu, 2)(\xi_a^i(\mu) - f)(\xi_b^j(\mu) - f) \tag{9}$$

(ii) Assign a connected/not-connected value to each ordered pair of neurons:

$$c_{ab}^{ij} = \begin{cases} 0 & \text{if } (z_{ab}^{ij} \leq \zeta_{ab} \text{ and } b = E) \text{ or} \\ & (-z_{ab}^{ij} \leq \zeta_{ab} \text{ and } b = I) \\ 1 & \text{otherwise} \end{cases}$$

where $\zeta_{ab} \equiv \phi^{-1}(1 - p_{ab})$ and ϕ^{-1} is the *inverse of the Gaussian cumulative distribution* (the probability of finding a z-score between $-\infty$ and ζ_{ab} is $1 - p_{ab}$). (iii) Provided c_{ab}^{ij} is 1:

$$W_{ab}^{ij} = \frac{\langle W_{ab} \rangle^2}{\sqrt{\langle W_{ab}^2 \rangle}} \exp\left(\sqrt{\ln \frac{\langle W_{ab}^2 \rangle}{\langle W_{ab} \rangle^2}} \ \phi^{-1}(y_{ab}^{ij}) \right) \tag{10}$$

[1] In (Mongillo et al., 2018), the terms $\epsilon_a^i(\mu, 1)\epsilon_b^j(\mu, 2)$ are substituted by a single term $\epsilon_{ab}^{ij}(\mu)$, defined to take values 0 and 1 with probability 1/2 each and also fulfill the condition $\epsilon_{ba}^{ji}(\mu) = 1 - \epsilon_{ab}^{ij}(\mu)$. This is done so that if one pattern contributes (positively or negatively) to the synaptic strength of connection (a, b) it does not participate in that of the connection (b, a) and, ultimately, W_{ab}^{ij} and W_{ba}^{ji} are uncorrelated. Due to computational limitations, we have instead defined two independent vectors $\epsilon(\mu, 1)$ and $\epsilon(\mu, 2)$, with each component set to 1 with probability $1/\sqrt{2}$, so that their product is 1 with probability 1/2 and yet the correlation is minimized (although not 0). It should be noted how z_{ab}^{ij} and z_{ba}^{ji} see their correlation reduced because $\epsilon_a^i(\mu, 1)\epsilon_b^j(\mu, 2)$ and $\epsilon_b^j(\mu, 1)\epsilon_a^i(\mu, 2)$ are independent.

with

$$y_{aE}^{ij} = \frac{1}{c_{aE}} \int_{\zeta_{aE}}^{z_{ab}^{ij}} Dz \ , \ y_{aI}^{ij} = \frac{1}{c_{aE}} \int_{z_{ab}^{ij}}^{\zeta_{aI}} Dz \qquad (11)$$

where Dz is the Gaussian measure $Dz = 1/\sqrt{2\pi}e^{z^2}dz$

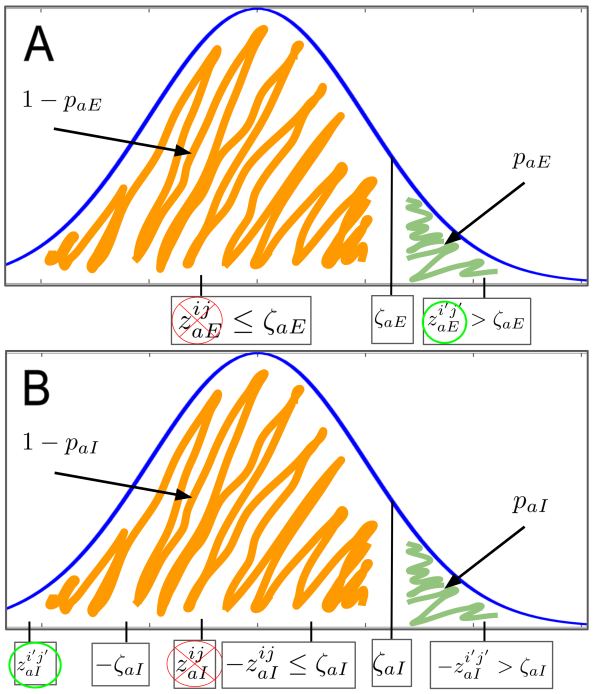

Figure 2: Schematic representation of Hebbian terms distribution in equation (9). Blue line corresponds to the probability density function (y axis) of z-score values (x axis). The green area represents the imposed probability of connection for a particular matrix W_{ab}, and orange its complement. This is by construction, as $\zeta_{ab} \equiv \phi^{-1}(1-p_{ab})$ In the x axis one can see how different values are passed to the final matrix (green circle) or not (red cross) according to the rules presented in step (ii). The blue line corresponds to an idealized Gussian probability density function of z_{ab}^{ij}. Proportions should not be taken into consideration as the figure is included for illustrative purposes.

The reasoning behind this particular procedure is as follows: step (i) generates a distribution of terms that, in the limit $P \to \infty$, is normal with mean 0 and variance 1; at the same time, these terms follow Hebbian statistics that ensure bigger terms between neurons having more patterns in common; because z_{ab}^{ij} is normally distributed, step (ii) ensures that the sparseness of the matrix corresponds to the original probabilities p_{ab} (c_{ab}^{ij} is 1 with probability $(1-(1-p_{ab})) = p_{ab}$) ; the only terms that survive are the most Hebbian for excitatory connections ($b = E$) and the most anti-Hebbian for inhibitory presynaptic input ($b = I$), which is achieved by taking the most negative tail of the distribution in the case of inhibitory synapses ($c_{aI} = 1$ when $z_{aI}^{ij} \leq -\zeta_{aI}$); step (iii) guarantees that the synaptic efficacies of connected neurons are sampled from a log-normal distribution (as for the random network), with the exception that the normal variable, in this case $\phi^{-1}(y_{ab}^{ij})$, takes higher values with higher probability when the corresponding Hebbian term has a greater magnitude (through equation (11)).

Implementation

Network Simulation

We have used the Brian2 (Stimberg et al., 2019) Python package for simulating our network. This package provides with an intuitive and powerful way of building neural network models, that achieves top performance by generating C++ code on the back-end to actually run the models.

Listing 1 NeuronGroup object in Brian2.

```
1  from brian2 import *
2  eqs = '''
3  dv/dt = -(v - H)/tau :volt(unless refractory)
4  H : volt
5  '''
6  all_neurons = NeuronGroup(N, eqs,
7      threshold = 'v>v_theta',
8      reset = 'v=v_reset',
9      refractory = tau_ref,
10     method = 'exact')
```

The network dynamics are given to the simulator with a multi-line string that is later fed to the NeuronGroup constructor. The NeuronGroup object is the core of the simulation and captures the total number of neurons and their dynamics. As can be seen in Listing 1, the object constructor accepts as direct arguments the condition for firing (*threshold*), as well as the reset voltage (*reset*) and refractory period (*refractory*). One can also choose the ODE resolution method; because our model consists on an exponential decay, we can set the method to be *exact*, so that the corresponding exponential increment is computed at every time-step. It should be noted that the model (*eqs*) differs from equation (4): there is no synaptic input term. Because the model is assumed to be a Spiking Neural Network, Brian2 manages spiking processes and their corresponding dynamics (which is done through *Synapses* objects).
In this sense, listing 2 shows how to integrate in the model the input dynamics associated to equation (5), by simply specifying that a neuron voltage should be incremented by the corresponding synaptic efficacy every time a presynaptic neuron fires (line 3). In lines 1 and 2 *subgroups* of the main NeuronGroup are

created, so that one can specify the particular parameters of each population (in our case inhibitory and excitatory). In line 3 a *Synapses* object is created, thus generating projections from (in this case) *ex_neurons* to *in_neurons*. It is also needed to effectively connect (line 5) the subgroups by specifying the distributions or conditions synapses should meet. According to the model presented previously, for example, E→I connections happen with a probability $p_{EI} = 0.3$.

Listing 2 Subgroups and Synapses in Brian2

```
1 in_neurons = all_neurons[first_inh:last_inh]
2 ex_neurons = all_neurons[first_exc:last_exc]
3 S_IE = Synapses(ex_neurons, in_neurons,
4     'w : volt', on_pre ='v += w')
5 S_IE.connect(p = 0.3)
```

Listing 3 Subgroups (populations) and Synapses generation in Brian2

```
envelopes1 =
    np.array([((1.5 - 0.15*i)*pattern0
    + (0.25 + 0.075*i)*ones(N)
    if i < 10 else (ones(N)))
    for i in range(80)])
#envelopes2 similarly for pattern1
#envelopes3 similarly for pattern2
envelopes=[*envelopes1, *envelopes2,
    *envelopes3]
stimulus = TimedArray(envelopes, dt=25*ms)
            ###
eqs = '''
dv/dt = -(v - H_ext*stimulus(t, i))/tau
'''
```

Weights Computation

Weight computation is primarily done in Julia (Bezanson et al., 2017), after seeing some computational limitations for certain matrix operations in Python (the connectivity matrices are of the order of almost a billion elements). Although it is a general-purpose language, Julia is gaining popularity as a scientific-computing language due to is high-level and intuitive syntax (similar to Python) and extreme efficiency (comparable to C). The matrices are stored in a sparse format to reduce its size (approximately from 5GB to 1GB).

Memory Retrieval

The weights generated as described in the previous section have a certain set of patterns stored in them. Hebbian learning ensures that those states of the network defined as the neurons of a pattern being active become attractor states. As such, when the network finds itself in a state in which a significant amount of the neurons belonging to a specific pattern are active, the network suffers an acceleration in the direction of the pattern. Once there, it remains oscillating around it unless external input is provided to the system, thus allowing it to visit the transient unstable states that it would find in its trajectory towards a new stable point of the activity space.
In our experiments, we implement the process of memory retrieval by selecting some of the $P = 2000$ patterns stored in the connectivity matrices. Then, we create a *TimedArray* object in Brian2, which allows us to define variables whose value varies from one time-bin to another. We start providing 1.75 of external input to the neurons belonging to one of the patterns and 0.25 of it to those not participating in it. Then, these values are linearly restored to 1 in the interval of 250 ms. External input is then kept constant during 1750 ms, until another memory is retrieved similarly. This process allows to test the capability of the network to retrieve a pattern when presented with an input that resembles the associated attractor state. In listing 3 one can see the how the different *envelopes* are defined (each envelope is a matrix with dimensions number of neurons times number of time bins) and then stored as a time dependent stimulus that can be accessed by the model during the simulation (note how $stimulus(t, i)$ depends on both the time t and the neuron i).

Code Availability

The code for reproducing results and figures in this paper can be accessed in (Albesa et al., 2021).

Results

Random Network Setup

Obtaining a network that evolves in an asynchronous regime resulted to be less trivial that initially expected. For the network to be stable, voltages (and subsequently firing rates) must follow certain distributions, otherwise, it can undergo high-amplitude oscillations that result in unrealistic values.
The initial distribution of voltages was chosen to be uniform between the resting potential and the threshold voltage. To understand this, one must consider the trajectories that each neuron follows at the beginning: exponential towards threshold, with a time constant τ_m. A uniform distribution across voltages, thus, results in a non-uniform (in time) distribution of firings, causing *batches* of neurons to fire simultaneously.
This problem was tackled by calculating the voltage trajectories of the neurons, and imposing the firings

to be equidistant in time. To do this, one can consider what is the evolution of a leaky integrator in the presence of an external drive:

$$v_a^i(t) = H_a^{(ext)} + (v_{a0}^i - H_a^{(ext)}) \exp(-t/\tau_m) \quad (12)$$

The time needed to fire (to reach v_θ) can be found by setting the voltage to the threshold potential and solve for t:

$$T_a^i = \frac{1}{\tau_m} ln\Big[\frac{v_{a0}^i - H_a}{v_\theta - H_a}\Big] \quad (13)$$

To guarantee that firings have uniform interspike intervals (at least at the beginning), one can choose a neuron $(a, 0)$ to start at a value $v_{a0}^0 = v_{a0}$. The time to fire T_a^0 will be given by equation (13) . If now one lets the initial voltage of neuron $(a, 1)$ be

$$v_{a0}^1 = H_a^{(ext)} + (v_{a0} - H_a^{(ext)}) \exp\Big(-\frac{T_a^0}{\tau_m N_a}\Big) \quad (14)$$

it is clear (equation (12)) that its initial voltage is precisely that of neuron $(a, 0)$ after a time T_a^0/N_a. One can subsequently define initial voltages to be:

$$v_{a0}^i = H_a^{(ext)} + (v_{a0} - H_a^{(ext)}) \exp\Big(-i\frac{T_a^0}{\tau_m N_a}\Big) \quad (15)$$

and it is immediate to see how all neurons in population a will fire (from $i = (N-1)$ to $i = 0$) every $t = T_a^0/N_a$.

An additional necessary ingredient not explicitly mentioned in (Mongillo et al., 2018) is that, during refractoriness, cells not only can not fire again, but also are not affected by recurrent input (after firing, voltage is *clamped* at resting potential during a time τ_{ref}). This was included in the model after revising (Feng, 2003), which includes this condition as necessary for deriving the equations presented in original mean-field analysis.

Figure 3 presents the first set of direct results on the network simulation, and illustrates the typically observed behavior of the membrane potential and the firing rate distribution for each type of cell. Irregular and Poisson-like action potentials can be observed in figure 3A, thus reproducing the basic features of cortical dynamics and mimicking asynchronous biological firings. As the total time of the simulation was 1 second, the number of spikes can be interpreted as the firing rate, which in both cases matches the typical values inferred from the distribution shown in figure 3B (it should be noted that the vertical lines were added manually to show the actual firing times). Figure 3A is also very illustrative of the underlying *stochastic* process that describes voltage dynamics (the equations for simulation are, nevertheless, deterministic, but the resulting behaviour can very well be described in stochastic terms). In these conditions, the membrane potential dynamics have two sources of

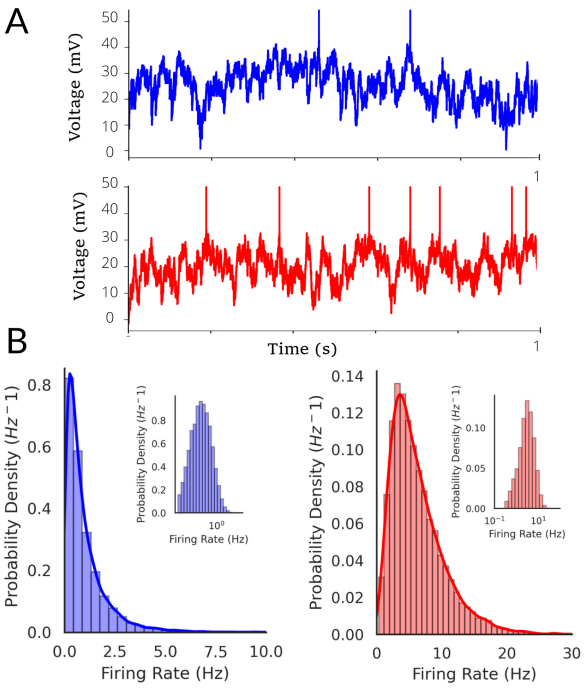

Figure 3: A, Plots of the membrane potential traces for one randomly selected excitatory neuron (blue) and one randomly selected inhibitory neuron (red) for 1 second, with vertical lines indicating the spike times. B, Histogram plots show the distribution of firing rates in the excitatory neuron population (blue) and the inhibitory neuron population (red) for a simulation of 17 minutes, on a linear scale (large figures) and on a logarithmic scale (small figures).

drive: one coming from external input (with variance zero, as it is constant) and the other from recurrent connections. Neurons, thus, experience a subthreshold voltage average around which they are perturbed by the white noise inherited from the imperfect counterbalance of excitatory and inhibitory populations. These types of dynamics have been extensively studied in the context of Ornstein-Uhlenbeck processes, which describe random walks with a tendency to be re-centered around a mean. While theoretical results from the original are not reproduced in this paper, it is worth mentioning that it is precisely these simple stochastic dynamics that allow for mean-field analysis of the network (a high dimensional system such as this would otherwise be virtually intractable).

As presented in the introduction, the firing of neurons can be understood as a noise-driven escape process, which depends exponentially on the input received. As the firings from presynaptic neurons are only weakly correlated, and the number of recurrent connections is very high, the time-averaged input a cell accepts can be considered normal. The two processes (exponential noisy escape and normal input)

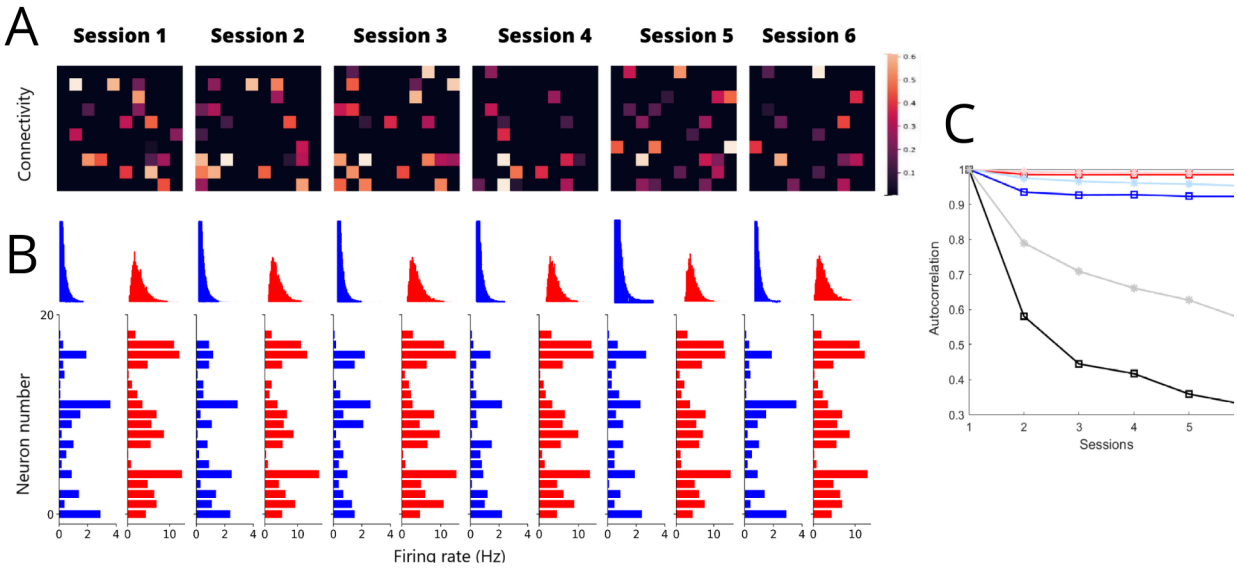

Figure 4: A, Connectivity matrices of ten randomly selected pairs of excitatory neurons, where the colour indicates the strength of synapses (black represents unconnected pairs). B, Top: distributions of firing rates of excitatory (blue) and inhibitory (red) neurons (where session 1 is same as in figure 3). Bottom: firing rates of randomly selected 20 excitatory (blue) and 20 inhibitory (red) neurons. C, Autocorrelogram of E→E connectivity matrices (black) and firing-rate vectors of excitatory (blue) and inhibitory (red) neurons, including original study's results for E→E connectivity matrices (grey) and firing-rate vectors of excitatory (light blue) and inhibitory (light red) neurons.

combined have as outcome the firing rate log-normal distribution observed in figure 3B.
Inhibitory neuron population demonstrates a higher firing rate compared to the excitatory population (7.8Hz versus 1.64 Hz respectively). As there are less neurons in the inhibitory population, their firing rates must increase in order to compensate for the excitatory recurrent input.
To achieve curves as smooth as those presented in (Mongillo et al., 2018) the simulated time was 17 minutes, as opposed to the 3 minutes reported in the original work; this indicates that, although the temporal average of firings is similar, in our simulations there is a greater variance in time.

Volatility of Spines and Network Activity

We aimed to study and test the significance of previously mentioned volatility of E→E connectivity in cortical network models. Experimentally observed data of spine changes was used to assess changes in E→E connectivity. Figures 4 shows how the different distributions in the networks vary across sessions. Figure 4A presents six heatmaps corresponding to 10x10 submatrices of the E→E connectivity at each sessions. Each non-zero entrance corresponds to a single spine, which is updated according to spine formation, elimination and change in the size. In figure 4B one can see how, despite notable changes in excitatory synapses, the activity of both E and I populations is spared from one session to another. Not only the firing rate distributions are maintained (figure 4B, top), but also the activity of neurons at the single-cell level (figure 4B, bottom).
To further test the extent of the stability of the network to a rewiring of the excitatory connections, figure 4C shows the autocorrelogram of both the E→E matrices and activity vectors of each session with the ones corresponding to session 1. These were obtained by computing the scaled dot product of the activity vectors and flattened matrices at each session with those corresponding to session 1. In the same figure, we show the results obtained in the original work (lighter color). While the relative decrease in correlation of the activity vectors compared to that of connectivity matrices is similar in both our results and the ones obtained originally, the variation of connectivity across sessions seems to be higher in our case. Howbeit, although this indicates a possible dissimilarity in the procedure to generate the E→E matrices from the *in vivo* dataset, from both one can confirm that the network activity is highly resilient to a redistribution in excitatory connections.
The original paper includes a very illuminative mental picture that helps understand the reason behind this. From the perspective of a postsynaptic cell with two cells projecting into it, rewiring effect should be

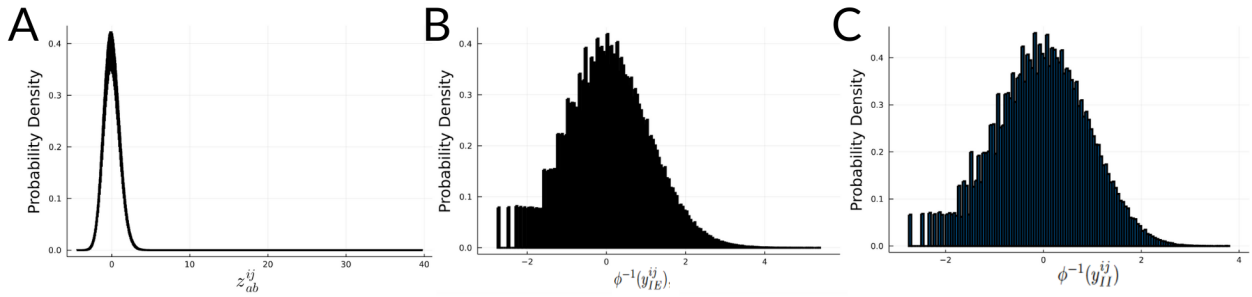

Figure 5: Hebbian terms distributions obtained in the process of constructing the Structured Network connectivity matrices. A, Probability density function of z_{ab}^{ij}. Probability density function of $\phi^{-1}(y_{ab}^{ij})$ terms following Hebbian (illustrated with $\phi^{-1}(y_{IE}^{ij})$, B) and anti-Hebbian (illustrated with $\phi^{-1}(y_{II}^{ij})$, C) rules.

little if either both presynaptic neurons have a similar firing rate or similar synaptic strength. Following this reasoning, the wider and more variant is the distribution of efficacies or firing rates across a population, the higher the impact of rearranging its outgoing weights. Network activity should therefore be more resistant to excitatory rewiring, providing that the firing rate of this population is much more narrowly distributed.

Structured Network Setup

In figure 5, we aimed to show that z_{ab}^{ij} and the different $\phi^{-1}(y_{ab}^{ij})$ follow normal distributions, as it is assumed in the methods and a necessary condition for the final efficacies distribution to have the same statistics as in the random network. A quasi-Gaussian distribution can be appreciated across each of the subplots. Apart from the natural stochastic departure from ideal normality, one can see a clear asymmetry in all the density functions. This is inherited from equation (9), where the summative terms can be -0.01, 0.81 and -0.09 with probabilities 0.81, 0.01 and 0.18 (respectively) which results in greater variance if one considers terms over the mean than terms under the mean. Despite the central limit theorem pushing the distribution to be Gausssian, and the number of samples being considerably large, it is not enough to completely neutralize this intrinsic asymmetry.

Rewiring and Memory

The goal of this third experiment was to test the ability of the network to store memory patterns (in a short-term memory manner) as well as its robustness to synaptic rewiring. Figure 6 illustrates the process of memory retrieval described in the Methods. One can see on 6A how different patterns are successfully retrieved after presenting the network with the appropriate stimulus. This correlation between the network activity and the presented pattern is demonstrated by the observed over chance overlap,

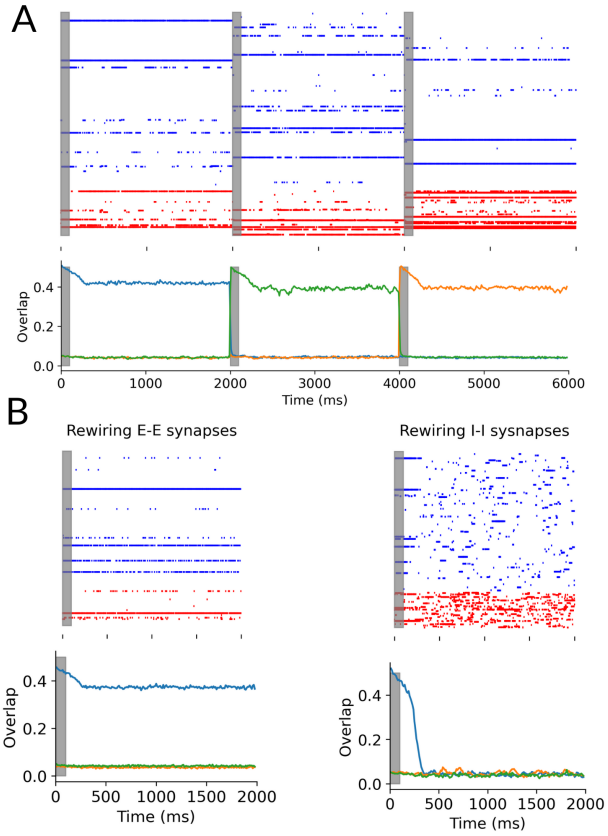

Figure 6: Memory pattern recall and the effect of synaptic rewiring of the model with 2000 embedded memory patterns. Grey bars indicate when pattern-like external drive was applied to the network (both figures). A, Raster plot (firing times) of 400/100 excitatory/ inhibitory arbitrarily chosen neurons (top). Overlap (scaled dot product) as a function of time between the firing rate vectors (averaged over 10ms time bins) and the binary pattern vector (bottom). The green, orange and blue coloured lines depict the overlap for 3 different memories. B, The same as figure A after rewiring the E→E (left) and I→I (right) connectivity.

even after the pattern-like input vanishes (6A bottom). Also, by looking at the raster plots in 6A (top) the traces left by the neurons belonging to the pattern can be appreciated.
As for the second part of the experiment, the E→E and the I→I connections were changed to see their importance in storing memories (6B). While the rewiring of the E→E synapses had no effect in pattern retrieval, the memory patterns still could be recalled when I→I connections got rewired.
These findings match the theoretical predictions of mean-field analysis regarding the storage capacity of the network, which demonstrates that inhibitory connections are able to store more memory pattern than excitatory ones.

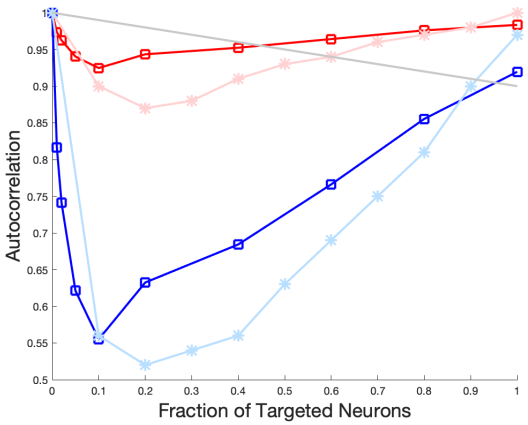

Figure 7: Autocorrelogram of excitatory (red) and inhibitory (blue) firing rate vector for different fractions of targeted neurons. Results from original work are also included: excitatory population (light blue), inhibitory population (light red) and E→E matrix (grey).

Excitatory plasticity reconsidered

The last simulation included tries to reconcile the weak relevance that results presented so far seem to grant to excitatory connections with the major role in memory and learning that has traditionally been associated with it. To do so, we have followed the same approach as in the original work, by letting a small cluster (or subpopulation) of the excitatory neurons to suffer e a small increase in their excitatory recurrent inputs. In particular, we have taken a fraction f of the excitatory neurons *target neurons* and projected every excitatory neuron with a probability of 0.04 to the target neurons, thus resulting in an increase in the total probability of E→E connection, which goes from 20 to 24% (this is slightly different that as in the original paper, where the extra connection probability was of 0.05 because it did not include already projecting neurons, but final number of connections is kept constant). Figure 7 shows the outcome of our simulations, together with the original paper results. As in (Mongillo et al., 2018), one can observe a rapid decrease in activity correlation that then recovers in an approximately linear manner. We have nevertheless obtained one major discrepancy with regarding the minimum of the curve (that in the originally was found to be 0.2 and for us is 0.1). A natural proposal to explain this would be our outlined little methodological departure, which had not other reason that being much more simple to implement in Brian2.

Discussion

The model replicates the higher and more heterogeneous firing rates of inhibitory population, compared to excitatory, as obtained in the original work. This implies the domination of inhibition in cortical activity.
While the model itself does not predict a predominant computational role of inhibitory neurons, this is due to its particular parameter values. As these were adjusted according to experimental data, and as pointed out in the original paper, results for values representing different contexts, such as brain region, or animal studied, could be different.
The fact that cortical activity is strongly defined by inhibitory synapses can explain its stability regardless of highly dynamic excitatory connections, which can be inferred from observed spine volatility. This significance of inhibition is also observed in memory encoding and retrieval.
The less relevant role that these results could give to excitatory connectivity becomes explicitly false in light of results shown in figure 7. When the rewiring of excitatory synapses is not random, and instead targets a small subpopulation, the effect on the network activity can be highly significant.
Using spine volatility to investigate inhibitory neuron populations has not always been considered, and so these results may seem surprising. However this is consistent with previous research that has been done in this area, in that the network appears robust to the volatility of excitatory synapses, as long as this is not able to disrupt the balance between the two populations. Development of a balance between excitation and inhibition is key for the modulation of neural circuits that are involved in sensory processing (Alitto & Dan, 2010), as well as learning, and memory (Moczulska et al., 2013). Imbalances between neuronal excitation and inhibition have been associated with many neurological disorders, with an example being Down's syndrome (Fernandez et al., 2007). It is thought that some deficits experienced in Down's syndrome are a result of excessive inhibition,

with cognitive functioning improving when inhibition is reduced.

Results from the original paper and our reproduced results indicate that inhibitory connectivity prevents long-term storage in volatile network models (reference to figure). These results further suggest the function of inhibition in stabilising excitatory connectivity. Studies have built upon this research and have suggested the function of homeostatic regulation of inhibition to establish and preserve computational organisation of excitation in mice using monocular deprivation (Ma et al., 2019).

Furthermore, both the original paper and our reproduced results showed that changes in inhibitory connectivity has a greater impact on the network as a whole compared to those in excitatory connections, indicating that inhibition is important for storing memories, and can prompt changes in excitatory connectivity. Additional research has proposed that inhibition is specifically important for storing reward-influenced memories, which can influence changes in excitatory plasticity (Wilmes & Clopath, 2019).

Moreover, studies have suggested context-dependent modulation and plasticity of inhibitory networks have significant contribution towards memory and learning, indicating that inhibition and excitation are not accurately balanced and that synaptic plasticity is in fact *multisynaptic*. (Herstel & Wierenga, 2021).

In the original research, results led to a specific proposal of how memories are imprinted in the brain by means of a two-stage process. This mechanism involves early excitatory plasticity increasing excitatory connections between an assembly of neurons that shapes the neural activity representing for the memory, in a way that very much resembles the methods followed to obtain figure 7. This first step is predicted to increase overall network activity, which triggers inhibitory plasticity to re-balance the network on a much slower timescale. It is proposed that long-term memories are encoded in the brain in the precise inhibitory connections that are formed in this second stage. Altogether, this reconciles the known role of spines and, in general, excitatory connections and glutamtatergic Hebbian learning with the robustness of networks to random rewiring of its excitatory units... As appealing as this theory of memory formation might seem, it makes some obvious experimental predictions that are not mentioned in the original paper. Glutamate-induced inhibitory long-term potentiation has been observed in hippocampal neurons (Chiu et al., 2019), consistent with the idea that there are activity-triggered plasticity mechanisms that induce inhibitory connections strengthening in a region that plays such an important role in memory as the hippocampus. Nevertheless, a future line of experimental research could include testing whether blocking inhibitory connections dynamics affects the consolidation of acquired memories. Not only should long-term memory be affected by such manipulations, this theory predicts that these should leave untouched the ability of a subject to learn and remember in the short-term.

While this provides an explanation for resilience of memories to spine volatility, there are other proposals that do not rely exclusively on inhibitory connections. In (Susman et al., 2019), memories are proposed to be encoded in the complex components of the eigenvalue modifications implemented by Hebbian plasticity. While different models of homeostatic plasticity (as one would expect for a stabilizing drive) erode the real part of the eigenvalues of the connectivity matrix, most of them are proven to leave intact (or little changed) its complex component.

We would like to also refer to our work in terms of the replication and reproduction of previously published results. While we have successfully reproduced all of the simulations included in the original paper, we have found certain differences in the precise values obtained for some of the experiments. However, we think it is important to point out that none of these dissimilarities is in contradiction with the conclusions that original results lead to. The sources of these deviations could well be the slight methodological differences that we have reported above: (i) the way to reduce correlation between opposite weights in the structured network and (ii) the generation of additional connections to reproduce figure 7. Nevertheless, this can only explain part of our discrepancies. For example, the notable differences in the E→E matrix autocorrelogram in figure 4C seem unlikely to be related to any of these changes.

References

Acker, D., Paradis, S., & Miller, P. (2019). Stable memory and computation in randomly rewiring neural networks. *Journal of neurophysiology*, *122*(1), 66–80.

Albesa, A., Chawla, D., Hird, E., & Anna, H. (2021). `https://github.com/albertalbesa/inhibitory-connectivity`. GitHub.

Alitto, H. J., & Dan, Y. (2010). Function of inhibition in visual cortical processing. *Current opinion in neurobiology*, *20*(3), 340–346.

Bailey, C. H., & Chen, M. (1988). Long-term memory in aplysia modulates the total number of varicosities of single identified sensory neurons. *Proceedings of the National Academy of Sciences*, *85*(7), 2373–2377.

Bencsik, N., Pusztai, S., Borbély, S., Fekete, A., Dülk, M., Kis, V., ... others (2019). Dendritic spine morphology and memory formation depend on postsynaptic caskin proteins. *Scientific reports*, *9*(1), 1–16.

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical

computing. *SIAM Review*, *59*(1), 65–98. doi: 10.1137/141000671

Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, *2*(1), 32–48.

Chen, J. L., Villa, K. L., Cha, J. W., So, P. T., Kubota, Y., & Nedivi, E. (2012). Clustered dynamics of inhibitory synapses and dendritic spines in the adult neocortex. *Neuron*, *74*(2), 361–373.

Chiu, C. Q., Barberis, A., & Higley, M. J. (2019). Preserving the balance: diverse forms of long-term gabaergic synaptic plasticity. *Nature Reviews Neuroscience*, *20*(5), 272–281.

Davis, M., Hithcock, J. M., Bowers, M. B., Berridge, C. W., Melia, K. R., & Roth, R. H. (1994). Stress-induced activation of prefrontal cortex dopamine turnover: blockade by lesions of the amygdala. *Brain research*, *664*(1-2), 207–210.

Dorostkar, M. M., Zou, C., Blazquez-Llorca, L., & Herms, J. (2015). Analyzing dendritic spine pathology in alzheimer's disease: problems and opportunities. *Acta neuropathologica*, *130*(1), 1–19.

Feng, J. (2003). *Computational neuroscience: a comprehensive approach*. CRC press.

Fernandez, F., Morishita, W., Zuniga, E., Nguyen, J., Blank, M., Malenka, R. C., & Garner, C. C. (2007). Pharmacotherapy for cognitive impairment in a mouse model of down syndrome. *Nature neuroscience*, *10*(4), 411–413.

Friedman, G. N., Johnson, L., & Williams, Z. M. (2018). Long-term visual memory and its role in learning suppression. *Frontiers in psychology*, *9*, 1896.

Froemke, R. C., Merzenich, M. M., & Schreiner, C. E. (2007). A synaptic memory trace for cortical receptive field plasticity. *Nature*, *450*(7168), 425–429.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Psychology Press.

Herstel, L. J., & Wierenga, C. J. (2021). Network control through coordinated inhibition. *Current Opinion in Neurobiology*, *67*, 34–41.

Kennedy, M. B. (2016). Synaptic signaling in learning and memory. *Cold Spring Harbor perspectives in biology*, *8*(2), a016824.

Lai, C. S. W., Franke, T. F., & Gan, W.-B. (2012). Opposite effects of fear conditioning and extinction on dendritic spine remodelling. *Nature*, *483*(7387), 87–91.

Loewenstein, Y., Kuras, A., & Rumpel, S. (2011). Multiplicative dynamics underlie the emergence of the log-normal distribution of spine sizes in the neocortex in vivo. *Journal of Neuroscience*, *31*(26), 9481–9488.

Loewenstein, Y., Yanover, U., & Rumpel, S. (2015). Predicting the dynamics of network connectivity in the neocortex. *Journal of Neuroscience*, *35*(36), 12535–12544.

Ma, Z., Turrigiano, G. G., Wessel, R., & Hengen, K. B. (2019). Cortical circuit dynamics are homeostatically tuned to criticality in vivo. *Neuron*, *104*(4), 655–664.

Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, *275*(5297), 213–215.

Milad, M. R., & Quirk, G. J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature*, *420*(6911), 70–74.

Moczulska, K. E., Tinter-Thiede, J., Peter, M., Ushakova, L., Wernle, T., Bathellier, B., & Rumpel, S. (2013). Dynamics of dendritic spines in the mouse auditory cortex during memory formation and memory recall. *Proceedings of the National Academy of Sciences*, *110*(45), 18315–18320.

Mongillo, G., Rumpel, S., & Loewenstein, Y. (2017). Intrinsic volatility of synaptic connections—a challenge to the synaptic trace theory of memory. *Current opinion in neurobiology*, *46*, 7–13.

Mongillo, G., Rumpel, S., & Loewenstein, Y. (2018). Inhibitory connectivity defines the realm of excitatory plasticity. *Nature neuroscience*, *21*(10), 1463–1470.

Pavlov, P. I. (2010). Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex. *Annals of neurosciences*, *17*(3), 136.

Pérez-Garci, E., Larkum, M. E., & Nevian, T. (2013). Inhibition of dendritic ca2+ spikes by gabab receptors in cortical pyramidal neurons is mediated by a direct gi/o-$\beta\gamma$-subunit interaction with cav1 channels. *The Journal of physiology*, *591*(7), 1599–1612.

Stimberg, M., Brette, R., & Goodman, D. F. (2019, August). Brian 2, an intuitive and efficient neural simulator. *eLife*, *8*, e47314. doi: 10.7554/eLife.47314

Susman, L., Brenner, N., & Barak, O. (2019). Stable memory with unstable synapses. *Nature communications*, *10*(1), 1–9.

Takeuchi, T., Duszkiewicz, A. J., & Morris, R. G. (2014). The synaptic plasticity and memory hypothesis: encoding, storage and persistence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1633), 20130288.

Tønnesen, J., & Nägerl, U. V. (2016). Dendritic spines as tunable regulators of synaptic signals. *Frontiers in psychiatry*, *7*, 101.

Van Vreeswijk, C., & Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, *274*(5293), 1724–1726.

Wilmes, K. A., & Clopath, C. (2019). Inhibitory microcircuits for top-down plasticity of sensory representations. *Nature communications*, *10*(1), 1–10.

Yu, W., & Lu, B. (2012). Synapses and dendritic spines as pathogenic targets in alzheimer's disease. *Neural plasticity*, *2012*.

Zuo, Y., Lin, A., Chang, P., & Gan, W.-B. (2005). Development of long-term dendritic spine stability in diverse regions of cerebral cortex. *Neuron*, *46*(2), 181–189.