Biol Cybern (2009) 101:81–102
DOI 10.1007/s00422-009-0319-4

ORIGINAL PAPER

Biological
Cybernetics

Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. I. Input selectivity–strengthening correlated input pathways

Matthieu Gilson · Anthony N. Burkitt · David B. Grayden · Doreen A. Thomas · J. Leo van Hemmen

Received: 23 October 2008 / Accepted: 13 May 2009 / Published online: 18 June 2009

Abstract Spike-timing-dependent plasticity (STDP) determines the evolution of the synaptic weights according to their pre- and post-synaptic activity, which in turn changes the neuronal activity. In this paper, we extend previous studies of input selectivity induced by STDP for single neurons to the biologically interesting case of a neuronal network with fixed recurrent connections and plastic connections from external pools of input neurons. We use a theoretical framework based on the Poisson neuron model to analytically describe the network dynamics (firing rates and spike-time correlations) and thus the evolution of the synaptic weights. This framework incorporates the time course of the post-synaptic potentials and synaptic delays. Our analysis focuses on the asymptotic states of a network stimulated by two homogeneous pools of "steady" inputs, namely Poisson spike trains which have fixed firing rates and spike-time correlations. The STDP model extends rate-based learning in that it can implement, at the same time, both a stabilization of the individual neuron firing rates and a slower weight specialization depending on the input spike-time correlations. When one input pathway has stronger within-pool correlations, the resulting synaptic dynamics induced by STDP are shown to be similar to those arising in the case of a purely feed-forward network: the weights from the more correlated inputs are potentiated at the expense of the remaining input connections.

Keywords Learning · STDP · Recurrent neuronal network · Spike-time correlation

M. Gilson (✉) · A. N. Burkitt · D. B. Grayden · D. A. Thomas
Department of Electrical and Electronic Engineering,
The University of Melbourne, Melbourne, VIC 3010, Australia
e-mail: mgilson@bionicear.org

M. Gilson · A. N. Burkitt · D. B. Grayden
The Bionic Ear Institute, 384-388 Albert St,
East Melbourne, VIC 3002, Australia

M. Gilson · A. N. Burkitt · D. B. Grayden · D. A. Thomas
NICTA, Victoria Research Lab, Melbourne, VIC 3010, Australia

J. L. van Hemmen
Physik Department (T35) and BCCN-Munich,
Technische Universität München,
85747 Garching bei München, Germany

1 Introduction

The brain performs many sophisticated everyday computational tasks, such as image and speech recognition, with a speed and precision unparalleled by present-day computers. How the brain achieves such a performance is interesting both to gain a better understanding of the brain and for the artificial applications inspired by it. Learning is central to understanding neuronal information processing and many of its aspects have been studied from the molecular level up to the behavioral level. The neurophysiological basis of learning consists of changes at the molecular level that take place at the "connection" site between two neurons (synapse). This 'synaptic plasticity' describes the strengthening (potentiation) or the weakening (depression) of the synaptic efficacy (or weight), which is related to the amplitude of the potential variation at the soma of the post-synaptic neuron in response to a pre-synaptic action potential (or incoming spike).

Recent studies have established the importance of the timing of individual spikes in synaptic plasticity (Gerstner et al. 1996). Such spike-timing-dependent plasticity (STDP) has been the subject of both theoretical (Gerstner et al. 1996; Kempter et al. 1999; Gütig et al. 2003; Burkitt et al. 2004; Meffin et al. 2006) and experimental (Markram et al. 1997; Bi and Poo 2001) research; for a review, see Morrison et al. (2008). Although experiments have shown that the effect of

STDP can depend upon triplets of spikes or the post-synaptic membrane potential (Sjöström et al. 2001, 2004), we will constrain our study to pairwise STDP in order to study its functional properties, knowing the limitations of the model. A number of analyses (Burkitt et al. 2004) have shown the relation between STDP and various firing-rate-based learning models. In addition to rate-based information, STDP captures the effect of spike-time correlations on short time scales (from ms to tens of ms) that are neglected by rate-based learning.

Previous studies of STDP have primarily investigated single neurons and the resulting implications for the development of network structure in feed-forward architectures (Kempter et al. 1999; Song et al. 2000; Gütig et al. 2003; Burkitt et al. 2004; Meffin et al. 2006). STDP has been observed in various areas of the central neuronal system (e.g., hippocampus and cortex), which show recurrent connections between local neurons. The study of the neuronal dynamics induced by STDP in recurrent architectures is therefore a crucial step in order to gain insight into learning in the brain. However, the study of STDP in such recurrent networks has only begun to be addressed, mainly using numerical simulation (Song and Abbott 2001; Senn et al. 2002; Wenisch et al. 2005; Morrison et al. 2007; Lubenov and Siapas 2008; Câteau et al. 2008; Kang et al. 2008). There exists only a few theoretical results for recurrent architectures (Karbowski and Ermentrout 2002; Masuda and Kori 2007) due to mathematical difficulties in evaluating the effect of feedback synaptic loops. Theory is nevertheless necessary in order to understand the role of the main players in the network: neuronal and synaptic mechanisms, input structure, connectivity topology and learning parameters.

In a recent paper (Burkitt et al. 2007), we have developed a framework for the analysis of STDP in recurrent networks with arbitrary topology subject to external stimulation. This framework describes how the network activity, viz., firing rates and spike-time correlations, determines the evolution of the weights that occurs on a much slower time scale than the synaptic activation mechanisms. It provides a *soluble* differential system of equations that allows us to predict the resulting development of structure within the network, in particular the asymptotic distribution of the firing rates and of the weights after a sufficiently long learning epoch (the emerged structure). The analysis of the weight dynamics was carried out only for a network with full recurrent connectivity and no external inputs.

The present paper is the first in a series that analyzes the more biologically interesting case in which a network is stimulated by external input neurons with a functional structure. It extends our previously developed framework to incorporate the effect of the post-synaptic response, which was simplified as a delta-function response by Burkitt et al. (2007). This series aims to study the unsupervised learning scheme generated by the weight dynamics due to STDP. This learning results, for e.g., in strengthened synaptic pathways or symmetry breaking of an initially homogeneous weight distribution at the mesoscopic scale, which can be related to functional self-organization (Kohonen 1982) of the neuronal network. Each paper focuses on a particular network configuration, as illustrated in Fig. 1, where STDP only modifies some of the connections (thick arrows) in the neuronal network (top circles) excited by two pools of input spike trains that have homogeneous within-pool firing rates and spike-time correlations (bottom circle, fill-in indicates correlation), an idea inspired by Kempter et al. (1999) and Gütig et al. (2003). Minimal assumptions are made about the network connectivity (partial or full) and input structure. The first four papers will focus on additive STDP (Gerstner et al. 1996; Kempter et al. 1999) while the fifth will consider weight-dependence in the learning scheme (van Rossum et al. 2000; Bi and Poo 2001; Gütig et al. 2003).

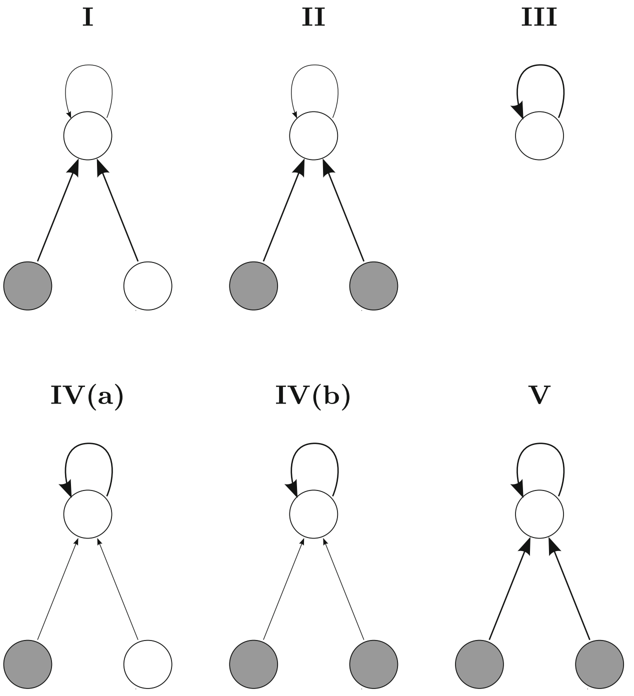


Fig. 1 Network configurations studied in the five papers in the series (denoted by *roman numerals*). *Top circles* represent the neuronal network and *bottom circles* the two input pools, for which *filled circles* indicate non-zero within-pool correlation. *Thick* (respectively *thin*) *arrows* indicate plastic (fixed) weights

After presenting the STDP model (Sect. 2.1) and neuron model (Sect. 2.2) used in this series of papers, we derive a dynamical system to describe the evolution of the *plastic* input weights while the recurrent connections are kept *fixed* (Sect. 2.3). We investigate the weight dynamics for a general network configuration (Sect. 3) and then focus on a specific topology, where the external inputs are divided into two homogeneous pools (Sect. 4).

2 Modeling learning and neuronal activity

2.1 Hebbian additive STDP

The STDP model describes the change in the synaptic weight due to the precise timing of spikes. We constrain our study to the contributions induced by single spikes and pairs of spikes. For two neurons in and out connected by a synapse $in \to out$ with weight J, the weight change δJ induced by a sole pair of pre- and post-synaptic spikes at t^{in} and t^{out}, respectively, is determined by three additive contributions:

$$\delta J = \eta \begin{cases} w^{\text{in}} & \text{at time } t^{\text{in}} \\ w^{\text{out}} & \text{at time } t^{\text{out}} \\ W(t^{\text{in}} - t^{\text{out}}) & \text{at time } \max(t^{\text{in}}, t^{\text{out}}). \end{cases} \quad (1)$$

The rate-based terms w^{in} (respectively, w^{out}) accounts for the effect of each pre-synaptic (post-synaptic) spike. The STDP window function W determines the correlation contribution for each pair of pre- and post-synaptic spikes in terms of the difference between their spike times $t^{\text{in}} - t^{\text{out}}$ (Gerstner et al. 1996; Kempter et al. 1999). Note that t^{in} is the time when the pre-synaptic spike effect reaches the post-synaptic site: it incorporates axonal propagation time on the pre-synaptic neuron and the neurotransmitter diffusion time. This feature was not considered by Burkitt et al. (2007) and is of importance, as shown in Sect. 2.3; see also Senn et al. (2002) and Lubenov and Siapas (2008). Likewise, t^{out} should incorporate a dendritic back-propagation delay but this will not be considered in this paper. Last, all these contributions are scaled by a learning parameter η, typically very small ($\eta \ll 1$), so that learning occurs very slowly compared to the other neuronal and synaptic mechanisms.

Figure 2 illustrates a typical choice for the function W. The left side, corresponding to the case of a pre-synaptic spike that occurs before the post-synaptic one ($t^{\text{in}} - t^{\text{out}} < 0$), has positive values of W, which induces potentiation. The right side, corresponding to the converse situation ($t^{\text{in}} - t^{\text{out}} > 0$), is negative and thus induces depression. This relates to Hebbian learning: “when a neuron consistently or significantly takes part in the firing of a target neuron, the efficacy of that connection, as one of the connections that excite the target neuron, is strengthened” (Hebb 1949). We also choose the particular value $W(0) = 0$, i.e., there is no contribution for a pair of pre- and post-synaptic spikes that occur at the same time, although $W(0)$ has no consequence in the dynamics.

In the present paper, we use the so-called additive version of STDP (Gerstner et al. 1996; Kempter et al. 1999), where the weight changes are independent of the current value of the weight J. This is in contrast to previous experimental, numerical and theoretical studies that suggested that weight-dependence should be taken into account (van Rossum et al. 2000; Bi and Poo 2001; Gütig et al. 2003; Morrison et al. 2007). One of our motivations is to keep the calculations as tractable as possible. This choice is also supported by the study for a single neuron by Gütig et al. (2003), whose models exhibit two distinct classes of behavior depending on a parameter that scales between the additive and multiplicative STDP: the weight distribution either becomes bimodal or remains unimodal for the same level of input correlation. In particular, additive STDP always splits the input weights, which corresponds to correct neuronal specialization for sufficient input correlations, but even occurs for uncorrelated inputs. Using the additive model, we focus on the specialization of the weights in the presence of input correlations, leaving aside their stabilization that we enforce by making use of explicit bounds on the weights (Kempter et al. 1999; Burkitt et al. 2007). The effect of this non-linearity upon the weight dynamics will be studied in paper V.

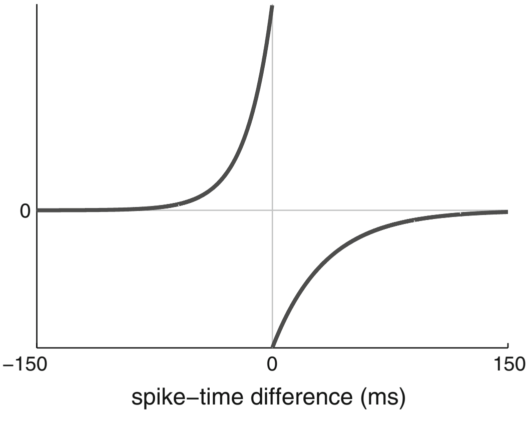


Fig. 2 Example of STDP window function W. It consists of one decaying exponential for potentiation (*left curve*) with characteristic time constant equal to 17 ms, and one for depression (*right curve*) with 34 ms. See Appendix D for details on the parameters

2.2 Poisson neuron model

In the Poisson neuron model (Kempter et al. 1999), the spiking mechanism of a given neuron i is approximated by an inhomogeneous Poisson process driven by an intensity function $\rho_i(t)$, in order to generate an output spike-time series $S_i(t)$, which can be represented as a sum of delta-functions or Dirac comb. One of the easiest ways to understand this model is to use discrete time, and at each time step Δt, the probability that the neuron fires is $\rho_i(t)\Delta t$. In addition, the probability that two or more spikes occur during Δt is $o(\Delta t)$, i.e., $o(\Delta t)/\Delta t \to 0$ when $\Delta t \to 0$, while events in disjoint intervals are independent.

The rate function $\rho_i(t)$ is to be related to the soma potential and it evolves over time according to the activity of its pre-synapses (indexed by k) that are excited by spike-time series

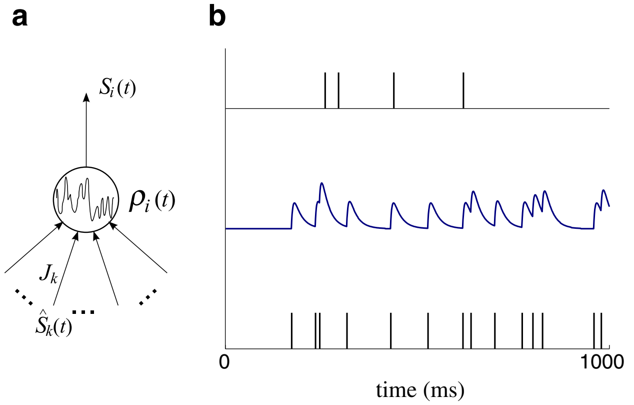


Fig. 3 Poisson neuron model. **a** Schematic view of the neuron. The soma (or cell body, *circle*) receives inputs from the synapses (*below*, indexed by k) and fires spikes that propagates on the axon (*above*), towards synaptic connections with other neurons (not represented). The flow of information is from *below* to *top*. **b** Illustration of the variation of $\rho_i(t)$ (*middle plot*) for a given succession of pre-synaptic spikes $\hat{S}_k(t)$ (*bottom spike train*) and the resulting output $S_i(t)$ (*top spike train*) over 1,000 ms

$\hat{S}_k(t)$, as shown in Fig. 3,

$$\rho_i(t) = \nu_0 + \sum_k \left[K_{ik}(t) \sum_n \epsilon\left(t - \hat{t}_{k,n} - \hat{d}_{ik}\right)\right]. \qquad (2)$$

The constant ν_0 is the spontaneous firing rate (identical for all neurons), which accounts for other pre-synaptic connections that are not considered in detail. Each spike of the kth spike train induces a variation of $\rho_i(t)$, namely the post-synaptic potential (PSP), that is determined by the synaptic weight K_{ik}, the post-synaptic response kernel ϵ, and the delay $\hat{d}_{ik}$. The kernel function ϵ models the time course of the PSP due to the current injected into the post-synaptic neuron for one single pre-synaptic spike; $\epsilon(t)$ is normalized to one: $\int \epsilon(t)\,\mathrm{d}t = 1$; and in order to preserve causality, we have $\epsilon(t) = 0$ for $t < 0$. The overall synaptic influx is the sum of the PSPs over all past spike times $\hat{t}_{k,n}$ (related to the kth synapse, and indexed by n) of the trains $\hat{S}_k(t)$. Note that we only consider positive weights here, i.e., excitatory synapses.

The Poisson neuron model is a coarse approximation of the activation mechanisms that take place in real neurons. However, we target the weight dynamics that is very slow compared to the neuronal activation dynamics. This separation of the time scales suggests that the activation mechanisms need not be too “realistic” in order to evaluate the evolution of the weights (Kempter et al. 1999; Gütig et al. 2003; Burkitt et al. 2007). Indeed, what impacts upon STDP is the increase of the probability of firing an output spike when receiving a PSP, which is qualitatively captured by this neuron model.

2.3 Dynamical system to model network activity

2.3.1 Overview

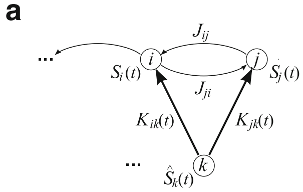
This study extends the framework presented by Burkitt et al. (2007), where the neuronal information contained in the spike trains is conveyed by firing rates and spike-time correlations (cf. Sect. 2.1). These variables of importance to describe the neuronal activity are linked in a dynamical system together with the synaptic weights. The derivation up to (14) summarizes and adapts results from previous work (Kempter et al. 1999; Burkitt et al. 2007); (16) is a new theoretical contribution. This framework is adapted to predict the evolution of the expectation values for the firing rates, correlations and weights depending on the learning and stimulation parameters.

We use the following mathematical assumptions (Kempter et al. 1999; Burkitt et al. 2007) in order to derive the dynamical system:

- the expectation values of the firing rates and pairwise covariances are constant in time for the external inputs, which is equivalent, here, to constant time-averaged firing rates and covariances for any realization of the spiking history;
- the separation of the time scales of the activation mechanisms and the learning dynamics, the latter happening on a slower time scale (adiabatic hypothesis);
- the expectation values of the firing rates and pairwise covariances are quasi-constant in time for the network neurons, i.e., they only vary due to the slow learning on the weights.

2.3.2 Description of the network

Let us consider a network of N Poisson neurons (indexed by $1 \leq i, j \leq N$) with fixed recurrent connections that is stimulated by M Poisson spike trains (a.k.a. external inputs or sources, indexed by $1 \leq k, l \leq M$) through plastic input connections, as illustrated in Fig. 4a. Typically both M and N are large. In addition to receiving synaptic input from the external sources, as shown for a sole neuron in Fig. 3, each neuron is also excited by other neurons via connections that may form feedback loops in the network, but without self-connections. The weight of the plastic connection from input k to neuron i is denoted by $K_{ik}(t)$ and the corresponding delay is $\hat{d}_{ik}$ (as defined in Sect. 2.2); respectively, J_{ij} and d_{ij} for the fixed connection from neuron j to neuron i; see Fig. 4b. Both input and recurrent synapses share the same PSP kernel ϵ. We will consider both fully- and partially-connected networks, where n^K denotes the number of input connections and n^J the number of recurrent connections. Partially-connected networks are generated by randomly



b

	external inputs	input-to-neuron	network neurons
firing rates	$\hat{\nu}_k(t)$		$\nu_i(t)$
pairwise correlations	$\hat{Q}_{kl}(t,u)$	$D_{ik}(t,u)$	
pairwise covariances	$\hat{C}_{kl}(t,u)$	$F_{ik}(t,u)$	
weights		$K_{ik}(t)$	J_{ij}
delays		$\hat{d}_{ik}$	d_{ij}

Fig. 4 Presentation of the network and notation. **a** Schematic representation of one of the M external inputs (*bottom circle*, indexed by $1 \leq k \leq N$) and two of the N network neurons (*top circles*, $1 \leq i, j \leq N$). The input connections have plastic weights $K_{ik}(t)$ (*thick arrows*) while the recurrent connections have fixed weights J_{ij} (*thin arrows*). The spike trains of the inputs and neurons are denoted by $\hat{S}_k(t)$ and $S_i(t)$, respectively. **b** The table shows the variables that describe the neuronal activity: time-averaged firing rates $\hat{\boldsymbol{\nu}}$ and $\boldsymbol{\nu}$; time-averaged correlations $\hat{Q}$ and D (Burkitt et al. 2007); time-averaged covariances $\hat{C}$ and F; and the variables related to the synaptic connections: weights K and J; delays $\hat{d}$ and d

assigning input-to-neuron and neuron-to-neuron connections. The term 'pool' will always refer to the external inputs, the term 'group' to the neurons.

Spikes are considered to be instantaneous events. We define $\hat{S}_k(t)$ as the spike-time series (Dirac comb) of the external input k; its value is zero except at the times when a spike is fired and the spike train is described as a sum of Dirac delta-functions. If there is a spike in a given small time interval $[t, t+\delta t]$, then

$$\int_t^{t+\delta t} \hat{S}_k(t')\, \mathrm{d}t' = 1, \tag{3}$$

where δt is "small" compared to the time scale of other neuronal mechanisms (ϵ, delays, etc.) so that there is no harm in considering $\hat{S}_k$ as approximate delta-functions, which is what they are in reality. The spike-time series $S_i(t)$ for network neuron i is defined similarly.

2.3.3 Slow weight evolution

We now derive a learning equation to describe the evolution of the input weights due to STDP according to the activities of the pre- and post-synaptic neurons for each synapse. For a small time interval $[t, t+\delta t]$, the change in the input weight $K_{ik}(t)$ described in (1) can be expressed using the pre- and post-synaptic spike trains (Kempter et al. 1999)

$$\delta K_{ik}(t) = \eta \int_t^{t+\delta t} \left[w^{\mathrm{in}}\, \hat{S}_k(t' - \hat{d}_{ik}) + w^{\mathrm{out}}\, S_i(t') \right] \mathrm{d}t' + \eta \int_{(t',u)\in\mathcal{I}(t)} W(u)\, S_i(t')\, \hat{S}_k(t' - \hat{d}_{ik} + u)\, \mathrm{d}u\, \mathrm{d}t'. \tag{4}$$

We assimilate the axonal delay described in Sect. 2.1 to $\hat{d}_{ik}$ defined in Sect. 2.2: $\hat{d}_{ik}$ then accounts for the axonal propagation and the diffusion of neurotransmitters and we neglect the dendritic delay compared to them. Thus, $\hat{S}_k(t' - \hat{d}_{ik})$ is the delayed time series of the pre-synaptic spikes and the time difference at the synaptic site between the pre- and the post-synaptic spikes (at respective times t^{pre} and t^{post} at the somas of both neurons) is $u = t^{\mathrm{pre}} + \hat{d}_{ik} - t^{\mathrm{post}}$. The domain of integration $\mathcal{I}(t)$ is the subset $(t', u) \in \mathbb{R}^2$ satisfying the three conditions

$$\begin{aligned} & t' \leq t + \delta t\,; \\ & t' - \hat{d}_{ik} + u \leq t + \delta t\,; \\ & t \leq t' \quad \text{or} \quad t \leq t' - \hat{d}_{ik} + u. \end{aligned} \tag{5}$$

The first two lines require that the spikes occur before $t + \delta t$, and the last line that at least one of them is in the time interval $[t, t + \delta t]$.

The change in weights over many independent trials (repetitions) is equivalent to averaging over a single long trial of length T. This self-averaging property of the learning requires the learning rate η to be small (van Hemmen 2001); then T can be chosen to be long compared to the time scale of the neuronal and synaptic activation mechanisms, but small compared to η^{-1} (separation of time scales). Typically, T is of the order of seconds (or tens of seconds) for synaptic mechanisms with characteristic times of tens of ms. This allows us to choose η such that δK_{ik} in (4) is at most a hundredth or a thousandth of the weight upper bound; for $\eta = 5 \times 10^{-7}$ (Appendix D), the effective learning epoch is of the order of tens of minutes. The ensemble average over the resulting random process is denoted by the angular brackets $\langle \cdots \rangle$. The rate of change for the expectation value of the external weight $\dot{K}_{ik}(t)$ is approximated by the temporal average of the summation of all $\langle \delta K_{ik}(t) \rangle$, i.e., the ensemble average taken of (4), over a time interval of duration T. This time-averaging allows the bounds of integration of t' in (4) to be slightly modified with good approximation in order to obtain (Kempter et al. 1999)

$$\dot{K}_{ik}(t) \simeq \frac{\eta}{T} \int_{t-T}^{t} \left[w^{\mathrm{in}} \left\langle \hat{S}_k(t' - \hat{d}_{ik}) \right\rangle + w^{\mathrm{out}} \left\langle S_i(t') \right\rangle \right] \mathrm{d}t' + \frac{\eta}{T} \int W(u) \left[\int_{t-T}^{t} \left\langle S_i(t')\, \hat{S}_k(t' - \hat{d}_{ik} + u) \right\rangle \mathrm{d}t' \right] \mathrm{d}u \tag{6}$$

The two first terms of the rhs of (6) involving w^{in} and w^{out} give the time-averaged firing rates of the pre- and post-synaptic spike trains, $\hat{\nu}_k(t)$ and $\nu_i(t)$, respectively; see Fig. 4b for an overview of the network variables. The last term involves the time-averaged pairwise correlation (Kempter et al. 1999; Burkitt et al. 2007)

$$D_{ik}(t,u) := \frac{1}{T}\int_{t-T}^{t} \left\langle S_i(t')\,\hat{S}_k(t'+u)\right\rangle \mathrm{d}t' \tag{7}$$

and this expression is convolved with the STDP window function $W(u)$ shifted by the delay $\hat{d}_{ik}$, which is embodied by the coefficient

$$D_{ik}^{W}(t) := \int_{-\infty}^{+\infty} W(u) D_{ik}(t, u-\hat{d}_{ik})\,\mathrm{d}u. \tag{8}$$

In order to incorporate $\hat{d}_{ik}$, this correlation coefficient has been modified compared to that used by Burkitt et al. (2007).

The separation of time scales between the "fast" neuronal and synaptic activation mechanisms on the one hand, and the "slow" learning dynamics ($\eta \ll 1$) on the other hand allows us to capture the evolution of the network activity. Under this assumption, the neuron firing rates $\nu_i(t)$ can be expressed in terms of the weights $K_{ik}(t)$ and J_{ij} and the input firing rates $\hat{\nu}_k(t)$; and likewise for the covariance coefficients $D_{ik}^{W}(t)$ with the input covariance coefficient $\hat{C}_{kl}(t,u)$ in Fig. 4b. This concept is used in the remainder of this section to rewrite (6) as a dynamical system of the general form

$$\dot{K}_{ik}(t) = \eta\,\mathbb{F}\left[K_{ik}(t), J_{ij}, \nu_0, \hat{\nu}_k(t), \hat{C}_{kl}(t,u)\right]. \tag{9}$$

This system of matrix equations is then used to predict the asymptotic evolution of the weights $K_{ik}(t)$, depending on the input parameters $\hat{\nu}_k(t)$ and $\hat{C}_{kl}(t,u)$.

2.3.4 Definition of the state variables for the network

The time-averaged firing rate $\nu_i(t)$ for neuron i corresponds to the duration T defined in Sect. 2.3.3 (Kempter et al. 1999; Burkitt et al. 2007)

$$\nu_i(t) := \frac{1}{T}\int_{t-T}^{t} \left\langle S_i(t')\right\rangle \mathrm{d}t'\ , \tag{10}$$

where $\langle S_i(t)\rangle$ is the instantaneous firing rate averaged over the randomness. The firing rate $\hat{\nu}_k(t)$ for input k is defined similarly.

For fixed input weights and steady inputs, the network activity as a stochastic process is actually ergodic, which implies that $\frac{1}{T}\int_{t-T}^{t} S_i(t')\,\mathrm{d}t' \simeq \langle S_i(t)\rangle = \text{const.}$ to a good approximation for large T. Therefore, we could simply define $\nu_i(t)$ as $\frac{1}{T}\int_{t-T}^{t} S_i(t')\,\mathrm{d}t'$ in (10), when the weights vary very slowly compared to T. We keep the notation of (10) to comply with the formalism developed by Kempter et al. (1999); Burkitt et al. (2007).

Instead of the correlation D_{ik} defined in (7), we use the neuron-to-input time-averaged covariance F_{ik} and define the following covariance coefficient F_{ik}^{Ψ} similar to D_{ik}^{W} in (8)

$$\begin{aligned} F_{ik}(t,u) &:= D_{ik}(t,u) - \nu_i(t)\,\hat{\nu}_k(t),\\ F_{ik}^{\Psi}(t) &:= \int_{-\infty}^{+\infty} \Psi(u) F_{ik}(t, u-\hat{d}_{ik})\,\mathrm{d}u \\ &= D_{ik}^{\Psi}(t) - \widetilde{\Psi}\,\nu_i(t)\,\hat{\nu}_k(t), \end{aligned} \tag{11}$$

where Ψ is a given kernel function and $\widetilde{\Psi} = \int \Psi(u)\,\mathrm{d}u$ its integral value; D_{ik}^{Ψ} is defined similar to D_{ik}^{W} with Ψ instead of W. These two formulas are usually defined for stationary second-order stochastic processes, which requires in particular constant instantaneous firing rates $\langle S_i(t)\rangle$ and $\langle \hat{S}_k(t)\rangle$ (steady inputs and fixed weights); we used $\hat{\nu}_k(t) \simeq \hat{\nu}_k(t+u)$ for u in the range of the STDP window function W.

Likewise, the time-averaged covariance $\hat{C}_{kl}$ and covariance coefficient $\hat{C}_{kl}^{\Psi}$ between inputs k and l are defined in the following way, instead of the correlation $\hat{Q}_{kl}$ in Fig. 4b used by Burkitt et al. (2007),

$$\begin{aligned} \hat{C}_{kl}^{\Psi}(t) &:= \int_{-\infty}^{+\infty} \Psi(u)\hat{C}_{kl}(t,u)\,\mathrm{d}u,\\ \hat{C}_{kl}(t,u) &:= \frac{1}{T}\int_{t-T}^{t} \left\langle \hat{S}_k(t')\,\hat{S}_l(t'+u)\right\rangle \mathrm{d}t' \\ &\quad - \frac{1}{T}\int_{t-T}^{t} \left\langle \hat{S}_k(t')\right\rangle \left\langle \hat{S}_l(t'+u)\right\rangle \mathrm{d}t'. \end{aligned} \tag{12}$$

Note that the input covariance coefficients $\hat{C}_{kl}^{\Psi}$ do not involve delays. As explained in Appendix A.1, the covariances $\hat{C}_{kl}$ (t,u) by convention do not incorporate the atomic (or point) discontinuity at $u = 0$ due to the autocorrelation of the stochastic point processes $\hat{S}_k$ for $k = l$, namely $\langle \hat{S}_k(t)\rangle\,\delta(u)$, where δ is the Dirac delta-function. This means that $\hat{C}$ represents the input correlation structure ("spiking information") but does not contain the autocorrelation intrinsic to the neuron model.

For the sake of simplicity, we use matrix notation in the remainder of the text: vectors $\boldsymbol{\nu}(t)$ and $\hat{\boldsymbol{\nu}}(t)$, and matrices $F^{W}(t)$, $\hat{C}^{W}(t)$, $K(t)$ and J. We now derive consistency equations to express $\boldsymbol{\nu}(t)$ and $F^{W}(t)$ in terms of the input parameters $\hat{\boldsymbol{\nu}}(t)$ and $\hat{C}^{W}(t)$, as well as the synaptic weights $K(t)$ and J.

2.3.5 *Short duration of the PSP kernel and of the recurrent delays*

Assuming that the weights $K(t)$ are quasi-stationary compared to the time scale of the PSP kernel ϵ and delays (Kempter et al. 1999), we take the ensemble average of (2) for neuron i to obtain

$$\begin{aligned}\langle S_i(t)\rangle &= \langle \rho_i(t)\rangle \\ &= \nu_0 + \sum_{j\neq i} J_{ij}\left\langle \epsilon * S_j(t-d_{ij})\right\rangle \\ &\quad + \sum_{k} K_{ik}(t)\left\langle \epsilon * \hat{S}_k(t-\hat{d}_{ik})\right\rangle,\end{aligned} \tag{13}$$

where $*$ denotes the convolution operation. Because T is very large compared to the neuronal time scale (ϵ and the delays), the integral of $\langle \epsilon * S_j(t-d_{ij})\rangle$ over the time interval $[t-T, t]$ can be approximated by the integral of $\langle S_j(t)\rangle$ over the same time interval (recall that $\int \epsilon(t)\,\mathrm{d}t = 1$). As a result we obtain the same matrix self-consistency equation as Burkitt et al. (2007) for the firing rates

$$\boldsymbol{\nu}(t) = \left[\mathbb{1}_N - J\right]^{-1}\left[\nu_0\,\mathbf{e} + K(t)\hat{\boldsymbol{\nu}}(t)\right], \tag{14}$$

where $\mathbb{1}_N$ is the identity matrix of size N and $\mathbf{e}$ is the column vector with N elements all equal to one ('$\mathbf{T}$' denotes the matrix transposition)

$$\mathbf{e} := [1, \ldots, 1]^{\mathbf{T}}. \tag{15}$$

To ensure the stability of the firing rates, the matrix of the recurrent weights J must have all eigenvalues with modulus strictly smaller than one (Burkitt et al. 2007).

We now derive a consistency equation similar to (14) for the neuron-to-input covariance F defined in (11). The case of non-identical delays could be rigorously dealt with using Fourier analysis (Hawkes 1971), as shown in Appendix A.2.4. However, this method does not lead to an easily tractable solution for arbitrary PSP kernel ϵ and distribution of delays. In the remainder of this paper, we consider the simplified case where all the recurrent delays are almost identical, i.e., $d_{ij} \simeq d$ for all connections $j \to i$, and likewise the input delays satisfy $\hat{d}_{ik} \simeq \hat{d}$ for all connections $k \to i$. The impact of the PSP kernel ϵ and of the recurrent delays d_{ij} can be evaluated when their two distributions are narrow in comparison to the width of the learning window W, as detailed in Appendix A.2.7, which gives

$$\begin{aligned}F^W(t) = \left[\mathbb{1}_N - J\right]^{-1} K(t) \\ \left\{\hat{C}^{W*\epsilon}(t) + [W*\epsilon](0)\,\mathrm{diag}(\hat{\boldsymbol{\nu}}(t))\right\}.\end{aligned} \tag{16}$$

The input covariance structure $\hat{C}$ is filtered by the PSP kernel ϵ to obtain the neuron-to-input covariance F, which affects, via $\hat{C}^{W*\epsilon}$, the effect of STDP embodied in F^W (Kempter et al. 1999; Sprekeler et al. 2007). As a comparison, Burkitt et al. (2007) neglected this effect and used $\hat{C}^W$ instead. The use of the covariance coefficient F instead of D (Burkitt et al. 2007) sheds a clearer light on the relationship between the neuron-to-input and the input-to-input correlation structures: the network connectivity operates on the input covariance $\hat{C}$ through the term $(\mathbb{1}_N - J)^{-1}\,K$. The following approximation is made in deriving these equations and its accuracy is illustrated in Fig. 5:

$$\int W(u-r)\epsilon(r-d)\,\mathrm{d}r \;\simeq\; W(u). \tag{17}$$

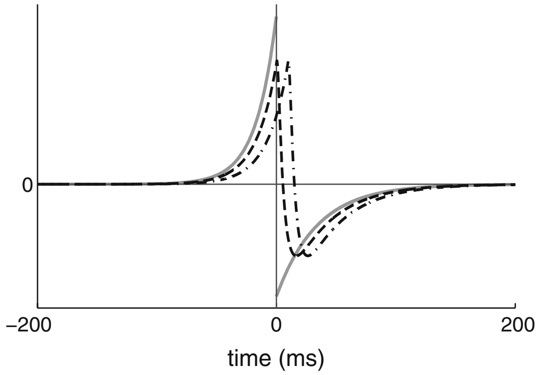


Fig. 5 Impact of the PSP kernel ϵ on learning with the STDP window function W. The *solid line* represents the function W and the *dashed line* its convolution $W*\epsilon_d$ with the PSP kernel ϵ delayed by $d = 0.4$ ms. Globally, the shape of $W*\epsilon$ is similar to that of W, but for small $u > 0$ we have $W(u) < 0$ whereas $W*\epsilon(u) > 0$. Also plotted in *dashed-dotted line* is $W*\epsilon_d$ for longer delay $d = 10$ ms: increasing d shifts the curve of $W*\epsilon_d$ to the right and increases the discrepancies between the two curves

2.3.6 *The equations describing the dynamical system*

In the limit of large networks ($N \gg 1$ and $M \gg 1$), we can ignore the effects due to autocorrelation, i.e., the term involving 'diag' in (16). The system of equations that describe the dynamics of the firing rates, covariance coefficients and weights reduces to

$$\boldsymbol{\nu} = (\mathbb{1}_N - J)^{-1}\left(\nu_0\,\mathbf{e} + K\,\hat{\boldsymbol{\nu}}\right), \tag{18a}$$

$$F^W = (\mathbb{1}_N - J)^{-1}\,K\,\hat{C}^{W*\epsilon}, \tag{18b}$$

$$\dot{K} = \Phi_K\left(w^{\mathrm{in}}\,\mathbf{e}\,\hat{\boldsymbol{\nu}}^{\mathbf{T}} + w^{\mathrm{out}}\,\boldsymbol{\nu}\,\hat{\mathbf{e}}^{\mathbf{T}} + \widetilde{W}\,\boldsymbol{\nu}\,\hat{\boldsymbol{\nu}}^{\mathbf{T}} + F^W\right). \tag{18c}$$

Time has been rescaled to remove η and the time variable t is omitted from all the vectors and matrices that evolve over time. The constant $\widetilde{W}$ denotes the integral of the STDP window function W

$$\widetilde{W} := \int_{-\infty}^{+\infty} W(u)\,\mathrm{d}u. \tag{19}$$

The column vectors $\hat{\mathbf{e}}$ and $\mathbf{e}$ have all elements equal to one (15). The projector Φ_K operates on the vector space of $N \times M$

matrices and nullifies the matrix components that correspond to non-existent connections in the network.

The superscript '$\mathbf{T}$' denotes the matrix transposition and $\mathbf{e}\,\hat{\boldsymbol{\nu}}^{\mathbf{T}}$ denotes a $N \times M$ matrix. Bra-ket notation could be used instead: for e.g., the ket $|\mathbf{a}\rangle = \mathbf{e}$ stands for a column vector and the bra $\langle\mathbf{b}| = |\mathbf{b}\rangle^{\mathbf{T}} = \hat{\boldsymbol{\nu}}^{\mathbf{T}}$ for a row vector; $|\mathbf{a}\rangle\,\langle\mathbf{b}| = \mathbf{e}\,\hat{\boldsymbol{\nu}}^{\mathbf{T}}$ is then a ket-bra matrix (the bra-kets and the ensemble average should not be confused).

2.3.7 Generation of the input spike trains

The inputs that stimulate the network are partitioned into a given number of pools, such that inputs from the same pool are correlated but independent of inputs from different pools. The firing rates of inputs within a pool are all equal to, say, $\hat{\nu}_0$. The positive within-pool correlation is generated so that, for any input, a given portion of its spikes occur at the same time as some other spikes within its pool, while the remainder occur at independent times (Gütig et al. 2003; Meffin et al. 2006).

The spikes from input k are selected from two homogeneous Poisson spike trains each of rate $\hat{\nu}_0$ such that the within-group correlation strength is $0 \leq \hat{c} \leq 1$ (Meffin et al. 2006). The first spike train is common to all inputs in the pool and generates the correlated events; distinct pools have different common reference spike trains. For a given input, the spikes are selected from this train with probability $\sqrt{\hat{c}}$, independently of other neurons in the pool. Thus, only a portion of all the neurons in the pool participate in each correlated event. The second spike train is the own independent train attached to each input and the spikes are selected from this train with probability $1 - \sqrt{\hat{c}}$. For each input k, we create a random variable $\hat{X}_k(t)$ that is one if there is an input spike at time t and zero otherwise. The correlation between the variables $\hat{X}_k(t)$ and $\hat{X}_l(t)$ corresponding to distinct sources from the same pool is given by (Gütig et al. 2003)

$$\frac{\mathrm{Cov}\left[\hat{X}_k(t), \hat{X}_l(t)\right]}{\sqrt{\mathrm{Var}\left[\hat{X}_k(t)\right]\,\mathrm{Var}\left[\hat{X}_l(t)\right]}} = \hat{c}. \tag{20}$$

Typically, we use *small* input correlations: $0 \leq \hat{c} \leq 10^{-1}$. Numerical simulation uses discrete time to generate the Poisson pulse trains (Appendix D).

In this way, we obtain input spike trains $\hat{S}_k(t)$ with "instantaneous" firing rates $\langle\hat{S}_k(t)\rangle = \hat{\nu}_0$ and pairwise covariances (for $k \neq l$)

$$\hat{C}_{kl}(t, u) \simeq \hat{c}\,\hat{\nu}_0\,\delta(u) \tag{21}$$

that are both constant in time. The latter follows since (20) implies $\mathrm{Cov}[\hat{S}_k(t), \hat{S}_l(t+u)] \simeq \hat{c}\,\hat{\nu}_0\,\delta(u)$, as defined by (48) in Appendix A.1. Inputs from the same pool are only correlated for $u = 0$, which we denote as 'delta-correlated' inputs. In this paper, as well as the rest of the series, we only consider delta-correlated inputs. It follows that, for inputs $k \neq l$,

$$\hat{C}^{W*\epsilon}_{kl} \simeq \hat{c}\,\hat{\nu}_0\,[W * \epsilon](0). \tag{22}$$

Due to the PSP kernel, delta-correlated inputs induce a non-trivial (richer) correlation structure F in the network according to (16). Under the Hebbian assumption that $W(u) > 0$ for $u < 0$ (cf. Sect. 2.1), $[W * \epsilon](0) = \int W(u)\epsilon(-u)\mathrm{d}u > 0$ and the matrix $\hat{C}^{W*\epsilon}$ has non-negative elements for delta-correlation. It also implies that the spike-triggering effect for F^W, involving diag in the rhs of (16), is always positive, as explained in Appendix A. This is similar to the case of a feed-forward architecture (Kempter et al. 1999).

2.3.8 Analysis of the system dynamics

Our aim is to investigate the steady states of the firing rates and weights, as well as their stability. The system of equations (18a–18c) describes the evolution of their expectation values, i.e., the first order of the stochastic process. In the remainder of this paper, we refer to this leading order as the *drift* of the dynamics, in comparison to *higher orders* of the stochastic process. Phenomena such as evolution of the weight variance rely upon higher-order stochastic mechanisms and are not captured by (18a–18c) but they can nevertheless be analyzed using this formalism (Kempter et al. 1999; Burkitt et al. 2007). Within the range of *small* correlation strengths (Sect. 2.3.7), we will discriminate between *weak* and *sufficiently strong* values when discussing their impact on the weight dynamics.

The term *mean* (applied to firing rates and weights) will refer to an average over the neurons, inputs, connections, etc. of the network (topological averaging), whereas *averaged* stands for time averaging, unless otherwise specified. The *homeostatic equilibrium* describes the situation where the mean firing rate and mean weight have reached an equilibrium, although individual firing rates and weights may continue to change. The expression *emergence of weight structure* will refer to the situation where the learning dynamics has imposed a specific weight structure on the network, i.e., further learning may cause individual weights to change but the qualitative character of the distribution (e.g., bimodal) will remain unchanged.

It is necessary to introduce bounds on the input weights in numerical simulation because of their tendency to diverge due to the competition induced by STDP (Kempter et al. 1999). This follows from our choice of additive STDP (cf. Sect. 2.1), which ensures that the equations remain as tractable as possible. In this way, we focus on the splitting of the weight distribution and leave aside the stabilization issue.

The simulation results presented in this paper were run using the neuron and learning parameters listed in Appendix D.

3 General case of learning input weights with fixed recurrent weights

In this section, we analyze the general solution of the dynamical system (18a–18c). We formulate stability conditions for the mean firing rate and weight (homeostatic equilibrium defined in Sect. 2.3.8), and study the asymptotic weight distribution through a fixed-point analysis. The evolution of the input weights is decomposed in order to understand how stability and specialization occur together.

3.1 Homeostatic equilibrium

The evolution equation of the mean input weight K_{av} is given by

$$\begin{aligned}\dot{K}_{\mathrm{av}} &= w^{\mathrm{in}}\hat{\nu}_{\mathrm{av}} + w^{\mathrm{out}}\nu_{\mathrm{av}} + \widetilde{W}\hat{\nu}_{\mathrm{av}}\nu_{\mathrm{av}} + F^{W}_{\mathrm{av}} \\ &= w^{\mathrm{in}}\hat{\nu}_{\mathrm{av}} + \frac{\nu_0\left(w^{\mathrm{out}} + \widetilde{W}\hat{\nu}_{\mathrm{av}}\right)}{1 - n^J_{\mathrm{av}}J_{\mathrm{av}}} \\ &\quad + n^K_{\mathrm{av}}K_{\mathrm{av}}\frac{\hat{\nu}_{\mathrm{av}}\left(w^{\mathrm{out}} + \widetilde{W}\hat{\nu}_{\mathrm{av}}\right) + \hat{C}^{W*\epsilon}_{\mathrm{av}}}{1 - n^J_{\mathrm{av}}J_{\mathrm{av}}}.\end{aligned} \tag{23}$$

The subscript 'av' denotes the mean-averaged variable over the network, i.e., when neglecting the discrepancies among the external inputs, the neurons or the connections, so that

$$\nu_{\mathrm{av}} = \frac{\nu_0 + n^K_{\mathrm{av}}K_{\mathrm{av}}\,\hat{\nu}_{\mathrm{av}}}{1 - n^J_{\mathrm{av}}J_{\mathrm{av}}}. \tag{24}$$

The constants $n^K_{\mathrm{av}} := n^K/N$ and $n^J_{\mathrm{av}} := n^J/N$ denote the mean number of pre-synaptic input and recurrent connections (respectively) in the network. Equation (23) is linear in K_{av}, which converges towards an equilibrium if and only if

$$\frac{\hat{\nu}_{\mathrm{av}}\left(w^{\mathrm{out}} + \widetilde{W}\hat{\nu}_{\mathrm{av}}\right) + \hat{C}^{W*\epsilon}_{\mathrm{av}}}{1 - n^J_{\mathrm{av}}J_{\mathrm{av}}} < 0. \tag{25}$$

We have $1 - n^J_{\mathrm{av}}J_{\mathrm{av}} > 0$ since the matrix J has a spectrum in the unit circle (to prevent firing rates from diverging), so the mean recurrent feedback $n^J_{\mathrm{av}}J_{\mathrm{av}}$ does not change the stability of the network. For weakly correlated inputs, the covariance coefficient $\hat{C}^{W*\epsilon}_{\mathrm{av}}$ is small compared to the mean stimulation firing rate $\hat{\nu}_{\mathrm{av}}$ (Kempter et al. 1999) and the previous stability condition reduces to

$$w^{\mathrm{out}} + \widetilde{W}\hat{\nu}_{\mathrm{av}} < 0. \tag{26}$$

Consequently, there are four situations to consider for the homeostatic equilibrium, depending on the mean input stimulation $\hat{\nu}_{\mathrm{av}}$:

(i) $\widetilde{W} < 0$ and $w^{\mathrm{out}} < 0$: stable whatever the value $\hat{\nu}_{\mathrm{av}}$;
(ii) $\widetilde{W} < 0$ and $w^{\mathrm{out}} > 0$: stable for $\hat{\nu}_{\mathrm{av}} < -w^{\mathrm{out}}/\widetilde{W}$;
(iii) $\widetilde{W} > 0$ and $w^{\mathrm{out}} < 0$: stable for $\hat{\nu}_{\mathrm{av}} > -w^{\mathrm{out}}/\widetilde{W}$;
(iv) $\widetilde{W} > 0$ and $w^{\mathrm{out}} > 0$: never stable for any value $\hat{\nu}_{\mathrm{av}}$.

The input stimulation can thus change the stability in some cases, unlike the recurrent feedback. We recall that $\widetilde{W} < 0$ in cases (i) and (ii) above leads to homeostatic stability of the learning dynamics when STDP modifies only the recurrent weights in a network with no external inputs (Burkitt et al. 2007; Gilson et al. 2009b). Case (iii) corresponds to a stability analysis already elsewhere described (Kempter et al. 1999). The simulation parameters used in this paper (see Appendix D) correspond to case (i).

As a consequence of (23), the asymptotic value of K_{av} is given by the fixed point (if it is stable)

$$K^*_{\mathrm{av}} = \frac{-1}{n^K_{\mathrm{av}}}\,\frac{\left(1 - n^J_{\mathrm{av}}J_{\mathrm{av}}\right)w^{\mathrm{in}}\hat{\nu}_{\mathrm{av}} + \nu_0\left(w^{\mathrm{out}} + \widetilde{W}\hat{\nu}_{\mathrm{av}}\right)}{\hat{\nu}_{\mathrm{av}}\left(w^{\mathrm{out}} + \widetilde{W}\hat{\nu}_{\mathrm{av}}\right) + \hat{C}^{W*\epsilon}_{\mathrm{av}}}. \tag{27}$$

Since we require the weights K to remain positive, the equilibrium is realizable only if the asymptotic value K^*_{av} is positive, which requires, similar to the case of a single neuron (Kempter et al. 1999),

$$w^{\mathrm{in}} > -\frac{\nu_0\left(w^{\mathrm{out}} + \widetilde{W}\hat{\nu}_{\mathrm{av}}\right)}{\hat{\nu}_{\mathrm{av}}\left(1 - n^J_{\mathrm{av}}J_{\mathrm{av}}\right)} > 0. \tag{28}$$

When the fixed point $K^*_{\mathrm{av}} < 0$ is stable, the input weights $K(t)$ will all become silent. The presence of strong recurrent feedback $n^J_{\mathrm{av}}J_{\mathrm{av}}$ can thus cause the homeostatic equilibrium to become non-realizable. This condition is also consistent with the stability analysis in the case of learning on the recurrent weights J with no external inputs, for which $w^{\mathrm{in}} \gg |w^{\mathrm{out}}|$ ensures the stability of individual firing rates (Burkitt et al. 2007; Gilson et al. 2009b).

From (27) and (24), the fixed point ν^*_{av} of the mean firing rate is given by

$$\nu^*_{\mathrm{av}} = \frac{-w^{\mathrm{in}}\hat{\nu}^2_{\mathrm{av}} + \nu_0\left(1 - n^J_{\mathrm{av}}J_{\mathrm{av}}\right)^{-1}\hat{C}^{W*\epsilon}_{\mathrm{av}}}{\hat{\nu}_{\mathrm{av}}\left(w^{\mathrm{out}} + \widetilde{W}\hat{\nu}_{\mathrm{av}}\right) + \hat{C}^{W*\epsilon}_{\mathrm{av}}}. \tag{29}$$

For weakly correlated inputs, it reduces to

$$\nu^*_{\mathrm{av}} \simeq -\frac{w^{\mathrm{in}}\hat{\nu}_{\mathrm{av}}}{w^{\mathrm{out}} + \widetilde{W}\hat{\nu}_{\mathrm{av}}}. \tag{30}$$

From (30), we see that the fixed point ν^*_{av} is an increasing function of the mean input firing rate $\hat{\nu}_{\mathrm{av}}$ when $w^{\mathrm{out}} < 0$ (decreasing otherwise).

3.2 Emergence of a weight structure

The learning equation (18c) can be rewritten as a linear differential matrix equation in K,

$$\dot{K} = \Phi_K\left[\left(\mathbb{1}_N - J\right)^{-1}KA + B\right] \tag{31}$$

with the two following matrices containing the input firing-rate and correlation structures

$$A := w^{\mathrm{out}}\hat{\boldsymbol{\nu}}\hat{\mathbf{e}}^{\mathbf{T}} + \widetilde{W}\hat{\boldsymbol{\nu}}\hat{\boldsymbol{\nu}}^{\mathbf{T}} + \hat{C}^{W*\epsilon},$$
$$B := w^{\mathrm{in}}\,\mathbf{e}\,\hat{\boldsymbol{\nu}}^{\mathbf{T}} + (\mathbb{1}_N - J)^{-1}\,\nu_0\mathbf{e}\left(w^{\mathrm{out}}\hat{\mathbf{e}}^{\mathbf{T}} + \widetilde{W}\hat{\boldsymbol{\nu}}^{\mathbf{T}}\right). \tag{32}$$

We denote by $\mathbb{M}_K$ the subspace of $\mathbb{R}^{N\times M}$ where the matrix K evolves, i.e., the vector subspace of matrices X such that $\Phi_K(X) = X$.

We examine the solution of (31), first, for the case of full input connectivity (Φ_K is the identity), which depends upon the invertibility of the matrix A. We then complete the general analysis for partial connectivity (and any matrix A), which will be illustrated though a specific network example in Sect. 4.

3.2.1 Full input connectivity and invertible A

If the matrix A is invertible, the solution of (31) is given by

$$K(t) = K(\infty) + \sum_{n\geq 0}\frac{t^n}{n!}(\mathbb{1}_N - J)^{-n}\,[K(0) - K(\infty)]\,A^n \tag{33}$$

with the fixed point

$$K(\infty) = -(\mathbb{1}_N - J)\,BA^{-1} \tag{34}$$

and $K(0)$ the initial weight matrix at $t = 0$. Note that without rescaling time, t would be replaced by $\eta\, t$ in (33). The weight stability of $K(\infty)$ is determined by the eigenvalues of A since the spectrum of $\mathbb{1}_N - J$ lies in the unit circle for the sake of bounded (non-diverging) firing rates. In the same way as with the homeostatic equilibrium, the presence of recurrent connections affects the asymptotic weight matrix $K(\infty)$, as well as the rate of convergence (or divergence) of $K(t)$. Note that the spike-triggering effect, which we neglected, only adds the diagonal matrix $[W * \epsilon](0)\,\mathrm{diag}(\hat{\boldsymbol{\nu}})$ to A.

The weight matrix $K(t)$ converges exponentially fast towards $K(\infty)$ when the eigenvalues of A have negative real parts. The fixed point $K(\infty)$ may then be attained depending on the weight bounds. On the contrary, if A has any eigenvalue with positive real part, some components of K diverge in the direction of the principal eigenvector until hitting the bounds. Then, the relative position of the initial conditions $K(0)$ compared to that of the fixed point $K(\infty)$ in $\mathbb{M}_K$ will determine the evolution of $K(t)$. A combination of stability and divergence gives interesting dynamics, as was shown by Kempter et al. (1999) for the single-neuron case: the first corresponds to partial equilibria (e.g., homeostatic) while the second can imply robust weight specialization through a splitting of their distribution.

3.2.2 Partial input connectivity and/or non-invertible A

The matrix A is not invertible whenever there are symmetries in the input pools and in the weights K; for e.g., in the case of homogeneous input pools. Details of the general analysis are provided in Appendix B. In summary, we can decompose the evolution of K in three subspaces defined using the null-spaces of the matrices A and B:

- an exponential evolution (convergence in the stable case) on a subspace where "A is invertible", similar to the solution given in (33);
- a zero drift on a subspace related, for e.g., to symmetries of the input pools and of the input connectivity, where higher stochastic orders of the weight dynamics have a significant effect;
- a constant drift that drives weights towards their bounds in a particular direction (this case corresponds in general to very specific parameter values and we ignore it).

When the network has symmetries, the weight drift in the first subspace can be studied using a reduction of dimensionality for $K(t)$ as explained in Appendix B.1. In the second subspace, the weight evolution is not constrained by STDP in a way that organizes the input weights and, consequently, does not correspond to learning of the input firing-rate and correlation structure. Higher-order effects can nevertheless be the source of organization in the network, as will be described in a companion paper that is devoted to symmetry breaking (Gilson et al. 2009a).

4 Network stimulated by two homogeneous input pools

We now illustrate the general analysis in Sect. 3, in particular how the input correlations determine the asymptotic weight structure, through a specific network example inspired by Kempter et al. (1999). The network is stimulated by external inputs that are divided into two homogeneous pools of the same size (indices $1 \leq k \leq M/2$ vs. $M/2 + 1 \leq k \leq M$), as illustrated in Fig. 1I. The analysis below is carried out for full input connectivity for the sake of simplicity, but simulations show corresponding cases with partial connectivity.

4.1 Reduction of dimensionality to study the weight drift

The vector $\hat{\boldsymbol{\nu}}$ and the matrix $\hat{C}^{W*\epsilon}$ that appear in A and B, cf. (32), can be expressed in terms of the M-column vector $\hat{\mathbf{e}}$, defined similarly to $\mathbf{e}$ in (15), and the M-column vector $\hat{\mathbf{h}}$, whose first $M/2$ elements are 1 and last $M/2$ elements are -1:

$$\hat{\mathbf{h}} := [1, \ldots, 1, -1, \ldots, -1]^{\mathbf{T}}. \tag{35}$$

Denoting the mean firing rates by $\bar{\hat{\nu}}_1$ and $\bar{\hat{\nu}}_2$ for each input pool and their correlation strengths by $\hat{c}_1$ and $\hat{c}_2$, we have

$$\begin{aligned}\hat{\boldsymbol{\nu}} &= \frac{\bar{\hat{\nu}}_1}{2}\,(\hat{\mathbf{e}}+\hat{\mathbf{h}}) + \frac{\bar{\hat{\nu}}_2}{2}\,(\hat{\mathbf{e}}-\hat{\mathbf{h}}),\\ \hat{C}^{W*\epsilon} &= \frac{\hat{c}_1\bar{\hat{\nu}}_1[W*\epsilon](0)}{4}\,(\hat{\mathbf{e}}+\hat{\mathbf{h}})(\hat{\mathbf{e}}+\hat{\mathbf{h}})^{\mathbf{T}}\\ &\quad + \frac{\hat{c}_2\bar{\hat{\nu}}_2[W*\epsilon](0)}{4}\,(\hat{\mathbf{e}}-\hat{\mathbf{h}})(\hat{\mathbf{e}}-\hat{\mathbf{h}})^{\mathbf{T}},\end{aligned} \tag{36}$$

where we have used (22). Substituting the above expressions into (32), we obtain the following special form for the matrices A and B

$$\begin{aligned}A &= \alpha\hat{\mathbf{e}}\hat{\mathbf{e}}^{\mathbf{T}} + \beta\hat{\mathbf{h}}\hat{\mathbf{e}}^{\mathbf{T}} + \gamma\hat{\mathbf{e}}\hat{\mathbf{h}}^{\mathbf{T}} + \kappa\hat{\mathbf{h}}\hat{\mathbf{h}}^{\mathbf{T}},\\ B &= \alpha'\mathbf{e}\hat{\mathbf{e}}^{\mathbf{T}} + \beta'\,(\mathbb{1}_N - J)^{-1}\,\mathbf{e}\hat{\mathbf{e}}^{\mathbf{T}}\\ &\quad + \gamma'\mathbf{e}\hat{\mathbf{h}}^{\mathbf{T}} + \kappa'\,(\mathbb{1}_N - J)^{-1}\,\mathbf{e}\hat{\mathbf{h}}^{\mathbf{T}},\end{aligned} \tag{37}$$

where the constants $\alpha, \beta, \gamma, \kappa, \alpha', \beta', \gamma', \kappa'$ absorb all the input and learning parameters,

$$\begin{aligned}\alpha &= w^{\text{out}}\frac{\bar{\hat{\nu}}_1+\bar{\hat{\nu}}_2}{2} + \widetilde{W}\frac{\left(\bar{\hat{\nu}}_1+\bar{\hat{\nu}}_2\right)^2}{4}\\ &\quad + [W*\epsilon](0)\frac{\hat{c}_1\bar{\hat{\nu}}_1+\hat{c}_2\bar{\hat{\nu}}_2}{4},\\ \beta &= w^{\text{out}}\frac{\bar{\hat{\nu}}_1-\bar{\hat{\nu}}_2}{2} + \widetilde{W}\frac{\bar{\hat{\nu}}_1^2-\bar{\hat{\nu}}_2^2}{4}\\ &\quad + [W*\epsilon](0)\frac{\hat{c}_1\bar{\hat{\nu}}_1-\hat{c}_2\bar{\hat{\nu}}_2}{4},\\ \gamma &= \widetilde{W}\frac{\bar{\hat{\nu}}_1^2-\bar{\hat{\nu}}_2^2}{4} + [W*\epsilon](0)\frac{\hat{c}_1\bar{\hat{\nu}}_1-\hat{c}_2\bar{\hat{\nu}}_2}{4},\\ \kappa &= \widetilde{W}\frac{\left(\bar{\hat{\nu}}_1-\bar{\hat{\nu}}_2\right)^2}{4} + [W*\epsilon](0)\frac{\hat{c}_1\bar{\hat{\nu}}_1+\hat{c}_2\bar{\hat{\nu}}_2}{4},\\ \alpha' &= w^{\text{in}}\frac{\bar{\hat{\nu}}_1+\bar{\hat{\nu}}_2}{2},\\ \beta' &= w^{\text{out}}\nu_0 + \widetilde{W}\nu_0\frac{\bar{\hat{\nu}}_1+\bar{\hat{\nu}}_2}{2},\\ \gamma' &= w^{\text{in}}\frac{\bar{\hat{\nu}}_1-\bar{\hat{\nu}}_2}{2},\\ \kappa' &= \widetilde{W}\nu_0\frac{\bar{\hat{\nu}}_1-\bar{\hat{\nu}}_2}{2}.\end{aligned} \tag{38}$$

The drift $\dot{K}(t)$ is clearly zero in the subspace orthogonal to both $\hat{\mathbf{e}}$ and $\hat{\mathbf{h}}$. The interesting values to predict are the mean input weights for each neuron, embodied in the vector $K\hat{\mathbf{e}}/M$, and the difference between the mean weights from the first and the second pools, contained in $K\hat{\mathbf{h}}/M$. This reduction of dimensionality explained in Appendix B.1 would still be valid for homogeneous partial input connectivity and input pools of different sizes. For the network configuration in Fig. 6, the equivalence classes for the input weights to describe the drift $\dot{K}(t)$ simply correspond to the mean input weights over each input pool and neuron group, such as $\bar{K}_{21}$ from pool $\hat{1}$ to group 2.

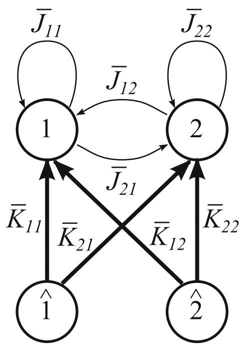


Fig. 6 Reduced weight matrices K and J for a network of two homogeneous groups of neurons (*top circles*) stimulated by two homogeneous input pools (*bottom circles*). The variable $\bar{K}_{11}$, for e.g., corresponds to the mean input weights from pool $\hat{1}$ to group 1. The drift of the input weights can be completely studied using a reduced system of equations with $\bar{K}$ and $\bar{J}$

In the following, we study the drift of the input weights using the two vectors $K\hat{\mathbf{e}}$ and $K\hat{\mathbf{h}}$, which evolve over time according to

$$\begin{aligned}K\dot{\hat{\mathbf{e}}} &= M\;(\mathbb{1}_N - J)^{-1}\;K\;\left(\alpha\hat{\mathbf{e}} + \beta\hat{\mathbf{h}}\right)\\ &\quad + M\alpha'\,\mathbf{e} + M\beta'\;(\mathbb{1}_N - J)^{-1}\,\mathbf{e},\end{aligned} \tag{39a}$$

$$\begin{aligned}K\dot{\hat{\mathbf{h}}} &= M\;(\mathbb{1}_N - J)^{-1}\;K\;\left(\gamma\hat{\mathbf{e}} + \kappa\hat{\mathbf{h}}\right)\\ &\quad + M\gamma'\,\mathbf{e} + M\kappa'\;(\mathbb{1}_N - J)^{-1}\,\mathbf{e}.\end{aligned} \tag{39b}$$

This reduced system evolves according to the eigenvalues of the matrix

$$A_{\text{r}} := \begin{pmatrix}\alpha & \beta\\ \gamma & \kappa\end{pmatrix}. \tag{40}$$

4.2 Firing-rate equilibrium for weak correlations

First, we constrain our study to weakly correlated inputs. In this case, we have $|\alpha| \gg |\beta|$ in (39a) and we can separate the evolution of $K\hat{\mathbf{e}}$ from that of $K\hat{\mathbf{h}}$. Namely,

$$\begin{aligned}\alpha \;&=\; \xi + (\hat{c}_1 + \hat{c}_2)\hat{\nu}_{\text{av}}\,[W*\epsilon](0)/4 \;\simeq\; \xi,\\ \xi \;&:=\; w^{\text{out}}\hat{\nu}_{\text{av}} + \widetilde{W}\hat{\nu}_{\text{av}}^2,\end{aligned} \tag{41}$$

and ξ is much larger in absolute value than β (as well as γ and κ) in A_{r}, cf. (38) and (40). It follows that $K\hat{\mathbf{e}}$ evolves much faster than $K\hat{\mathbf{h}}$, because $|\kappa| \ll |\alpha|$, in a similar way to the corresponding analysis for a single neuron (Kempter et al. 1999). We can thus consider $K\hat{\mathbf{e}}$ to be at its fixed point $K(\infty)\hat{\mathbf{e}}$ when stable, and then study the structure of the input weights through $K\hat{\mathbf{h}}$. The condition $\xi < 0$ ensures the stability of $K(t)\hat{\mathbf{e}}$, i.e., of the mean input weight for each neuron, and is the same as the condition in (26). The equilibrium of all individual firing rates at ν_{av}^* is equivalent to the stability of the mean input weight for each neuron at K_{av}^*.

In the remainder of Sect. 4, we consider small input correlations and require the stability of the firing rate for each neuron (embodied in $K\hat{\mathbf{e}}$) with an equilibrium value within the bounds, even if (41) is not strictly satisfied. Otherwise, input weights would end up clustered at a bound and the learning would then be null. When neglecting the inhomogeneities of J, the vector $K\hat{\mathbf{e}}$ can be approximated at the equilibrium by $K_{\mathrm{av}}\mathbf{e}$ at all times. However, discrepancies between the effective equilibrium value of $K\hat{\mathbf{e}}$ and the homeostatic equilibrium value $K^*_{\mathrm{av}}\mathbf{e}$ may occur depending on the correlation strengths, weight bounds or inhomogeneities in the network and initial conditions; see Fig. 7 for an example.

The evolution of $K\hat{\mathbf{h}}$ described by (39b) corresponds to the fixed point

$$K(\infty)\hat{\mathbf{h}} = -\frac{\gamma}{\kappa}\,K(\infty)\hat{\mathbf{e}} - \frac{\gamma'}{\kappa}\,\left(\mathbb{1}_N - J\right)\mathbf{e} - \frac{\kappa'}{\kappa}\,\mathbf{e} \qquad (42)$$

and it is determined by the sign of κ for the the stability as well as the respective positions of $K(0)\hat{\mathbf{h}}$ and $K(\infty)\hat{\mathbf{h}}$. We now elucidate the dynamics of $K\hat{\mathbf{h}}$ depending on the input parameters.

4.3 Two uncorrelated input pools with any firing rates

For uncorrelated inputs that have different firing rates, $\kappa = \widetilde{W}(\bar{\hat{\nu}}_1 - \bar{\hat{\nu}}_2)^2/4$, so it follows from (39b) that $K\hat{\mathbf{h}}$ is stable when $\widetilde{W} < 0$. The terms in the rhs of (42) are $\gamma/\kappa = 2\hat{\nu}_{\mathrm{av}}/(\bar{\hat{\nu}}_1 - \bar{\hat{\nu}}_2)$, $\gamma'/\kappa = 2w^{\mathrm{in}}/\widetilde{W}(\bar{\hat{\nu}}_1 - \bar{\hat{\nu}}_2)$ and $\kappa'/\kappa = 2\nu_0/(\bar{\hat{\nu}}_1 - \bar{\hat{\nu}}_2)$, so the vector elements of the fixed point $K(\infty)\hat{\mathbf{h}}$ have the same sign for homogeneous recurrent connectivity, which is given by

$$-2\frac{n^K_{\mathrm{av}}K^*_{\mathrm{av}}\hat{\nu}_{\mathrm{av}} + w^{\mathrm{in}}\left(1 - n^J_{\mathrm{av}}J_{\mathrm{av}}\right)/\widetilde{W} + \nu_0}{\bar{\hat{\nu}}_1 - \bar{\hat{\nu}}_2} = -\frac{2\left(1 - n^J_{\mathrm{av}}J_{\mathrm{av}}\right)w^{\mathrm{in}}}{\bar{\hat{\nu}}_1 - \bar{\hat{\nu}}_2}\;\frac{w^{\mathrm{out}}/\widetilde{W}}{w^{\mathrm{out}} + \widetilde{W}\hat{\nu}_{\mathrm{av}}}. \qquad (43)$$

We have used the expression of K^*_{av} in (27).

Since we required $K\hat{\mathbf{e}}$ to be stable, the conditions for homeostatic stability given in Sect. 3.1 must be satisfied: $w^{\mathrm{in}} > 0$ and $w^{\mathrm{out}} + \widetilde{W}\hat{\nu}_{\mathrm{av}} < 0$. Consequently, the sign of the vector elements of $K(\infty)\hat{\mathbf{h}}$ is the same as that of $w^{\mathrm{out}}/\widetilde{W}(\bar{\hat{\nu}}_1 - \bar{\hat{\nu}}_2)$. In both cases, where the fixed point $K(\infty)\hat{\mathbf{h}}$ is stable or unstable, depending on the sign of $\widetilde{W}$, the condition $w^{\mathrm{out}} < 0$ corresponds to the potentiation of the input weights coming from the pool with stronger firing rate. Recall that the same condition $w^{\mathrm{out}} < 0$ implies that the neuron firing rate ν_{av} increases with $\hat{\nu}_{\mathrm{av}}$ at the homeostatic equilibrium (cf. Sect. 3.1).

4.4 The two input pools have correlations and the same input firing rate

We now consider the special case where both input pools have the same firing rate equal to $\hat{\nu}_0$ and the vector $\hat{\boldsymbol{\nu}} = \hat{\nu}_0\hat{\mathbf{e}}$ is homogeneous. We thus have $\gamma' = \kappa' = 0$ in (42), and the fixed point for $K\hat{\mathbf{h}}$ reduces to

$$K(\infty)\hat{\mathbf{h}} = -\frac{\gamma}{\kappa}\,K(\infty)\hat{\mathbf{e}} \simeq -\frac{\hat{c}_1 - \hat{c}_2}{\hat{c}_1 + \hat{c}_2}\,n^K_{\mathrm{av}}\,K^*_{\mathrm{av}}\,\mathbf{e}\,, \qquad (44)$$

where K^*_{av} is given in (27). This fixed point is always unstable since $\kappa = [W * \epsilon](0)\hat{\nu}_0(\hat{c}_1 + \hat{c}_2)/4 > 0$, cf. (38), similar to the case of a single neuron (Kempter et al. 1999). This holds since $[W * \epsilon](0) > 0$ (see Sect. 2.3.7) and the correlation strengths $\hat{c}_1$ and $\hat{c}_2$ are positive. All the elements of the vector $K(\infty)\hat{\mathbf{h}}$ have the opposite sign to $\hat{c}_1 - \hat{c}_2$ when the equilibrium of the mean input weight for each neuron is realizable, viz., the vector $K(\infty)\hat{\mathbf{e}}$ has positive elements. It follows that the fixed point $K(\infty)\hat{\mathbf{h}}$ is determined by the balance between the input correlation strengths.

The instability will lead $K\hat{\mathbf{h}}$ to evolve in the opposite direction to the fixed point $K(\infty)\hat{\mathbf{h}}$. As a result, if the network starts with random initial input weights such that $K(0)\hat{\mathbf{h}} \simeq 0$, the weights coming from the input pool with stronger correlation will be potentiated compared to the weights from the other pool, as illustrated in Fig. 8a. This is similar to that seen in the case of feed-forward architecture with $\widetilde{W} > 0$ described by Kempter et al. (1999). It actually holds whatever the sign of $\widetilde{W}$ and the recurrent connections do not qualitatively change this behavior. When the initial conditions correspond to $K(0)\hat{\mathbf{h}} \neq 0$ but not too far from 0, then regardless of their initial specialization the input weights evolve into the "naturally" expected distribution, as shown in Fig. 7a. Note that the homeostatic equilibrium did not hold asymptotically (discrepancy between the thin solid and dotted lines) in that simulation.

However, if the input weights are initially already specialized such as $\bar{K}_{11}(0) \gg \bar{K}_{12}(0)$ then, despite $\hat{c}_2 > \hat{c}_1 = 0$, the initial "wrong" specialization may be preserved, contrary to the expected potentiation of the weights from the more correlated input pool, as illustrated in Fig. 7b.

The specific case $K(\infty)\hat{\mathbf{h}} = 0$, which occurs for e.g., when the two pools have the same correlation strength $\hat{c}_1 = \hat{c}_2$, will not be studied here and is left for a subsequent companion paper. When at least one of the input pools has correlation, the specific case where the matrix A_{r} in (40) is not invertible almost always leads to an unstable mean over the two input pools of the input weights for each neuron, i.e., $K(\infty)\hat{\mathbf{e}}$ will diverge to the bounds. This case is not interesting for learning and will not be considered further.

4.5 Distinct firing rates for the two correlated input pools

For a general choice of learning and input parameters, we have $\gamma' \neq 0$ and $\kappa' \neq 0$, and the signs of the vector elements of the fixed point $K(\infty)\hat{\mathbf{h}}$ in (42) depend on a complex relationship between the two mean input firing rates ($\bar{\hat{\nu}}_1$ and $\bar{\hat{\nu}}_2$) and mean correlation strengths ($\hat{c}_1$ and $\hat{c}_2$). Neglecting the inhomogeneities in the network, we approximate

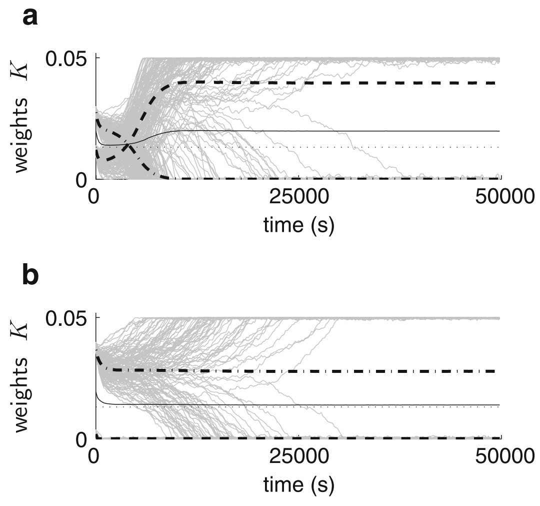


Fig. 7 Comparison between the weight evolution of different initial conditions for the same network configuration. The network consisted of $N = 100$ neurons and two pools of $M = 100$ inputs each, for the topology described in Fig. 1I. The partial input and recurrent connectivity were randomly generated with probability 30%. Input pool $\hat{2}$ had correlation ($\hat{c}_2 = 0.1$) while pool $\hat{1}$ had none; the firing rates were $\bar{\hat{\nu}}_1 = \bar{\hat{\nu}}_2 = 30$ Hz. The two plots show the evolution of individual weights (*grey bundle*); the mean weight $K_{\rm av}$ from the simulation (*thin solid line*); the analytically-predicted equilibrium value $K^*_{\rm av}$ (*thin dotted line*); the two simulated mean weights $\bar{K}_{11}$ (*thick dashed-dotted line*) from the uncorrelated pool $\hat{1}$ and $\bar{K}_{12}$ (*thick dashed line*) from the correlated pool $\hat{2}$. **a** For an initial distribution corresponding to the means $\bar{K}_{11}(0) = 0.028$ and $\bar{K}_{12}(0) = 0.012$, STDP inverted the weight distribution to potentiate $\bar{K}_{12}$ (*thick dashed line*) eventually. The homeostatic equilibrium held momentarily and then broke down. **b** When starting with means $\bar{K}_{11}(0) = 0.038$ and $\bar{K}_{12}(0) \simeq 0.002$, the initial weight distribution was not inverted by STDP. The weights corresponding to $\bar{K}_{12}$ (*thick dashed line*) are barely visible at zero in the plot. The homeostatic equilibrium held satisfactorily throughout the simulation, which determined the equilibrium value of $\bar{K}_{11}$ (*thick dashed-dotted line*)

$K(\infty)\hat{\mathbf{e}} \simeq n^K_{\rm av}K^*_{\rm av}\mathbf{e}$ and $(\mathbb{1}_N - J)\mathbf{e} \simeq (1 - n^J_{\rm av}J_{\rm av})\mathbf{e}$ in (42), in order to obtain

$$K(\infty)\hat{\mathbf{h}} \simeq -\frac{n^K_{\rm av}K^*_{\rm av}\gamma + (1 - n^J_{\rm av}J_{\rm av})\gamma' + \kappa'}{\kappa}\,\mathbf{e}. \tag{45}$$

If the input firing rates are very different, the γ' and κ' terms may dominate the γ term in (45), and hence $K(\infty)\hat{\mathbf{h}}$ depends on the input firing rates and not the input correlation strengths. On the other hand, we can obtain an approximate condition on the mean parameters to ensure that the input correlation strengths determine the sign of the numerator of (45), namely

$$\left|\frac{\hat{c}_1\bar{\hat{\nu}}_1 - \hat{c}_2\bar{\hat{\nu}}_2}{\bar{\hat{\nu}}_1 - \bar{\hat{\nu}}_2}\right| > h\left(\hat{\nu}_{\rm av}, \hat{C}^{W*\epsilon}_{\rm av}\right), \tag{46}$$

where the function h is defined by (74) in Appendix C. The formula is more interesting qualitatively than quantitatively: this condition is satisfied when the difference between the correlation strengths $\hat{c}_1 - \hat{c}_2$ is sufficiently large for given input firing rates $\bar{\hat{\nu}}_1$ and $\bar{\hat{\nu}}_2$, which can always be obtained when the difference $\bar{\hat{\nu}}_1 - \bar{\hat{\nu}}_2$ is not too large in absolute value. The recurrent connectivity generally affects such a balance between the input firing rates and correlations.

Under the qualitative condition of small discrepancies between the input firing rates, all the vector elements have the same sign given by

$$\mathrm{sgn}\left[K(\infty)\hat{\mathbf{h}}\right] = \mathrm{sgn}\left[-\frac{\gamma}{\kappa}K(\infty)\hat{\mathbf{e}}\right] = \mathrm{sgn}\left[\hat{c}_2 - \hat{c}_1\right]. \tag{47}$$

In this situation, the input weights from the more correlated input pool will be potentiated, as illustrated in Fig. 9 for partial input and recurrent connectivity, $\bar{\hat{\nu}}_1 > \bar{\hat{\nu}}_2$ and $\hat{c}_1 < \hat{c}_2$. The corresponding simulation with uncorrelated inputs would lead to a potentiation of the input weights from the first input pool since we have used $w^{\rm out} < 0$ (see Sect. 4.3). In that simulation, because of the inhomogeneities in J, the firing rates ended up stable though not clustered near that value, as shown in Fig. 9a; the homeostatic equilibrium for the weights was not strictly satisfied, cf. the discrepancies between the black thin solid line compared to the thin dashed line in Fig. 9b.

Figure 8b illustrates the dependence of the asymptotic weight specialization upon the input firing rates and correlation strengths with the same learning parameters: stronger input correlation is necessary to select a correlated input pool when its firing rate is smaller than that of the uncorrelated pool. Compared to the case where the two input pools have the same firing rate in Fig. 8a, the border (dotted line) between the potentiation of pool $\hat{1}$ and pool $\hat{2}$ is shifted to the left.

4.6 Extension to several homogeneous input pools

The analysis in Sect. 4 can be generalized to the case of an arbitrary number m of homogeneous input pools. It is possible to construct $m - 1$ vectors, $\hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_{m-1}$, in a similar way to $\hat{\mathbf{h}}$ above, in order to form an orthogonal basis together with $\hat{\mathbf{e}}$ in which A and B can be expressed in a form analogous to (37). For m pools of the same size, the vectors $\hat{\mathbf{h}}_1, \ldots, \hat{\mathbf{h}}_{m-1}$ can be constructed using the mth root of unity ($\hat{\mathbf{e}}$ corresponding to one), and the decomposition of K using this basis can be obtained using the discrete Fourier transform. The details are left to the reader.

5 Discussion

We have presented a mathematical framework in Sect. 2.3 to analyze the learning dynamics in a neuronal network with fixed recurrent connections and input connections whose weights are modified by STDP. The model uses the Poisson neuron model (Kempter et al. 1999) and can account

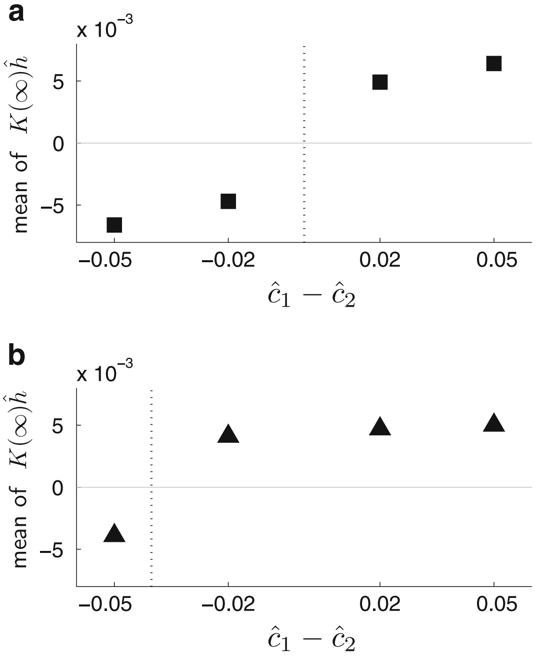


Fig. 8 Asymptotic weight specialization dependence upon the difference between the input correlation strengths; influence of the firing rates. The network corresponded to Fig. 1I with 30%-random partial connectivity for both input and recurrent connections; the weights were initially homogeneous with respective means $K_{\mathrm{av}}(0) = 0.02$ and $J_{\mathrm{av}} = 0.015$, and ±10% spread; both input and recurrent delays were randomly chosen with mean 7 ±2 ms. In each case, the four simulations corresponded to, respectively, $\hat{c}_1 = 0.05$ and $\hat{c}_2 = 0$; $\hat{c}_1 = 0.02$ and $\hat{c}_2 = 0$; $\hat{c}_1 = 0$ and $\hat{c}_2 = 0.02$; and $\hat{c}_1 = 0$ and $\hat{c}_2 = 0.05$. The plotted points represent the mean of $K(\infty)\hat{\mathbf{h}}$ over the neurons at the end of the simulation: it is positive if the network specialized to pool $\hat{1}$ and negative for pool $\hat{2}$. The *dotted line* indicates the demarcation border between the specializations to pool $\hat{1}$ and pool $\hat{2}$, estimated using similar calculations to those for (45). **a** For equal input firing rates $\hat{\bar{\nu}}_1 = \hat{\bar{\nu}}_2 = 30$ Hz, the difference $\hat{c}_1 - \hat{c}_2$ determines the specialization scheme and the squares have the same sign as $\hat{c}_1 - \hat{c}_2$. **b** When the firing rates $\hat{\bar{\nu}}_1 = 40$ Hz and $\hat{\bar{\nu}}_2 = 30$ Hz, the correlation strength $\hat{c}_2$ required for the network to specialize to pool $\hat{2}$ is higher: weak correlation strength still leads to the selection of pool $\hat{1}$ (*triangle* on the *right* of the demarcation *dotted line* with $\hat{c}_1 = 0$ and $\hat{c}_2 = 0.02$)

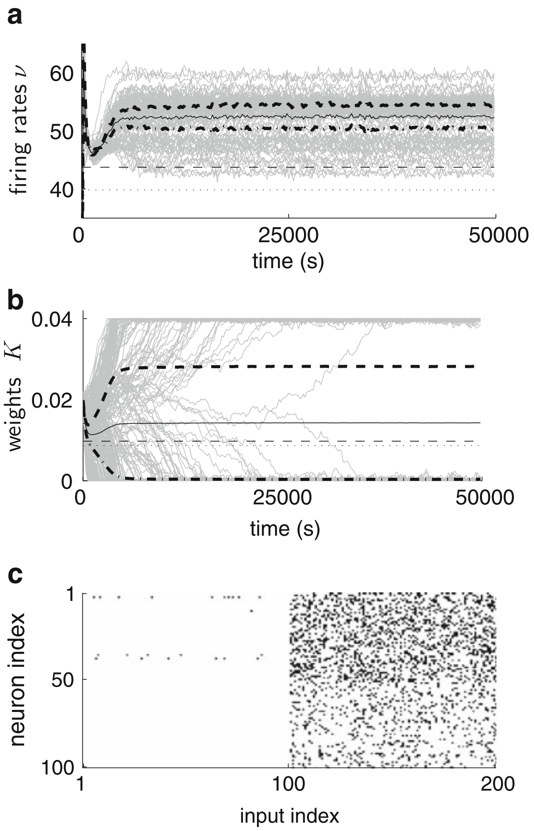


Fig. 9 Weight evolution for unbalanced correlations. The network of $N = 100$ neurons was stimulated by two pools of $M/2 = 100$ inputs each, with partial input and recurrent connectivity (30%). The input weights were initially homogeneous around the mean value 0.02 (±10%) while the recurrent weights were inhomogeneous with lumped feedback $\bar{J}_{11} = 0.45$, $\bar{J}_{12} = 0$, $\bar{J}_{21} = 0.9$ and $\bar{J}_{22} = 0.45$; see Fig. 6. The input firing rates and correlations were set to $\hat{\bar{\nu}}_1 = 35$ Hz, $\hat{\bar{\nu}}_2 = 30$ Hz, $\hat{c}_1 = 0.05$ and $\hat{c}_2 = 0.1$, respectively. **a** The firing rates ν (*grey bundle*; mean ν_{av} in *black thin solid line*) eventually stabilized not far from the predicted value for the homeostatic equilibrium in (29) (*thin dashed line*; the *dotted line* shows the corresponding value for uncorrelated inputs); the means for each half of the network are plotted ($\bar{\nu}_1$ in *thick dashed-dotted line* and $\bar{\nu}_2$ in *thick dashed line*). **b** The input weights K individually diverged (*grey bundle*) while the mean weight K_{av} (*thin solid line*) first converged towards and then stabilized close to the homeostatic equilibrium value (*thin dashed line*; the *thin dotted line* stands for uncorrelated inputs); see (27). The weights from the more correlated pool $\hat{2}$ ($\bar{K}_{12}$ in *thick dashed line*) became potentiated compared to those from pool $\hat{1}$ ($\bar{K}_{11}$ in *thick dashed-dotted line*). **c** Emerged structure in the weight matrix K. Darker pixels represent potentiated weights. Generally only weights coming from the more correlated input pool became potentiated (input indices #101 to #200, *right side*)

for any arbitrary connectivity topology and input structure, such as the case of two input pools with within-pool correlation. It incorporates the effect of the post-synaptic response, in this way extending previous work (Burkitt et al. 2007); however, dendritic delays (Senn et al. 2002) are ignored. For richer input signals than the delta-correlated pools (Sect. 2.3.7) considered in this paper, the post-synaptic response may play an important role (Sprekeler et al. 2007), which is captured by the theory developed in Sect. 2.3. The evolution of the input weights for slow learning is described by a dynamical system, which is analyzed in terms of fixed point and stability in order to predict the asymptotic behavior of the weights. This framework targets network dynamics beyond the mean-field approach in order to study the emergence of a network structure due to external stimulation.

In the case of learning input connections while the recurrent weights are kept fixed, homeostatic stability of both the firing rates and weights can be obtained for a wide range of learning parameters. The stability condition (26) was derived for weak input correlations, but it proved to be sufficient beyond the limitation of weak correlation strengths. Stability for additive STDP requires that the effect of a single pre-synaptic spike increases the weight. In numerical simulations we

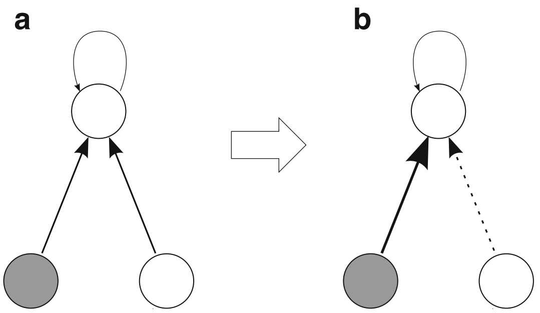


Fig. 10 Schematic representation of the input weight distribution (**a**) before and (**b**) after learning: potentiation of the input weights (*very thick arrow*) from the input pool with stronger within-pool correlation (*filled bottom circle*) at the expense of the weights from the less correlated pool (*open bottom circle*) that become depressed (*dashed arrow*). Other features are described in Fig. 1

chose $w^{\mathrm{in}} > 0$, $w^{\mathrm{out}} < 0$ and $\widetilde{W} < 0$, which leads to homeostatic stability whatever the input firing rate. This is in agreement with earlier numerical studies of integrate-and-fire (IF) neurons in feed-forward networks (Song et al. 2000) and recurrent networks (Song and Abbott 2001).

While the homeostatic equilibrium is satisfied, the individual weights exhibit a diverging behavior, indicating strong competition between them. For two correlated input pools with firing rates in the same range (but not necessarily identical), this generally results in selecting the input connections coming from the more correlated pool (Sect. 4), as illustrated in Fig. 10 for the case where only one input pool has correlations. Both the convergence of the mean input weight and the specialization (asymptotic bimodal weight distribution) are exponentially fast, the latter occurring on a slower time scale for weak input correlations. The value of the learning rate η was chosen to obtain a convergence towards the homeostatic equilibrium in hundreds of seconds, similar to Burkitt et al. (2007), and a development of a weight structure in tens of thousands of seconds (i.e., hours). Similar results were obtained with faster learning rates ($\eta = 10^{-5}$). Our results show that, even for small learning rates, the combination of equilibrium and diverging behavior leads to the emergence of a weight structure.

When starting with an initial homogeneous distribution of input weights, the presence of the fixed recurrent connections does not qualitatively change this robust specialization of the weights compared to a purely feed-forward architecture (Kempter et al. 1999). This behavior is obtained whatever the shapes of the PSP kernel ϵ and the STDP window function W provided STDP is "Hebbian" (cf. Sect. 2.1). Short durations of both ϵ and the recurrent delays were required for the analysis but similar conclusions at the mesoscopic network scale held when varying these parameters (Fig. 8). An exception to this expected behavior occurs for initial conditions in which the weights are already dramatically specialized in the "wrong" way or large difference between input firing rates (under certain conditions on the learning parameters); then STDP does not always select the more correlated input pathway, as shown in Fig. 7b and Fig. 8b.

The specific case where the two input pools have the same firing rate and the same correlation strength (Gütig et al. 2003) was not considered in this paper. In order to keep the analysis tractable, we chose the additive version of STDP (Kempter et al. 1999), which requires bounds on the weights for numerical simulation. Similar conclusions to those presented here hold for non-additive STDP in the case of sufficiently strong input correlations when starting from homogeneous initial input weights, provided the STDP model induces effective competition upon the weights unlike the model proposed by van Rossum et al. (2000). Comparison of the splitting of the weights between additive STDP and other versions (van Rossum et al. 2000; Gütig et al. 2003; Meffin et al. 2006; Morrison et al. 2007) is left for a subsequent paper.

In the generalized case where the network receives inputs from more than two pools (with small within-pool spike-time correlations), the following behaviors are expected:

- When input pools have different firing rates, the weights coming from pools with higher mean firing rate will be potentiated when $w^{\mathrm{out}} < 0$, resulting in a redistribution of the weights (equilibrium if $\widetilde{W} < 0$).
- For sufficiently large input spike-time correlation, the input weights diverge in a way that results in effectively splitting the sets of weights from each input pool.
- Weights coming from input pools with stronger spike-time correlations are potentiated.

Note that the number of input pools to be selected (which remains to be analyzed in future work) would depend upon both on the mean weight equilibrium value and on the weight bounds.

Our results show the importance of spike-time correlations in generating a structure among synaptic weights, in agreement with previous studies (Kempter et al. 1999; Gütig et al. 2003; Song et al. 2000; Song and Abbott 2001). The role played by spike-time correlations in the encoding of neuronal information is still under debate. It was shown, however, that the spiking dynamics of IF neurons, either isolated or within networks, are sensitive to the correlation structure of their inputs (Salinas and Sejnowski 2002; Moreno-Bote et al. 2008). A better understanding of the interplay between the learning and spiking dynamics is a promising way to provide insight into the encoding of neuronal information. Further study involving richer correlation structure and their link to neuronal synchrony also requires the synaptic delays to be incorporated in more detail (Senn et al. 2002; Lubenov and Siapas 2008). The case of learning on the recurrent connections will be the focus of a subsequent companion paper, which builds upon the framework developed in the present paper.

Acknowledgments The authors are greatly indebted to Chris Trengove, Sean Byrnes, Hamish Meffin, Michael Eager and Paul Friedel for their constructive comments. They are also grateful to Iven Mareels, Konstantin Borovkov, Dragan Nesic and Barry Hughes for helpful discussions. MG is funded by scholarships from the University of Melbourne and from NICTA. MG also benefited from an enjoyable stay at the Physik Department (T35) of the Technische Universität München. LvH gratefully acknowledges a most enjoyable stay at the Department of Electrical and Electronic Engineering at the University of Melbourne. LvH is partially supported by the BCCN–Munich. Funding is acknowledged from the Australian Research Council (ARC Discovery Project #DP0771815) and The Bionic Ear Institute.

Appendix A: Derivation of the covariance consistency equation

In this appendix, we derive the self-consistency equations for the covariance coefficients presented in Sect. 2.3.5, which leads to (16). This analysis includes the spike-triggering effects induced by the autocorrelation of the external inputs and of the neurons (Kempter et al. 1999), which were sometimes neglected (Burkitt et al. 2007). In addition, we incorporate the fine-timing effects such as delays and the time course of the PSP response.

A.1 Definition of the external input covariance

In (12) the following definition for the covariance between two external inputs k and l is used

$$\mathrm{Cov}[\hat{S}_k(t), \hat{S}_l(t+u)] \\ := \left\langle \hat{S}_k(t)\hat{S}_k(t+u)\right\rangle - \left\langle \hat{S}_k(t)\right\rangle\left\langle \hat{S}_l(t+u)\right\rangle. \tag{48}$$

We assume that the inputs $\hat{S}_k$ are second-order stationary processes, which means that these functions $\mathrm{Cov}[\hat{S}_k(t), \hat{S}_l(t+u)]$ are constant in t. As Hawkes (1971), we take the convention that all the $\mathrm{Cov}[\hat{S}_k(t), \hat{S}_l(t+u)]$ are continuous at $u = 0$, which means that they do not include the atomic discontinuity for $u = 0$ and $k = l$ due to the autocorrelation of the processes $\hat{S}_k$. We refer to 'complete covariance' for the second moment that includes the extra contribution $\langle \hat{S}_k(t)\rangle\delta(u)$ for each pair $k = l$, where δ is the Dirac delta function and $\langle \hat{S}_k(t)\rangle$ the constant firing rate.

This convention aims to discriminate between the intrinsic covariance resulting from autocorrelation (always present even for uncorrelated inputs) and the correlation structure that encodes spike synchronization. For uncorrelated inputs, the matrix $\hat{C}(t, u)$ defined in (12) satisfies $\hat{C}(t, u) = 0$ for all $u \in \mathbb{R}$. Therefore, it can be related to the spike-timing information conveyed by the external inputs and encoded in their covariance. However, in the derivation of the self-consistency covariance equations (16), we incorporate terms related to the autocorrelation of the external inputs in order to assess their impact on the learning dynamics.

A.2 Neuron-to-input covariance F^W

The expression for the time-average covariance coefficient F_{ik} in (11) arises from the definition

$$\mathrm{Cov}[S_i(t), \hat{S}_k(t+u)] \\ := \left\langle S_i(t)\,\hat{S}_k(t+u)\right\rangle - \langle S_i(t)\rangle\left\langle \hat{S}_k(t+u)\right\rangle. \tag{49}$$

We consider $\langle \hat{S}_k(t)\rangle$ to be constant in time, which implies that the instantaneous firing rate $\langle S_i(t)\rangle$ for neuron i is quasi-constant due to the slow variation of the weights. The last term in the above expression reduces to $\nu_i(t)\hat{\nu}_k(t)$ in (11).

A.2.1 Evaluation of the covariance using the past spiking history

The analysis presented here is based upon that of Hawkes processes (Hawkes 1971). Hawkes processes are stationary second-order processes, and the stationary property strictly holds here for fixed weights $K_{ik}(t)$ (J_{ij} being constant), which we assume in the remainder of this appendix (their dependence upon t will be suppressed here).

The pairwise neuron-to-input correlation $\langle S_i(t)\hat{S}_k(t+u)\rangle$ can be evaluated using the same "stochastic expansion" as Kempter et al. (1999) and Burkitt et al. (2007, Sect. 3.4). It consists of using the definition of the intensity function $\rho_i(t)$ (cf. (2)) and depends on the past activity of the external inputs and of the neurons

$$\left\langle S_i(t)\,\hat{S}_k(t+u)\right\rangle = \left\langle \rho_i(t)\,\hat{S}_k(t+u)\right\rangle. \tag{50}$$

However, this equality hides effects induced by autocorrelation, which arise since $\rho_i(t)$ has an implicit dependence upon $\hat{S}_k$.

A.2.2 Spike-triggering effect

The expression of $\rho_i(t)$ in (2) can be written as

$$\rho_i(t) = \nu_0 + \sum_j J_{ij}\left(\epsilon * S_j\right)(t - d_{ij}) \\ + \sum_l K_{il}\left(\epsilon * \hat{S}_l\right)(t - \hat{d}_{il}). \tag{51}$$

When $l = k$, an extra contribution in (50) due to the autocorrelation of the external input k needs to be taken into account, since this term is defined not to be included in $\mathrm{Cov}[\hat{S}_k(t), \hat{S}_l(t')]$ for $k = l$ and $t = t'$ (cf. (48)), as discussed in Appendix A.1. When substituting (51) into (50), the term corresponding to $k = l$ is

$$\int \epsilon(r)\,\hat{S}_k(t-\hat{d}_{ik}-r)\,\hat{S}_k(t+u)\,\mathrm{d}r \tag{52}$$

and taking the ensemble average $\langle\cdots\rangle$ leads to

$$\int \epsilon(r)\left\langle \hat{S}_k(t-\hat{d}_{ik}-r)\,\hat{S}_k(t+u)\right\rangle \mathrm{d}r \\ + \int \epsilon(r)\left\langle \hat{S}_k(t+u)\right\rangle \delta(u+r+\hat{d}_{ik})\,\mathrm{d}r, \tag{53}$$

where δ denotes the Dirac delta function. The second term of (53) is the spike-triggering effect: each pre-synaptic spike from input k induces an extra contribution due to the autocorrelation of input k. The integral in r and the ensemble average brackets $\langle\cdots\rangle$ were swapped (Fubini theorem) in order to obtain (53), and ϵ can be taken out of the angular brackets since it is a (deterministic) function. The spike-triggering effect occur for $t-r-\hat{d}_{ik} = t+u$, i.e., $r+u+\hat{d}_{ik} = 0$, and it reduces to

$$\epsilon(-u-\hat{d}_{ik})\left\langle \hat{S}_k(t+u)\right\rangle. \tag{54}$$

Taking this spike-triggering effect into account, (50) becomes

$$\left\langle S_i(t)\,\hat{S}_k(t+u)\right\rangle \\ = \nu_0\left\langle \hat{S}_k(t+u)\right\rangle \\ + \sum_j J_{ij}\int_{-\infty}^{+\infty} \epsilon(r)\left\langle S_j(t-r-d_{ij})\,\hat{S}_k(t+u)\right\rangle \mathrm{d}r \\ + \sum_l K_{il}\int_{-\infty}^{+\infty} \epsilon(r)\left\langle \hat{S}_l(t-r-\hat{d}_{il})\hat{S}_k(t+u)\right\rangle \mathrm{d}r \\ + K_{ik}\,\epsilon(-u-\hat{d}_{ik})\left\langle \hat{S}_k(t+u)\right\rangle \tag{55}$$

A.2.3 Time-averaging

We substitute the equalities (55) and (13) in (49) to express $\mathrm{Cov}[S_i(t'),\,\hat{S}_k(t'+u)]$

$$\mathrm{Cov}[S_i(t'),\,\hat{S}_k(t'+u)] \\ = \sum_j J_{ij}\int \epsilon(r)\,\mathrm{Cov}[S_i(t'-r-d_{ij}),\,\hat{S}_k(t'+u)]\,\mathrm{d}r \\ + \sum_l K_{il}\int \epsilon(r)\,\mathrm{Cov}[\hat{S}_l(t'-r-\hat{d}_{il}),\,\hat{S}_k(t'+u)]\,\mathrm{d}r \\ + K_{ik}\,\epsilon(-u-\hat{d}_{ik})\left\langle \hat{S}_k(t'+u)\right\rangle. \tag{56}$$

Then we integrate in t' over the time interval $[t-T,t]$ with T much larger than the time scale of the activation mechanisms, so that we can neglect the impact of the two changes of variables $t' \to t'+r+d_{ij}$, $t' \to t'+r+\hat{d}_{il}$ and $t' \to t'-u$, for the terms in the rhs of (56), respectively, as Kempter et al. (1999).

$$\int \mathrm{Cov}[S_i(t'),\,\hat{S}_k(t'+u)]\,\mathrm{d}t' \\ = \sum_j J_{ij}\iint \epsilon(r)\mathrm{Cov}[S_i(t'),\,\hat{S}_k(t'+u+r+d_{ij})]\,\mathrm{d}r\,\mathrm{d}t' \\ + \sum_l K_{il}\iint \epsilon(r)\mathrm{Cov}[\hat{S}_l(t'),\,\hat{S}_k(t'+u+r+\hat{d}_{il})]\,\mathrm{d}r\,\mathrm{d}t' \\ + K_{ik}\int \epsilon(-u-\hat{d}_{ik})\left\langle \hat{S}_k(t')\right\rangle \mathrm{d}t'. \tag{57}$$

Ignoring the slight modifications caused by the changes of variables in t', the integration bounds are $t' \in [t-T,t]$ and $r \in \mathbb{R}$. After the further changes of variables $r \to r-d_{ij}$ and $r \to r-\hat{d}_{il}$, we obtain

$$F_{ik}(t,u) = \sum_j J_{ij}\int \epsilon(r-d_{ij})F_{jk}(t,u+r)\,\mathrm{d}r \\ + \sum_l K_{il}\int \epsilon(r-\hat{d}_{il})\hat{C}_{lk}(t,u+r)\,\mathrm{d}r \\ + K_{ik}\,\epsilon(-u-\hat{d}_{ik})\,\hat{\nu}_k(t), \tag{58}$$

where we have used the time-averaged firing rates and covariances defined in (10), (11) and (12).

A.2.4 Use of the Fourier transform

The Fourier operator $\mathcal{F}$ between the domains of u and ω for a given function f ($\mathbf{i}$ is the complex root of -1) is given by

$$\mathcal{F}f(\omega) := \int_{-\infty}^{+\infty} f(u)\,\exp(-\mathbf{i}\omega u)\,\mathrm{d}u. \tag{59}$$

We evaluate the Fourier transform $\mathcal{F}F(\omega)$ of $F(t,u)$ using matrix notation and (58), for fixed t,

$$\mathcal{F}F(\omega) = \underline{J}(\omega)\,\mathcal{F}\epsilon(-\omega)\,\mathcal{F}F(\omega) \\ + \underline{K}(\omega)\,\mathcal{F}\epsilon(-\omega)\,\mathcal{F}\hat{C}(\omega) \\ + \underline{K}(\omega)\,\mathcal{F}\epsilon(-\omega)\ \mathrm{diag}\left(\hat{\boldsymbol{\nu}}\right), \tag{60}$$

where we defined $\underline{K}_{ik}(\omega) := K_{ik}\exp(\mathbf{i}\hat{d}_{ik}\omega)$ and $\underline{J}_{ij}(\omega) := J_{ij}\exp(\mathbf{i}d_{ij}\omega)$; $\mathrm{diag}(X)$ is the diagonal matrix whose diagonal elements are the vector X.

A.2.5 Sharp distribution of delays

In order to simplify the expressions for $\underline{K}$ and $\underline{J}$ in (60), we now assume that all the recurrent delays are identical ($d_{ij} = d$) and all the input delays are identical ($\hat{d}_{ik} = \hat{d}$). This is a good approximation for sharp distributions of each type of delay. To obtain $\mathcal{F}F^W(\omega)$ ($F^W(t)$ is given in (11)), we multiply (60) by $\exp(-\mathbf{i}\hat{d}\omega)\,\mathcal{F}W(-\omega)$,

$$\begin{aligned}\mathcal{F}F^W(\omega) &= \underline{J}(\omega)\,\mathcal{F}\epsilon(-\omega)\,\mathcal{F}F^W(\omega)\\ &\quad + K\,\mathcal{F}\left(W*\epsilon\right)(-\omega)\,\mathcal{F}\hat{C}(\omega)\\ &\quad + K\,\mathcal{F}\left(W*\epsilon\right)(-\omega)\;\mathrm{diag}\left(\hat{\boldsymbol{\nu}}\right), \qquad (61)\end{aligned}$$

where $\underline{K}(0) = K$. The expression for $\mathcal{F}F^W(\omega)$ is

$$\begin{aligned}&\mathcal{F}F^W(\omega) = \left[\mathbb{1}_N - \underline{J}(\omega)\,\mathcal{F}\epsilon(-\omega)\right]^{-1}\\ &\left[K\,\mathcal{F}\hat{C}^{W*\epsilon}(\omega) + \mathcal{F}\left(W*\epsilon\right)(-\omega)\;K\;\mathrm{diag}\left(\hat{\boldsymbol{\nu}}\right)\right], \qquad (62)\end{aligned}$$

similar to that given in Hawkes (1971, Eq. (21)).

Expanding the inverse $[\mathbb{1}_N - \underline{J}(\omega)\,\mathcal{F}\epsilon(-\omega)]^{-1}$ in a power series and taking the inverse Fourier transform of (62), $F^W(t)$ can be rigorously expressed

$$\begin{aligned}F^W(t) &= \sum_{n\geq 0} J^n\;K\;\hat{C}^{W*\epsilon*\epsilon_d^{\{n\}}}(t)\\ &\quad + \sum_{n\geq 0}\left[W*\epsilon*\epsilon_d^{\{n\}}\right](0)\;J^n\;K\;\mathrm{diag}\left(\hat{\boldsymbol{\nu}}(t)\right), \qquad (63)\end{aligned}$$

where $\epsilon_d(t) := \epsilon(t-d)$ and $\epsilon_d^{\{n\}}$ is the n^{th} iterated self-convolution of ϵ_d. We use the convention $W*\epsilon*\epsilon_d^{\{0\}} = W*\epsilon$. The two series are well defined for any PSP kernel ϵ provided all eigenvalues of the weight matrix J have a modulus strictly less than one. Note that the spike-triggering effect is of order M^{-1} compared to the remainder of the synaptic influx for each neuron (in the case of full input connectivity), embodied by the presence of $\mathrm{diag}(\hat{\boldsymbol{\nu}})$ in the last term of (63).

A.2.6 Impact of synaptic mechanisms on the covariance structure

When incorporating the effect of the PSP kernel ϵ and of the recurrent delay d, the input covariance $\hat{C}(t,u)$ in (63) is convolved with $W*\epsilon*\epsilon_d^{\{n\}}$ and not W alone. This implies that the separation between depression and potentiation for $W*\epsilon$ is slightly shifted to the right compared to that of W, as illustrated in Fig. 5. Consequently, an input spike that arrives almost immediately after the neuron fires does not cause depression but rather potentiation.

For homogeneous delta-correlated inputs with correlation strength $\hat{c}_{\mathrm{av}}$ and firing rate $\hat{\nu}_{\mathrm{av}}$ (cf. Sect. 2.3.7), we have

$$\hat{C}^{W*\epsilon*\epsilon_d^{\{n\}}} = \hat{c}_{\mathrm{av}}\;\hat{\nu}_{\mathrm{av}}\left[W*\epsilon*\epsilon_d^{\{n\}}\right](0). \qquad (64)$$

Then, $\left[W*\epsilon*\epsilon_d^{\{n\}}\right](0) > 0$ for all $n \geq 0$ provided $W(u) \geq 0$ for $u < 0$, as described in Sect. 2.1. This means that delta-correlated inputs always induce non-zero correlation coefficients $\hat{C}^{W*\epsilon}$ and thus a non-zero correlation structure F^W in the network. This provides a finer approximation than in Burkitt et al. (2007) and contrasts with the predictions stated in that paper, i.e., uncorrelated inputs will induce no correlation structure within the network. Note that the spike-triggering effect is always positive.

The terms of the series for $n \geq 1$ in (63) arise due to the recurrent connections. Only a finite number of terms are non-zero, since $\epsilon_d^{\{n\}}$ vanishes uniformly when $n \to \infty$ provided ϵ is not a Dirac delta function (i.e., has a finite time course), and the series reduces to a polynomial in J.

A.2.7 Short-duration PSPs and short recurrent delays

In general, the expression in (63) is not tractable. However, we can approximate the solution $\mathcal{F}F^W$ in (62) by making the further assumptions that d is small compared to the time scale of W, and that ϵ has a short time course compared to that of W. This implies that

$$\underline{J}(\omega)\,\mathcal{F}\epsilon(-\omega) \;\simeq\; \underline{J}(0)\,\mathcal{F}\epsilon(0) \;=\; J\,, \qquad (65)$$

which leads to the expression for $F^W(t)$ in (16), and corresponds to the analysis and the simulations in the main body of this paper.

Under these assumptions, the approximation of (65) used to derive (16) is equivalent to the following approximation in (61)

$$\begin{aligned}\mathcal{F}F^{W*\epsilon_d}(\omega) &= \exp(\mathbf{i}d\omega)\mathcal{F}\epsilon(-\omega)\mathcal{F}F^W(\omega)\\ &\simeq \mathcal{F}F^W(\omega). \qquad (66)\end{aligned}$$

In the time domain, this corresponds to

$$\int W(u-r)\epsilon(r-d)\,\mathrm{d}r \;\simeq\; W(u). \qquad (67)$$

The discrepancies are illustrated in Fig. 5, where W is represented by a solid line and $W*\epsilon_d$ by a dashed line. This approximation may be the source of the small discrepancies observed when comparing analytical solutions with numerical simulations.

A.2.8 Long recurrent delays

When d is large compared to the time scale of W, such that $\left[W*\epsilon*\epsilon_d^{\{n\}}\right](0) = 0$ for $n \geq 1$, only the first term of the series for $n = 0$ remains in (63), which becomes

$$F^W(t) \;=\; K\,\hat{C}^{W*\epsilon}(t) \;+\; [W*\epsilon](0)\;K\;\mathrm{diag}\left(\hat{\boldsymbol{\nu}}(t)\right). \qquad (68)$$

In this case, F^W does not depend on the recurrent weights J and has the same expression as in the case for a feed-forward architecture, i.e., $J = 0$ (Kempter et al. 1999; Sprekeler et al. 2007).

Appendix B: Analysis of the drift of K due to STDP with fixed J

We present in this appendix the main arguments about the general solution of the differential equation (31) that describes the evolution of input weights $K(t)$. The results are summarized in Sect. 3.2.

B.1 Symmetries of the inputs and reduction of dimensionality for K

Here, we decompose the space $\mathbb{M}_K$, in which $K(t)$ evolves according to (31), depending on the symmetries of the input pools and input connectivity. We want to reduce the dimensionality of the matrix $K(t)$ in order to eliminate the subspaces within which the drift $\dot{K}(t) = 0$ always, in order to focus on the complementary subspace where the drift is meaningful and leads to the development of a structure in $K(t)$. We constrain this section to full input connectivity (Φ_K is the identity and $\mathbb{M}_K = \mathbb{R}^{N\times M}$) but the results can also be applied to the case of partial connectivity, after the transform detailed in Appendix B.3.

For each symmetry of the input pools and input connectivity, say inputs #$\hat{1}$ and #$\hat{2}$ belong to the same input pool and are interchangeable, we can construct a M-column vector $\hat{\mathbf{u}}_D := [1, -1, 0, \ldots, 0]^{\mathbf{T}}$ such that $A\hat{\mathbf{u}}_D = 0$ and $B\hat{\mathbf{u}}_D = 0$, which leads to $\dot{K}(t)\hat{\mathbf{u}}_D = 0$ always whatever the value of $K(t)\hat{\mathbf{u}}_D$. This implies that the value of $K\hat{\mathbf{u}}_D$ is not constrained by the drift of the dynamics determined by (31). Furthermore, higher stochastic orders of the weight dynamics may affect this value without any effect on the drift of K. The ith element of the column vector $K\hat{\mathbf{u}}_D$ corresponds to the difference $K_{i1} - K_{i2}$ between the weights from these two inputs. A displacement of K along this sole direction in $\mathbb{M}_K$, i.e., modifying $K_{i1} - K_{i2}$ and preserving all other matrix components in a suitable basis, consists in a redistribution between the weights K_{i1} and K_{i2}.

In order to study the drift $\dot{K}(t)$, we can thus define equivalence classes $\bar{K}$ of the matrices K in $\mathbb{M}_K$ modulo such redistributions of weights that do not impact the drift. In other words, matrices K belonging to the same class $\bar{K}_0$ have the same drift $\dot{K}_0$. The drift of $K(t)$ is completely captured by the evolution of $\bar{K}(t)$ in the "reduced" vector space of the equivalence classes determined by the symmetries. The evolution of K within classes $\bar{K}$ is only due to higher orders of the stochastic processes. A similar reduction can be performed when the neurons and recurrent weights also have symmetries, as described in Fig. 6.

In the reduced space, equivalence classes $\bar{K}$ lump the weights that correspond to symmetries. Taking the example above with #$\hat{1}$ and #$\hat{2}$, $\bar{K}$ is only concerned about the sum $K_{i1} + K_{i2}$ for all indices i, not the difference $K_{i1} - K_{i2}$; we thus reduce $\mathbb{M}_K$ by M components. Using matrix notation, we focus on $K\hat{\mathbf{u}}_S$ with $\hat{\mathbf{u}}_S := [1, 1, 0, \ldots, 0]^{\mathbf{T}}$ and not $K\hat{\mathbf{u}}_D$, as defined above. Generalizing to the case of many symmetries, the elements of $\bar{K}$ can be taken equal to the mean input weights (instead of the sums of weights) averaged over the considered inputs and neurons, when several neurons are involved.

Such a reduction of dimensionality can also be applied in the case where the parameters within an input pool or a neuron group are not strictly identical, but where they are sharply distributed and the connectivity can be considered homogeneous up to some "noise" in the parameters. Then the equivalence classes correspond to the respective mean weights.

B.2 General evolution for full input connectivity

Let us consider the evolution of the drift $\dot{K}(t)$ described in (31) when the matrix A defined in (32) is non-invertible. We also assume full input connectivity: the matrix K evolves in the space $\mathbb{M}_K = \mathbb{R}^{N\times M}$. We show how the structure of A and B determines the evolution of the matrix $K(t)$. We first assume that A is diagonalizable and the space $\mathbb{R}^M$ can be decomposed into the direct sum of three subspaces of eigenvectors $\hat{\mathbf{u}}$.

By restricting A to the quotient space obtained by factoring out the null-space of A, we can use the same formula as (33), since the restriction of A is invertible on the quotient space ($A\hat{\mathbf{u}} \neq 0$ in this subspace; take $\hat{\mathbf{u}}_S$ in Appendix B.1 for e.g.). The restriction of the weight matrix K thus converges or diverges according to the (non-zero) eigenvalues of the restriction of A. In this case, K will evolve subject to constraints determined by A and B and the network will learn the input firing-rate and correlation structures. For example, the case of one stable and one unstable fixed points for two components of K is illustrated in Fig. 11.

We now examine the behavior of the weights in the intersection of the two null-spaces of A and B, i.e., eigenvectors $\hat{\mathbf{u}}$ such that $A\hat{\mathbf{u}} = 0$ and $B\hat{\mathbf{u}} = 0$, which implies $\dot{K}\hat{\mathbf{u}} = 0$. Such vectors $\hat{\mathbf{u}}$ exist, for e.g., when the network has symmetries, cf. $\hat{\mathbf{u}}_D$ in Appendix B.1. In this subspace, only higher orders of the stochastic dynamics (cf. Sect. 2.3.8) drive the evolution of the weights K and the value of $K\hat{\mathbf{u}}$ can be arbitrary; it depends in particular on the initial conditions. Changing the value of $K\hat{\mathbf{u}}$ corresponds to a redistribution of the strengths of the weights that has no impact on the weight structure.

Finally, for any eigenvector $\hat{\mathbf{u}}$ such that $A\hat{\mathbf{u}} = 0$ and $B\hat{\mathbf{u}} \neq 0$, we have $\dot{K}\hat{\mathbf{u}} = B\hat{\mathbf{u}} = \text{const}$. Thus, $K\hat{\mathbf{u}}$ will grow linearly in time until the weights hit the bounds. This situation, however, corresponds to very specific values of the input and learning parameters. For example, from (32) the choice $\hat{\mathbf{v}} = -w^{\text{out}}\hat{\mathbf{e}}/\widetilde{W}$ with uncorrelated inputs ($\hat{C}^W = 0$) gives $A = 0$ and $B = w^{\text{in}}\hat{\mathbf{e}}\hat{\mathbf{v}}^{\mathbf{T}} \neq 0$, when $w^{\text{in}} \neq 0$. We do not investigate this case any further.

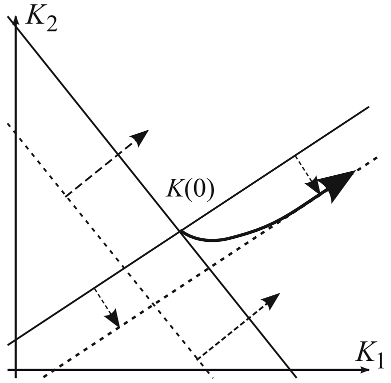


Fig. 11 Example of evolution of K in two dimensions. The direction of each *thick solid line* is determined by $\hat{\mathbf{u}}_1$ and $\hat{\mathbf{u}}_2$, respectively, their intersection corresponds to $K(0)$. The *thick dotted lines* correspond to one stable fixed point $K(\infty)\hat{\mathbf{u}}_1$ (*dashed arrows* pointing towards the *dotted line*) and one unstable fixed point $K(\infty)\hat{\mathbf{u}}_2$ (*dashed arrows* pointing away from the *dotted line*). Learning causes the value of $K(t)$ to reach the line corresponding to $K\hat{\mathbf{u}}_1 = K(\infty)\hat{\mathbf{u}}_1$ while pushing it away from the intersection of the *dashed lines* (*thick arrow*) until reaching the upper bound of K_1

In general, A is not diagonalizable and the decomposition above is not as simple; for e.g., the image of A can intersect with its null-space. But the set of diagonalizable matrices is dense in $\mathbb{M}_K$ and thus the previous analysis can be extended to all matrices in $\mathbb{M}_K$, because the behavior of K qualitatively depends on the eigenvalues of A only.

B.3 Partial input connectivity

We now look at the dynamics of $K(t)$ for partial input connectivity, when Φ_K nullifies some terms related to non-existing connections; the space $\mathbb{M}_K$ is then a strict subspace of $\mathbb{R}^{N\times N}$. Instead of considering K a matrix, we take it as a column vector $\breve{K}$ indexed by the duplet (i,k) such that the connection $k \to i$ exists and we omit the elements nullified by Φ_K. Equation (31) becomes

$$\dot{\breve{K}} = L\breve{K} + \breve{B}, \tag{69}$$

where $\breve{B}$ is the column vector constructed from B in the same way as $\breve{K}$ from K, and L is a square matrix of dimension $n^K \times n^K$ defined by its elements indexed by $\{(i,k)(i',k')\}$

$$L_{\{(i,k)(i',k')\}} := \tilde{J}_{ii'} A_{k'k}, \tag{70}$$

where $\tilde{J} := (\mathbb{1}_N - J)^{-1}$. This equation can be analyzed in the same way as in Appendix B.2, with a basis of column vectors $\breve{\mathbf{u}}$ instead of $\hat{\mathbf{u}}$. The cases where $\breve{\mathbf{u}}^{\mathbf{T}} L$ and $\breve{\mathbf{u}}^{\mathbf{T}} \breve{B}$ are zero or non-zero are to be considered as above (note the transposition '$\mathbf{T}$'). It follows that the evolution of $\breve{K}$ can be decomposed into evolution within three subspaces as in Appendix B.2.

When A is invertible, a generalization of (33) can be written as

$$K(t) = K(\infty) + \sum_{n\geq 0} \frac{t^n}{n!} \tilde{K}_n \tag{71}$$

with

$$\begin{aligned}
\tilde{K}_{n+1} &:= \Phi_K \left[(\mathbb{1}_N - J)^{-1} \; \tilde{K}_n \; A \right], \\
\tilde{K}_0 &:= K(0) - K(\infty), \\
K(\infty) &= -\Phi_K \left[(\mathbb{1}_N - J) \; \Phi_K\,(B) \; A^{-1} \right].
\end{aligned} \tag{72}$$

Appendix C: Dependence of the fixed point $K(\infty)\hat{\mathbf{h}}$ upon input correlation

Here, we derive a condition on the input correlation strengths $\hat{c}_1$ and $\hat{c}_2$ such that the sign of the elements of the fixed point $K(\infty)\hat{\mathbf{h}}$ in (45) is determined by the balance between the correlations $\hat{c}_1$ and $\hat{c}_2$ and not by that of the firing rates $\bar{\hat{\nu}}_1$ and $\bar{\hat{\nu}}_2$. Recall that this sign determines the evolution of K, i.e., which input pathway is potentiated by learning, for homogeneous initial input weights.

We focus on the role of the correlation strengths $\hat{c}_1$ and $\hat{c}_2$ in the numerator $n^K_{\text{av}} K^*_{\text{av}} \gamma + (1 - n^J_{\text{av}} J_{\text{av}})\gamma' + \kappa'$. Multiplying the numerator by the denominator of K^*_{av} in (27) gives

$$\begin{aligned}
&\left[\left(1 - n^J_{\text{av}} J_{\text{av}}\right) w^{\text{in}} \hat{\nu}_{\text{av}} + \nu_0 \left(w^{\text{out}} + \widetilde{W} \hat{\nu}_{\text{av}}\right) \right] \\
&\quad \times \left[\widetilde{W} \hat{\nu}_{\text{av}} \frac{\bar{\hat{\nu}}_1 - \bar{\hat{\nu}}_2}{2} + [W * \epsilon](0) \frac{\hat{c}_1 \bar{\hat{\nu}}_1 - \hat{c}_2 \bar{\hat{\nu}}_2}{4} \right] \\
&\quad - \left[(1 - n^J_{\text{av}} J_{\text{av}}) w^{\text{in}} \frac{\bar{\hat{\nu}}_1 - \bar{\hat{\nu}}_2}{2} + \widetilde{W} \nu_0 \frac{\bar{\hat{\nu}}_1 - \bar{\hat{\nu}}_2}{2} \right] \\
&\quad \times \left[\hat{\nu}_{\text{av}} \left(w^{\text{out}} + \widetilde{W} \hat{\nu}_{\text{av}}\right) + \hat{C}^{W*\epsilon}_{\text{av}} \right] \\
&= \frac{\bar{\hat{\nu}}_1 - \bar{\hat{\nu}}_2}{2} \hat{C}^{W*\epsilon}_{\text{av}} \left[-\left(1 - n^J_{\text{av}} J_{\text{av}}\right) w^{\text{in}} - \widetilde{W} \hat{\nu}_0 \right] \\
&\quad + \frac{\hat{c}_1 \bar{\hat{\nu}}_1 - \hat{c}_2 \bar{\hat{\nu}}_2}{4} [W * \epsilon](0) \\
&\quad \times \left[\left(1 - n^J_{\text{av}} J_{\text{av}}\right) w^{\text{in}} \hat{\nu}_{\text{av}} + \nu_0 \left(w^{\text{out}} + \widetilde{W} \hat{\nu}_{\text{av}}\right) \right],
\end{aligned} \tag{73}$$

where we have used the means $\hat{\nu}_{\text{av}} = (\bar{\hat{\nu}}_1 + \bar{\hat{\nu}}_2)/2$ and $\hat{C}^{W*\epsilon}_{\text{av}} = [W * \epsilon](0)(\hat{c}_1 \bar{\hat{\nu}}_1 + \hat{c}_2 \bar{\hat{\nu}}_2)/4$; the expressions for γ, γ' and κ' are given in (38).

If the difference $\hat{c}_1 \bar{\hat{\nu}}_1 - \hat{c}_2 \bar{\hat{\nu}}_2$ dominate the rhs of (73), then the correlation strengths $\hat{c}_1$ and $\hat{c}_2$ determine the sign of the fixed point $K(\infty)\hat{\mathbf{h}}$. This condition can be rewritten

$$\begin{aligned}
&\left| \frac{\hat{c}_1 \bar{\hat{\nu}}_1 - \hat{c}_2 \bar{\hat{\nu}}_2}{\bar{\hat{\nu}}_1 - \bar{\hat{\nu}}_2} \right| \\
&\quad > \frac{2\hat{C}^{W*\epsilon}_{\text{av}}}{[W * \epsilon](0)} \left| \hat{\nu}_{\text{av}} + \frac{w^{\text{out}} \nu_0}{\left(1 - n^J_{\text{av}} J_{\text{av}}\right) w^{\text{in}} + \widetilde{W} \nu_0} \right|^{-1}.
\end{aligned} \tag{74}$$

Table 1 Table of simulation parameters

Time step	10^{-4} s
Simulation duration	10^{5} s
Input Poisson spike trains	
Firing rates	$\hat{\nu}_{\rm av} = 30$–35 Hz
Correlation strength	$\hat{c}_{\rm av} = 0$–0.1
Poisson neurons	
Instantaneous firing rate	$\nu_0 = 5$ Hz
Synapses	
Rise time constant	$\tau_A = 1$ ms
Decay time constant	$\tau_B = 5$ ms
Mean of recurrent delays	$d = 3$ ms
Spread of recurrent delays	± 1 ms
Mean of input delays	$\hat{d} = 7$ ms
Spread of input delays	± 1 ms
STDP	
Learning parameter	$\eta = 5 \times 10^{-7}$
Pre-synaptic rate-based coeff.	$w^{\rm in} = 4$
Post-synaptic rate-based coeff.	$w^{\rm out} = -0.5$
Potentiation time constant	$\tau_P = 17$ ms
Potentiation scaling coefficient	$c_P = 15$
Depression time constant	$\tau_D = 34$ ms
Depression scaling coefficient	$c_D = 10$

Appendix D: Simulation parameters

The results in this paper were obtained using discrete-time numerical simulation and the parameters listed in Table 1, unless stated otherwise. The STDP window function W is given by

$$W(u) = \begin{cases} c_P \exp(u/\tau_P) & \text{for } u < 0 \\ -c_D \exp(-u/\tau_D) & \text{for } u > 0. \end{cases} \tag{75}$$

The PSP kernel ϵ is defined by

$$\epsilon(t) = \begin{cases} \frac{\exp(t/\tau_B) - \exp(t/\tau_A)}{\tau_B - \tau_A} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0. \end{cases} \tag{76}$$

The synaptic weights are not normalized, but defined such that the sum of the pre-synaptic weights for each neuron is of the order of one. This implies that the effective rate of change per second for the weights is roughly three orders of magnitude below (10^{-3}) their upper bound. These parameters are in the same range as those used in previous studies (Kempter et al. 1999; Burkitt et al. 2007).

References

Bi GQ, Poo MM (2001) Synaptic modification by correlated activity: Hebb's postulate revisited. Annu Rev Neurosci 24:139–166

Burkitt AN, Meffin H, Grayden DB (2004) Spike-timing-dependent plasticity: The relationship to rate-based learning for models with weight dynamics determined by a stable fixed point. Neural Comput 16(5):885–940

Burkitt AN, Gilson M, van Hemmen JL (2007) Spike-timing-dependent plasticity for neurons with recurrent connections. Biol Cybern 96(5):533–546

Câteau H, Kitano K, Fukai T (2008) Interplay between a phase response curve and spike-timing-dependent plasticity leading to wireless clustering. Phys Rev E 77(5):051909

Gerstner W, Kempter R, van Hemmen JL, Wagner H (1996) A neuronal learning rule for sub-millisecond temporal coding. Nature 383(6595):76–78

Gilson M, Burkitt AN, Grayden DB, Thomas DA, van Hemmen JL (2009a) Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks II: input selectivity—symmetry breaking. doi:10.1007/s00422-009-0320-y

Gilson M, Burkitt AN, Grayden DB, Thomas DA, van Hemmen JL (2009b) Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks III: Partially connected neurons driven by spontaneous activity. Preprint

Gütig R, Aharonov R, Rotter S, Sompolinsky H (2003) Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. J Neurosci 23(9):3697–3714

Hawkes AG (1971) Point spectra of some mutually exciting point processes. J Roy Statist Soc Ser B 33(3):438–443

Hebb DO (1949) The organization of behavior: a neuropsychological theory. Wiley, NY

van Hemmen JL (2001) Theory of synaptic plasticity. In: Moss F, Gielen S (eds) Handbook of biological physics, vol 4: neuro-informatics and neural modelling. Elsevier, Amsterdam, pp 771–823

Kang S, Kitano K, Fukai T (2008) Structure of spontaneous UP and DOWN transitions self-organizing in a cortical network model. PLoS Comput Biol 4(3):e1000022

Karbowski J, Ermentrout GB (2002) Synchrony arising from a balanced synaptic plasticity in a network of heterogeneous neural oscillators. Phys Rev E 65(3):031902

Kempter R, Gerstner W, van Hemmen JL (1999) Hebbian learning and spiking neurons. Phys Rev E 59(4):4498–4514

Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biol Cybern 43(1):59–69

Lubenov EV, Siapas AG (2008) Decoupling through synchrony in neuronal circuits with propagation delays. Neuron 58(1): 118–131

Markram H, Lubke J, Frotscher M, Roth A, Sakmann B (1997) Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. J Physiol (Lond) 500(2):409–440

Masuda N, Kori H (2007) Formation of feedforward networks and frequency synchrony by spike-timing-dependent plasticity. J Comput Neurosci 22(3):327–345

Meffin H, Besson J, Burkitt AN, Grayden DB (2006) Learning the structure of correlated synaptic subgroups using stable and competitive spike-timing-dependent plasticity. Phys Rev E 73(4):041911

Moreno-Bote R, Renart A, Parga N (2008) Theory of input spike auto- and cross-correlations and their effect on the response of spiking neurons. Neural Comput 20(7):1651–1705

Morrison A, Aertsen A, Diesmann M (2007) Spike-timing-dependent plasticity in balanced random networks. Neural Comput 19(6):1437–1467

Morrison A, Diesmann M, Gerstner W (2008) Phenomenological models of synaptic plasticity based on spike timing. Biol Cybern 98(6):459–478

van Rossum MCW, Bi GQ, Turrigiano GG (2000) Stable Hebbian learning from spike timing-dependent plasticity. J Neurosci 20(23):8812–8821

Salinas E, Sejnowski TJ (2002) Integrate-and-fire neurons driven by correlated stochastic input. Neural Comput 14(9):2111–2155

Senn W, Schneider M, Ruf B (2002) Activity-dependent development of axonal and dendritic delays, or, why synaptic transmission should be unreliable. Neural Comput 14(3):583–619

Sjöström PJ, Turrigiano GG, Nelson SB (2001) Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. Neuron 32(6):1149–1164

Sjöström PJ, Turrigiano GG, Nelson SB (2004) Endocannabinoid-dependent neocortical layer-5 LTD in the absence of postsynaptic spiking. J Neurophysiol 92(6):3338–3343

Song S, Abbott LF (2001) Cortical development and remapping through spike timing-dependent plasticity. Neuron 32(2):339–350

Song S, Miller KD, Abbott LF (2000) Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. Nat Neurosci 3(9):919–926

Sprekeler H, Michaelis C, Wiskott L (2007) Slowness: An objective for spike-timing-dependent plasticity? PLoS Comput Biol 3(6):1136–1148

Wenisch OG, Noll J, van Hemmen JL (2005) Spontaneously emerging direction selectivity maps in visual cortex through STDP. Biol Cybern 93(4):239–247