Emergence and long-term maintenance of modularity in plastic networks of spiking neurons

Raphaël Bergoin,[1,2,3,4,a)] Alessandro Torcini,[5] Gustavo Deco,[2,3,6] Mathias Quoy,[1,7] and Gorka Zamora-López[2,3]

[1)] *ETIS, UMR 8051, ENSEA, CY Cergy Paris Université, CNRS, 6 Av. du Ponceau, 95000 Cergy-Pontoise, France.*

[2)] *Center for Brain and Cognition, Pompeu Fabra University, Barcelona, Spain.*

[3)] *Department of Information and Communication Technologies, Pompeu Fabra University, Barcelona, Spain.*

[4)] *Institute of Neural Information Processing, Center for Molecular Neurobiology (ZMNH), University Medical Center Hamburg-Eppendorf (UKE), 20251 Hamburg, Germany.*

[5)] *Laboratoire de Physique Théorique et Modélisation, UMR 8089, CY Cergy Paris Université, CNRS, 2 Av. Adolphe Chauvin, 95032 Cergy-Pontoise, France.*

[6)] *Instituciò Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluis Companys 23, 08010 Barcelona, Spain.*

[7)] *IPAL, CNRS, 1 Fusionopolis Way #21-01 Connexis (South Tower), Singapore 138632, Singapore.*

In the last three decades it has become clear that cortical regions, interconnected via white-matter fibers, form a modular and hierarchical network. This type of organization, which has also been recognized at the microscopic level in the form of interconnected neural assemblies, is typically believed to support the coexistence of segregation (specialization) and integration (binding) of information. A fundamental open question is to understand how this complex structure can emerge in the brain. Here, we made a first step to address this question and propose that adaptation to various inputs could be the key driving mechanism for the formation of structural assemblies. To test this idea, we develop a model of quadratic integrate-and-fire spiking neurons, trained to stimuli targetting distinct sub-populations. The model is designed to satisfy several biologically plausible constraints: (i) the network contains excitatory and inhibitory neurons with Hebbian and anti-Hebbian spike-timing-dependent plasticity (STDP); and (ii) neither the neuronal activity nor the synaptic weights are frozen after the learning phase. Instead, the network is allowed to continue firing spontaneously while synaptic plasticity remains active. We find that only the combination of the two inhibitory STDP sub-populations allows for the formation of stable modular organization in the network, with each sub-population playing a distinct role. The Hebbian sub-population controls for the firing rate, while the anti-Hebbian mediates pattern selectivity. After the learning phase, the network activity settles into an asynchronous irregular resting-state—resembling the behaviour typically observed *in-vivo* in the cortex. This post-learning activity also displays spontaneous memory recalls, which are fundamental for the long-term consolidation of the learned memory items. The model here introduced can represent a starting point for the joint investigation of neural dynamics, connectivity and plasticity.

[a)] Electronic mail: raphael.bergoin@gmail.com

INTRODUCTION

Numerous studies have shown that brain's connectivity follows a modular and a hierarchical organization at different spatial and functional scales. From an operative point of view, this type of architecture facilitates the coexistence of segregation and integration of information[1–3]: neuronal circuits or brain regions associated to a specific function are densely connected with each other[4–6], while long-range connections and network hubs allow for the integration (or binding) of different information[2,7,8]. A crucial open question in brain connectivity is to understand how such modular and hierarchical organization could naturally emerge as a consequence of the functional needs of the nervous system, following biologically plausible constraints and mechanisms.

Synaptic plasticity appears as the natural mechanism that can drive the brain organization[9–11] since the relation between learning and network organization is inherently associated with the notion of semantic memory, where correlated information or functions share common structural substrates[12–14]. The encoding, consolidation and long-term preservation of memories require various neural mechanisms such as neurochemical changes and synaptic plasticity[15–17]. These changes are modulated by the dynamical activity of interconnected neurons via spike-timing-dependent plasticity (STDP)[18]. For example, sleep is characterized by sequences of hippocampal activations linked to replays of memories that promote memory consolidation[19–21]. During the wake, short and random events of partial synchronous activation of neural sub-networks are related both to the spontaneous recalls (or retrieval)[22] and to the consolidation[23–25] of learned memories. Although memory recalls have been reported in several studies of neural models[26–30], these are generally associated with working memory, do not necessarily occur spontaneously and their relationship with memory consolidation still remains unclear. Moreover, it also remains an open question how the typical cortical dynamics, that is asynchronous and irregular, despite being replicated in several theoretical and computational studies[31–34], can coexist with these spontaneous recalls.

In this paper, we investigate the formation of modular architecture in plastic networks of spiking neurons promoted by learning to selective stimuli, resembling input-driven segregation of (sensory) information. Furthermore, we explore the mechanisms underlying the long-term maintenance of these memories in a plastic environment. In particular, we aim at clarifying the role played by spontaneous memory recalls for the consolidation of the learned structures. Therefore, we introduce a network model of excitatory and inhibitory spiking neurons subject to STDP satisfying a few biologically relevant constraints. On one side, we assume that the network is composed by 80% (20%) excitatory (inhibitory) neurons, as commonly observed in the mammalian cortex[35]. On the other side, in contrast to common practice in learning models, whose neural activity and plasticity are artificially frozen once the training is terminated, we allow the neurons to continue firing spontaneously while the synaptic plasticity remains active after the learning phase.

In the proposed model we show that for the correct emergence and maintenance of modular structures, the network should be composed of two inhibitory neuronal sub-populations, each playing a distinct functional role. One sub-population subjected to Hebbian-STDP (controlling for the firing activity) and another population following anti-Hebbian-STDP (mediating memory selectivity). After the learning phase the model activity settles into an asynchronous irregular state, as typically observed during *in-vivo* recordings of the brain activity at rest[36]. This post-learning activity also exhibits the occurrence of transient events of partial synchrony, associated to the learned items. As we show, these spontaneous memory recalls are crucial for the long-term consolidation of the stored memories by promoting the reinforcement of the underlying connectivity. Moreover, we show that the maximal memory capacity of the model is related to the number of inhibitory neurons present in the network. Finally, if the model is trained to overlapping memories, it gives rise to hub neurons displaying mixed selectivity, which can represent the seeds for the hierarchical organization.

EMERGENCE OF MODULAR CONNECTIVITY DRIVEN BY LEARNING TO SELECTIVE STIMULI

We consider networks of excitatory and inhibitory Quadratic Integrate-and-Fire (QIF)[37] neurons, pulse coupled via exponentially decaying post-synaptic potentials and in the presence of STDP. The connections involving pre-synaptic excitatory neurons are subject to Hebbian STDP, while those involving pre-synaptic inhibitory neurons can be of two types: either Hebbian or anti-Hebbian STDP. The introduction of two different sub-populations of inhibitory neurons is justified on one side by the large variety of sub-classes of GABAergic cells present in the brain[38,39] and on another side by studies highlighting the fundamental role played by feedback[39,40] and feedforward inhibition[39,41] in neural dynamics.

First, we will investigate the necessary conditions for the emergence of modular assemblies induced by learning to selective stimuli. In particular, as a first example we will compare the efficacy of Hebbian and anti-Hebbian plasticity applied to inhibitory neurons in a network subject to two external stimuli. As a second aspect, we will investigate the role of spontaneous recalls in order to consolidate and maintain both the learned memory items and the underlying modular connectivity. To clarify this aspect we will perturb the synaptic connectivity matrix and we will examine to which extent spontaneous recalls are able to regenerate the original structure induced by the training. As a last point, we will generalize the model to account for an increasing

number of stored memories and for the presence of overlapping assemblies, thus showing that more complex architectures can develop and be maintained over time.

We begin by investigating the necessary conditions for the emergence of modular structure as induced by training to external stimuli. As a general set-up, we consider a heterogenous network of $N = 100$ QIF neurons, each receiving weak and independent Gaussian noise. In particular, neurons with labels in the interval $i \in [0 : 79]$ are excitatory and those labelled with $i \in [80 : 99]$ are inhibitory. We consider normally distributed excitabilities across neurons leading to spontaneous firing frequencies in the range $[0, 8]$ Hz. Furthermore, the synaptic weights w_{ij} from the j-th pre-synaptic neuron to the i-th post-synaptic one are subject to STDP, with weights bounded in the interval $w_{ij} \in [0, 1]$ for excitatory pre-synaptic neurons and in the interval $w_{ij} \in [-1, 0]$ for inhibitory ones. A detailed description of the network model as well as of the implementation of the STDP rules is reported in Methods.

The stimulation protocol consists in three stages as illustrated in Fig. 1A : an initial resting phase followed by a learning period and a final consolidation phase. The simulations are performed by ensuring the respect of some biologically realistic conditions: (i) the networks are allowed to continue their spontaneous activity after the learning phase; (ii) the adaptation of the synaptic weights is always active throughout the whole simulation i.e., before, during and after the learning.

Synaptic weights are randomly initialized with positive (negative) values for excitatory (inhibitory) synapses. The system is left to relax for five seconds in the absence of external stimuli. During this stage the network stabilizes into an asynchronous state with the neurons firing irregularly at low frequencies, as shown in the raster plots corresponding to the time interval $[0 : 5]$ seconds in Fig. 1. During the learning phase two independent stimuli are applied, each targetting a different (non-overlapping) neuronal population, as shown in Fig. 1A. This is done in order to mimic the segregation of (sensory) information being projected to nearby but separate neuronal populations, and to study the role of this segregation for the emergence of modular neuronal architectures. In particular, one stimulus targets population P_1 consisting of the first half of excitatory and inhibitory neurons labelled as $\mathrm{E}_1 := \{i \in [0 : 39]\}$ and $I_1 := \{i \in [80 : 89]\}$. The second stimulus targets population P_2 consisting of the second half of excitatory and inhibitory neurons, $E_2 := \{i \in [40 : 79]\}$ and $I_2 := \{i \in [90 : 99]\}$.

The two stimuli are applied for a time duration of one second, randomly alternating between the two populations P_1 and P_2. During each stimulation period, a constant external positive current is applied to the selected population for 800 ms. The stimulus triggers the target neurons to fire at about 50 Hz. After 800 ms the external current is turned off and the network is left to relax for other 200 ms in order to prevent temporal correlations when alternating between stimulated populations. This protocol is repeated 35 times for a total of 35 seconds. Once the training phase is finished, the network is allowed to evolve freely for 20 seconds in the absence of stimuli. During this phase, however, the synaptic adaptation remains active thus affecting the stabilization of the learned patterns.

Role of Hebbian and anti-Hebbian learning for the emergence of modular connectivity

To explore the role played by the inhibitory neurons in the learning process, we study three different scenarios: (a) all the inhibitory neurons follow anti-Hebbian STDP, (b) all the inhibitory neurons follow Hebbian learning, and (c) a mixed situation where 50% are Hebbian and 50% anti-Hebbian.

(a) *Anti-Hebbian inhibition.* In the case with only anti-Hebbian inhibitory plasticity, the network develops into a *winner-takes-all* architecture promoting the competition of the two sub-populations P_1 or P_2 on the other one, as shown in Fig. 1B. In order to describe the mechanism leading to this architecture, imagine that population P_1 is stimulated, consequently all its neurons become active and fire with high frequency. Therefore all the connections $\mathrm{E}_1 \to \{\mathrm{E}_1, \mathrm{I}_1\}$ are reinforced due to the Hebbian nature of the pre-synaptic excitatory neurons. At the same time, all the inhibitory synapses $\mathrm{I}_1 \to \{\mathrm{E}_1, \mathrm{I}_1\}$ are weakened because they are anti-Hebbian. During the stimulation of P_1, the neurons in populations P_1 and P_2 are far from being mutually synchronized, therefore the connections $\mathrm{E}_1 \to \{\mathrm{E}_2, \mathrm{I}_2\}$ are weakened while $\mathrm{I}_1 \to \{\mathrm{E}_2, \mathrm{I}_2\}$ reinforce. The random alternation of the stimuli to populations P_1 and P_2 induces a gradual emergence of a modular structure, as visible in the connectivity (weight) matrices at times $t = 20$ sec and $t = 40$ sec in Fig. 1B.

The resulting architecture promotes the competition between the two excitatory populations E_1 and E_2 and the alternating prevalence of one of them. For example, stimulation to E_2 would activate the companion inhibitory neurons through the *feedforward* connections $(\mathrm{E}_2 \to \mathrm{I}_2)$ which, in turn, would shut down all neurons in P_1 via the strong $\mathrm{I}_2 \to \{\mathrm{E}_1, \mathrm{I}_1\}$ connections. This is visible in the raster plot in Fig. 1B : as the training progresses ($t = 5 - 40$ sec), the neurons of the stimulated population fire at high frequency but the neurons of the non-stimulated one are silenced.

(b) *Hebbian inhibition.* In the presence of only Hebbian inhibitory neurons the network also develops a modular organization, however in this case the two populations become disconnected from each other, Fig. 1C. Analogously to the previous case, the stimulation of one population (e.g., P_1) results in the strengthening of the intra excitatory connections (increase of $\mathrm{E}_1 \to \{\mathrm{E}_1, \mathrm{I}_1\}$

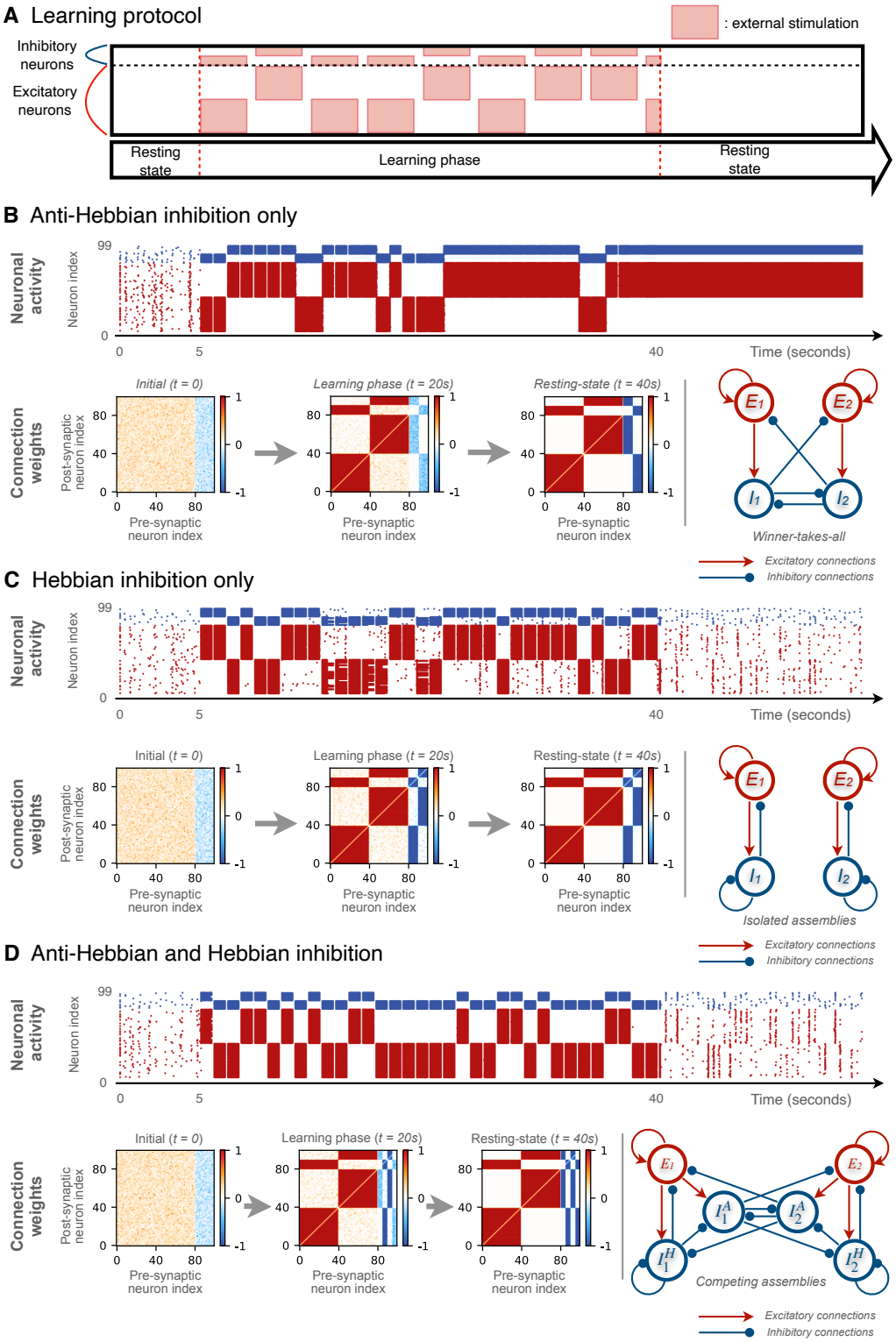


FIG. 1. **Emergence of modular connectivity by learning to two selective stimuli in spiking networks. (A)** Experimental protocol consists of the stimulation of two non-overlapping neuronal populations of QIF neurons with plastic synapses. Networks are made of 80% of excitatory and 20% of inhibitory neurons. Stimuli are presented in temporal alternation. Results of the performed numerical experiments are reported for a network with all anti-Hebbian inhibitory neurons **(B)**; with all Hebbian inhibitory neurons **(C)**; with 50% anti-Hebbian and 50% Hebbian inhibitory neurons **(D)**. Raster plots display the firing times of excitatory (red dots) and inhibitory (blue dots) neurons during the simulations. Matrices represent the temporal evolution of the connection weights at different times: $t = 0$s (random initialization of the weights), $t = 20$s (middle of the learning phase) and $t = 40$s (end of the learning phase). The color denotes if the connection is excitatory (red), inhibitory (blue) or absent (white) and the color gradation the strength of the synaptic weight. The final configuration of the connection weights is shown schematically on the right in each case.

synaptic weights) and the weakening of the inter excitatory connections across different populations (decrease of $E_1 \rightarrow \{E_2, I_2\}$ weights). However, the Hebbian inhibitory plasticity now induces the strengthening of the internal inhibitory connections $I_1 \rightarrow \{E_1, I_1\}$ and the weakening of the inhibitory synapses across populations. The disconnection of the two populations happens gradually during the learning phase. As shown in the raster plot in Fig. 1C, during the initial stimulation epochs, the stimulated population shuts down the activity of the non-stimulated population. But as the training proceeds, the two populations detach one from the other and the non-stimulated population begins to display a low firing resting-state activity.

(*c*) *Anti-Hebbian and Hebbian inhibition.* In the last case with mixed anti-Hebbian and Hebbian inhibitory neurons, the network also develops two modules, but, as shown in Fig. 1D, the resulting connectivity is a combination of the two configurations observed in the previous cases. The Hebbian neurons form self $E-I$ loops within each population (i.e. $E_1 - I_1^H$ and $E_2 - I_2^H$). This internal feedback inhibition avoids that the excitatory neurons fire at too large rates. Meanwhile, the anti-Hebbian inhibitory neurons form lateral, feedforward connections which shut down the firing of the other population, in other terms the sub-population I_1^A (I_2^A) inhibits all neurons in P_2 (P_1).

Resting-state network dynamics after learning

So far, we have shown that selective stimulation to distinct populations consistently gives rise to modular networks and that the resulting configuration depends on the type of plasticity affecting the inhibitory neurons. However, the relevance or plausibility of a learning model should also require that the network exhibits a biologically meaningful dynamical behaviour after the training.

(*a*) *Anti-Hebbian inhibition.* In the simulations with anti-Hebbian inhibitory STDP, the neuronal activity in the post-learning stage ($t > 40$ sec) is dominated by one of the two populations. In Fig. 1B, P_2 is active and P_1 is silent. But this randomly changes over realizations. In general, once the training stage is finished, all neurons start to fire spontaneously driven by the background Gaussian noise. The excitatory sub-population that attains a larger level of internal activity earlier is the one that wins. Due to the lack of internal feedback inhibition ($I_1 \nrightarrow E_1$ or $I_2 \nrightarrow E_2$) the winning excitatory sub-population fires rapidly and the companion inhibitory neurons (strongly stimulated) suppress the activity of the other population: e.g. if the neurons in E_2 fire faster than those in E_1 the sub-population I_2 inhibits both E_1 and I_1.

(*b*) *Hebbian inhibition.* In the case with Hebbian inhibitory neurons, we found that the two populations P_1 and P_2 become independent, each forming a *feedback E–I loop*. The presence of the internal feedback inhibition allows the two populations to settle into a resting-state behaviour after the training that is characterized by a low firing frequency of approximately 1.0 Hz, as visible in the raster plot of Fig 1C, for $t > 40$ sec. Interestingly, brief events of internal partial synchrony of the two populations are also observed. While in the case with only anti-Hebbian inhibition such a synchronized event would trigger the excitatory neurons to permanently increase their firing, here the presence of the internal feedback inhibition ($I_1 \rightarrow E_1$ and $I_2 \rightarrow E_2$) avoids the constant synchronization of excitatory neurons, while it keeps their activity at low frequency.

(*c*) *Anti-Hebbian and Hebbian inhibition.* The post-learning behaviour in the mixed scenario is very similar to the Hebbian case, as shown in Fig. 1D. However, in this case populations P_1 and P_2 are not independent but they inhibit each other. As a consequence, an event of partial synchrony occurring in one of the two modules temporarily shuts down the other module, avoiding their mutual synchrony. Given that the plasticity remains active during this post-learning spontaneous activity, the connection weights continue to be updated. As we will discuss in the following, the occurrence of these spontaneous synchronization events plays a relevant role in the maintenance of the learned memories.

In summary, despite modular organization of the connectivity was found to emerge in the three considered scenarios, only the combination of anti-Hebbian and Hebbian inhibitory neurons resulted in a network satisfying all the desired biologically plausible properties. Anti-Hebbian inhibition alone led to a network with unrealistic post-learning dynamics. Hebbian inhibition alone gives rise to a biologically meaningful resting-state behaviour but at the cost of splitting the network into two disconnected populations. Therefore, in the following we will limit to consider the mixed scenario, with both anti-Hebbian and Hebbian inhibition.

Supplementary Secs. S1, S2 and S3 repeat the numerical experiments presented so far but for modified protocols, corroborating the robustness of the results. The examined situations include cases where (i) some of the neurons are not trained, (ii) at every stimulation epoch random subsets of neurons are targetted, and (iii) the intensity of the stimulation current is randomly fluctuating.

SPONTANEOUS RECALLS SUPPORT CONSOLIDATION AND LONG-TERM MAINTENANCE OF MEMORIES

Given that in our model plasticity is not frozen after the training—but it is left active afterwards—we now investigate the potential role of the spontaneous events

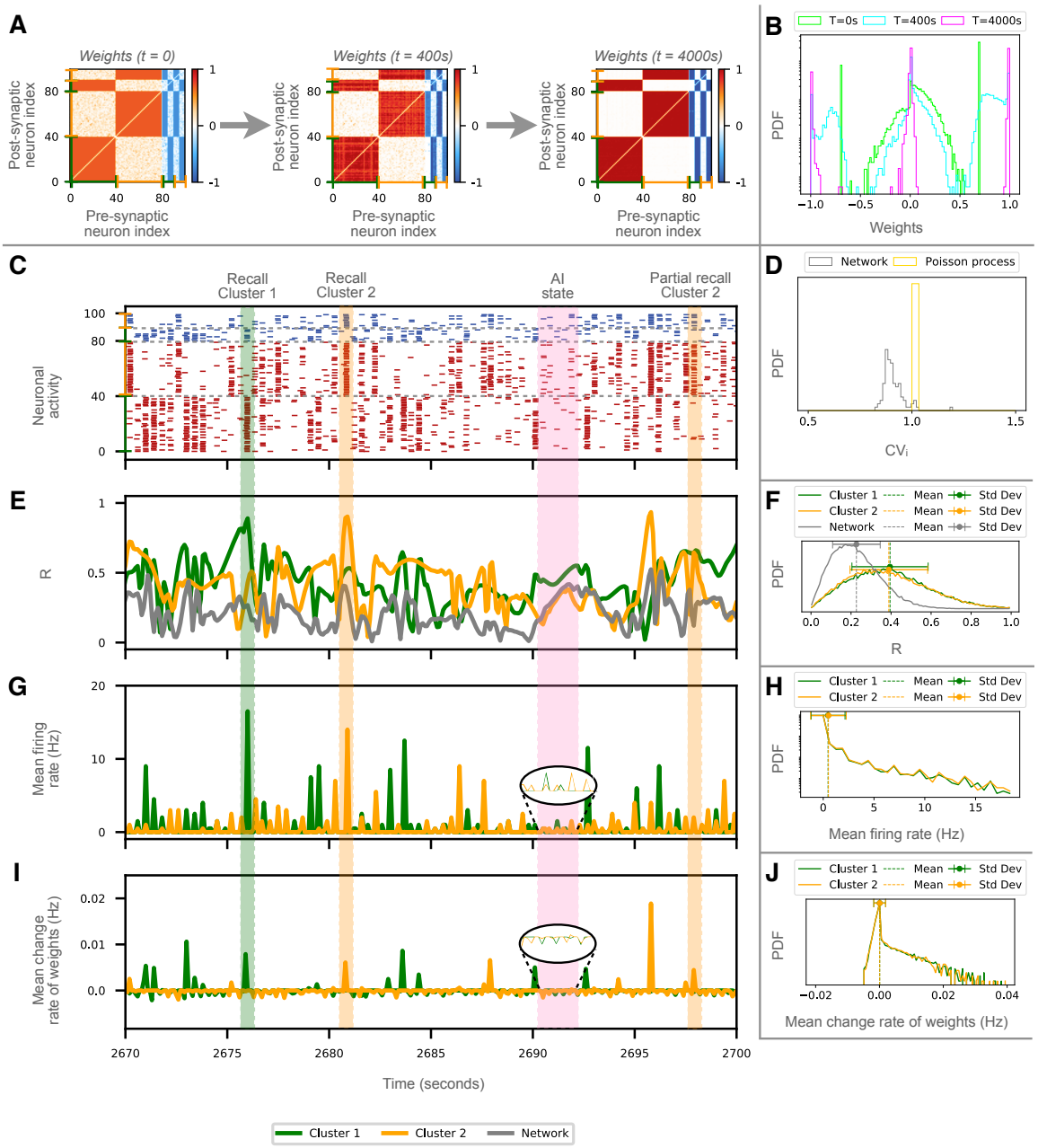


FIG. 2. **Consolidation of imperfectly learned memories.** **(A)** Temporal evolution of the connectivity matrix during spontaneous activity in the absence of stimulation. Initial connectivity with two unfinished modules at $t = 0$ is reinforced over time. The excitatory (inhibitory) connections are marked as in Fig. 1. **(B)** Evolution of the distribution of link weights (probability density functions, PDFs, in a linear-logarithmic scale) in the connectivity matrices at $t = 0$s (light green), $t = 400$s (cyan) and $t = 4000$s (magenta). **(C-J)** Evolution of the network activity and various metrics in absence of stimulation for a sample of 30 seconds. Some spontaneous recalls are highlighted, for population P_1 (green shade) and P_2 (orange shade). Pink shadow marks an epoch of asynchronous irregular firing activity without recalls. **(C)** Raster plot with excitatory (inhibitory) neurons marked in red (blue). **(D)** PDFs of the coefficient of variations CV_i for all the neurons during the entire simulation (grey) and for a homogeneous Poisson process (yellow). **(E)** Instantaneous Kuramoto order parameters R for the networks (gray) and for neurons in population P_1 (green) and P_2 (orange), and their corresponding PDFs over the entire simulation **(F)**. **(G)** Temporal evolution of the mean firing rates for populations P_1 and P_2, and **(H)** their corresponding frequency distributions (in linear-logarithmic scale) showing a peak at 2 Hz and long time tail. **(I)** Instantaneous change rates of synaptic weights in both populations P_1 and P_2 over time and their PDF (in linear-logarithmic scale) **(J)** over the entire simulation time, showing the prevalence of positive weight changes (reinforcement) with respect to negative ones (depression).

observed during the post-learning resting-state for the consolidation and maintenance of the memories. We will refer to these events as *memory recalls*. First, we will show that the recalls facilitate the completion of imperfectly learned memories and, second, we will study their role for the regeneration of memories that have been partially lost.

Memory consolidation

In order to mimic a hypothetical scenario in which the training stage would stop before completion, we prepared an initial weight matrix representing an imperfectly learned connectivity structure induced by two non overlapping stimuli. An example of this weight matrix is shown at $t = 0$ in Fig. 2A. Specifically, the intra-modular synaptic weights are set to $w_E = 0.7$ for excitatory and to $w_I = -0.7$ for inhibitory connections. The inter-modular excitatory and inhibitory connections are chosen randomly from a flat distribution with mean 0 and standard deviation 0.15.

Once fixed the initial weight matrix, the activity of the network is then left evolve spontaneously, driven only by the background Gaussian noise without any external stimulation. However, we observe that the modular organization of the network is reinforced: the intra-modular synaptic weights of the two populations are strengthened and the inter-modular synapses are weakened, as shown by the weight matrices displayed in Fig. 2A and the corresponding weight distributions in Fig. 2B at $t = 400$ sec and $t = 4000$ sec.

The mechanism allowing for the completion of the connectivity pattern is summarised in Figs. 2C-J. Panel C shows a 30 seconds sample of activity of the network. This is characterised by an asynchronous irregular evolution with low firing spontaneous activity joined to occasional spontaneous recalls occurring at random times. This observation is confirmed by the fact that the background activity is characterized by a distribution of the coefficient of variations CV_i of the neurons in the interval $[0.8 : 1.0]$ (panel D, grey distribution) lying close to a Poisson process (in yellow), thus resembling the irregular activity observed in the cortex *in vivo*[42]. The synchronization order parameter of the network R fluctuates in time with values around 0.2 (panel F, gray distribution), meaning that the network is far from being synchronized. As a matter of fact, for an asynchronous network of $N = 100$ neurons, due to finite size fluctuations the expected value of the order parameter will be not far from the one here measured, namely $R \simeq 1/\sqrt{N} = 0.1$. Finally, the distribution of firing rates displays a main peak around 2 Hz and an exponential tail reaching at most 20 Hz (panels G and H).

During spontaneous recalls, the transient increase in synchrony and firing rates of a subset of the neurons (in one of the two clusters) triggers the activation of their synaptic adaptation (panel I), reshaping the synaptic weights and completing the formation of the modular patterns. The specific neurons participating in a recall changes from event to event but they tend to involve a majority of neurons of either population P_1 or P_2; see for example the events highlighted by green and orange shadows. The recalls coincide with transient peaks of the order parameter and of the mean firing rates of the populations involved (panels E and G, green lines). Consequently, P_1 and P_2 are internally more synchronized than the network average, with their order parameters fluctuating around 0.4 (panel F), and a broad distribution of firing rates (panel H). The similarity between the PDFs for populations P_1 and P_2 evidences that recalls occur—on average—with the same probability in both populations. Over time, the net effect of recalls is that the intra-modular synapses between neurons of the same population become reinforced while the inter-modular connections are weakened.

At this point, we remind that our model incorporates a natural, slow tendency to forget the value of the synaptic weights. Therefore, the question arises of how often should spontaneous recalls happen—and how strong should they be—in order to avoid the forgetting of the learned memories. In the example considered here, the weights between the neurons of a population undergo a natural depression during epochs of the resting-activity without recalls (pink shadow in Fig. 2C). For the used parameters, a forgetting term of $f = 0.1$ seconds and an average spontaneous firing rate of 2 Hz, we estimate that a period of $\simeq 12$ seconds of asynchronous firing would be required to forget the contribution of a single recall event (for more details on the calculus see Supplementary Sec. S4). The dominance of the reinforcement during recalls with respect to depression during the asynchronous irregular epochs is confirmed by the asymmetry in the distribution of positive weight change rates over negative ones, Fig. 2J.

In summary, we can conclude that, in our model, spontaneous recalls happening during resting-state activity allow for the consolidation of imperfectly learned memories as well as for their long-term maintenance against natural forgetting.

Regeneration of damaged excitatory and inhibitory connectivity

Empirical observations have shown that excitatory synapses are more volatile than inhibitory ones[43,44], leading to the hypothesis that reorganization of excitatory connections might be associated with short-term plasticity while inhibitory adaptation might support the long-term maintenance of those memories[45–48]. We now test the plausibility of this idea in our model.

We consider a first case, where the weights of the excitatory synapses (w_{EE} and w_{IE}) are initially randomized between 0 and 1, but the values of the weights of the inhibitory synapses coincide with those obtained as after

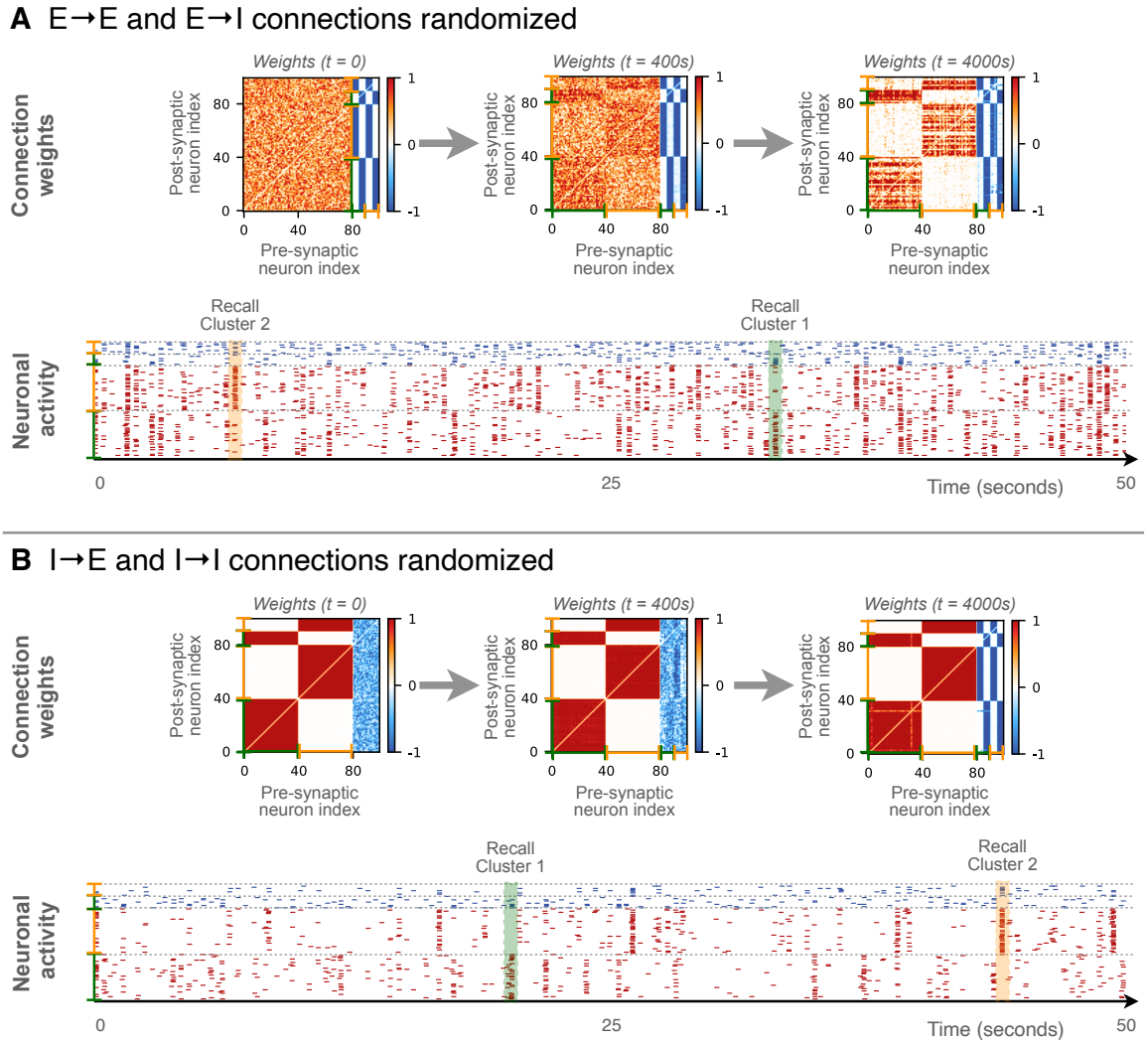


FIG. 3. **Recovery of damaged modular structure by spontaneous recalls.**
A Recovery experiment starting from randomized excitatory **(A)** or inhibitory **(B)** synaptic connections. In both cases, we study the partial recovery of the original modular organization of the synaptic weights over time. This recovery is mediated in both cases by the emergence of spontaneous recalls, —transient events of partial synchronization between neurons associated with one of the two originally stored memories, highlighted in green for cluster 1 and orange for cluster 2.

the training stage, see the weight matrix for $t = 0$ in Fig. 3A. The network is then let to evolve spontaneously for 4000 seconds, driven only by the background Gaussian noise. At the end of the simulation, the excitatory connections have partly recovered the original modular pattern (see the weight matrix for $t = 4000s$ in Fig. 3A). This recovery is clearly mediated by the spontaneous recalls as illustrated in the raster plot in Fig. 3A. Whenever a sufficiently large group of excitatory neurons (associated to one of the previously learned memories) fire in a short time window, this event activates the inhibitory neurons associated to the corresponding memory items and triggers a partial recall via the feedback and feedforward mechanism previously explained. Consequently, the excitatory synapses involved in the recalled memory are reinforced while all other connections are weakened.

If the experiment is repeated but randomizing only the inhibitory weights (w_{EI} and w_{II}, the corresponding weight matrix is shown at $t = 0$ in Fig. 3B), the memorized pattern is almost fully recovered. These results show that the conservation of either the excitatory or the inhibitory connections allows for the recovery of the memories in our model, although the recovery of inhibitory synapses seems more robust.

The main reason for this difference is related to the fact that when the excitatory synapses are preserved, the network displays a more sparse activity than with initially randomized excitatory connections. Since in the latter case there are more interactions among excitatory neurons, this gives rise to a sort of background activity involving most of the excitatory neurons. Therefore one finally obtains a weight matrix with a more

disordered structure (see in Fig. 3A weight matrix at $t = 4000$ s). Another reason of this difference is related to the asymmetric 80-20% E–I ratio, leading to notably different spontaneous firing frequencies—and recall probabilities—when either the excitatory or the inhibitory synapses are randomized.

GENERALIZATION OF THE MODEL TO MORE COMPLEX SITUATIONS

So far, we have illustrated the results for networks trained with two stimuli, applied at separate neuronal sub-populations. We will now show that the model can be generalized. Firstly, the model will be trained to an increasing number of memories, this in order to estimate its memory capacity. Secondly, the model will be trained to overlapping memory items, a process that will lead to the emergence of hub neurons.

Memory capacity of the model

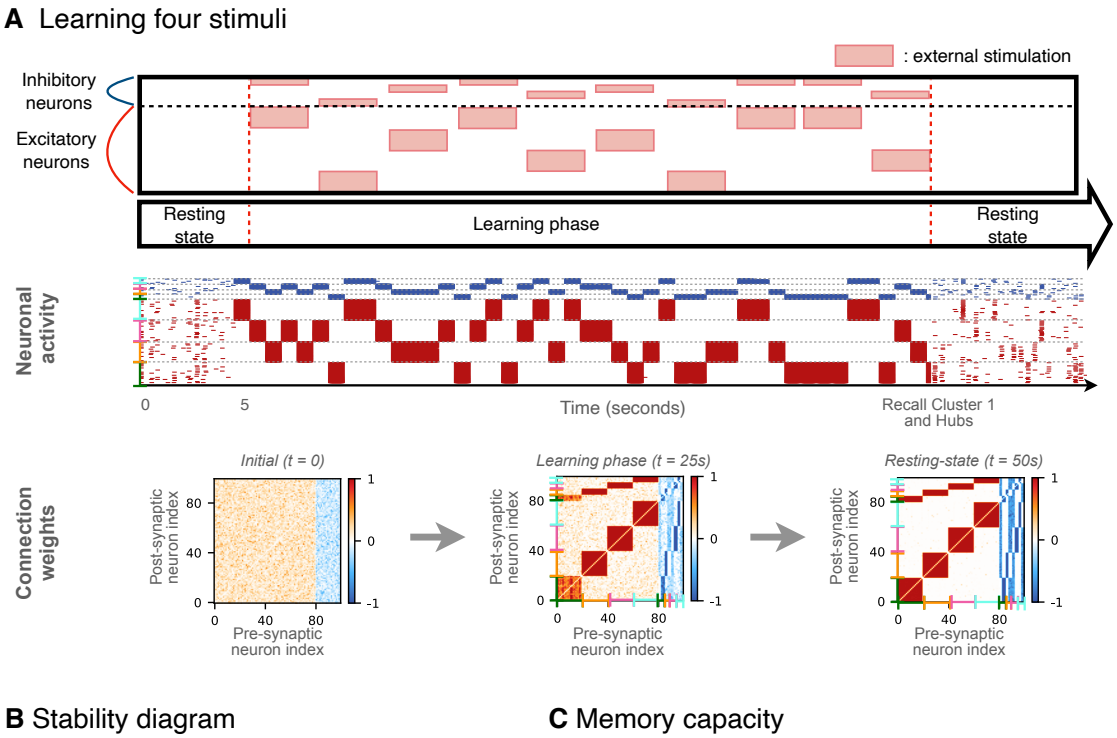
Firstly, we generalise the experimental protocol reported in Fig. 1D by training the network to $M = 4$ non-overlapping stimuli, as shown in Fig. 4A. We observe the same qualitative results as in the case with 2 stimuli, the only difference is that now four modules emerge in the connectivity matrix. Also in the present situation, the four neuronal assemblies display spontaneous recalls during the post-learning resting-state phase. A minor difference is that now the 20 inhibitory neurons split into 8 sub-populations, each made of $2-3$ anti-Hebbian and Hebbian neurons. This opens the question of the memory capacity of the network, which seems limited by the number of available inhibitory neurons.

In order to measure the memory capacity, we examine the potential of the network to maintain an increasing number of memories while also varying the ratio of excitatory to inhibitory neurons. Specifically, connectivity matrices were initialized containing an arbitrary number M of memory items (modules), from $M = 0$ to $M = 100$, and by varying the number of inhibitory neurons from $N_I = 0$ to $N_I = 100$. The size of the network is maintained constant to $N = 100$, hence $N_E = 100 - N_I$. A network storing M memory items is considered stable if (i) it displays asynchronous irregular firing and (ii) all the memory modules exhibit independent spontaneous recalls. A network with M memories is considered unstable if at least one of the modules does not exhibit recalls and therefore disappears in the long term.

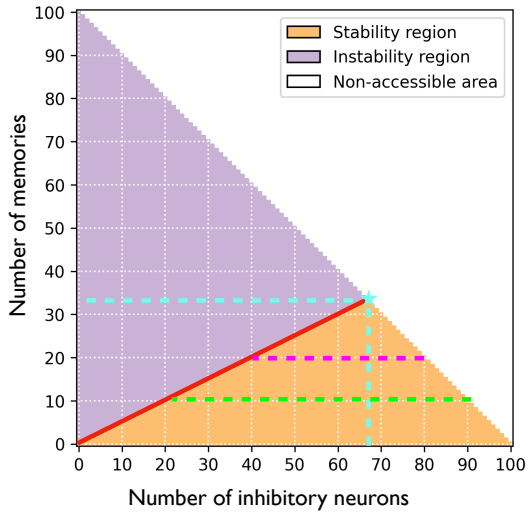
The results are summarised in the stability diagram in Fig. 4B. The red line separating the stable and the unstable regimes evidences that at least $2M$ inhibitory neurons are needed to hold M memories. This result implies that the network can learn and stabilize a maximum of $M^\star = N/3 = 33$ memories (marked by a cyan star in Fig. 4B). In particular, this maximal capacity corresponds to the case with $N_I = 66$ inhibitory neurons, 33 of them anti-Hebbian and 33 Hebbian. In sub-optimal scenarios, we find that for a given number of memory items $M < M^\star$, different E/I ratios can guarantee the long-term maintenance of the memories (connectivity patterns). In fact, we have found that the minimal condition for a memory item to be robustly encoded is that the memory should be associated to at least one excitatory neuron, one Hebbian inhibitory and one anti-Hebbian inhibitory neuron. For example, to store $M = 20$ memories (magenta dashed line in Fig. 4B) the network needs to contain at least 20 excitatory neurons and 40 inhibitory neurons (20 Hebbian and 20 anti-Hebbian). All the possible E/I ratios respecting these constraints can give rise to configurations in which the $M = 20$ memories are robustly stored. This amounts to have the N_E/N_I ratio ranging from $N_E/N_I = 20/80 = 0.25$ to $60/40 = 1.5$. For the case with $M = 10$ memories (green dashed line) the interval of acceptable ratios widens from $N_E/N_I = 10/90 = 0.111$ to $80/20 = 4$.

The analysis of a few test cases reported in Supplementary Sec. S5 for $M = 4$ memory items confirms the validity of the upper limit $M = 2N_I$. In particular, we studied the following situation: (i) each memory pattern is associated to one Hebbian inhibitory and one anti-Hebbian inhibitory neuron, (ii) one memory pattern is missing an anti-Hebbian inhibitory neuron, and (iii) one memory item is missing one Hebbian inhibitory neuron.

The flexible relation between memory capacity and E/I ratios found in our model invites to question why would nature favour the specific E/I ratios observed in the mammalian brains, and whether these ratios might be related to their memory storage. To explore this question we computed the hypothetical memory capacities of the human and mouse brains, extrapolating the predictions from our model. Moreover, given that different parts of the brain are made of distinct number of neurons and E/I ratios, we extended the calculations to distinguish the cortex, the cerebellum and a few subcortical regions. The results are shown in Fig. 4C. The 80/20 excitatory to inhibitory ratio—which is mostly representative for the cerebral and cerebellar cortices—implies a hypothetical memory capacity of $M = 0.1N$, well below the theoretical maximal capacity of $M^\star = 0.33N$. Still, given that the cortex and the cerebellum contain most of the neurons in the brain, in absolute terms this would imply that the human cortex and cerebellum have memory capacities of $M^{ctx} \approx 1.6 \times 10^9$ and $M^{cer} \approx 7.0 \times 10^9$ respectively, and $M^{ctx} \approx 2.2 \times 10^6$ and $M^{cer} \approx 10.5 \times 10^6$ in mice. Rather surprisingly, we find the lowest memory capacity in the Hippocampus, with $M = 0.050N$ and $M = 0.025N$ in humans and mice due to a reduced fraction of inhibitory neurons. The amygdala lies among the regions with largest memory capacity, $M = 0.15N$ and $M = 0.2N$ for humans and mice. This particularity of the amygdala could be related to the fact that this area is involved in the formation of emotional mem-



B Stability diagram



C Memory capacity

Brain area	Species	Number of neurons N	E/I ratio	Hypothetical memory capacity
Cortex	Human	16B	80/20	0.1N
	Mouse	22M	80/20	0.1N
Cerebellum	Human	69B	80/20	0.1N
	Mouse	42M	90/10	0.05N
Hippocampus	Human	48M	90/10	0.05N
	Mouse	3M	95/5	0.025N
Amygdala	Human	13M	70/30	0.15N
	Mouse	117K	60/40	0.2N
Thalamus	Human	20M	60/40	0.2N
	Mouse	1M	80/20	0.1N
Basal ganglia	Human	415M	10/90	0.1N
	Mouse	2M	20/80	0.2N
Striatum	Human	105M	5/95	0.05N
	Mouse	1.7M	20/80	0.2N
Globus pallidus	Human	711K	10/90	0.1N
	Mouse	250K	10/90	0.1N

FIG. 4. **Memory capacity of the model. (A)** Training the network to $M = 4$ non-overlapping stimuli. The green, orange, pink and cyan brackets represent clusters 1, 2, 3 and 4. Results are qualitatively the same as for the example with two memory items. **(B)** Stability diagram of the model for a network of $N = 100$ neurons. The red line $M = N_I/2$ separating the stability (orange) and instability (purple) regions, represents an upper limit for the number of inhibitory neurons ($N_I = 2M$) needed to maintain M independent memory items. The number of excitatory neurons in each realization is $N_E = 100 - N_I$ delimiting the non-accessible areas (white region) of the diagram. The star symbol marks the maximal memory capacity of the network, corresponding to $M^\star = 33$ items with $N_I^\star = 66$ inhibitory neurons (cyan lines, 33 anti-Hebbian and 33 Hebbian). Magenta and light-green lines highlight the ranges of number of inhibitory neurons (and therefore of the $N_I/N_E = N_I/(100-N_I)$ ratios) that allow the stabilization of $M = 20$ and $M = 10$ memory items, respectively. **(C)** Table comparing the number of neurons and E/I ratios of different parts of the human and mouse brain and their hypothetical memory capacity obtained by extrapolating the results of our model.

ories[49]. Nevertheless, the amygdala contains few neurons as compared to other regions, hence its absolute capacity would be definitely smaller.

Training to overlapping memories

We finish the investigation of the model by considering the case in which two stimuli entrain an overlapping set of excitatory neurons and explore the possibility that these could encode for more than one stimulus. This would correspond to a "mixed selectivity" scenario as it is often found in neurons of the prefrontal cortex[50,51]. During the training stage, populations P_1 and P_2 are now allowed to share eight excitatory neurons, see Fig. 5A. Furthermore, stimulation to P_1 and P_2 is strictly alternated in order to facilitate the formation of the connections, instead of selecting randomly between P_1 and P_2, as done in the previous simulations.

The results in Fig. 5B are similar to the previous ones only that besides the formation of two clusters, now a set of structural *hub neurons* emerge in the connectivity matrix. These excitatory hubs are strongly connected to both modules. They are weakly affected by the feedforward anti-Hebbian inhibition (since these hubs belong to both populations), but are notably affected by the feedback Hebbian inhibition of the two populations. Figure 5C shows a schematic diagram of the initial and of the emergent connectivities. Regarding the post-learning resting phase, see raster plot of Fig. 5B, the network displays again a stable low-frequency asynchronous firing activity but with richer spatio-temporal patterns than in the previous examples. The spontaneous recalls are more varied now with events composed of: (i) synchronous spiking of one population including the hubs (event shaded in green), (ii) synchronous spiking of the population without the hubs (event shaded in orange), and (iii) synchronous spiking of the hub neurons alone (event shaded in pink). Supplementary Sec. S6 shows similar results for a network trained to $M = 4$ overlapping stimuli.

In conclusion, the introduced model can be generalized to account for neurons that admit persistent mixed selectivity. This shows that, besides modular organization, the model can also incorporate centralised hierarchical organization which is a necessary ingredient for integration. In previous analyses based on phase oscillatory neuronal models, the mixed selectivity was only a transient phenomenon disappearing on the long run[52].

SUMMARY AND DISCUSSION

The architecture of brain's connectivity, the dynamics of neural populations and the learning capacities of neural networks are three fundamental topics of computational neuroscience. However, these topics are too often investigated separately one from another. Our aim was to develop a learning model that exhibits the emergence of relevant architectural features but which remains *alive* in all phases. Meaning that neither the dynamical activity nor the plasticity evolution were frozen once the training is finalized. Indeed, we have been able to develop a model displaying realistic firing patterns after the learning phase while the synaptic plasticity remains active—as it is the case in the brain *in-vivo*.

To achieve these objectives, we have introduced a network of excitatory and inhibitory QIF neurons with plastic synapses. By targetting stimuli to distinct subpopulations—mimicking the segregated projections of different features into early sensory layers—the network developed a stable modular connectivity. Moreover, by allowing the stimuli to target also neurons in overlapping sub-populations, we observed the emergence of hub-like neurons displaying mixed selectivity. Furthermore, the learned connectivity patterns were robust against catastrophic forgetting over time despite plasticity remained active after the learning phase. As a matter of fact, we found that the key factor to guarantee the reinforcement and the long-term maintenance of the memories was the spontaneous occurrence of transient memory recalls in the post-learning neural dynamics. These events occurred at random times on top of an otherwise asynchronous and irregular background neuronal activity. In particular, the recalls acted as short, punctuated boosts to the synaptic adaptation which allowed the memory patterns to persist against natural forgetting.

The role of feedback and feedforward inhibition

Although inhibitory neurons account for only around 20% of neural cells in the human brain, GABAergic interneurons represent the majority of sub-classes present in the brain[38]. There is substantial evidence that inhibitory cells are regulated by different types of synaptic plasticity[53,54] promoting, for example, the creation of feedforward[39,41] and feedback inhibitory circuits[39,40]. However, whether a direct relation exists between the type of plasticity and the various functionalities of inhibitory neurons is yet largely to be clarified[55]. Accordingly, we opted for exploring the behaviour of the model under Hebbian and anti-Hebbian inhibition. We found that for our model to fulfill all the expected requirements, neuronal populations of both Hebbian and anti-Hebbian inhibitory neurons were needed, with each population developing a differential function (see Fig. 1D).

During the training, the Hebbian inhibitory neurons formed internal feedback loops (with the excitatory neurons targetted by the same stimulus, Figs. 1C,D) while the anti-Hebbian neurons formed lateral feedforward connections (across the populations that were targetted by different stimuli, Figs. 1B,D). As a consequence, the Hebbian feedback inhibition turned responsible for controlling the firing rate of the populations, which is crucial for the stabilization of the network activity into a realistic

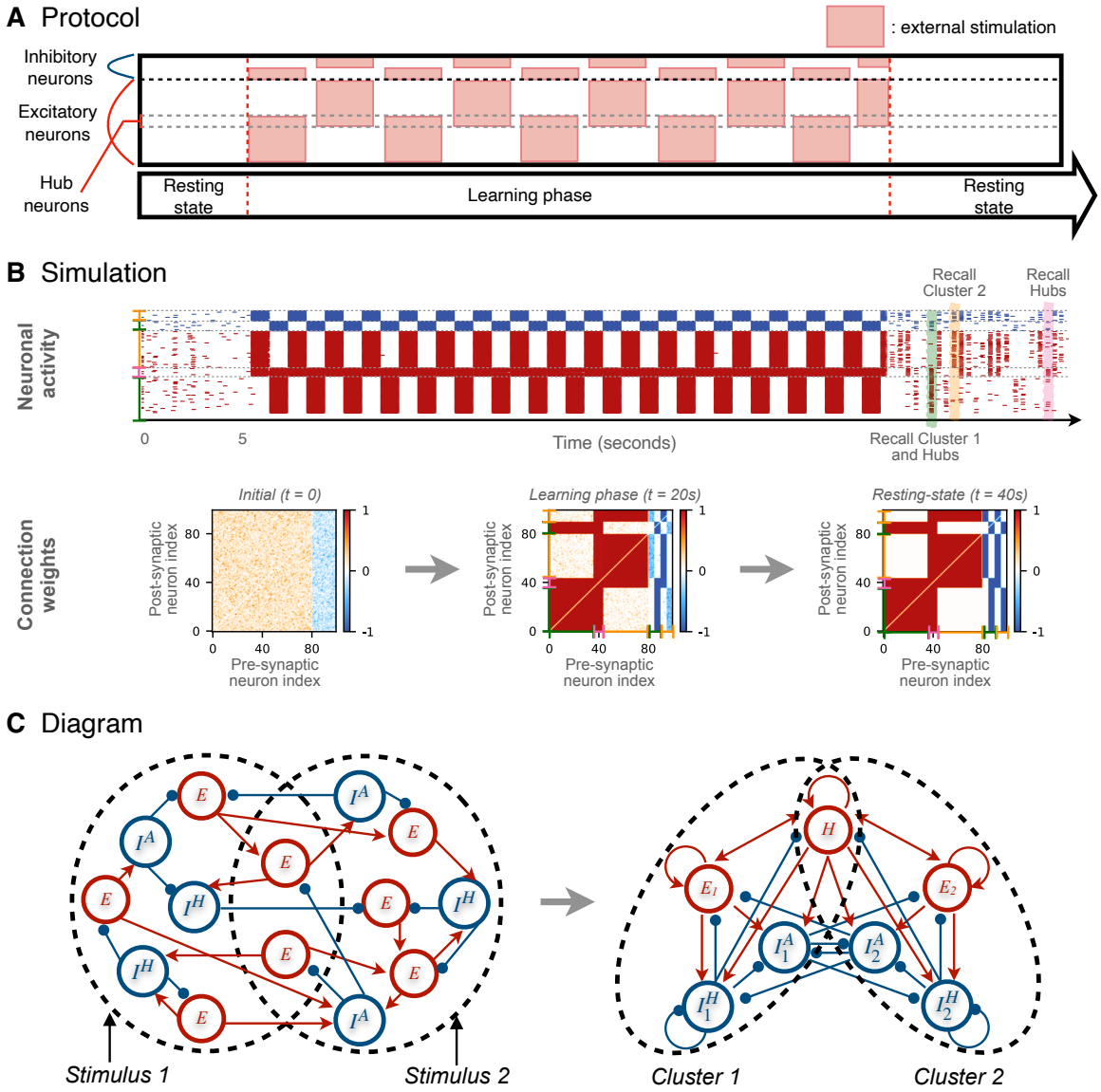


FIG. 5. **Training the network to overlapping stimuli.** **(A)** Experimental protocol for a network of $N = 100$ neurons trained to $M = 2$ stimuli that share 8 excitatory neurons. **(B)** Simulation and learning results. Connectivity matrices show the evolution of the synaptic weights leading to the emergence of two modules which overlap over 8 hub neurons. The raster plot shows the simulated neuronal activity of the network during the initial resting phase, the learning stage and the post-learning period. The activity during the post-learning phase is characterized by a variety of spontaneous recall events involving P_1 neurons and the hubs (green shadow), P_2 neurons without the hubs (orange shadow) and the hubs alone (pink shadow). **(C)** Schematic diagrams representing the formation of the synaptic connectivity with two stable populations of neurons and overlapping hubs. Each population is composed at least of a population of excitatory neurons E_x and a population of hub neurons H (in red), one population of Hebbian inhibitory neurons I_x^H and one population of anti-Hebbian inhibitory neurons I_x^A (in blue) ($x = 1, 2$). Dashed circles identify groups of neurons admitting synchronization events (memory recalls).

asynchronous irregular dynamics[23,56] and to prevent abnormal behaviour, e.g., the pathological high frequency firing observed in Fig. 1B. Meanwhile, the anti-Hebbian feedforward inhibition mediates the competition and selective activation across populations by, e.g., allowing the stimulated population (the stored memory item) to silence the activity of the other populations, representing other memory items.

These results resonate with recent empirical and modeling observations. Lagzi, et al.[25] showed via *in-vitro* experiments that parvalbumin-expressing (PV) and somatosatin-expressing (SOM) interneurons of the mouse frontal cortex follow symmetric and asymmetric Hebbian STDP respectively. Subsequent modeling suggested PV neurons to mediate homeostasis in excitatory activity, while SOM neurons build lateral inhibitory connections providing competition between excitatory assemblies. Guy et al.[55] performed *in-vivo* measurements of

the responses of PV, somatostatin-expressing (SST) and vasoactive intestinal peptide-expressing (VIP) interneurons at the mouse barrel cortex and found differentiated angular selectivity functionality for each neuron type. In particular, whisker stimulation evoked direction-selective inhibition in a majority of SST interneurons.

Previous models have compared the function of Hebbian and anti-Hebbian plasticities[23,57–59] but these were mainly investigated separately. Here, we have studied their joint impact on the adaptation and the dynamical behaviour of a neuronal network, allowing us to naturally associate both types of plasticity with given functions. In our model the feedback and feedforward inhibitory circuits emerged spontaneously—through adaptation—and were not imposed a priori. In the light of our results, we conclude that the coexistence of Hebbian and anti-Hebbian inhibition is decisive for the formation of stable memory-related structural modules (assemblies) and for the onset of spontaneous memory recalls in the neuronal activity.

The relevance of spontaneous recalls

The consolidation and long-term preservation of memories involve neurochemical changes that are modulated by the dynamical activity of the neurons and their neighbours[15–17]. Cortical and hippocampal activity during REM and non-REM sleep is characterized by low and high frequency firing events[60,61] that are relevant for the consolidation of daytime experiences[19–21,62–65]. The cortical activity during resting awake is typically characterised by asynchronous irregular dynamics, facilitating the integration of various sensory inputs and the formation of new associations between different information sources[66,67]. Additionally, short and random events of partial synchronous activation—occurring on top of the irregular firing background—are associated to spontaneous memory recalls (or retrieval)[22] and to the consolidation[23–25] of learned memories.

Once the learning phase has finalized, the dynamical behaviour of our model resembles that of awake resting-state *in-vivo* (Figs. 2C-J). It shows both a persistent asynchronous irregular activity with low firing ($\approx 0.0-5$ Hz) and punctuated events of transient synchrony of higher rates ($\approx 8-15$ Hz). The rules of STDP are effective models of the relation between the dynamical activity of neurons and the underlying biochemical processes leading to synaptic plasticity[18]. STDP stresses that these biochemical changes—either for depression or potentiation—take effect primarily when neurons fire concurrently within short time windows. In consequence, if the activity of a group of neurons is uncorrelated over time, the memories they form tend to be slowly forgotten. We have shown that this slow forgetting is compensated by the occurrence of spontaneous (partial or complete) recalls of the previously learned memories, happening at random times. These recalls act as short boosts reinforcing the patterns of synaptic connectivity formed during the learning, and thus constitute an autonomous mechanism for memory consolidation during rest. These results emphasize the importance of considering dynamical models of learning which remain *alive* after the training—instead of freezing their activity and plasticity—in order to understand the processes leading to memory consolidation and long-term maintenance.

It shall also be stressed that the coexistence of asynchronous irregular firing and memory recalls is an emergent dynamical property of our model, instead of being an enforced behaviour by the addition of internal mechanisms for frequency adaptation or by the application of external modulation. In a previous study with phase oscillators[52], the post-learning activity at rest showed a partially synchronous state similar to a limit cycle. Here the network remains stable in its asynchronous regime while the recalls spontaneously *pop out* at random times and order. This behaviour naturally results from the interplay between the underlying structural organization—shaped by the learning process—and the background activity eliciting the transitions, somehow similar to the noise-driven state switching found in neural activity[68]. These results evidence the benefit of studying connectivity, dynamics and learning simultaneously, as they represent different faces of interdependent phenomena.

The memory capacity of the model

Empirical and theoretical studies have associated inhibitory plasticity with the long-term storage of memories[45–48]. From these indications one could deduce the idea that the (long-term) memory capacity of the brain might be related to the quantity of inhibitory neurons. From the results of the numerical experiments displayed in Fig. 4B, we showed that in our model the number of Hebbian and anti-Hebbian inhibitory neurons determines the number of stable neural assemblies that the network admits. Subsequently, we concluded that the maximal storage capacity of a network of N neurons to be $M^\star = N/3$ memory items, which can only be achieved if 66% of the neurons in the network were inhibitory. Despite the maximal capacity $M^\star$ is much higher than that of the Hopfield model ($\simeq 0.14N$[69,70] while violating Dales principle), the requirement that the E/I ratio should be 1/2 (with 33% excitatory and 66% inhibitory neurons) seem not to be realistic in the brain. In particular, in our model the maximal capacity for a network with 20% of inhibitory neurons—as in the human cortex—will be approximately $0.1N$.

An overview of the known variability of E/I ratios across different parts of the human brain (Table of Fig. 4C) reveals that smaller regions such as the amygdala, thalamus, striatum and globus pallidus tend to have a higher proportion of inhibitory neurons than the cortex or the cerebellum. Extrapolating the results for the memory capacity from our model, it leads us to specu-

late that regions with fewer resources for memory storage (in terms of total number neurons), may compensate by allowing for a more favourable E/I ratio. Similarly, the mouse admits a more favourable ratio than humans in some areas such as the basal ganglia, striatum, amygdala or the hippocampus. While this interpretation is at the moment largely hypothetical, it may reflect a biological strategy for resource optimisation. Furthermore, it shall be noted that memory capacity is further affected by the complexity of the memories; with simple ones requiring less neurons to be stored and complex memories each requiring more neurons. Hence, the empirically observed E/I ratios may also reflect a compromise between the number of memories and the complexity of the information that needs to be stored.

Real neurons generally respond to and encode for multiple inputs and memories[71]. This mixed selectivity plays a crucial role in complex cognitive tasks allowing the brain to simultaneously represent and integrate multiple sources of information[50,72]. In addition to retaining segregated memories, our model also exhibits the possibility of learning and recalling complex memories admitting mixed selectivity, see Fig. 5. The neurons supporting mixed selectivity can be considered as hub neurons[73] connecting between the stored clusters and facilitating the ability to transmit and integrate information[1,74]. The presence of hub neurons does not increase the memory capacity of a network but instead, they allow for more complex information to be encoded by establishing associations between memories.

Limitations and outlook

Despite the model was developed to respect several biological constrains and to satisfy some realistic behavioural aspects, other choices and limitations could be implemented in future refinements. The brain is characterized sparse connectivity[75], but in our model the networks were initialised assuming *all-to-all* connectivity matrices with random weights. We omitted possible evolutionary aspects and assumed that only synaptic plasticity to be responsible for shaping the structure of the connectivity. Sensory systems organize information spatially with neighbouring neurons representing similar response properties and forming, e.g., retinotopic, tonotopic or somatotopic maps[76–78]. However, in the present model the neurons form a graph with no spatial embedding. It would be thus a natural step to extend the model in the future with spatially distributed neurons, following biologically representative spatial constraints.

The long-range white matter connections between distant cortical regions are mainly formed by axonal bundles of excitatory pyramidal neurons[79,80], although some GABAergic cells have also been found to project across different brain regions[73,81]. When our model was trained such that stimuli targetted separate neuronal populations, excitatory neurons formed local connections and only the anti-Hebbian inhibitory neurons developed "long-range" projections to other clusters. While this architecture is consistent with the lateral inhibition found in circuits for selectivity and decision-making[55,82,83], it is not representative of the excitatory long-range white matter fibers in the cortex. However, in the model, cross-modular excitatory connections appeared when the stimuli acted on overlapping populations, leading to the formation of excitatory hubs. While the reasons for the organization of excitatory and inhibitory cross-modular connectivities requires further investigation, extensions of the model here presented might be relevant to explore those mechanisms. In fact, both the organization of micro-/macroscopic connectivities and the development of short-/long-term memories, might be governed by distinct rules which could be separately implemented in the model.

METHODS

This section describes the spiking neuronal network model governing the dynamics of the neurons and the learning rules used for the adaptation of synaptic weights, as well as the microscopic and macroscopic indicators employed to characterize the network states and dynamics in the paper.

Spiking neuronal network model

We consider a network of QIF neurons, pulse coupled via exponentially decaying post-synaptic potentials and in the presence of STDP. Unless otherwise specified, the network will be composed of 80% (20%) excitatory (inhibitory) neurons as usually observed in the human cortex[84]. Depending on the plasticity rules controlling the synaptic strengths of the connections, three neural subpopulations can be identified depending on the nature of the pre-synaptic neurons:

- excitatory neurons subject to asymmetric Hebbian STDP;
- inhibitory neurons subject to symmetric Hebbian STDP;
- inhibitory neurons subject to symmetric anti-Hebbian STDP.

The evolution of the membrane potential V_i of each neuron ($i = 1, ..., N$) is described by the following ordinary differential equation:

$$\tau_m \dot{V}_i = V_i^2(t) + \eta_i + g_e S_i^e(t) + g_{hi} S_i^{hi}(t) + g_{ai} S_i^{ai}(t) + I_i(t) + \xi_i(t), \quad (1)$$

where $\tau_m = \tau_0$ ms (with $\tau_0 = 20$) is the membrane time constant; $\eta_i \sim \mathcal{N}(0.0, (\pi\tau_0)^2)$ are the neuronal excitabilities, chosen to have an average neuronal firing rate at

rest of around 1 Hz, $I_i(t) = \{0, (50\pi\tau_0)^2\}$ are the external DC currents, leading the neurons to fire around 50 Hz whenever stimulated and $\xi_i(t) \sim \mathcal{N}(0.0, (4\pi\tau_0)^2)$ is Gaussian additive noise tuned to induce a firing variability of $\simeq 4$ Hz. Whenever the membrane potential V_i reaches infinity, a spike is emitted and V_i is reset to $-\infty$. In the absence of synaptic coupling, external DC current and noise, the QIF model displays excitable dynamics for $\eta_i < 0$, while for positive η_i, it behaves as an oscillator with period $T_i^{(0)}/\tau_m = \pi/\sqrt{\eta_i}$.

The synaptic dynamics is mediated by the global synaptic strengths : $g_e = 100$, $g_{hi} = 400$ and $g_{ai} = 200$ for the excitatory, the Hebbian and anti-Hebbian inhibitory neurons, respectively. Finally, the evolution of the excitatory, Hebbian inhibitory and anti-Hebbian inhibitory synaptic currents $S_i^e(t)$, $S_i^{hi}(t)$ and $S_i^{ai}(t)$ is given by

$$\tau_d^e \dot{S_i^e} = -S_i^e + \frac{\tau_d^e}{N_e} \sum_{j=1}^{N_e} \sum_n w_{ij}(t)\delta(t - t_j^{(n)}), \qquad (2)$$

$$\tau_d^i \dot{S_i^{hi}} = -S_i^{hi} + \frac{\tau_d^i}{N_{hi}} \sum_{j=1}^{N_{hi}} \sum_n w_{ij}(t)\delta(t - t_j^{(n)}), \qquad (3)$$

$$\tau_d^i \dot{S_i^{ai}} = -S_i^{ai} + \frac{\tau_d^i}{N_{ai}} \sum_{j=1}^{N_{ai}} \sum_n w_{ij}(t)\delta(t - t_j^{(n)}), \qquad (4)$$

where $\tau_d^e = 2$ ms ($\tau_d^i = 5$ ms) are the exponential time decay for the excitatory (inhibitory) post-synaptic potentials; $w_{ij}(t)$ are the plastic coupling weights from neuron j towards neuron i, whose dynamics is described in the following and $t_j^{(n)}$ is the n-th spike time of the j-th pre-synaptic neuron, and $\delta(t)$ is the Dirac delta function.

We consider a network of N all-to-all connected neurons without self-connections and $N = N_e + N_{hi} + N_{ai} = 100$, where $N_e = 80$, $N_{hi} = 10$ and $N_{ai} = 10$ are the number of excitatory, Hebbian inhibitory and anti-Hebbian inhibitory neurons, respectively.

We integrate Eqs. (1), (2), (3) and (4) via the Euler method with time step dt (see Table I for its value). Whenever $V_i(t)$ overcomes a threshold value $V_p = 10$, we approximate the n-th firing time of neuron i as $t_i^{(n)} = t + \frac{1}{V_i}\tau_m$, where $\frac{1}{V_i}\tau_m$ corresponds to the time needed to reach $+\infty$ from the value V_p. Furthermore, the neuron is reset to $V_r = -10$ and held to such value for a time interval $\frac{2}{V_i}\tau_m$ seconds corresponding to the time needed for the neuron to reach $V_i(t) = \infty$ from V_p and to return to V_r after being reset to $-\infty$[29].

The choice of parameters sets the neurons into a low firing state with a relatively large variability ranging from 0.0 to 8 Hz, analogous to those observed in the cortex at rest. The external drives $\{I_i(t)\}$ turn the neuronal activity into high frequency oscillations in the γ range of approximately 50 Hz, typical of high-order perceptual activity[85,86].

STDP rules

The time evolution of the synaptic weights $w_{ij}(t)$ is controlled by STDP rules which depend on the time difference $\Delta t = t_i - t_j$ between the last spikes of the post-synaptic neuron i and the pre-synaptic neuron j[18,87].

The potentiation of the synapses is controlled by the plasticity function

$$\Lambda^+(\Delta t) = \begin{cases} \Lambda(\Delta t), & \text{if } \Lambda(\Delta t) \geq 0, \\ 0, & \text{if } \Lambda(\Delta t) < 0, \end{cases} \qquad (5)$$

while their depression is controlled by

$$\Lambda^-(\Delta t) = \begin{cases} 0, & \text{if } \Lambda(\Delta t) \geq 0, \\ \Lambda(\Delta t), & \text{if } \Lambda(\Delta t) < 0, \end{cases} \qquad (6)$$

where $\Lambda(\Delta t)$ entering in Eqs. (5) and (6) depends on the nature of the pre-synaptic neuron.

For the pre-synaptic excitatory neurons we use an asymmetric Hebbian STDP function $\Lambda_e(\Delta t)$[87]. This function affects the evolution of the weights in a causal way: when the pre-synaptic neuron j emits a spike before (after) the post-synaptic neuron i, the synaptic weight w_{ij} increases (decreases)[88]. The function $\Lambda_e(\Delta t)$ is depicted in Fig. 6A, and its explicit expression reads as

$$\Lambda_e(\Delta t) = \begin{cases} A_+ e^{-\frac{\Delta t}{\tau_+}} - A_- e^{-\frac{4\Delta t}{\tau_+}} - f, & \text{for } \Delta t \geq 0, \\ A_+ e^{\frac{4\Delta t}{\tau_-}} - A_- e^{\frac{\Delta t}{\tau_-}} - f, & \text{for } \Delta t < 0, \end{cases} \qquad (7)$$

with the time constants $\tau_+ = 0.02$ sec and $\tau_- = 0.05$ sec, the amplitudes $A_+ = 5.296$ and $A_- = 2.949$. The term $f = 0.1$ models a natural slow forgetting of the memories[89,90] by causing a small but constant depression of the weights.

For pre-synaptic Hebbian inhibitory neurons, we use a symmetric Hebbian STDP function $\Lambda_{hi}(\Delta t)$[23,39,41,58,59,91], which takes the form of a Ricker wavelet function (or Mexican hat), potentiating (depressing) weights of neurons spiking in a correlated (uncorrelated) way. $\Lambda_{hi}(\Delta t)$ is shown in Fig. 6B and it reads as

$$\Lambda_{hi}(\Delta t) = A\left[1 - \left(\frac{\Delta t}{\tau}\right)^2\right] e^{\frac{-\Delta t^2}{2\tau^2}} - f, \qquad (8)$$

with the time constant $\tau = 0.1$ sec, the amplitude $A = 3$ and the forgetting term $f = 0.1$.

For pre-synaptic anti-Hebbian inhibitory neurons we use a symmetric anti-Hebbian STDP function $\Lambda_{ai}(\Delta t)$ [39,40,54,57,59,92] which corresponds to a reverse Ricker wavelet function (or reverse Mexican hat), potentiating (depressing) weights of neurons spiking in a uncorrelated (correlated) way. The function $\Lambda_{ai}(\Delta t)$ is shown in Fig. 6C and it takes the following expression:

$$\Lambda_{ai}(\Delta t) = -A\left[1 - \left(\frac{\Delta t}{\tau}\right)^2\right] e^{\frac{-\Delta t^2}{2\tau^2}} + f, \qquad (9)$$

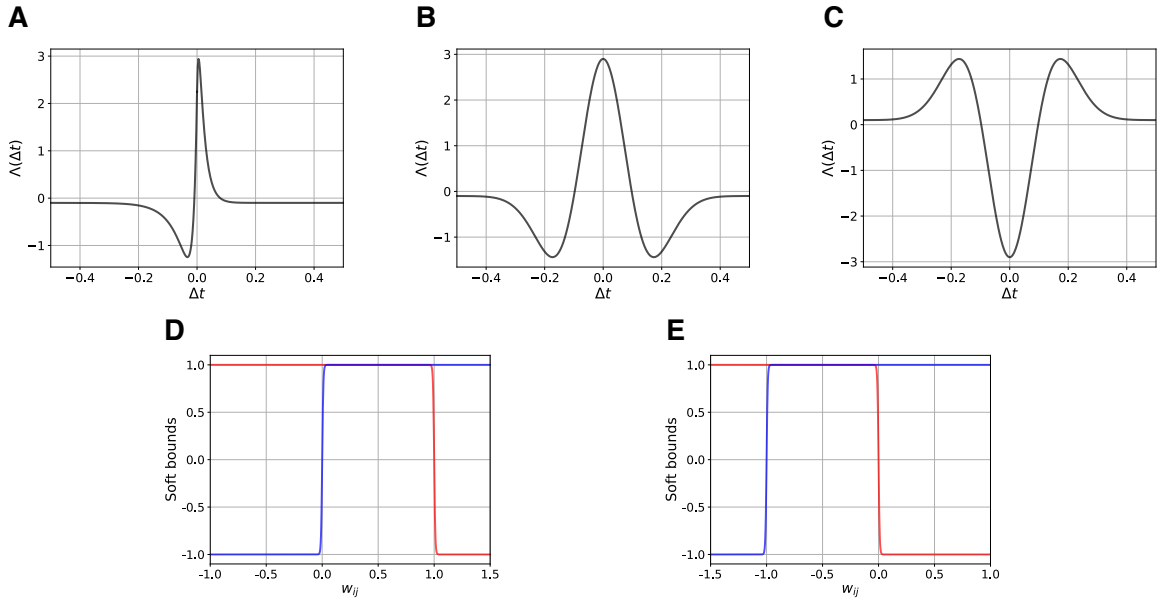


FIG. 6. **Plasticity and soft bound functions.** Plasticity functions: **(A)** Hebbian asymmetric STDP $\Lambda_e(\Delta t)$ (Eq. (7)); **(B)** Hebbian symmetric STDP $\Lambda_{hi}(\Delta t)$ (Eq. (8)); **(C)** anti-Hebbian symmetric STDP $\Lambda_{ai}(\Delta t)$ (Eq. (9)). Soft bound functions for excitatory **(D)** and inhibitory **(E)** neurons. The bounds for potentiation are shown in red and for depression in blue.

with the time constant $\tau = 0.1$ sec and the amplitude $A = 3$. In this case, the forgetting term $f = 0.1$ allows to have a constant small potentiation (the rule being anti-Hebbian) of the weights whatever the spike timing difference.

Adaptation of the synaptic weights

In this sub-section we explain in details the evolution of the synaptic weights. In particular, when the pre-synaptic neuron j or the post-synaptic neuron i spikes at time t, the weight w_{ij} is updated according to the following equations:

$$w_{ij}(t^+) = w_{ij}(t^-) + \Delta w_{ij}, \tag{10}$$

with

$$\Delta w_{ij} = \frac{1}{\tau_l}(-1)^{a_q}[\tanh(\lambda(w_q^l - w_{ij})) * \Lambda_q^+(\Delta t) + \tanh(\lambda(w_{ij} + w_q^u)) * \Lambda_q^-(\Delta t)], \tag{11}$$

where q denotes if the pre-synaptic neuron is excitatory $q = e$ or Hebbian (anti-Hebbian) inhibitory $q = hi$ ($q = ai$). For excitatory (inhibitory) neurons we set $w_q^l = 1$ ($w_q^l = 0$) and $w_q^u = 0$ ($w_q^u = 1$), thus ensuring that the excitatory (inhibitory) couplings are defined within the interval $w_{ij} \in [0 : 1]$ ($w_{ij} \in [-1 : 0]$). Moreover, $a_e = 2$ and $a_{hi} = a_{ai} = 1$ thus, for inhibitory synapses, the plasticity functions $\Lambda^+(\Delta t)$ and $\Lambda^-(\Delta t)$ are inverted and multiplied by -1 since potentiation (depression) of inhibitory weights makes them converge towards -1 (0). Furthermore, $\tau_l = 0.2$ sec is the learning time scale for the adaptation, while the parameter $\lambda = 100$ controls the slope of the soft bound function $\tanh(\lambda x)$.

These non-linear functions, depicted in Figs. 6D-E, model the saturation of the synaptic weights $|w_{ij}|$ between 0 and 1[93]. The saturation is obtained by allowing for synaptic depression (potentiation) dominating the potentiation (depression) term for large (small) values of the weights[18,94]. The steep slope of these functions (controlled by the parameter λ) guarantees a dynamical evolution of the synaptic weights even for large values of w_{ij}[95]. In this sense, the functions here used are at the limit between a soft and a hard bound.

We note that the firing activity of the neurons directly impacts the adaption rate and that the synaptic weights are always subject to adaptation, unlike in conventional artificial neural networks.

All network parameters are summarized in Table I.

Microscopic and macroscopic indicators

In this sub-section, we define the indicators employed to characterize the network dynamics at a microscopic and macroscopic level.

Microscopic Indicators The behaviour of single neurons was quantified by their instantaneous firing rates $\nu_j(t)$ defined as

$$\nu_j(t) = \frac{n_j^{sp}(t)}{T} \ . \tag{12}$$

Here, $n_j^{sp}(t)$ is the number of spikes emitted by neuron j (i.e. the spike count) in the time interval $[t : t + T]$. For

TABLE I. **Parameters for the network of QIF neurons.**

Parameters	Values
N	100
N_e	80
N_{hi}	10
N_{ai}	10
g_e	100
g_{hi}	200
g_{ai}	400
η	$\mathcal{N}(0.0,\ (\pi\tau_0)^2)$
$I(t)$	$\{0, (50\pi\tau_0)^2\}$
$\xi(t)$	$\mathcal{N}(0.0,\ (4\pi\tau_0)^2)$
V_p	10
V_r	-10
τ_0	0.02
τ_m	0.02 sec
τ_d^e	0.002 sec
τ_d^i	0.005 sec
τ_l	0.2 sec
τ_+	0.02 sec
τ_-	0.05 sec
τ	0.1 sec
dt	0.001 sec
A_+	5.296
A_-	2.949
A	3
f	0.1
λ	100

a population of N_p neurons, their instantaneous activity can be characterized as follows

$$\nu_p(t) = \frac{1}{N_p}\sum_{j=1}^{N_p}\nu_j \ . \tag{13}$$

The firing rate variability has been measured in terms of the coefficient of variation $CV_j = \frac{\sigma_j}{\mu_j}$ where μ_j is the mean interspike interval (ISIs) of the neuron j and σ_j its standard deviation. For a perfectly periodic firing $CV_j = 0$, while for a Poissonian process $CV_j = 1$.

Macroscopic Indicators The degree of synchronization in the network was quantified by the complex Kuramoto order parameter[96]

$$Z(t) = R(t)\mathrm{e}^{i\Phi(t)} = \frac{1}{N}\sum_{j=1}^{N}\mathrm{e}^{i\theta_j(t)} \ , \tag{14}$$

where $R(t)$ ($\Phi(t)$) represents the modulus (phase) of the macroscopic indicator. The modulus R is employed to characterize the level of phase synchronization in the network: $R > 0$ ($R = 1$) for a partially (fully) synchronized network, while $R \simeq \mathcal{O}(1/\sqrt{N})$ for an asynchronous dynamics due to finite size effects.

To associate a continuous phase $\theta_j(t) \in [0 : 2\pi]$ to the spiking activity of neuron j, we proceed in the following way:

$$\theta_j(t) = 2\pi\frac{(t - t_j^{(n)})}{(t_j^{(n+1)} - t_j^{(n)})} \qquad t_j^{(n)} \le t \le t_j^{(n+1)}, \tag{15}$$

with $t_j^{(n)}$ the n-th firing time of neuron j.

The mean rate of variation of the synaptic weights was estimated as

$$K(t) = \frac{1}{N*(N-1)}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{[w_{ij}(t+\Delta t) - w_{ij}(t)]}{\Delta t} \ , \tag{16}$$

where $w_{ij}(t)$ and $w_{ij}(t+\Delta t)$ are the synaptic coupling weights from neuron j to i at times t and $t+\Delta t$ respectively, $\Delta t = 0.1$ sec. The normalization term is $N*(N-1)$ since we consider an all-to-all connected network without autapses. The parameter $K(t)$ takes positive (negative) values for an overall increase (decrease) in the weight connectivity.

DATA AVAILABILITY

All code written in support of this publication is publicly available at `https://github.com/rbergoin/QIF-neurons-with-3-synaptic-plasticities/`.

ACKNOWLEDGMENTS

This work was supported (R.B., G.D. and G.Z.L.) by the European Union's Horizon 2020 Framework Programme for Research and Innovation under the Specific [Grant Agreement No. 945539 (Human Brain Project SGA3)] and by an EUTOPIA funding [EUTOPIA-PhD-2020-0000000066 - NEUROAI]. A.T. received financial support by the Labex MME-DII [Grant No. ANR-11-LBX-0023-01] (together with M.Q.), and by the ANR Project ERMUNDY [Grant No. ANR-18-CE37-0014] all part of the French program "Investissements d'Avenir". G.D. is supported by the Spanish national research project [ref. PID2019-105772GB-I00/AEI/10.13039/501100011033] funded by the Spanish Ministry of Science, Innovation, and Universities (MCIU). M.Q. is also partially supported by CNRS through the IPAL lab in Singapore. The authors thank Matthieu Gilson for useful discussions.

REFERENCES

[1] Zamora-López G, Zhou C, Kurths J. Cortical hubs form a module for multisensory integration on top of the hierarchy of cortical networks. Frontiers in Neuroinformatics. 2010; p. 1.

[2] Zamora-López G, Zhou C, Kurths J. Exploring brain function from anatomical connectivity. Frontiers in Neuroscience. 2011;5:83.

[3]Sporns O. Network attributes for segregation and integration in the human brain. Current Opinion in Neurobiology. 2013;23(2):162–171.
[4]Scannell JW, Blakemore C, Young MP. Analysis of connectivity in the cat cerebral cortex. Journal of Neuroscience. 1995;15(2):1463–1483.
[5]Hilgetag CC, Burns GA, O'Neill MA, Scannell JW, Young MP. Anatomical connectivity defines the organization of clusters of cortical areas in the macaque and the cat. Philosophical Transactions of the Royal Society of London Series B: Biological Sciences. 2000;355(1393):91–110.
[6]Meunier D, Lambiotte R, Bullmore ET. Modular and hierarchically modular organization of brain networks. Frontiers in Neuroscience. 2010;4:200.
[7]Senden M, Deco G, De Reus MA, Goebel R, Van Den Heuvel MP. Rich club organization supports a diverse set of functional network configurations. Neuroimage. 2014;96:174–182.
[8]Bertolero MA, Yeo BTT, D'Esposito M. The modular and integrative functional architecture of the human brain. Proceedings of the National Academy of Sciences. 2015;112(49):E6798–E6807.
[9]Damicelli F, Hilgetag CC, Hütt MT, Messé A. Topological reinforcement as a principle of modularity emergence in brain networks. Network Neuroscience. 2019;3(2):589–605.
[10]Miehl C, Onasch S, Festa D, Gjorgijeva J. Formation and computational implications of assemblies in neural circuits. Journal of Physiology. 2022;601(15):3070–3090.
[11]Manz P, Memmesheimer RM. Purely STDP-based assembly dynamics: Stability, learning, overlaps, drift and aging. PLoS Computational Biology. 2023;19(4):e1011006.
[12]Zamora-López G, Russo E, Gleiser PM, Zhou C, Kurths J. Characterizing the complexity of brain and mind networks. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2011;369(1952):3730–3747.
[13]Stella F, Cerasti E, Si B, Jezek K, Treves A. Self-organization of multiple spatial and context memories in the hippocampus. Neuroscience & Biobehavioral Reviews. 2012;36(7):1609–1625.
[14]Russo E, Treves A. Cortical free-association dynamics: distinct phases of a latching network. Physical Review E. 2012;85(5):051920.
[15]Bliss TV, Collingridge GL. A synaptic model of memory: long-term potentiation in the hippocampus. Nature. 1993;361(6407):31–39.
[16]McGaugh JL. Memory–a century of consolidation. Science. 2000;287(5451):248–251.
[17]Malenka RC, Bear MF. LTP and LTD: an embarrassment of riches. Neuron. 2004;44(1):5–21.
[18]Sjöström J, Gerstner W. Spike-timing dependent plasticity. Scholarpedia. 2010;5(2):1362. doi:10.4249/scholarpedia.1362.
[19]Steriade M. Impact of network activities on neuronal properties in corticothalamic systems. Journal of Neurophysiology. 2001;86(1):1–39.
[20]Stickgold R. Sleep-dependent memory consolidation. Nature. 2005;437(7063):1272–1278.
[21]Theodoni P, Rovira B, Wang Y, Roxin A. Theta-modulation drives the emergence of connectivity patterns underlying replay in a network model of place cells. Elife. 2018;7:e37388.
[22]Gu Y, Gong P. The dynamics of memory retrieval in hierarchical networks. Journal of Computational Neuroscience. 2016;40(3):247–268.
[23]Vogels TP, Sprekeler H, Zenke F, Clopath C, Gerstner W. Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. Science. 2011;334(6062):1569–1573.
[24]Carrillo-Reid L, Yang W, Bando Y, Peterka DS, Yuste R. Imprinting and recalling cortical ensembles. Science. 2016;353(6300):691–694.
[25]Lagzi F, Bustos MC, Oswald AM, Doiron B. Assembly formation is stabilized by Parvalbumin neurons and accelerated by Somatostatin neurons. BioRxiv. 2021; p. 2021–09.
[26]Amit DJ, Brunel N. Global spontaneous activity and local structured (learned) delay period activity in cortex. Cerebral Cortex. 1997;7:237–252.
[27]Brunel N, Wang XJ. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. Journal of Computational Neuroscience. 2001;11:63–85.
[28]Mongillo G, Barak O, Tsodyks M. Synaptic theory of working memory. Science. 2008;319(5869):1543–1546.
[29]Taher H, Torcini A, Olmi S. Exact neural mass model for synaptic-based working memory. PLoS Computational Biology. 2020;16(12):e1008533.
[30]Boscaglia M, Gastaldi C, Gerstner W, Quian Quiroga R. A dynamic attractor network model of memory formation, reinforcement and forgetting. PLoS Computational Biology. 2023;19(12):e1011727.
[31]Van Vreeswijk C, Sompolinsky H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. Science. 1996;274(5293):1724–1726.
[32]Brunel N. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. Journal of Computational Neuroscience. 2000;8:183–208.
[33]Vogels TP, Abbott LF. Signal propagation and logic gating in networks of integrate-and-fire neurons. Journal of Neuroscience. 2005;25(46):10786–10795.
[34]Destexhe A. Self-sustained asynchronous irregular states and up–down states in thalamic, cortical and thalamocortical networks of nonlinear integrate-and-fire neurons. Journal of Computational Neuroscience. 2009;27:493–506.
[35]Alreja A, Nemenman I, Rozell CJ. Constrained brain volume in an efficient coding model explains the fraction of excitatory and inhibitory neurons in sensory cortices. PLoS Computational Biology. 2022;18(1):e1009642.
[36]Softky WR, Koch C. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. Journal of Neuroscience. 1993;13(1):334–350.
[37]Ermentrout GB, Kopell N. Parabolic bursting in an excitable system coupled with a slow oscillation. SIAM Journal on Applied Mathematics. 1986;46(2):233–253.
[38]Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. Nature Reviews Neuroscience. 2008;9(7):557–568.
[39]Wu YK, Miehl C, Gjorgjieva J. Regulation of circuit organization and function through inhibitory synaptic plasticity. Trends in Neurosciences. 2022;45(12):884–898.
[40]Lamsa KP, Heeroma JH, Somogyi P, Rusakov DA, Kullmann DM. Anti-Hebbian long-term potentiation in the hippocampal feedback inhibitory circuit. Science. 2007;315(5816):1262–1266.
[41]Lamsa K, Heeroma JH, Kullmann DM. Hebbian LTP in feedforward inhibitory interneurons and the temporal fidelity of input discrimination. Nature Neuroscience. 2005;8(7):916–924.
[42]Perkel DH, Gerstein GL, Moore GP. Neuronal spike trains and stochastic point processes: II. Simultaneous spike trains. Biophysical journal. 1967;7(4):419–440.
[43]De Felipe J, Marco P, Fairén A, Jones E. Inhibitory Synaptogenesis in Mouse Somatosensory Cortex. Cerebral Cortex. 1997;7(7):619–634.
[44]Rubinski A, Ziv NE. Remodeling and tenacity of inhibitory synapses: relationships with network activity and neighboring excitatory synapses. PLoS Computational Biology. 2015;11(11):e1004632.
[45]Karunakaran S, Chowdhury A, Donato F, Quairiaux C, Michel C, Caroni P. PV plasticity sustained through D1/5 dopamine signaling required for long-term memory consolidation. Nature Neuroscience. 2016;19(3):454–466.
[46]Barron HC, Vogels TP, Behrens TE, Ramaswami M. Inhibitory engrams in perception and memory. Proceedings of the National Academy of Sciences. 2017;114(26):6666–6674.
[47]Mongillo G, Rumpel S, Loewenstein Y. Inhibitory connectivity defines the realm of excitatory plasticity. Nature Neuroscience.

2018;21(10):1463–1470.
[48]Giorgi C, Marinelli S. Roles and transcriptional responses of inhibitory neurons in learning and memory. Frontiers in Molecular Neuroscience. 2021;14:689952.
[49]McGaugh JL. The amygdala modulates the consolidation of memories of emotionally arousing experiences. Annual Review of Neuroscience. 2004;27:1–28.
[50]Rigotti M, Barak O, Warden MR, Wang XJ, Daw ND, Miller EK, et al. The importance of mixed selectivity in complex cognitive tasks. Nature. 2013;497(7451):585–590.
[51]Parthasarathy A, Herikstad R, Bong JH, Medina FS, Libedinsky C, Yen SC. Mixed selectivity morphs population codes in prefrontal cortex. Nature Neuroscience. 2017;20(12):1770–1779.
[52]Bergoin R, Torcini A, Deco G, Quoy M, Zamora-López G. Inhibitory neurons control the consolidation of neural assemblies via adaptation to selective stimuli. Scientific Reports. 2023;13(1):6949.
[53]Kullmann DM, Moreau AW, Bakiri Y, Nicholson E. Plasticity of inhibition. Neuron. 2012;75(6):951–962.
[54]Koch G, Ponzo V, Di Lorenzo F, Caltagirone C, Veniero D. Hebbian and anti-Hebbian spike-timing-dependent plasticity of human cortico-cortical connections. Journal of Neuroscience. 2013;33(23):9725–9733.
[55]Guy J, Möck M, Staiger JF. Direction selectivity of inhibitory interneurons in mouse barrel cortex differs between interneuron subtypes. Cell Reports. 2023;42(1).
[56]Politi A, Torcini A. A robust balancing mechanism for spiking neural networks. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2024;34(4):041102. doi:10.1063/5.0199298.
[57]Földiák P. Forming sparse representations by local anti-Hebbian learning. Biological Cybernetics. 1990;64(2):165–170.
[58]Luz Y, Shamir M. Balancing feed-forward excitation and inhibition via Hebbian inhibitory synaptic plasticity. PLoS Computational Biology. 2012;8(1):e1002334.
[59]Kleberg FI, Fukai T, Gilson M. Excitatory and inhibitory STDP jointly tune feedforward neural circuits to selectively propagate correlated spiking activity. Frontiers in Computational Neuroscience. 2014;8:53.
[60]Torao-Angosto M, Manasanch A, Mattia M, Sanchez-Vives MV. Up and down states during slow oscillations in slow-wave sleep and different levels of anesthesia. Frontiers in Systems Neuroscience. 2021;15:609645.
[61]Wang JY, Heck KL, Born J, Ngo HVV, Diekelmann S. No difference between slow oscillation up-and down-state cueing for memory consolidation during sleep. Journal of Sleep Research. 2022;31(6):e13562.
[62]Squire LR, Alvarez P. Retrograde amnesia and memory consolidation: a neurobiological perspective. Current Opinion in Neurobiology. 1995;5(2):169–177.
[63]Marzano C, Ferrara M, Mauro F, Moroni F, Gorgoni M, Tempesta D, et al. Recalling and forgetting dreams: theta and alpha oscillations during sleep predict subsequent dream recall. Journal of Neuroscience. 2011;31(18):6674–6683.
[64]Dudai Y. The restless engram: consolidations never end. Annual Review of Neuroscience. 2012;35:227–247.
[65]Eichenlaub JB, Nicolas A, Daltrozzo J, Redouté J, Costes N, Ruby P. Resting brain activity varies with dream recall frequency between subjects. Neuropsychopharmacology. 2014;39(7):1594–1602.
[66]Buzsaki G, Draguhn A. Neuronal oscillations in cortical networks. Science. 2004;304(5679):1926–1929.
[67]Fries P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends in Cognitive Sciences. 2005;9(10):474–480.
[68]Rolls ET, Deco G. The noisy brain: stochastic dynamics as a principle of brain function. Oxford University Press; 2010.
[69]Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences. 1982;79(8):2554–2558.
[70]Amit DJ, Gutfreund H, Sompolinsky H. Storing infinite numbers of patterns in a spin-glass model of neural networks. Physical Review Letters. 1985;55(14):1530.
[71]Quiroga RQ, Reddy L, Kreiman G, Koch C, Fried I. Invariant visual representation by single neurons in the human brain. Nature. 2005;435(7045):1102–1107.
[72]Johnston WJ, Palmer SE, Freedman DJ. Nonlinear mixed selectivity supports reliable neural computation. PLoS Computational Biology. 2020;16(2):e1007544.
[73]Bonifazi P, Goldin M, Picardo MA, Jorquera I, Cattani A, Bianconi G, et al. GABAergic hub neurons orchestrate synchrony in developing hippocampal networks. Science. 2009;326(5958):1419–1424.
[74]Van Den Heuvel MP, Sporns O. Rich-club organization of the human connectome. Journal of Neuroscience. 2011;31(44):15775–15786.
[75]Douglas RJ, Martin KA. Mapping the matrix: the ways of neocortex. Neuron. 2007;56(2):226–238.
[76]Mountcastle VB. Modality and topographic properties of single neurons of cat's somatic sensory cortex. Journal of Neurophysiology. 1957;20(4):408–434.
[77]Kaas JH. Topographic maps are fundamental to sensory processing. Brain Research Bulletin. 1997;44(2):107–112.
[78]Kaas JH, Collins CE. The organization of sensory cortex. Current Opinion in Neurobiology. 2001;11(4):498–504.
[79]Petreanu L, Mao T, Sternson SM, Svoboda K. The subcellular organization of neocortical excitatory connections. Nature. 2009;457(7233):1142–1145.
[80]Zhang S, Xu M, Kamigaki T, Hoang Do JP, Chang WC, Jenvay S, et al. Long-range and local circuits for top-down modulation of visual cortex processing. Science. 2014;345(6197):660–665.
[81]Caputi A, Melzer S, Michael M, Monyer H. The long and short of GABAergic neurons. Current Opinion in Neurobiology. 2013;23(2):179–186.
[82]Gerstner W, Kistler WM, Naud R, Paninski L. Neuronal dynamics: From single neurons to networks and models of cognition. Cambridge University Press; 2014.
[83]Maass W. On the computational power of winner-take-all. Neural Computation. 2000;12(11):2519–2535.
[84]Abeles M. Corticonics: Neural circuits of the cerebral cortex. Cambridge University Press; 1991.
[85]Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J. Stimulus specificity of phase-locked and non-phase-locked 40 Hz visual responses in human. Journal of Neuroscience. 1996;16(13):4240–4249.
[86]Herrmann CS. Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. Experimental Brain Research. 2001;137:346–353.
[87]Bi G, Poo M. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. Journal of Neuroscience. 1998;18(24):10464–10472.
[88]Carlson KD, Richert M, Dutt N, Krichmar JL. Biologically plausible models of homeostasis and STDP: stability and learning in spiking neural networks. In: The 2013 International Joint Conference on Neural Networks (IJCNN). IEEE; 2013. p. 1–8.
[89]Wixted JT. The psychology and neuroscience of forgetting. Annual Review of Psychology. 2004;55:235–269.
[90]Hardt O, Nader K, Nadel L. Decay happens: the role of active forgetting in memory. Trends in Cognitive Sciences. 2013;17(3):111–120.
[91]Perez Y, Morin F, Lacaille JC. A hebbian form of long-term potentiation dependent on mGluR1a in hippocampal inhibitory interneurons. Proceedings of the National Academy of Sciences. 2001;98(16):9401–9406.
[92]Plumbley MD. Efficient information transfer and anti-Hebbian neural networks. Neural Networks. 1993;6(6):823–833.
[93]Jin Y, Li P. AP-STDP: A novel self-organizing mechanism for efficient reservoir computing. In: 2016 International Joint Con-

ference on Neural Networks (IJCNN). IEEE; 2016. p. 1158–1165.
[94]Van Rossum MC, Shippi M, Barrett AB. Soft-bound synaptic plasticity increases storage capacity. PLoS Computational Biology. 2012;8(12):e1002836.
[95]Gilson M, Fukai T. Stability versus neuronal specialization for STDP: long-tail weight distributions solve the dilemma. PloS ONE. 2011;6(10):e25339.
[96]Kuramoto Y. Chemical oscillations, waves, and turbulence. Courier Corporation; 2003.

Supplementary Information for:

Emergence and long-term maintenance of modularity in plastic networks of spiking neurons

Raphaël Bergoin[1,2,3,4*], Alessandro Torcini[5], Gustavo Deco[2,3,6], Mathias Quoy[1,7], Gorka Zamora-López[2,3]

1 ETIS, UMR 8051, ENSEA, CY Cergy Paris Université, CNRS, 6 Av. du Ponceau, 95000 Cergy-Pontoise, France.
2 Center for Brain and Cognition, Pompeu Fabra University, Ramón Trias Fargas 25-27, 08005 Barcelona, Spain.
3 Department of Information and Communication Technologies, Pompeu Fabra University, Roc Boronat 138, 08018 Barcelona, Spain.
4 Institute of Neural Information Processing, Center for Molecular Neurobiology (ZMNH), University Medical Center Hamburg-Eppendorf (UKE), 20251 Hamburg, Germany.
5 Laboratoire de Physique Théorique et Modélisation, UMR 8089, CY Cergy Paris Université, CNRS, 2 Av. Adolphe Chauvin, 95032 Cergy-Pontoise, France.
6 Instituciò Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluis Companys 23, 08010 Barcelona, Spain.
7 IPAL, CNRS, 1 Fusionopolis Way #21-01 Connexis (South Tower), Singapore 138632, Singapore.

* raphael.bergoin@gmail.com

S1. UNTRAINED GROUP OF NEURONS.

This alternative protocol is analogous to the one of the numerical experiment reported in Fig. 1D of the main text. The only difference lies in the fact that a group of excitatory neurons is never stimulated and so it is untrained. The results obtained are described in Supplementary Fig. S1. On the one hand, we obtain analogous results with the formation of two modular structures in the weighted connectivity associated to spontaneous recalls of the two different memories during the dynamical evolution as shown in raster plot. On the other hand, neurons of the untrained group are weakly connected among them in accordance with the absence of stimulation and are decoupled from the other clusters while receiving anti-Hebbian inhibition from them. As a result, these neurons spike in a totally asynchronous and irregular way, without impacting the dynamics of the rest of the network.

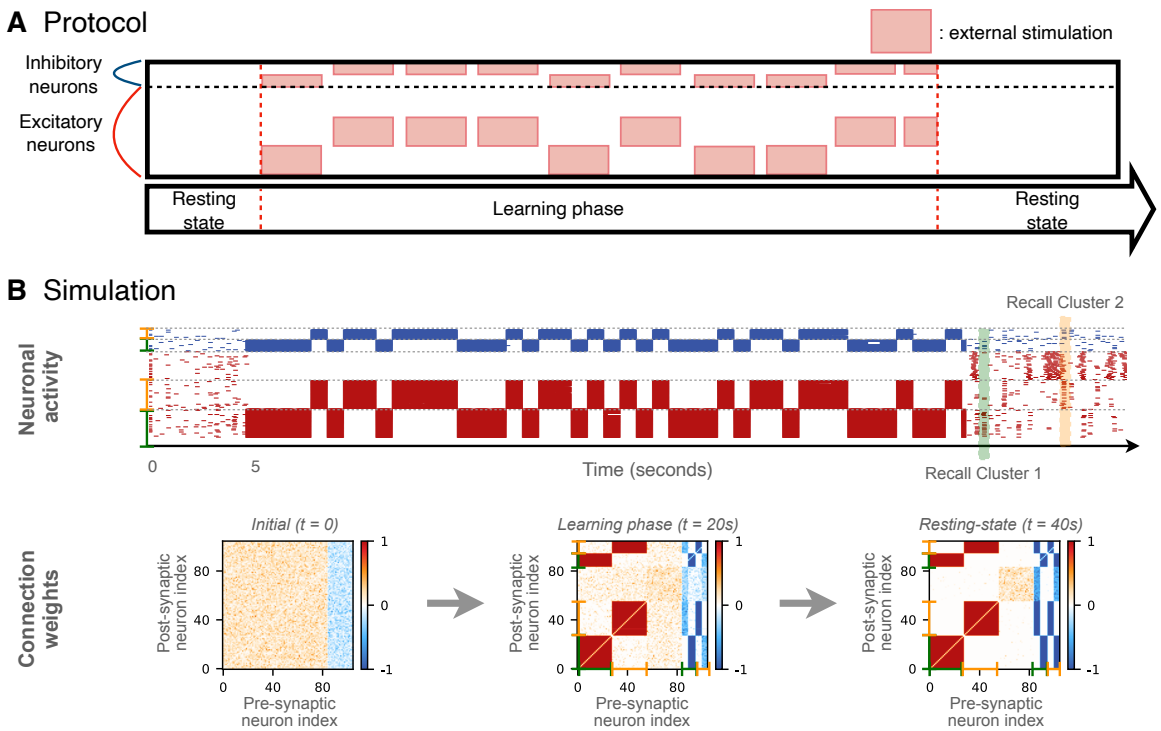


FIG. S1. **Learning of 2 stimuli with an untrained sub-population. (A)** Stimulation protocol for a network of $N = 105$ neurons entrained to $M = 2$ stimuli with an untrained group of excitatory neurons. **(B)** Simulation and learning results. Connectivity matrices show the evolution of the synaptic weights leading to the emergence of two modules and the decoupling of the unstimulated sub-population. The raster plot shows the simulation for the three stages: initial resting phase, entrainment stage and the post-learning neuronal activity characterized by spontaneous recall events of P_1 neurons (green shadow) and P_2 neurons (orange shadow).

S2. RANDOMLY STIMULATED NEURONS WITHIN EACH POPULATION.

This alternative protocol is analogous to that of the numerical experiment shown in Fig. 1D of the main text. The only difference lies in the fact that when a population is selected during learning, a random number of neurons in the excitatory population (with a probability of 0.5) is stimulated. The results obtained are described in Supplementary Fig. S2. The direct consequence is that the two modular structures in the weighted connectivity are less well formed compared to the original experiment. Nevertheless, the clusters remain decoupled with the same feedforward and feedback inhibition described in the main text. Only the weights within the clusters appear have more random values. Therefore, although the spontaneous recalls of the two memory items are present in the dynamics as shown in the raster plot, they appears to be somewhat sparser and less synchronized. We assume that this is largely due to the fact that the connections to and from the inhibitory neurons are incomplete. In Fig. 3A of the main text, the randomness of the E-E connections does not impact the spontaneous recalls. This highlights the need to reach a convergence of the weights linked to inhibition to correctly memorize the items. Nevertheless, this experiment also shows that even by partially learning memory items at each stimulation period leads the entire original memory item to be somehow retrieved and learned.

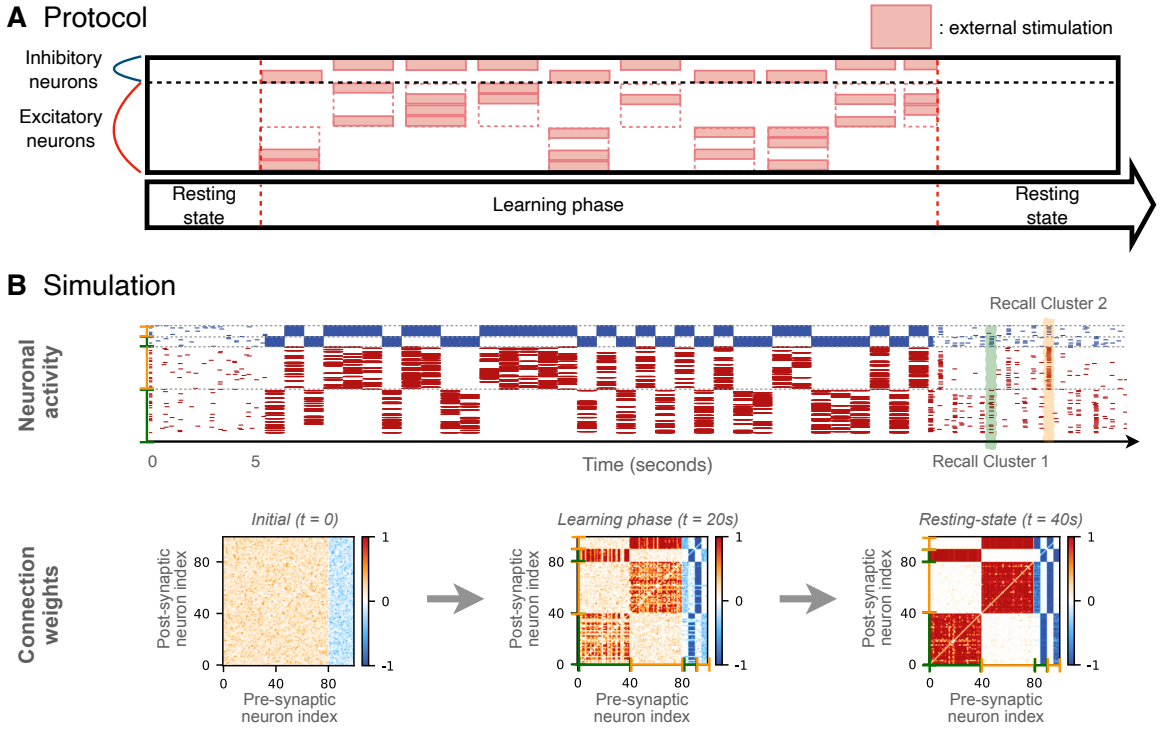


FIG. S2. **Learning of 2 stimuli with randomly stimulated neurons within each distinct population. (A)** Stimulation protocol for a network of $N = 100$ neurons entrained to $M = 2$ stimuli, neurons within each population being randomly selected. **(B)** Simulation and learning results. Connectivity matrices show the evolution of the synaptic weights leading to the emergence of two modules. The raster plot shows the simulation for the three stages: initial resting phase, entrainment stage and the post-learning neuronal activity characterized by spontaneous recall events of P_1 neurons (green shadow) and P_2 neurons (orange shadow).

S3. RANDOM STIMULATION VALUES.

This alternative protocol is again analogous to that of experiment of Fig. 1D of the main text. The only difference lies in the fact that when a population is stimulated (excitatory and inhibitory neurons), the neurons within it receive inputs of random amplitude (i.e. inducing firing activity between 50 and 100 Hz). The results obtained are described in Supplementary Fig. S3. We observe very similar results to those in Supplementary Fig. S2. However in this case, the weights connections seem stronger than in the previous experiment. As a result, the spontaneous recalls appear to be more visible in the dynamics of raster plot. In addition to the conclusions made in the previous study, this experience allows us to conclude that the spatio-temporal correlations of the applied inputs are more impactful on the formation of the modular structures than the intensities of these same inputs. Indeed, the firing frequency induced by the inputs applied to the neurons necessarily has an impact on the encoding of information and the construction of memory as can be seen here. Nevertheless, compared with the previous experiment where the neurons received the same inputs but at times that were not necessarily correlated, the structure is much better learned in the current case.

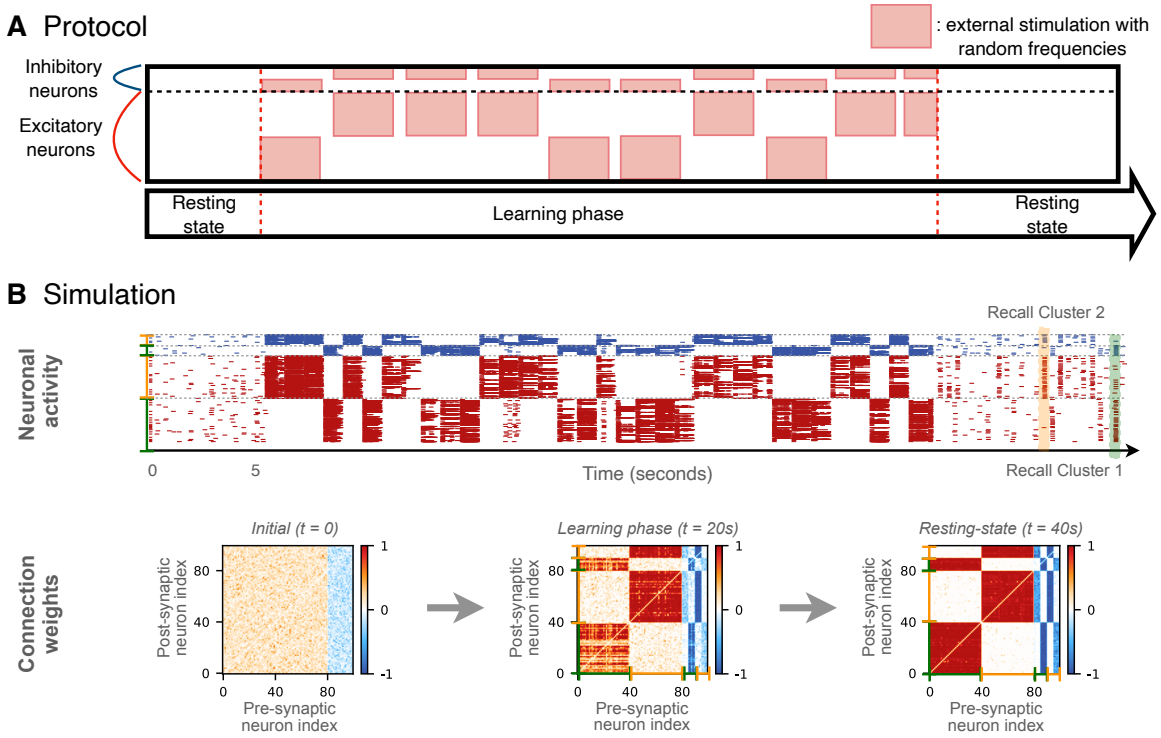


FIG. S3. **Learning of 2 stimuli with random frequencies. (A)** Stimulation protocol for a network of $N = 100$ neurons entrained to $M = 2$ stimuli of random amplitude. **(B)** Simulation and learning results. Connectivity matrices show the evolution of the synaptic weights leading to the emergence of two modules. The raster plot shows the simulation for the three stages: initial resting phase, entrainment stage and the post-learning neuronal activity characterized by spontaneous recall events of P_1 neurons (green shadow) and P_2 neurons (orange shadow).

S4. ESTIMATION OF THE TIME NEEDED TO FORGET A RECALL EVENT.

The rate of change of the synaptic weights depends on several factors: the current value of the weight $w_{ij}(t)$, the number of neurons spiking at a given time t, and the temporal precision of their spikes. Therefore, it is non-trivial to quantify the precise amplitude of the change of weights at all times. If a synapse is already at its maximum weight capacity, a potentiation (due to a recall) has almost no impact whereas the absence of recalls leads to more impactful depression (see soft bound functions in Methods, Figs. 6D and E).

We can derive an approximation for the impact of a recall on a synaptic weight—compared to a forgetting epoch—by inspecting closer the plasticity function (see Fig.6A). For uncorrelated spikes during an asynchronous irregular firing epoch (i.e. for $|\Delta t| > 0.5$), the weights depress with an amplitude of $\Lambda(\Delta t) \approx -0.1$ due to the forgetting term. On the other hand, if the spikes are fully synchronized (i.e. $\Delta t = 0$), $\Lambda(\Delta t) = A_+ - A_- - f = 2.347$. Thus, by considering an average weight of $w_{ij} = 0.5$ (where $\tanh(\lambda(1 - w_{ij})) = \tanh(\lambda w_{ij}) \approx 1$) Eq. (10) becomes:

$$\tau_l[w_{ij}(t^+) - w_{ij}(t^-)] \simeq \begin{cases} 2.347, & \text{for synchronized spikes during a recall} \ , \\ -0.1, & \text{for uncorrelated spikes during AI state} \ . \end{cases} \tag{S1}$$

The ratio of these two (absolute) values for potentiation and depression is $2.347/0.1 = 23.47$ and it gives an estimate of the number of uncorrelated spikes that would be required to loose the acquired potentiation. By considering an average firing rate of 2 Hz for the uncorrelated firings, this implies that a period of $\simeq 12$ seconds would be required to forget the contribution of a single recall event.

S5. STABILITY OF FOUR STRUCTURAL MODULES.

The reported numerical experiments analyse the limiting cases concerning the stability of four structural modules in absence of any stimulation. In Supplementary Fig. S4A, we consider the case where each population contains only one Hebbian and one anti-Hebbian inhibitory neuron (i.e. a total of $N_I = 2 \times 4 = 8$ inhibitory neurons). This arrangement corresponds to the upper limit for the number of inhibitory neurons needed to maintain 4 independent memory items, represented by the red line in Fig. 4B of the main text. We observe that these conditions are sufficient for each cluster to present distinct spontaneous recall in neuronal activity.

In Supplementary Fig. S4B, we break this limit by allocating only an Hebbian inhibitory neuron to the population P_1. We find that even if other populations are correctly recalled, memory recall of cluster 1 can occur at similar instant to that of the other clusters. Indeed, during recall, population P_1 does not inhibit the activity of the other populations, letting them activate. This has a direct effect on the consolidation process, where simultaneous recall of two memories patterns tends to induce their structural merging.

In Supplementary Fig. S4C, we perform an opposite test, assigning only an anti-Hebbian inhibitory neuron to the population P_1. In that case, we observe a very short period of spontaneous activity with recalls from the other population. However, once population P_1 becomes active, it totally dominates the others, inhibiting them. These results are very similar to those observed in Fig. 1B. Indeed the absence of feedback inhibition provided by Hebbian inhibitory neuron, prevents P_1 activity to be regulated. Ultimately, this has no direct impact on the long-term maintenance of memory items in the weight matrix since their activity is suppressed. Nevertheless, this causes issues in the processing of stored information, since the network is blocked in an abnormal state.

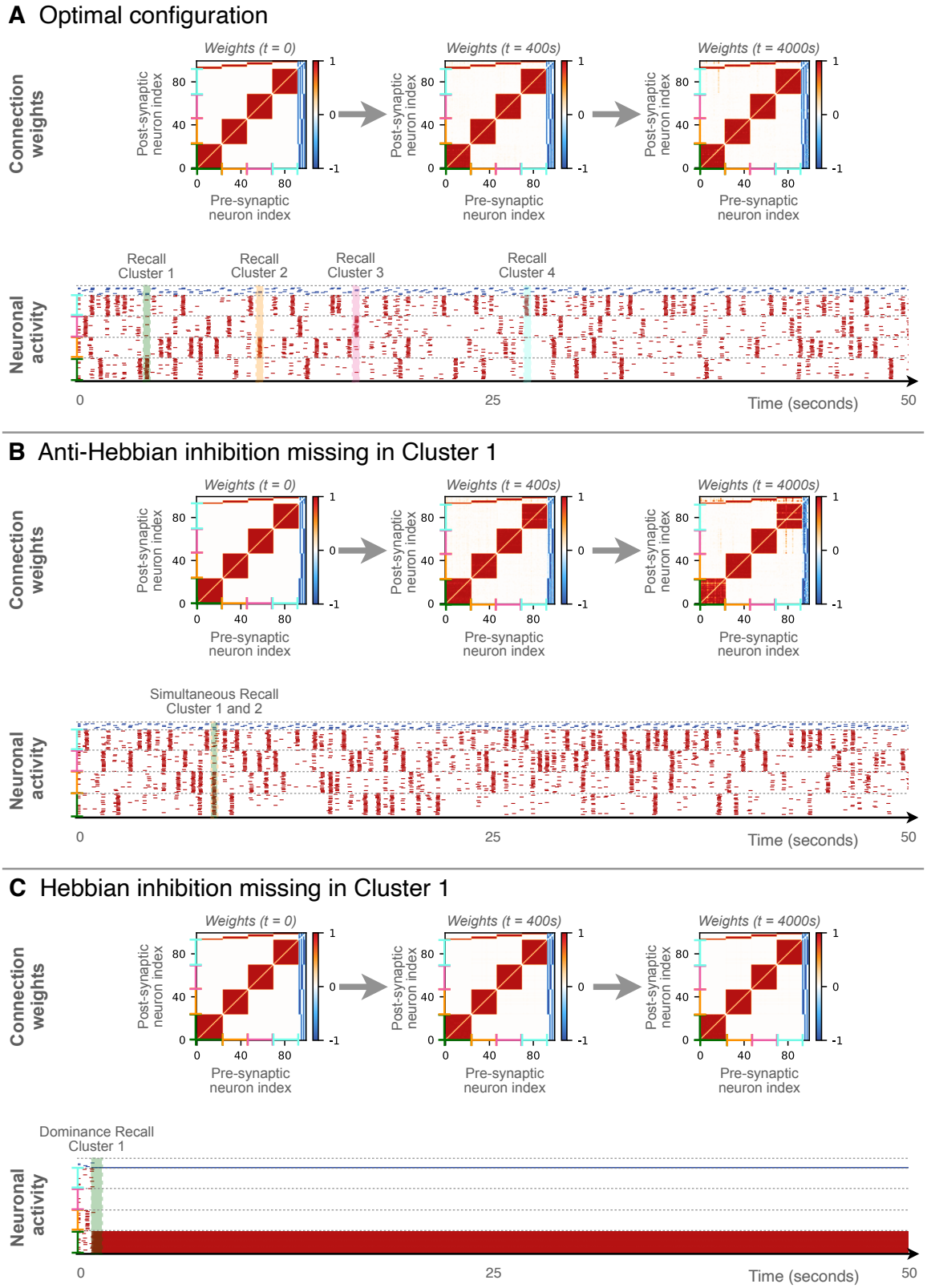


FIG. S4. **Evolution of a network initially made of four structural modules in absence of any stimulation. (A)** Network of 92 excitatory neurons and 8 inhibitory neurons (one Hebbian and one anti-Hebbian in each cluster). **(B)** Network of 93 excitatory neurons and 7 inhibitory neurons (only an Hebbian in cluster 1). **(C)** Network of 93 excitatory neurons and 7 inhibitory neurons (only an anti-Hebbian in cluster 1). In each case, we study the stability of the organization from a structural and dynamical point of view. In each panel, the connectivity matrices show the evolution of the synaptic weights and the raster plot shows the neuronal activity. The green, orange, pink and cyan brackets and shadows represent clusters 1, 2, 3 and 4.

S6. FOUR OVERLAPPING STIMULI.

In this alternative protocol, we reproduce the experiment of Fig. 5 of the main text but considering $M = 4$ stimuli which share 8 neurons. The results obtained are described in Supplementary Fig. S5. We obtain similar results as in main text with the formation of four modules in the connectivity matrix as expected, together with the formation of hubs which are here connected (incoming and outgoing connections) with the four clusters. Regarding the dynamics in raster plot, the resting-state activity is comparable that observed in Fig. 5. We identify different types of spontaneous recalls involving one of the four clusters alone, recalls of one cluster accompanied by the hubs, or recall events of hubs alone. This experiment again highlights the richness of the different dynamics that the network can display and maintain for a certain period of time.

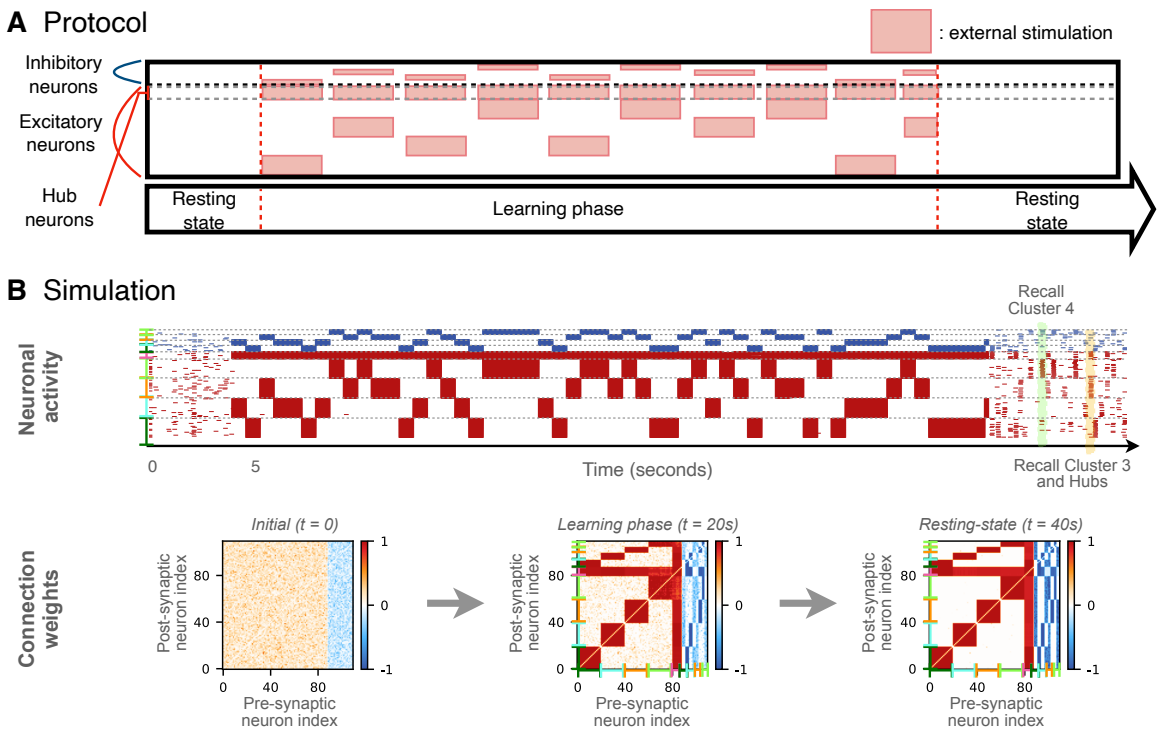


FIG. S5. **Learning of 4 overlapping stimuli.** **(A)** Stimulation protocol for a network of $N = 110$ neurons entrained to $M = 4$ stimuli that share 8 excitatory neurons. **(B)** Simulation and learning results. Connectivity matrices show the evolution of the synaptic weights leading to the emergence of four modules which overlap over 8 hub neurons. The raster plot shows the simulation for the three stages: initial resting phase, entrainment stage and the post-learning neuronal activity characterized by a variety of spontaneous recall events as of P_3 neurons with hubs (orange shadow) or P_4 neurons without hubs (green shadow).