Purely STDP-based assembly dynamics: stability, learning, overlaps, drift and aging

Paul Manz [1], Raoul-Martin Memmesheimer [1,*],

1 Neural Network Dynamics and Computation, Institute of Genetics, University of Bonn, Bonn, Germany

* rm.memmesheimer@uni-bonn.de

Abstract

Memories may be encoded in the brain via strongly interconnected groups of neurons, called assemblies. The concept of Hebbian plasticity suggests that these assemblies are generated through synaptic plasticity, strengthening the recurrent connections within select groups of neurons that receive correlated stimulation. To remain stable in absence of such stimulation, the assemblies need to be self-reinforcing under the plasticity rule. Previous models of such assembly maintenance require additional mechanisms of fast homeostatic plasticity often with biologically implausible timescales. Here we provide a model of neuronal assembly generation and maintenance purely based on spike-timing-dependent plasticity (STDP) between excitatory neurons. It uses irregularly and stochastically spiking neurons and STDP that depresses connections of uncorrelated neurons. We find that assemblies do not grow beyond a certain size, because temporally imprecisely correlated spikes dominate the plasticity in large assemblies. Assemblies in the model can be learned or spontaneously emerge. The model allows for prominent, stable overlap structures between static assemblies. Further, assemblies can drift, particularly according to a novel, transient overlap-based mechanism. Finally the model indicates that assemblies grow in the aging brain, where connectivity decreases.

Author summary

It is widely assumed that memories are represented by ensembles of nerve cells that have strong interconnections with each other. It is to date not clear how such strongly interconnected nerve cell ensembles form, persist, change and age. Here we show that already a basic rule for activity-dependent synaptic strength plasticity can explain the learning or spontaneous formation and the stability of assemblies. In particular, it is not necessary to explicitly keep the overall total synaptic strength of a neuron nearly constant, a constraint that was incorporated in previous models in a manner inconsistent with current experimental knowledge. Furthermore, our model achieves the challenging task of stably maintaining many overlaps between assemblies and generating the experimentally observed drift of memory representations. Finally, the model predicts that if the number of synaptic connections in the brain decreases, as observed during aging, the size of the neuron ensembles underlying memories increases. This may render certain memories in the aging brain more robust and prominent but also less specific.

Introduction

A widely used model of long-term memory posits that items are stored in the brain by strongly interconnected neuronal assemblies [1–3]. A memory item is represented by a group of neurons that coactivate upon memory recall. The assembly structure allows for associative recall from an incomplete input: due to the strong interconnections, activation of a part of the neurons in an assembly can trigger reactivation of the entire assembly and thus a recall of the full memory. The assemblies may each be created through experience-dependent plasticity or they may already form during development [4]. In the latter case, memories may form by connecting the pre-existing assemblies to appropriate input and output neurons [5]. The theory of the formation and maintenance of neuronal assemblies has been studied in much detail in previous works. The creation of new memories is commonly modeled using Hebbian plasticity [6–13]: if a set of neurons is co-activated, Hebbian plasticity increases the strength of their mutual connections leading to the formation of what is called a Hebbian assembly. However, memory networks in the brain also show ongoing spontaneous and irregular activity. If plasticity still takes place during this activity, it should not interfere with the existing memory assemblies – otherwise memories would have implausibly short lifespans. Hebbian assemblies can be self-reinforcing under plasticity since their interconnectedness leads to higher correlations in the activities, which in turn leads to potentiation of the intra-assembly weights. Models of assembly maintenance, however, found that fast homeostatic plasticity was needed in addition to Hebbian learning. This introduces competition between synapses and prevents pathological growth of assemblies and exploding activity [5, 7–9, 11–13]. Homeostatic plasticity has been observed in experiments, but it is much slower than Hebbian plasticity and does therefore not suffice to prevent runaway potentiation [14–18] (see, however, [19] for a different view and [10, 20] for a small timescale implementation of homeostasis via inhibitory STDP).

Experiments indicate that distinct memory assemblies have a fraction of shared neurons, i.e. neurons that are part of both assemblies [21–23]; the size of these overlaps appears to correspond to the strength of the associations between the concepts encoded by the assemblies. Previous models of assemblies stabilized by recurrent synaptic plasticity and fast homeostatic normalization usually do not show prominent overlaps [5, 7–9, 12]. An example of a network with weight plasticity, structural plasticity, multiple synapses per connection and short-term depression that can store two strongly overlapping assemblies was given in [24]. We will explore whether our purely STDP-based model can maintain overlaps.

Another topic of interest in the study of memory networks is whether they can generate the representational drift observed in recent experiments [25]. Such drift has recently been modeled by drifting assemblies, which spontaneously exchange neurons with each other, leading to a gradual resorting of the whole network [5]. The network model incorporated fast homeostatic normalization to stabilize the assemblies. We will explore whether also our purely STDP-based model can exhibit drifting assemblies.

Finally, aging and other conditions are associated with a decrease in overall connectivity [26, 27]. We will therefore explore how the assemblies in our networks adapt to such a decrease.

The paper is structured as follows: We initially introduce the model of spiking neurons and STDP and describe existing analytical approximations for the time-averaged weight dynamics. As a first result, we then show spontaneous assembly formation. We obtain an analytical approximation of the weight growth in different assembly sizes to obtain an understanding of the numerically observed assembly formation. Next we study whether assemblies can be learned by correlated external input. The subsequent section shows that our networks can stably maintain overlapping

assemblies. We thereafter examine whether our model can be set up to exhibit representational drift. Finally, we investigate the dependence of assembly sizes on network sparsity and relate our results to effects of aging on the brain.

Materials and methods

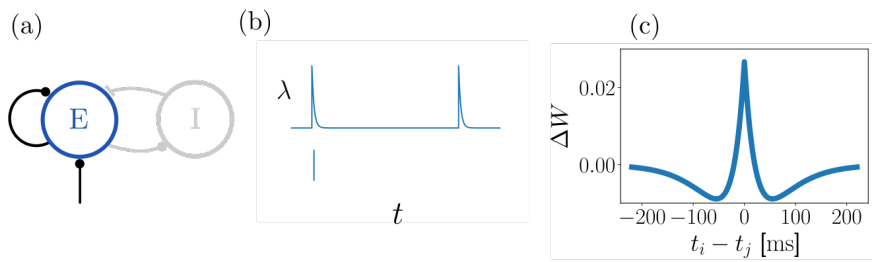

Fig 1. Network, neuron and plasticity model. (a): A network consists of recurrently connected excitatory neurons, which generate irregular, stochastic spiking. It may be interpreted as the excitatory population of a neural network where excitatory and inhibitory neurons generate a balanced state [28] of irregular spiking. Thus, a balancing inhibitory neuron population is implicitly contained in our model. (b): Poisson model of stochastic single neuron spiking: each neuron is characterized by its instantaneous rate $\lambda(t)$ (upper subpanel), which depends on incoming spikes and determines the probability of emitting a spike (lower subpanel). (c): Strength of STDP updates as a function of the time difference between pre- and postsynaptic spikes ($\mu = 1$).

Poisson neurons

Neural networks in the mammalian brain typically generate irregular and apparently largely stochastic spiking. To guarantee that our model networks generate similar spiking activity, we consider networks of stochastic linear Poisson (or "Hawkes") model neurons [9, 29–31]. Such networks are additionally analytically well tractable. We explicitly model excitatory neurons. Since the irregular activity in biological neural networks is likely due to a balanced state, where excitatory and inhibitory inputs to each neuron balance [28, 32–36], our model implicitly incorporates inhibitory neurons, see Fig. 1.

The spiking activity of each neuron i is an inhomogeneous Poisson process whose time-dependent instantaneous spike rate (intensity) $\lambda_i(t) = \langle S_i(t) \rangle$ given input spike trains $S_k(t') = \sum_m \delta(t' - t_k^m)$ up to time t and weights W_{ik} is

$$\lambda_i(t) = \lambda_i^0 + \sum_k W_{ik} \sum_m a(t - t_k^m). \tag{1}$$

The angular brackets here denote trial-averaging with fixed input spike trains and weights to neuron i. We use an exponentially decaying synaptic kernel

$$a(t) = \frac{\Theta(t)}{\tau_\mathrm{s}} \mathrm{e}^{-\frac{t}{\tau_\mathrm{s}}}, \tag{2}$$

where $\Theta(t)$ is the Heaviside distribution, and a constant external drive λ^0.

It is now useful to introduce quantities that are trial-averaged over the entire spiking network dynamics. The trial-averaged instantaneous rate (intensity) of neuron i is $r_i(t) = \langle \lambda_i(t) \rangle$, where angular brackets now denote the trial-averaging over the network

dynamics. For static W with spectral radius less than 1, the trial-averaged rate dynamics relax to a fixed point where the vector $r(t)$ is constant,

$$r = (\mathbb{1} - W)^{-1} \lambda^0, \tag{3}$$

see [9, 30, 37, 38]. In addition, one can analytically compute the corresponding stationary cross-correlation functions $C_{ij}(\tau) = \langle S_i(t+\tau) S_j(t) \rangle$ of pairs of neurons i, j at arbitrary t; in the frequency domain they read

$$\tilde{C}(\omega) = 2\pi\delta(\omega) r r^T + \left(\mathbb{1} - \tilde{a}(\omega) W\right)^{-1} D \left(\mathbb{1} - \tilde{a}(-\omega) W^T\right)^{-1}, \tag{4}$$

where we used matrix notation and $D_{ij} = \delta_{ij} r_i$ [9, 30, 38]. The Fourier transform of a function $g(t)$ is defined as $\tilde{g}(\omega) = \int_{-\infty}^{\infty} \mathrm{d}t \, \mathrm{e}^{-i\omega t} g(t)$. The first right hand side term in Eq. 4 is the correlation of independent Poisson spike trains with rate r, the second one reflects the interdependence of evoked spikes in the network. If the scale of the STDP updates is sufficiently small, one can assume that the weight dynamics are quasistationary with respect to neuronal dynamics. Then Eqs. 3,4 still approximately hold true despite W changing over time due to plasticity.

Spike-timing-dependent plasticity

We consider networks of spiking neurons with pair-based STDP, i.e. the change of synaptic strength depends on the time lags of pairs of pre- and postsynaptic spikes. The characteristics of the plasticity function crucially determine how synaptic strengths evolve in a network. Networks with symmetric plasticity functions can establish a structure of neuronal assemblies whereas plasticity functions with a large antisymmetric part tend to establish feedforward chains of connectivity; see, however, [10, 20] for networks with asymmetric STDP maintaining assemblies and [12] for a triplet STDP rule that forms assemblies despite an asymmetric two-spike interaction part. In the present work, we consider symmetric plasticity functions because of their simplicity and analytical tractability, building on previous theoretical work that has employed them [5, 9]. Symmetric STDP has been found in the CA3 region of the rodent hippocampus [39], i.e. in a region that is assumed to serve as an associative memory network and to store assemblies [1]. Recently, near-symmetric STDP has also been observed in the primary motor cortex of primates [40].

The change induced by a spike pair with time lag t is given by the STDP function F,

$$F(t) = \mu \left(A_\mathrm{p} \mathrm{e}^{-\frac{|t|}{\tau_p}} + A_\mathrm{d} \mathrm{e}^{-\frac{|t|}{\tau_d}} \right), \tag{5}$$

with $A_\mathrm{p} > 0$, $A_\mathrm{d} < 0, \left|A_\mathrm{p}\right| > |A_\mathrm{d}|$ and $\tau_\mathrm{p} < \tau_\mathrm{d}$, see Fig. 1c. We consider the overall scaling factor μ as the learning rate of the STDP function. μ is redundant and could be absorbed into A_p and A_d, whose magnitudes also determine the scale of the STDP function. We keep μ because it facilitates the formulation of task conditions with the same STDP shape but different amplitudes. For the analytical treatment of our plasticity rule we set $\mu = 1$. The assumption of small STDP scale to ensure a separation of timescales between plasticity and neuronal dynamics then becomes equivalent to assuming small A_p and A_d. In all our simulations the parameters are chosen such that the integral $\int_{-\infty}^{\infty} \mathrm{d}t \; F(t)$ over F is negative.

At each spike time, plasticity acts additively on the pre- and postsynaptic weights of the spiking neuron, with amplitudes given by F. Specifically, at a spike time t_{k_i} of the postsynaptic neuron there is a jump-like change in W_{ij} of $\sum_{t_{k_j} \leq t_{k_i}} F(t_{k_i} - t_{k_j})$. At a

spike time t_{j_k} of the presynaptic neuron, W_{ij} jumps by $\sum_{t_{k_i} \leq t_{k_j}} F(t_{k_i} - t_{k_j})$. This can be compactly written as

$$\frac{\mathrm{d}}{\mathrm{d}t} W_{ij}(t) = \sum_{t_{k_i}, t_{k_j} \leq t} \left(F(t_{k_i} - t_{k_j}) \delta(t - t_{k_i}) + F(t_{k_i} - t_{k_j}) \delta(t - t_{k_j}) \right). \tag{6}$$

Here and henceforth we assume $i \neq j$; there is no self-interaction in our networks, $W_{ii} = 0$. We further stipulate that no weight can become negative or exceed a maximum value $\hat{w}$ due to STDP: if a synapse would be depressed below 0, it will be set to 0 instead; if a synapse would be potentiated to a value beyond $\hat{w}$, it will be set to $\hat{w}$.

In the regime of quasistationary weight dynamics the time-averaged drift in synaptic efficacy, $\overline{\Delta W_{ij}}$, can be approximated by

$$\overline{\Delta W_{ij}} := \frac{1}{T} \int_t^{t+T} \mathrm{d}t' \; \frac{\mathrm{d}}{\mathrm{d}t'} W_{ij}(t') = \int_{-\infty}^{\infty} \mathrm{d}\tau \; C_{ij}(\tau) F(\tau), \tag{7}$$

where $C_{ij}(\tau)$ are again the correlations of the spike trains of neurons i and j [29]. We note that $\overline{\Delta W_{ij}}$ has the dimensionality of one over time. Plancherel's theorem yields

$$\overline{\Delta W_{ij}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega \; \tilde{C}_{ij}(\omega) \tilde{F}(-\omega), \tag{8}$$

with the Fourier transforms of the correlation function and of the STDP window. Inserting the correlation function $\tilde{C}_{ij}(\omega)$ of Eq. 4 gives

$$\begin{aligned} \overline{\Delta W_{ij}} &= f_0 r_i r_j \\ &\quad + \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega \; \sum_k \left[(\mathbb{1} - \tilde{a}(\omega) W)^{-1} \right]_{ik} r_k \left[(\mathbb{1} - \tilde{a}(-\omega) W^T)^{-1} \right]_{kj} \\ &\qquad \times \tilde{F}(-\omega), \end{aligned} \tag{9}$$

see [9], where

$$f_0 = \tilde{F}(0) = \int_{-\infty}^{\infty} \mathrm{d}t \; F(t) = 2 \left(A_\mathrm{p} \tau_\mathrm{p} + A_\mathrm{d} \tau_\mathrm{d} \right). \tag{10}$$

It is often useful to expand Eq. 4 into a power series with respect to W [9,30]. Inserting the series into the right hand side of Eq. 8 (or directly expanding Eq. 9) results in a series expansion for $\overline{\Delta W_{ij}}$, [9]:

$$\overline{\Delta W_{ij}} = f_0 r_i r_j + \sum_{\alpha, \beta} f_{\alpha\beta} \sum_m r_m (W^\alpha)_{im} (W^\beta)_{jm}. \tag{11}$$

The terms of the sum encode contributions from motifs in which a source neuron affects post- and presynaptic neurons via a chain of α and β connections, respectively (the same connection may be counted more than once). If $\alpha = 0$ ($\beta = 0$) the post(pre)synaptic neuron itself is the source. The coefficients $f_{\alpha\beta}$ contain integrals over the Fourier transform of the STDP window and powers of the Fourier transform of the synaptic kernel function,

$$f_{\alpha\beta} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega \; \tilde{F}(-\omega) \tilde{a}(\omega)^\alpha \tilde{a}(-\omega)^\beta. \tag{12}$$

	N	$\hat{w}$	τ_s	A_p	A_d	τ_p	τ_d	μ	λ_0
Fig. 1c				0.08	-0.0533	0.025 s	0.05 s		
Fig. 2	120	0.04	0.01 s	0.08	-0.0533	0.025 s	0.05 s	0.07	0.15 Hz
Fig. 3a			0.01 s	0.08	-0.0533	0.025 s	0.05 s		0.15 Hz
Fig. 3b	150		0.01 s	0.08	-0.0533	0.025 s	0.05 s	0.04	0.15 Hz
Fig. 4	80	0.026	0.01 s	0.08	-0.066	0.035 s	0.05 s	0.07	0.15 Hz
Fig. 5b	120	0.017	0.01 s	0.08	-0.044	0.022 s	0.055 s	0.02	0.15 Hz
Fig. 6a	60	0.0158	0.01 s	0.08	-0.044	0.022 s	0.055 s	0.021	0.15 Hz
Fig. 6b	60	0.024	0.01 s	0.08	-0.0533	0.025 s	0.05 s	0.045	0.15 Hz
Fig. 7	190	0.018	0.01 s	0.08	-0.042	0.026 s	0.065 s	0.017	0.15 Hz
Fig. 8a	72	0.056	0.01 s	0.08	-0.0533	0.025 s	0.05 s		0.2 Hz
Fig. 8b	120	0.024	0.01 s	0.08	-0.0533	0.025 s	0.05 s	0.05	0.15 Hz
Fig. 9a		0.058	0.01 s	0.08	-0.0533	0.025 s	0.05 s		0.15 Hz
Fig. 9b	120	0.058	0.01 s	0.08	-0.0533	0.025 s	0.05 s	0.07	0.15 Hz
Fig. 9c	150	0.067	0.01 s	0.08	-0.0533	0.025 s	0.05 s		0.2 Hz

Table 1. List of parameters for the simulations in the referenced figures

Simulation parameters

We have used the following parameters in the network simulations for the figures below.

Fig. 2: The network starts with an initially random connectivity where each weight is independently and uniformly drawn from $[0, 0.25\hat{w}]$.

Fig. 3b: The networks start with an initially random connectivity where each weight is independently and uniformly drawn from $[0, 0.20\hat{w}]$. We simulate for 10 different initial conditions and then take the median of all final assemblies.

Fig. 4: The initial weight matrix has an assembly of neurons 1-20 interconnected with $W_{ij} = \hat{w}$ and background connectivity as in Fig. 2. During the stimulation period neurons 21-40 are given input from a single source of Poisson spiking with $\lambda = 4.18\,\mathrm{Hz}$ and input weights $w_{\mathrm{in}} = 10\hat{w}$.

Fig. 6a: The neuron with index 10 has spontaneous rate $\lambda_0 = 0.08\,\mathrm{Hz}$.

Fig. 6b: The neuron with index 10 has spontaneous rate $\lambda_0 = 0.01\,\mathrm{Hz}$.

Fig. 8a: The initial weight matrix is as in Fig. 2 but with weights drawn from $[0, 0.15\hat{w}]$.

Fig. 8b: Every 3×10^5 s a neuron changes its spontaneous rates with probability $p = 0.03$ to $\lambda_0' = 0.03\,\mathrm{Hz}$ for 3×10^5 s. For the last 3×10^5 s of the simulation all spontaneous rates are kept at $\lambda_0 = 0.15\,\mathrm{Hz}$.

Results

Spontaneous assembly formation

We first simulate the model described above with initially unstructured weight matrix and without structured external stimulation, in order to test its capability of spontaneous assembly formation and subsequent maintenance. We find that for appropriately chosen parameters in the plasticity function the network weights indeed converge towards a structure with segregated assemblies of a characteristic size.

The mechanism underlying the increase of weights between future assembly neurons is well known: Initially basically randomly coincident spiking leads to strengthening of some weights. By directly and indirectly coupling neurons, these weights induce more near-simultaneous spiking, which further strengthens them. The positive feedback loop leads to weights that increase until they reach $\hat{w}$ [41]. Neurons with little coupling do

not display similar near-simultaneous spiking such that their weights grow more slowly, or even shrink due to long-term depression. It has been shown that if the summed weight strength to and from a neuron are additionally normalized by fast homeostasis, there can be spontaneous emergence of assemblies [5, 9, 12, 13]. The reason is that the homeostatic normalization lets synapses compete, such that more slowly growing ones are suppressed. Once an imbalance of connectivity and thus first assembly-like structures occur, the weight increase of intra-assembly synapses prior to normalization is stronger, due to the stronger co-spiking of the connected neurons. The normalization then suppresses the inter-assembly connections and the assembly structure consolidates.

While the activity-dependent plasticity-based mechanisms underlying spontaneous assembly formation remain in our networks, there is no fast homeostatic normalization that could introduce competition between synapses. Then, why do not all weights tend to $\hat{w}$? In other words: why does the network not turn into one big assembly?

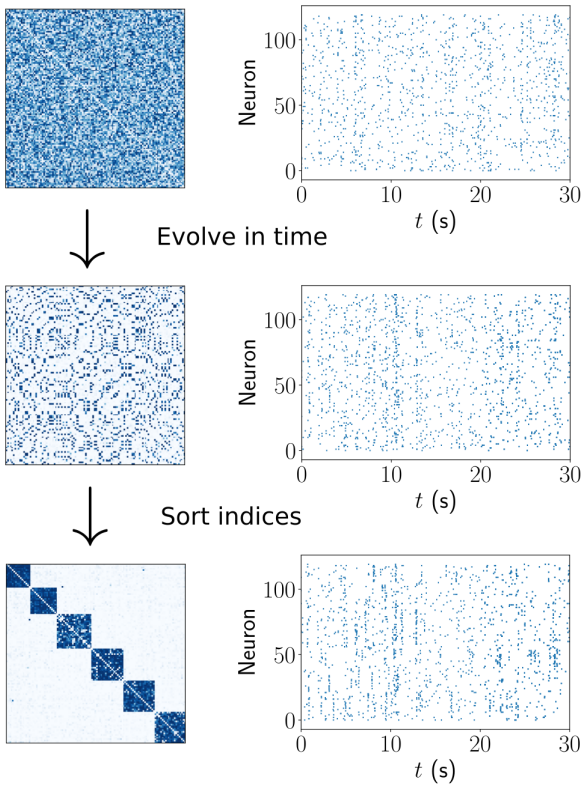

Fig 2. Spontaneous assembly formation. Several assemblies (strongly interconnected ensembles of neurons) emerge in a network with initially random connectivity, due to spontaneous activity.

Plasticity in homogeneously connected assemblies

In this section, we will argue that the assemblies in our model grow if their size is small and that their final size is limited. We explain the latter by the depression dominance of the learning rule and imprecisely correlated spiking dominating the plasticity in large assemblies. To model assembly weight dynamics, we consider the special case of an isolated assembly of N neurons that is homogeneously connected with synaptic

strengths $\hat{w}$. We compute the time-averaged weight change

$$\overline{\Delta W_{ij}} := \frac{1}{T}\int_t^{t+T} \mathrm{d}t' \; \frac{\mathrm{d}}{\mathrm{d}t'} W_{ij}(t') \tag{13}$$

(cf. Eq. 7) for weights within this type of assembly, disregarding the clipping at $\hat{w}$. This indicates whether and how vigorously weights that fall below $\hat{w}$ are restored towards it. Further, it indicates how successfully further neurons are recruited, as recruitment relies on increasing the weights to new neurons once they are randomly increased. We expect that the assembly size is limited by values of N above which $\overline{\Delta W_{ij}}(N)$ is negative: Larger assemblies will lose neurons due to the depression of weights, to other, smaller assemblies that tend to potentiate their connections with these neurons. This yields as an estimate for the size N of assemblies the solution of $\overline{\Delta W_{ij}}(N) = 0$. The assumption that an assembly loses neurons already if another assembly can potentiate its connections to them more yields as preferred assembly size the N that maximizes $\overline{\Delta W_{ij}}(N)$.

To analytically obtain $\overline{\Delta W_{ij}}$ as a function of the assembly size, we first compute from Eq. 3 the average rate of an assembly neuron

$$r = \lambda_0 \sum_{m=0}^{\infty} ((N-1)\hat{w})^m = \frac{\lambda_0}{1-(N-1)\hat{w}} = \frac{\lambda_0}{1-\rho}, \tag{14}$$

where $\rho = (N-1)\hat{w}$ is the branching parameter, which gives the average number of spikes that are directly induced by a single spike in the assembly. N is restricted to $N < 1 + \frac{1}{\hat{w}}$ because for larger values there is no stationary network state with finite firing rates. This is reflected by the divergence of the geometric series in Eq. 14. In other words, a stable network requires $\rho < 1$ [31, 42, 43]. Eqs. 4, 14 and 8 yield

$$\begin{aligned}\overline{\Delta W_{ij}}(N) = &\frac{2\lambda_0^2(A_\mathrm{p}\tau_\mathrm{p} + A_\mathrm{d}\tau_\mathrm{d})}{(1-(N-1)\hat{w})^2} + \\ &+ \frac{\lambda_0 A_\mathrm{p}\tau_\mathrm{p}\hat{w}\left[(2-(N-2)\hat{w})\tau_\mathrm{p} + (2-(N-2)\hat{w}-(N-1)\hat{w}^2)\tau_\mathrm{s}\right]}{(1+\hat{w})(1-(N-1)\hat{w})^2(\tau_\mathrm{s}+(1+\hat{w})\tau_\mathrm{p})(\tau_\mathrm{s}+(1-(N-1)\hat{w})\tau_\mathrm{p})} + \\ &+ \frac{\lambda_0 A_\mathrm{d}\tau_\mathrm{d}\hat{w}\left[(2-(N-2)\hat{w})\tau_\mathrm{d} + (2-(N-2)\hat{w}-(N-1)\hat{w}^2)\tau_\mathrm{s}\right]}{(1+\hat{w})(1-(N-1)\hat{w})^2(\tau_\mathrm{s}+(1+\hat{w})\tau_\mathrm{d})(\tau_\mathrm{s}+(1-(N-1)\hat{w})\tau_\mathrm{d})},\end{aligned} \tag{15}$$

see SI S.1 for details. The first right hand side term covers the impact of uncorrelated pre- and postsynaptic spiking with rates r_j and r_i on the synaptic strengths ($C_{ij}(\tau) = r_i r_j$ for uncorrelated spiking). The second and third terms, which are proportional to A_p and A_d, respectively, cover the weight potentiation and depression due to correlated spiking.

We now study under which conditions the weights of small assemblies tend to increase. For this we assume small $N \geq 2$, such that together with $\hat{w} \ll 1$, which holds for the parameters that we employ throughout our article, also $N\hat{w} \ll 1$. Eq. 15 then simplifies to

$$\begin{aligned}\overline{\Delta W_{ij}}(N) &\approx 2\lambda_0^2(A_\mathrm{p}\tau_\mathrm{p} + A_\mathrm{d}\tau_\mathrm{d}) + \frac{2\lambda_0 A_\mathrm{p}\tau_\mathrm{p}\hat{w}\left(\tau_\mathrm{p}+\tau_\mathrm{s}\right)}{(\tau_\mathrm{s}+\tau_\mathrm{p})^2} + \frac{2\lambda_0 A_\mathrm{d}\tau_\mathrm{d}\hat{w}\left(\tau_\mathrm{d}+\tau_\mathrm{s}\right)}{(\tau_\mathrm{s}+\tau_\mathrm{d})^2} \\ &= 2\lambda_0^2(A_\mathrm{p}\tau_\mathrm{p} + A_\mathrm{d}\tau_\mathrm{d}) + \frac{2\lambda_0 A_\mathrm{p}\tau_\mathrm{p}\hat{w}}{\tau_\mathrm{s}+\tau_\mathrm{p}} + \frac{2\lambda_0 A_\mathrm{d}\tau_\mathrm{d}\hat{w}}{\tau_\mathrm{s}+\tau_\mathrm{d}}.\end{aligned} \tag{16}$$

We require the right hand side to be larger than zero. To interpret this criterion, we rewrite it in terms of the integrals of the potentiating part $2A_\mathrm{p}\tau_\mathrm{p}$ and of the depressing

part $2A_\mathrm{d}\tau_\mathrm{d}$ of the STDP window and multiply by τ_s,

$$\left(\lambda_0\tau_\mathrm{s} + \frac{\hat{w}}{1+\tau_\mathrm{p}/\tau_\mathrm{s}}\right) 2A_\mathrm{p}\tau_\mathrm{p} > \left(\lambda_0\tau_\mathrm{s} + \frac{\hat{w}}{1+\tau_\mathrm{d}/\tau_\mathrm{s}}\right) 2\left|A_\mathrm{d}\right|\tau_\mathrm{d}. \tag{17}$$

The integrals $2A_\mathrm{p}\tau_\mathrm{p}$ and $2A_\mathrm{d}\tau_\mathrm{d}$ characterize the overall strengths of potentiation and depression. $\lambda_0\tau_\mathrm{s}$ and $\hat{w}$ weight the strength of spontaneous, always uncorrelated spiking within a synaptic time constant against the induced spiking by an input spike. $\frac{1}{1+\tau_\mathrm{x}/\tau_\mathrm{s}}$ describes how effectively the time constants allow induced spikes to translate to plasticity: Since we have $\hat{w} \ll 1$ and small N, low order connectivity motifs are most relevant for spike generation. These generate pre- and postsynaptic spikes with a small temporal delay on the order of τ_s. The evoked potentiation or depression is most effective if $2A_\mathrm{x}\tau_\mathrm{x}$ is concentrated to a narrow peak, i.e. τ_x is small. The effect saturates if τ_x falls below τ_s, because τ_s limits the precision of evoked correlated spiking.

For the parameters used in our paper, Eq. 17 is satisfied. It does not hold, however, for overly small $\hat{w}$, see our discussion in SI S.1 and Fig. S3. For the parameters used in our paper, although the overall strength of depression is larger than that of potentiation, the narrower peak of the latter ($\tau_p < \tau_d$) implies that potentiation will be more efficiently triggered by short-range spike correlations and therefore dominate in small networks where firing rates are lower and lower order connectivity motifs are more important.

To show that the size of assemblies is limited, we neglect N's discrete nature and consider the asymptotic behavior of $\overline{\Delta W_{ij}}(N)$ with N approaching $1+\frac{1}{\hat{w}}$. For this we insert $N = 1+\frac{1}{\hat{w}}$ in Eq. 15 except in the factor that leads to divergence,

$$\begin{aligned}
\overline{\Delta W_{ij}}(N) &\sim \left(\frac{2\lambda_0^2(A_\mathrm{p}\tau_\mathrm{p}+A_\mathrm{d}\tau_\mathrm{d})}{(1-(N-1)\hat{w})^2} + \frac{\lambda_0\hat{w}\left(A_\mathrm{p}\tau_\mathrm{p}+A_\mathrm{d}\tau_\mathrm{d}\right)}{\tau_\mathrm{s}(1+\hat{w})(1-(N-1)\hat{w})^2}\right) && (18)\\
&= f_0\left(\frac{\lambda_0^2}{(1-(N-1)\hat{w})^2} + \frac{\lambda_0\hat{w}}{2\tau_\mathrm{s}(1+\hat{w})(1-(N-1)\hat{w})^2}\right) && (19)\\
&= f_0\left(1+\frac{\hat{w}}{2\lambda_0\tau_\mathrm{s}(1+\hat{w})}\right)\frac{\lambda_0^2}{(1-(N-1)\hat{w})^2}, && (20)
\end{aligned}$$

where we used Eq. 10 to obtain the second line. The result implies

$$\lim_{N\to 1+\frac{1}{\hat{w}}} \overline{\Delta W_{ij}}(N) = -\infty \text{ if } f_0 < 0, \tag{21}$$

since the summands in the large bracket and the trailing factor are positive. Thus, if $f_0 < 0$ and there is an N for which $\overline{\Delta W_{ij}}(N)$ is positive, $\overline{\Delta W_{ij}}(N)$ will have a positive maximum and a zero at an N smaller than $1+\frac{1}{\hat{w}}$ (cf. also the phase diagram Fig. S3); we therefore expect limited assembly growth and assume $f_0 < 0$ throughout the article. In contrast, if $f_0 > 0$, $\overline{\Delta W_{ij}}(N)$ diverges to positive values for N where also the firing rate diverges, indicating that sufficiently large assemblies continue to grow until the network generates pathological activity.

The first terms in the large brackets of Eqs. 18-20 cover the impact of uncorrelated pre- and postsynaptic spiking. As the firing rates increase with N and $f_0 < 0$, the contribution of the STDP rule due to uncorrelated spiking tends to increasingly negative values. The fact that the impact of the second term, which encodes the effects of connectivity motifs on STDP (Eq. 11), also becomes negative for sufficiently large N is due to contributions from higher order motifs. We can see this by reconsidering

Eq. 11: For homogeneously connected assemblies, it simplifies to

$$\overline{\Delta W_{ij}} = f_0 r^2 + \frac{r}{N} \sum_{\alpha+\beta>0} f_{\alpha\beta} \left((N-1)^{\alpha+\beta} - (-1)^{\alpha+\beta} \right) \hat{w}^{\alpha+\beta} \tag{22}$$

$$= f_0 r^2 + \frac{r}{N} \sum_{k=1}^{\infty} f_k \left((N-1)^k - (-1)^k \right) \hat{w}^k, \tag{23}$$

where

$$f_k := \sum_{\alpha+\beta=k} f_{\alpha\beta}, \tag{24}$$

see SI S.2. Since $f_{\alpha\beta}$ (cf. Eq. 12) and thus f_k are independent of N, higher order terms in k grow in Eq. 23 faster with N than lower order ones. Therefore higher order motifs become more relevant for the weight change the larger N is. We find for high orders that

$$\lim_{k\to\infty} f_k = \frac{f_0}{2\tau_\mathrm{s}}, \tag{25}$$

see SI S.2 and Fig. S1; therefore high order motifs induce synaptic depression. This depression is impactful: For our sets of parameters, $\lambda_0 \tau_\mathrm{s}$, the average number of spontaneous spikes generated within the duration of a synaptic current, is considerably smaller than $\frac{\hat{w}}{2}$, half the average number of spikes generated by an incoming spike at a neuron (see Table 1). Further, $\hat{w}$ is considerably smaller than 1. The large bracket in Eq. 20 thus tells us that the imprecisely correlated spiking due to higher order motifs and not the overall increase in spike rate in the assemblies is mostly responsible for the decrease of $\overline{\Delta W_{ij}}(N)$ for $N \to 1 + \frac{1}{\hat{w}}$.

Importantly, the relation Eq. 25 holds generally, for any plasticity window F, not only for the considered symmetric one. We explain it as follows: A term with specific exponents α and β in Eqs. 11, 22 covers the contribution of a specific connectivity motif to the spike correlation. This motif consists of a common presynaptic neuron that is separated from neurons j and i by a chain of β and α connections, respectively [9], see Fig. 3c for an illustration. Higher orders of $k = \alpha + \beta$ thus encode the effects of long cascades of spiking activity in an assembly. The longer these cascades get, the more spread-out and Gaussian the temporal distribution of their impacts becomes, due to the summation of inter-spike-intervals and their jitters, see Fig. S1. (We note that since the considered model is linear, one can indeed attribute the generation of each spike to a precursor spike or to the external drive.) Eq. 24 homogeneously sums over motifs with different connection chain lengths to the pre- and the postsynaptic neuron. For sufficiently large k the evoked spike time differences are thus equidistant, broad, overlapping Gaussians over the STDP window, which leads to the dependence on the integral of the STDP function f_0, as for uncorrelated pre- and postsynaptic spiking.

The broadening of temporal spike correlations for $N \to 1 + \frac{1}{\hat{w}}$ and its impact on $\overline{\Delta W_{ij}}(N)$ can also be observed from the interdependence term of $\tilde{C}(\omega)$ (cf. Eq. 4) without expanding it. SI S.2 (last paragraph) shows that it tends in the time domain to a two-sided exponential with smaller and smaller decay time constant. In the limit there is no more decay; the correlation function becomes flat. This implies together with the flat time domain correlation function of Poisson spike trains (cf. the first term of Eq. 4) and Eq. 7 that the weight change $\overline{\Delta W_{ij}}(N)$ only depends on the integral f_0 of F.

Fig. 3a shows $\overline{\Delta W_{ij}}(N)$ as in Eq. 15 for different values of $\hat{w}$: $\overline{\Delta W_{ij}}(N)$ is maximal for a particular assembly size, then crosses zero and becomes increasingly negative. We confirmed these results using simulations of homogeneous assemblies, in which the weight updates that would occur due to STDP without clipping are tracked but not applied and then averaged over time. In Fig. 3b we show the typical assembly sizes that

emerge for different values of $\hat{w}$ in network simulations of spontaneous assembly formation. We find dense but also sparser assemblies where some internal connections are weak or basically missing, see, for instance, the third assembly in Fig. 2. For these we expect the maxima and zero crossings of $\overline{\Delta W_{ij}}$ to be at larger sizes compared to our estimates with fully connected assemblies. Consistent with this, we observe that sparser assemblies in our simulations tend to be larger. In Fig. 3b, we correct for the effect of missing connections by taking the sum of the intra-assembly connections of an assembly, w_{sum}, and computing the size $N_{\text{corrected}}$ that a homogeneously connected assembly with the same w_{sum} and weights $\hat{w}$ would have,

$$N_{\text{corrected}} = \frac{1 + \sqrt{1 + 4w_{\text{sum}}/\hat{w}}}{2}. \tag{26}$$

The corrected assembly sizes closely match the ones predicted from the zeros of $\overline{\Delta W_{ij}}(N)$, while the prediction from the maxima of $\overline{\Delta W_{ij}}(N)$ is farther off. This indicates that the assemblies grow until further growth would result in a decrease of the intra-assembly weights. The corrected assembly sizes lie below and scale differently than the upper limit for fully connected assemblies, $N = 1 + \frac{1}{\hat{w}}$. This is consistent with the theoretical prediction that assemblies reaching this size would quickly lose intra-assembly weight because $\overline{\Delta W_{ij}}(N) = -\infty$.

The positive part of the $\overline{\Delta W_{ij}}(N)$ curve shows that for our parameters the potentiating STDP part dominates for small and intermediate N. It reflects the spontaneous formation of assemblies due to precise spike correlations as in previous work and their tendency to grow bigger for small N, see SI S.1. We note that Eq. 15 may suggest to take the limit of $\hat{w} \to 0$ in order to simplify it, or to assume that the dynamical quantity N goes to infinity, which requires that the system parameter $\hat{w}$ becomes infinitesimally small to keep the rates finite for all admissible N (cf. Eq. 14). $\hat{w} \to 0$, however, implies that $\overline{\Delta W_{ij}}(N)$ becomes negative for all N, since the rate-rate-interaction term dominates, see SI S.1, last paragraph, and Fig. S3. Therefore our theory predicts that no assemblies form in this limit.

If we truncate our theory to correlations evoked by direct weight connections and common presynaptic neurons [44] (i.e. to the terms $f_0, f_{10}, f_{01}, f_{11}$ in Eq. 11), $\overline{\Delta W_{ij}}(N)$ also has a maximum, crosses zero and then tends to negative infinity. This decrease is a purely rate-based effect: it occurs because the magnitude of the rate-rate interaction term proportional to f_0 increases quadratically with the firing rate (the other terms are positive and linear in the rate). We compare two examples of $\overline{\Delta W_{ij}}(N)$ from the truncated theory with their full theory counterparts in Fig. S2. For the larger value of $\hat{w}$, the maximum and zero of the truncated theory curve occur at higher values of N compared to the full theory. In contrast, leaving out the rate-based term $f_0 r_i r_j$ in Eq. 11 has only little impact. For the smaller value of $\hat{w}$ both modifications lead to shifts that are significant and similar; the curve without rate based term starts to drop faster than the truncated one only beyond the zero. These observations show the importance of imprecisely correlated spiking for limiting the assembly growth through the higher order motif- and the rate-based term in Eq. 11. They are consistent with the dominance of the motif-based term in the asymptotics Eq. 20, which is more pronounced for larger $\hat{w}$.

We conclude that the basic mechanism of spontaneous assembly formation due to activity-dependent plasticity, which we described in the previous section, remains in our model. For the limitation of assembly size, however, we uncover a novel mechanism: temporally imprecisely correlated spikes dominate the plasticity in large assemblies and do not allow them to grow further. The imprecision results from the long pathways between a spike and the spikes that it indirectly generates. The depressing effect of the imprecisely correlated spikes is covered by the negative rate-rate interaction term and by the negative total contribution of higher order motifs; both are proportional to f_0.

Eq. 15 includes their impact and quantitatively describes how the size of spontaneously emerging assemblies remains limited in our model.

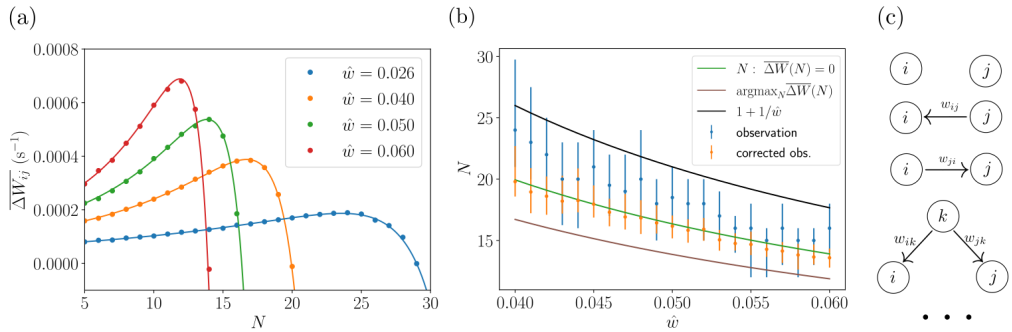

Fig 3. Explanation of assembly formation. (a): Time-averaged weight change (tracked, not applied) for a synapse in an assembly (fully connected, no other connections, all weights fixed at $\hat{w}$) as a function of its size N, for different values of $\hat{w}$. Solid lines: theoretical prediction, dots: numerical results from fixed homogeneous assemblies. (b): Sizes of spontaneously forming assemblies in network simulations with different $\hat{w}$ (blue dots: medians, error bars: first and third quartiles) and corresponding sizes of assemblies with homogeneous coupling of strength $\hat{w}$ computed using a correction for sparseness (orange). The lower continuous curves show analytical predictions for the sizes of homogeneously connected assemblies obtained from Eq. 15, namely the solutions of $\overline{\Delta W_{ij}}(N) = 0$ (green) and $\text{argmax}_N(\overline{\Delta W_{ij}}(N))$ (brown). The upper curve displays the analytical upper limit of the homogeneously connected assembly sizes, $N = 1 + \frac{1}{\hat{w}}$ (black). (c): Illustration of zeroth, first and second order contributions to the time-averaged weight change for a given rate and weight configuration. Circles and arrows represent neurons and synaptic connections, respectively.

Storing new assemblies

Next we show how assemblies can be learned via correlated feedforward input. Neurons recruited for a new assembly may previously have only weak connections, i.e. they may belong to a background of neurons in front of which assemblies exist. Alternatively recruited neurons may already be part of other assemblies. In our simulations, during the learning phase, each neuron that is to be recruited receives Poisson spike input from the same source. This stimulates the neurons to spike near-simultaneously, such that the weights between them grow. Fig. 4 displays the resulting weight and spiking dynamics in a network which prior to stimulation hosts one assembly in front of a background of weakly coupled neurons. Before time $t = 0$s, the single assembly is stably stored in the network, despite the ongoing plasticity. At $t = 0\,\text{s}$ a group of background neurons receives correlated stimulation for $T_{\text{stim}} = 180\,\text{s}$. This leads to the formation of a rudimentary assembly. After the stimulation has ended, its synapses grow over a longer period of time until a fully connected assembly is reached, where all interconnections have synaptic strength close to $\hat{w}$. The remaining background neurons, which did not receive correlated stimulation in the beginning, do not form assemblies. The stimulation furthermore does not affect the already existing assembly and neither does this assembly interfere with the formation of the new assembly. Fig. S6 shows another example of assembly learning; there again mostly background neurons are recruited, but also one neuron that is already part of two pre-existing assemblies.

Assembly creation requires that there is net potentiation $\overline{\Delta W_{ij}} > 0$ of the weights between the stimulated neurons. Using this condition and assuming that initially the

weights between assembly neurons are zero, Eq. 11 yields an approximate analytical lower bound for the required strength of the external input given the input rate, see SI S.3 for details. Numerical simulations confirm this bound and show that noise allows assemblies to form already for smaller weights, see Fig. S4.

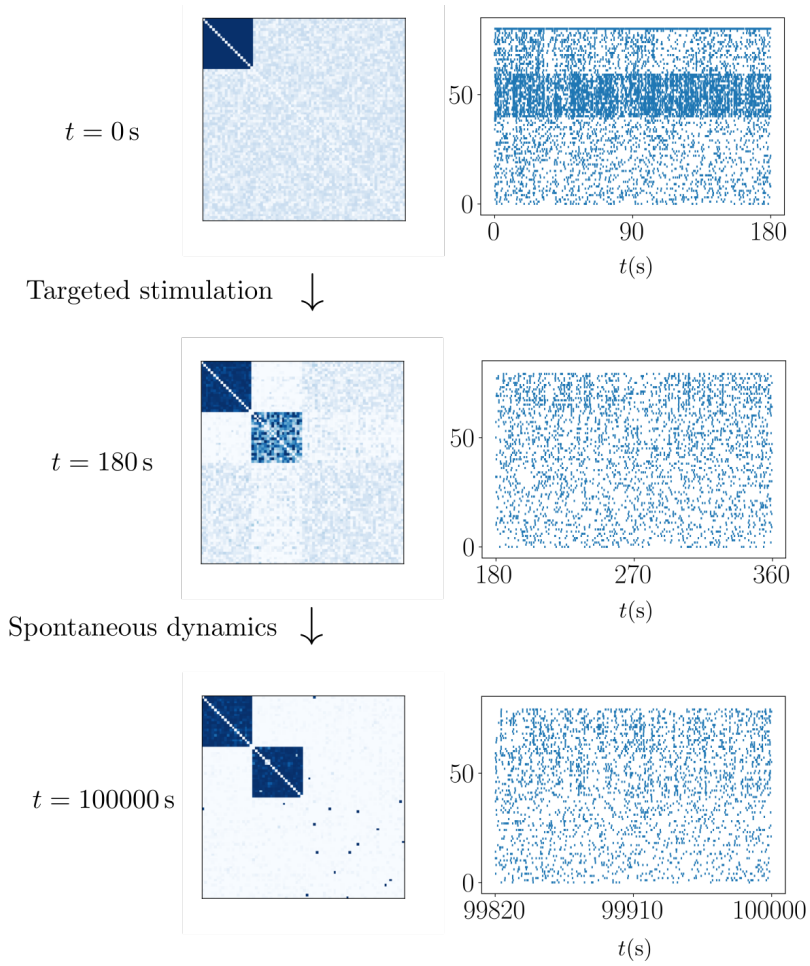

Fig 4. Learning a new assembly in a network starting with a single assembly and otherwise weak random connectivity. Initially, correlated external input stimulates the neurons of the new assembly, which leads to an increase of their interconnecting weights and thereby to the formation of a rudimentary assembly. The stimulation is then turned off and the network evolves on its own; the new assembly becomes fully connected.

Overlapping Assemblies

Experiments indicate that neurons can code for more than one memory item [21–23]. For assembly models of memory this implies that neurons can be part of more than one assembly, see Fig. 5a, i.e. assemblies have overlaps [45]. Such overlaps may encode associations between memories.

We find that for appropriate parameters, our networks can stably store assemblies with some overlaps, see Fig. 5b for an example weight matrix. However, we also observe that overlaps present a challenge for our models: overlapping assemblies either tend to merge or overlap neurons tend to disconnect from one of the assemblies they are a part of, dissolving the overlap. The first case occurred especially when the sizes of the

overlapping assemblies were significantly smaller than the sizes that would lead to $\overline{\Delta W_{ij}}(N) = 0$. We hypothesize that in this case a feedback loop arises: overlaps induce correlations between the assemblies, which facilitates the formation of additional overlaps until the assemblies have completely fused. The other case occurred when the sizes of the two assemblies were close to $\overline{\Delta W_{ij}}(N) = 0$. We hypothesize that here the higher firing rate of an overlap neuron and the resulting increased synaptic depression imply that overlaps become more likely to disappear due to fluctuations in the weight dynamics.

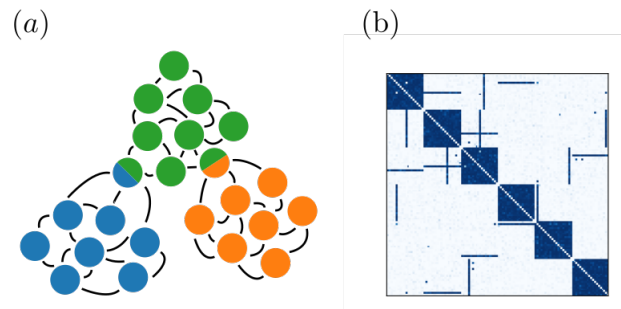

Fig 5. Overlapping assemblies. (a): Illustration of overlaps. (b): Weight matrix of a network with overlapping assemblies. The base assemblies (diagonal blocks) are each 20 neurons in size.

The hypotheses about the mechanism underlying the problems with overlap suggest two solutions how networks may solve them: First, if the synaptic long-term depression that occurs at high rates limits the connectivity of neurons, a neuron with a significantly lower spontaneous rate may be able to connect to multiple assemblies at the same time. In Fig. 6 we show how neurons with lower spontaneous firing rate join a second assembly. In panel (a) a neuron with lower spontaneous rate is initially already partially connected to another assembly. This partial connection then causes enough correlation with the second assembly to induce the completion of the connections. In this case both the initial partial connection and the lower λ_0 were necessary to create this overlap. For an even lower λ_0, an overlap emerged spontaneously, with a randomly selected assembly, through the inherent stochasticity of the dynamics, as shown in Fig. 6b.

We note that we use a simple analytical rule to determine the spontaneous rates of candidate neurons for the formation of overlaps. To obtain it, we consider two fully connected assemblies of size N. We assume that the total firing rate of an overlap neuron, once it is connected to both assemblies, equals the firing rates of the typical neurons, which are only part of a single assembly. If there is an overlap, the firing rates in one assembly will, via the overlap neuron, have a small influence on firing rates in the other assembly. If we neglect this influence, typical neurons' rates are given by Eq. 14, $r = \frac{\lambda_0}{1-(N-1)\hat{w}}$. The excess rate $r - \lambda_0 = \frac{\lambda_0}{1-(N-1)\hat{w}} - \lambda_0$ beyond the spontaneous rate λ_0 is thereby due to input from the other neurons in the assembly. An overlap neuron x receives twice the input of a typical neuron, because it is part of two assemblies. Therefore its excess beyond its spontaneous rate is $2\left(\frac{\lambda_0}{1-(N-1)\hat{w}} - \lambda_0\right)$. If we call its baseline rate $\lambda_{x,0}$, we obtain for the total rate of an overlap neuron

$$r_x = \lambda_{x,0} + \frac{2\lambda_0}{1-(N-1)\hat{w}} - 2\lambda_0. \tag{27}$$

Requiring $r_x \overset{!}{=} r = \frac{\lambda_0}{1-(N-1)\hat{w}}$ results in

$$\lambda_{x,0} = \lambda_0 \frac{1-2(N-1)\hat{w}}{1-(N-1)\hat{w}}. \tag{28}$$

We find that with this spontaneous rate, spontaneous overlap formation is possible and that firing rates for overlap neurons are in good agreement with our predictions, see Fig. S5. The exact threshold for possible spontaneous assembly formation is, however, also dependent on appropriate learning rates and the randomness in the network.

A second way of sustaining overlaps is by having most or all neurons be part of more than one assembly. In this case fusion of two assemblies appears less likely, because all assemblies and neurons are similarly saturated. Indeed, in the brain, we expect most or

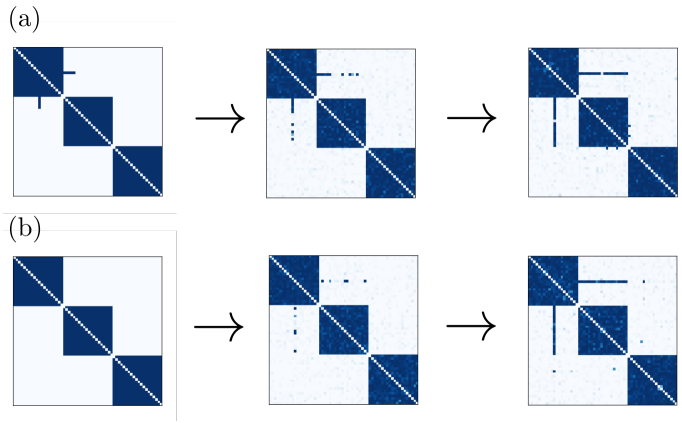

Fig 6. Creation of overlaps. (a): A partial overlap self-completes over time due to plasticity. (b): An overlap emerges for a neuron with lower spontaneous firing rate.

all neurons to be part of more than one assembly [21, 45]. To test whether our networks are capable of sustaining similarly prominent overlap structures, we consider a network of intertwined assemblies in which each neuron is part of exactly two assemblies. We choose equally sized assemblies of size n_{A} and the overlap structure such that for any given assembly each of its neurons is shared with a different assembly. In other words, the overlap between any two assemblies consists of one neuron or a fraction of $1/n_{\mathrm{A}}$ of each assembly. This implies that there are in total $n_{\mathrm{A}} + 1$ assemblies and that the setup requires a network with $N = n_{\mathrm{A}}(n_{\mathrm{A}} + 1)/2$ neurons. Notably, a network with the same number and size of assemblies, where each neuron is only part of one assembly (no overlap), would require twice as many, $(n_{\mathrm{A}} + 1)n_{\mathrm{A}}$, neurons. Fig. 7 shows that this structure is stable under the STDP rule over long timescales. The assemblies have size $n_{\mathrm{A}} = 19$; the typical (see the numerical results below) overlap between two assemblies is thus $1/n_{\mathrm{A}} = 1/19 \approx 5.3\%$. Random correlations cause additional, extra-assembly connections to appear (more precisely: to strengthen) and intra-assembly connections to disappear (to weaken). To demonstrate that these deviations of individual weights are transient, we have tracked in Fig. 7b and c the strengthening of extra-assembly weights and the weakening of intra-assembly weights, respectively. The upper panel of Fig. 7b displays the sum of the extra-assembly weights over time, indicating that there is no positive drift, i.e. no overall increase of weights that do not belong to the stored assembly structure over time. The lower panel displays the Pearson correlations of the extra-assembly weights over time with the extra-assembly weights observed at five reference time points. With increasing distance to the reference time point, the correlations decay to chance level showing that no persistent pattern of strengthened extra-assembly weights emerges. Fig. 7c analogously displays the differences between the maximal (optimal) and the actual intra-assembly weights. The upper panel shows that there is no overall increase of this difference, i.e. no overall decay of the assembly structure. The lower panel shows that the patterns of weakened intra-assembly weights are transient.

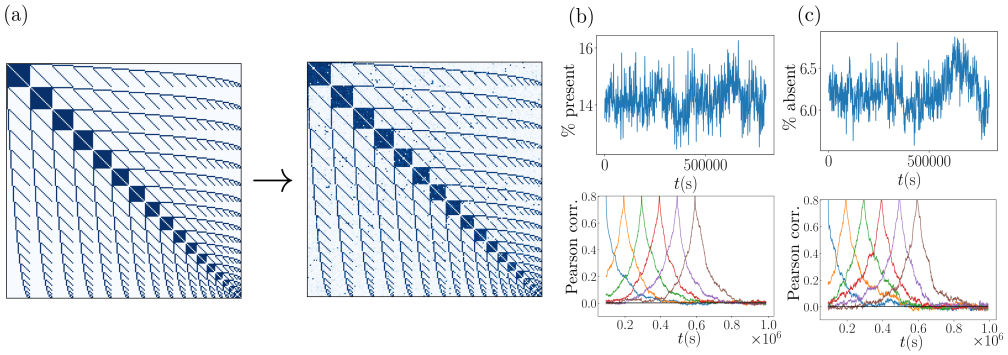

Fig 7. Intertwined assembly network. (a): Initial weight matrix (left hand side) and weight matrix after simulation over 1×10^6 s (right hand side). (b, upper): Sum of the extra-assembly weights, that is of those weights that are zero initially and if the assembly structure is optimally realized. The sum is given as a fraction of the total initial weights. (b, lower): Correlations of the extra-assembly weights with those at six different reference points in time. (c,upper): Sum of the absent intra-assembly weights, that is, the sum of distances of intra-assembly weights from their initial and optimal value $\hat{w}$. The sum is given as a fraction of the total initial weights. (c,lower): As in (b,lower) for absent intra-assembly weights.

Drifting assemblies

Experiments have shown that memory representations need not consist of the same neurons over time but can in fact exchange neurons without affecting behavior [25], a phenomenon called representational drift. It may occur because memory assemblies drift, by gradually exchanging neurons between each other [5]. The gradual exchange implies that at each point in time, each assembly is present and unambiguously identifiable by following the course of its evolution from the beginning. In the following, we show that our model networks can give rise to drifting assemblies. The drift happens due to two alternative mechanisms: (i) Neuron exchange between assemblies due to high weight plasticity noise, as in [5], and (ii) formation of temporal overlaps due to modulations in the spontaneous spike rate. Whether drift occurs due to mechanism (i) is chiefly determined by the learning rate μ: Fig. 8a displays the assembly dynamics in two networks with different values of μ, while all other parameters are kept the same. The network with smaller μ has stationary assemblies. In contrast, the network with larger μ exhibits drifting assemblies. Specifically, in this network the ensemble of neurons forming an assembly completely changes over time (Fig. S7a). Simultaneously, one can track the identity of an assembly by comparing its constituent neurons over short time intervals; the neurons forming it at one time can be unambiguously matched to the neurons forming it shortly thereafter, since the difference between those two ensembles is still small (Fig. S7b). We can explain the occurrence of this type of drift as an effect of fluctuations in the weight dynamics. The time-averaged weight dynamics as described by $\overline{\Delta W_{ij}}$ reinforce connections between neurons in the same assembly while suppressing connections between neurons of different assemblies. However, large enough fluctuations in the weight dynamics may nevertheless cause neurons to lose connections within their assembly and form connections to other assemblies. The size of these fluctuations is governed by μ, and if they are sufficiently large there is a finite probability that a neuron switches assemblies, see [5] for a detailed discussion. If μ is too large, the strong fluctuations prevent assemblies from forming at all.

In the second mechanism switching of assemblies by neurons is a two-step process. In the first step a neuron with a sufficiently low λ_0 spontaneously forms a connection to

an additional assembly, as in Fig. 6b. Then, when λ_0 increases again this neuron loses its connections to one of the assemblies. The neuron can thereby leave either of the assemblies it is connected to – if it loses its connections to the assembly that it was originally connected to, it has switched assemblies. If this happens sufficiently often for many neurons in the network, this also causes overall drift on a slow timescale, as shown in Fig. 8b.

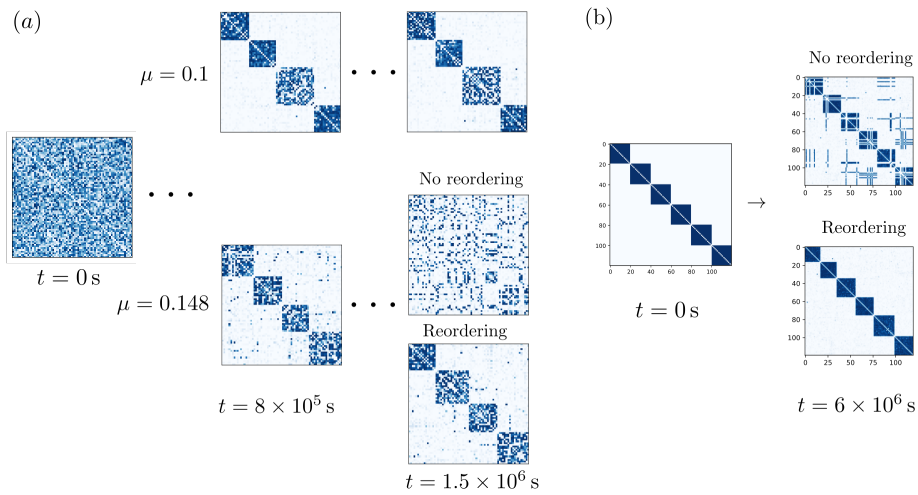

Fig 8. Drifting assemblies. (a): Drift through noisy spiking activity and resulting weight fluctuations. Assembly dynamics for lower ($\mu = 0.1$, upper part) and higher ($\mu = 0.145$, lower part) learning rate. The panel shows connection matrices W at different points in time. Initially, the weight matrices are random (left). After $t = 8 \times 10^5$ s assemblies have spontaneously emerged (middle). The neuron indices are sorted to reveal them (cf. Fig. 2). The simulation is thereafter continued, which shows that for lower learning rate the assemblies are static. In contrast, for higher learning rate the assemblies drift. Since the assemblies exchange neurons, the coupling matrix appears increasingly unstructured. Reordering of the indices, however, reveals that the assemblies are maintained at each point in time. (b): Drift through transient changes in intrinsic neuron properties and resulting transient overlaps. The spontaneous firing rates of neurons change on a slow timescale. This leads to the transient appearance of overlaps, cf. Fig. 6b. When an overlap vanishes, the neuron randomly decides for one of the assemblies, which leads to drifting assemblies.

Changing synaptic connectivity

We finally apply our assembly model to networks that undergo changes in their synaptic connectivity. This is motivated by processes related to aging: Anatomical studies have shown that the aging cortex is characterized by a decrease in the number of synaptic spines [26] and presynaptic terminals [27]. These changes may be interpreted as an increase in sparsity of connections between neurons and/or as an overall weakening of synaptic strengths (since connections between neurons often consist of multiple synapses). We can model the former by permanently setting a fraction of entries of the weight matrix, chosen at random, to zero. The latter can be modeled by decreasing $\hat{w}$. Fig. 3a shows the effect of decreasing $\hat{w}$: it shifts the zeros of $\overline{\Delta W_{ij}}(N)$ and thus the typical assembly size towards larger N. Increased sparsity similarly lowers the branching parameter for an assembly of a given size and thus also shifts the zeros of $\overline{\Delta W_{ij}}(N)$ and the characteristic assembly size towards larger N, see Fig. 9a. As a result,

spontaneously forming assemblies are larger in networks with higher sparsity, see Fig. 9c. In addition Fig. 9c suggests that smaller learning rates further increase the tendency to form larger assemblies. We observe that if the assembly sizes in a network are significantly smaller than the characteristic size predicted by Figs. 3a and 9a, assemblies will merge to form larger ones, see Fig. 9b. This represents a loss in memory capacity. Assuming that in the brain assemblies representing closely related memories merge (due to existing overlaps), during this process the overall memory content becomes less detailed and differentiated. At the same time the neuronal activity during recall increases due to larger assembly sizes indicating a less efficient use of neural resources.

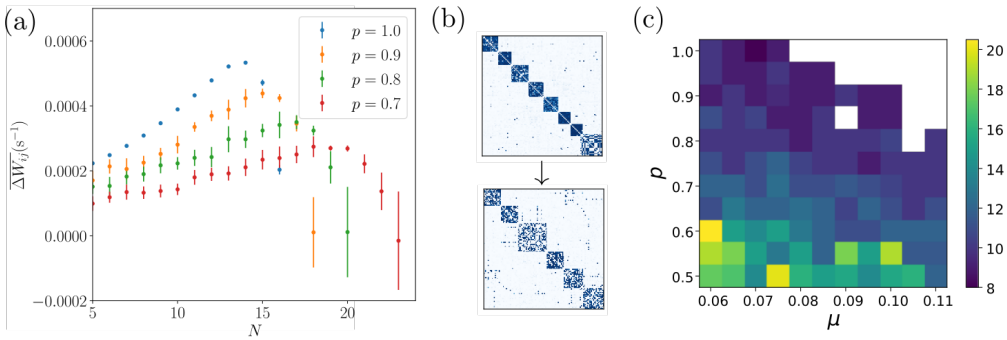

Fig 9. Effects of increasing network sparsity. (a): Time-averaged weight change like in Fig. 3a, of an assembly with intra-assembly connection probability p, as a function of its size for different values of p. Lower values of p lead to a shift of the zero crossing of $\overline{\Delta W_{ij}}(N)$ to larger N. Simulations with low $\overline{\Delta W_{ij}}$ converge slowly, which is reflected by large error bars. (b): An assembly structure becomes coarser (fewer, larger assemblies) as the connectivity sparsity increases. The panel shows the weight matrices of a network in which the probability p that a connection between neurons exists is decreased from $p = 1.0$ (upper subpanel, after spontaneous assembly emergence) to $p = 0.7$ (lower subpanel, after re-equilibration). (c): Dependence of the size of a typical assembly on the network connection probability p and the learning rate μ. The panel shows the median size of spontaneously forming assemblies in networks with random initial connectivity. White areas denote parameter regions for which activity becomes pathological, that is, firing rates diverge.

Discussion

We have studied assemblies in networks with plasticity that is purely spike-timing dependent. We find that the networks stably store assemblies, which may spontaneously emerge or be learned. Further, the assemblies can have overlaps, spontaneously drift and adapt to changing network properties.

In biological spiking neural networks the change of synaptic efficacies depends to a large extent on STDP, where depression and potentiation are a function of the time lags between pre- and postsynaptic spikes [46–48]. Assemblies may be kept up by the co-spiking of their member neurons, which strengthens their interconnections. If the total input to a neuron is not constrained, large assemblies generate large inputs to each member neuron and thereby highly reliably co-activate them. Furthermore, larger assemblies generate stronger input to other neurons and thereby tend to recruit them, such that these assemblies grow. This generates a positive feedback loop, which can without restraining mechanisms lead to excessive assembly growth. Previous work on assembly networks therefore usually added additional homeostatic plasticity that limits the total input strength to a neuron by fast normalization [5, 7–9, 11–13]. This curtails

the resulting excessive assembly growth. The homeostatic plasticity observed in neurobiological experiments is, however, much slower than that required to prevent runaway assembly growth [14–18].

In the present work, we therefore studied assemblies in networks with purely STDP-based plasticity. We find that our depression-dominated STDP-rule (the integral of the STDP window is negative) restricts the growth of assemblies by two mechanisms: On the one hand, the spike rate of the neurons in an assembly grows with the assembly size, which increases the depressing effect of the rate-based term of the time-averaged weight change. On the other hand, the contributions of the higher-order connectivity motifs to the time-averaged weight change become more significant the larger the assemblies are, because longer cascades of spikes become more likely. Further for sufficiently high order the appropriately summed contributions of the motifs are approximately proportional to the integral of the STDP window and thus negative. The contributions of the purely firing rate-based term and the higher-order motifs therefore reduce or even invert the tendency of weights to grow in large assemblies. As a consequence, neurons are more prone to leave a larger assembly and for example join another, smaller one, which induces larger growth of the interconnecting weights. A mechanism similar to the purely firing rate-based effect that we have described has been shown to stabilize the output spike rate of a single neuron receiving feedforward input, since a strong dominance of the rate-based term ultimately leads to the reversal of the sign of plasticity [49]. Another related work, [10], demonstrated the learning of static, non-overlapping assemblies in networks with STDP in the recurrent excitatory and the inhibitory-to-excitatory connections. These assemblies are maintained at least over several hours. The excitatory STDP is thereby balanced (the integral over the STDP window is about zero) such that the inhibitory STDP is effectively much faster. This yields rate homeostasis in the excitatory neurons. In our models the integral over the excitatory STDP curve is negative. Like rate homeostasis, this restricts the maximal average excitatory spike rates. In contrast to rate homeostasis, it also allows for smaller weights, for example in our networks with assembly and background neurons. We note that the fast weight homeostasis in [9] is implemented such that it also only constrains the maximal summed input and output weight of a neuron. We further note that highly diverse average spike rates are observed in biology [50].

We have investigated networks with symmetric plasticity function, motivated by past theoretical work [5, 9] and experimental findings [39, 40]. Experiments have, however, also found strongly asymmetric STDP windows and a dependence of plasticity on details of the neural activity beyond the temporal difference between pairs of spikes [47, 51, 52]. Recent theoretical studies showed assembly formation with such STDP rules and homeostasis [10, 12, 20]. Since the contributions of the higher order network motifs converge generally, for any STDP function, towards a multiple of the integral over the STDP function (cf. Eq. 25), we expect that the limitation of assemblies by our mechanism should also work rather generally. Prerequisites are irregular activity and an STDP function that allows for the growth of small assemblies and has negative integral. It will be important to see whether our mechanisms of assembly growth restriction indeed applies also to networks with strongly asymmetric and more biologically detailed STDP rules.

Our model networks generate irregular, probabilistic spiking activity. This is in agreement with experimentally observed irregular spiking in the cortex. In our models the irregularity of spiking is guaranteed by the usage of a Poisson spiking neuron model. In biological neural networks, it is usually assumed to result from the input fluctuation-driven spiking dynamics of individual neurons, which occur if there is an overall balance of excitatory and inhibitory input. If the excitatory synaptic weights are plastic, as in our model, this balance might be maintained by inhibitory

plasticity [6–8, 10]. In the future, it will be important to investigate our mechanisms of assembly growth restriction in networks with explicitly modeled excitatory and inhibitory populations of neurons. Individual neurons then should not intrinsically generate Poisson spiking, but for example be of the integrate-and-fire type.

We observe that our networks are able to maintain prominent overlap structures, where each neuron belongs to more than one assembly. The assembly structure is saturated and remains stable. In particular, additionally increased connections are sparse and transient. Previous related works have not shown similarly prominent overlap structures [5, 7–10, 12, 24], perhaps because the mostly assumed fast homeostatic normalization induces a stronger competition between the assemblies. This may force neurons to decide for one assembly. Networks with overlaps allow a more economic use of neural resources, in the sense that more assemblies of a specific size can be stored, in our example network twice as many as without overlaps. The overlap between assemblies in our simulations is about 5%. This agrees with the overlap estimated for assemblies representing associated concepts in the brain [45]. The overlap of randomly chosen assemblies is smaller, about 1%. Our findings raise a variety of future research questions. We have demonstrated highly prominent overlap structures with symmetric assembly patterns, where any two assemblies have an overlap and all assemblies as well as all overlaps have the same size. Numerical simulations indicate that we can still store similar patterns as these with a few overlaps missing. Future research should address whether and how assembly patterns may be stored that have a similar asymmetry and overlap pattern as estimated for the brain. Inhomogeneous neuron properties such as different spontaneous spike rates may help to achieve this. In this context, it will also be important to investigate how new assemblies can be stored with the right overlaps into a pattern of old ones. Finally, it is unclear whether symmetric or asymmetric assembly patterns with highly prominent overlaps may spontaneously emerge in neural networks. Narrowing down the focus, it will be interesting to study how the overlap between two assemblies evolves and under which conditions it becomes stable. A theoretical description of this evolution needs to account for the back-and-forth interaction between assemblies with dynamically changing overlap. Ref. [45] has recently studied interactions between fixed overlapping assemblies in non-plastic networks. On the microscopic level, our results suggest to investigate in more detail when neurons connect to two assemblies. To address this, one might analytically compute the probabilities of connection weights of a test neuron with two assemblies similar to the description of neuron switching between assemblies in [5]. In contrast to the effectively one-dimensional problem addressed in [5], the absence of homeostatic weight normalization will yield a two-dimensional reduced weight space for our model.

Our model networks can stably maintain assemblies in front of a background of neurons that are not part of any assembly. Such a scenario might be particularly relevant for early development when not many memories have been stored yet. Previous related works usually assume that the entire space is tiled by assemblies [5, 7–10, 12]. The reason may be similar as for the prominent overlap structures, namely that fast homeostatic plasticity has a strong tendency to force neurons into assemblies. Ref. [24] showed assemblies in front of a background of weakly connected neurons in networks with structural plasticity and multiple synapses per connection between neurons.

The assemblies in our networks can drift. Such assembly drift may explain the experimentally observed drift of memory representations in the brain [5, 25]. In our models, assemblies drift by exchanging neurons. In the first model this exchange originates from sufficiently large synaptic weight fluctuations due to noisy spiking, as in [5]. Besides STDP, ref. [5] had to assume the presence of fast weight homeostasis, which helps to keep the representational structure intact. We note that assemblies exchange neurons also in a recent model for spontaneous assembly formation in

zebrafish [13], which assumes fast homeostasis as well. The assemblies, however, break down and merge, and are therefore not drifting. In contrast, both our models for drifting assemblies require only STDP as plasticity mechanism. Our second model furthermore introduces a novel neuron exchange and thus drift mechanism: For high spontaneous rate, each neuron belongs to one assembly. If the spontaneous rate of a neuron in the network then transiently drops, the synaptic weights with another assembly increase, such that the neuron belongs to two assemblies. When the intrinsic rate recovers, the synaptic weights with one of the assemblies weaken. This may eliminate the strong weights with the assembly that the neuron originally belonged to and thus induce a switch to the other assembly. The observation suggests a new general mechanism for representational drift in the brain: Transient (or persistent) changes in single neuron properties may lead to changes in the synaptic weights. These in turn induce a change in the features represented by the neuron. The changes in the synaptic weights and the representation may thereby be much longer lasting than the changes in the intrinsic neuronal properties.

Since in the aging brain the overall connectivity decreases [26, 27], we have explored the impact of a such a decrease on the assemblies in our model networks. We observe that the size of assemblies is inversely related to the connection probability in the network. We expect that a similar relation can be observed rather generally, for example also in model networks where the assemblies are stabilized by fast homeostasis. In particular, sparser networks should, all other parameters being equal, lead to larger assemblies that are activated during recall. This is consistent with the observation that neural activity for the same task is stronger in aged individuals [53], where the neural networks are more sparsely connected. An increased assembly size and the resulting stronger activity during reactivation might also explain why episodic memories are experienced more vividly in elderly subjects [54]. In our model networks, we observe that assemblies merge to larger ones if networks become increasingly sparse. Such mergers might explain why episodic memories become less detailed in the aging brain [54, 55].

Acknowledgments

We thank Sven Goedeke for helpful comments on the manuscript.

References

1. Buzsáki G. Neural syntax: cell assemblies, synapsembles, and readers. Neuron. 2010;68(3):362–385.

2. Gerstner W, Kistler WM, Naud R, Paninski L. Neuronal dynamics: From single neurons to networks and models of cognition. Cambridge University Press; 2014.

3. Holtmaat A, Caroni P. Functional and structural underpinnings of neuronal assembly formation in learning. Nat Neurosci. 2016;19(12):1553.

4. Pietri T, Romano SA, Pérez-Schuster V, Boulanger-Weill J, Candat V, Sumbre G. The emergence of the spatial structure of tectal spontaneous activity is independent of visual inputs. Cell Rep. 2017;19(5):939–948.

5. Kalle Kossio YF, Goedeke S, Klos C, Memmesheimer RM. Drifting assemblies for persistent memory: Neuron transitions and unsupervised compensation. Proc Natl Adac Sci U S A. 2021;118(46).

6. 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.

7. Litwin-Kumar A, Doiron B. Formation and maintenance of neuronal assemblies through synaptic plasticity. Nat Commun. 2014;5(1):1–12.

8. Zenke F, Agnes EJ, Gerstner W. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. Nat Commun. 2015;6(1):1–13.

9. Ravid Tannenbaum N, Burak Y. Shaping neural circuits by high order synaptic interactions. PLoS Comput Biol. 2016;12(8):e1005056.

10. Ocker GK, Doiron B. Training and spontaneous reinforcement of neuronal assemblies by spike timing plasticity. Cereb Cortex. 2019;29(3):937–951.

11. Herpich J, Tetzlaff C. Principles underlying the input-dependent formation and organization of memories. Netw Neurosci. 2019;3(2):606–634.

12. Montangie L, Miehl C, Gjorgjieva J. Autonomous emergence of connectivity assemblies via spike triplet interactions. PLoS Comput Biol. 2020;16(5):e1007835.

13. Triplett MA, Avitan L, Goodhill GJ. Emergence of spontaneous assembly activity in developing neural networks without afferent input. PLoS Comput Biol. 2018;14(9):e1006421.

14. Huang YY, Colino A, Selig DK, Malenka RC. The influence of prior synaptic activity on the induction of long-term potentiation. Science. 1992;255(5045):730–733.

15. Turrigiano GG, Leslie KR, Desai NS, Rutherford LC, Nelson SB. Activity-dependent scaling of quantal amplitude in neocortical neurons. Nature. 1998;391(6670):892–896.

16. Ibata K, Sun Q, Turrigiano GG. Rapid synaptic scaling induced by changes in postsynaptic firing. Neuron. 2008;57(6):819–826.

17. Zenke F, Gerstner W, Ganguli S. The temporal paradox of Hebbian learning and homeostatic plasticity. Curr Opin Neurobiol. 2017;43:166–176.

18. Zenke F, Gerstner W. Hebbian plasticity requires compensatory processes on multiple timescales. Philos Trans R Soc Lond B Biol Sci. 2017;372(1715):20160259.

19. Tetzlaff C, Kolodziejski C, Timme M, Wörgötter F. Synaptic scaling in combination with many generic plasticity mechanisms stabilizes circuit connectivity. Front Comput Neurosci. 2011;5:47.

20. Lagzi F, Bustos MC, Oswald AM, Doiron B. Assembly formation is stabilized by Parvalbumin neurons and accelerated by Somatostatin neurons. bioRxiv. 2021;.

21. De Falco E, Ison MJ, Fried I, Quiroga RQ. Long-term coding of personal and universal associations underlying the memory web in the human brain. Nat Commun. 2016;7(1):1–11.

22. Rey HG, De Falco E, Ison MJ, Valentin A, Alarcon G, Selway R, et al. Encoding of long-term associations through neural unitization in the human medial temporal lobe. Nat Commun. 2018;9(1):1–13.

23. Rey HG, Gori B, Chaure FJ, Collavini S, Blenkmann AO, Seoane P, et al. Single neuron coding of identity in the human hippocampal formation. Curr Biol. 2020;30(6):1152–1159.

24. Fauth MJ, van Rossum MC. Self-organized reactivation maintains and reinforces memories despite synaptic turnover. Elife. 2019;8. doi:10.7554/elife.43717.

25. DeNardo LA, Liu CD, Allen WE, Adams EL, Friedmann D, Fu L, et al. Temporal evolution of cortical ensembles promoting remote memory retrieval. Nat Neurosci. 2019;22(3):460–469.

26. Feldman ML, Dowd C. Loss of dendritic spines in aging cerebral cortex. Anat Embryol (Berl). 1975;148(3):279–301.

27. Masliah E, Mallory M, Hansen L, DeTeresa R, Terry R. Quantitative synaptic alterations in the human neocortex during normal aging. Neurology. 1993;43(1 Part 1):192–192.

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

29. Kempter R, Gerstner W, Van Hemmen JL. Hebbian learning and spiking neurons. Phys Rev E. 1999;59(4):4498.

30. Pernice V, Staude B, Cardanobile S, Rotter S. How Structure Determines Correlations in Neuronal Networks. PLoS Comput Biol. 2011;7(5):e1002059. doi:10.1371/journal.pcbi.1002059.

31. Kalle Kossio YF, Goedeke S, van den Akker B, Ibarz B, Memmesheimer RM. Growing Critical: Self-Organized Criticality in a Developing Neural System. Phys Rev Lett. 2018;121:058301. doi:10.1103/PhysRevLett.121.058301.

32. Gerstein GL, Mandelbrot B. Random walk models for the spike activity of a single neuron. Biophys J. 1964;4(1):41–68.

33. Shadlen MN, Newsome WT. Noise, neural codes and cortical organization. Curr Opin Neurobiol. 1994;4(4):569–579.

34. Jahnke S, Memmesheimer RM, Timme M. Stable irregular dynamics in complex neural networks. Phys Rev Lett. 2008;100(4):048102.

35. Denève S, Machens CK. Efficient codes and balanced networks. Nat Neurosci. 2016;19:375–382. doi:10.1038/nn.4243.

36. Manz P, Goedeke S, Memmesheimer RM. Dynamics and computation in mixed networks containing neurons that accelerate towards spiking. Phys Rev E. 2019;100(4):042404.

37. Hawkes AG. Spectra of some self-exciting and mutually exciting point processes. Biometrika. 1971;58(1):83–90.

38. Hawkes AG. Point spectra of some mutually exciting point processes. J R Stat Soc Series B Stat Methodol. 1971;33(3):438–443.

39. Mishra RK, Kim S, Guzman SJ, Jonas P. Symmetric spike timing-dependent plasticity at CA3–CA3 synapses optimizes storage and recall in autoassociative networks. Nat Commun. 2016;7(1):1–11.

40. Yun R, Mishler JH, Perlmutter SI, Fetz EE. Paired stimulation for spike-timing dependent plasticity quantified with single neuron responses in primate motor cortex. bioRxiv. 2022;.

41. Dayan P, Abbott LF. Theoretical neuroscience: computational and mathematical modeling of neural systems. MIT press; 2005.

42. Harris TE, et al. The theory of branching processes. vol. 6. Springer Berlin; 1963.

43. Beggs JM, Plenz D. Neuronal avalanches in neocortical circuits. J Neurosci. 2003;23(35):11167–11177.

44. Ocker GK, Litwin-Kumar A, Doiron B. Self-Organization of Microcircuits in Networks of Spiking Neurons with Plastic Synapses. PLOS Computational Biology. 2015;11(8):e1004458. doi:10.1371/journal.pcbi.1004458.

45. Gastaldi C, Schwalger T, De Falco E, Quiroga RQ, Gerstner W. When shared concept cells support associations: Theory of overlapping memory engrams. PLoS Comput Biol. 2021;17(12):e1009691.

46. Dan Y, Poo Mm. Spike timing-dependent plasticity of neural circuits. Neuron. 2004;44(1):23–30.

47. Caporale N, Dan Y. Spike timing–dependent plasticity: a Hebbian learning rule. Annu Rev Neurosci. 2008;31:25–46.

48. Ziv NE, Brenner N. Synaptic tenacity or lack thereof: spontaneous remodeling of synapses. Trends Neurosci. 2018;41(2):89–99.

49. Kempter R, Gerstner W, Hemmen JLv. Intrinsic stabilization of output rates by spike-based Hebbian learning. Neural Comput. 2001;13(12):2709–2741.

50. Buzsáki G, Mizuseki K. The log-dynamic brain: how skewed distributions affect network operations. Nat Rev Neurosci. 2014;15(4):264–278. doi:10.1038/nrn3687.

51. Pfister JP. Triplets of Spikes in a Model of Spike Timing-Dependent Plasticity. Journal of Neuroscience. 2006;26(38):9673–9682. doi:10.1523/jneurosci.1425-06.2006.

52. Inglebert Y, Aljadeff J, Brunel N, Debanne D. Synaptic plasticity rules with physiological calcium levels. Proceedings of the National Academy of Sciences. 2020;117(52):33639–33648. doi:10.1073/pnas.2013663117.

53. Spaniol J, Grady C. Aging and the neural correlates of source memory: over-recruitment and functional reorganization. Neurobiol Aging. 2012;33(2):425–e3.

54. Folville A, Simons JS, D'Argembeau A, Bastin C. I remember it like it was yesterday: Age-related differences in the subjective experience of remembering. Psychon Bull Rev. 2021;doi:10.3758/s13423-021-02048-y.

55. Levine B, Svoboda E, Hay JF, Winocur G, Moscovitch M. Aging and autobiographical memory: Dissociating episodic from semantic retrieval. Psychol Aging. 2002;17(4):677–689. doi:10.1037/0882-7974.17.4.677.

Supporting information

S1 Appendix: Supporting analysis and figures