Biol Cybern (2009) 101:103–114
DOI 10.1007/s00422-009-0320-y

ORIGINAL PAPER

Biological Cybernetics

Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. II. Input selectivity—symmetry breaking

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

Received: 19 December 2008 / Accepted: 14 May 2009 / Published online: 18 June 2009

Abstract Spike-timing-dependent plasticity (STDP) is believed to structure neuronal networks by slowly changing the strengths (or weights) of the synaptic connections between neurons depending upon their spiking activity, which in turn modifies the neuronal firing dynamics. In this paper, we investigate the change in synaptic weights induced by STDP in a recurrently connected network in which the input weights are plastic but the recurrent weights are fixed. The inputs are divided into two pools with identical constant firing rates and equal within-pool spike-time correlations, but with no between-pool correlations. Our analysis uses the Poisson neuron model in order to predict the evolution of the input synaptic weights and focuses on the asymptotic weight distribution that emerges due to STDP. The learning dynamics induces a symmetry breaking for the individual neurons, namely for sufficiently strong within-pool spike-time correlation each neuron specializes to one of the input pools. We show that the presence of fixed excitatory recurrent connections between neurons induces a group symmetry-breaking effect, in which neurons tend to specialize to the same input pool. Consequently STDP generates a functional structure on the input connections of the network.

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

M. Gilson (✉) · A. N. Burkitt · D. B. Grayden · D. A. Thomas
Department of Electrical and Electronic Engineering,
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), BCCN Munich,
Technische Universität München,
85747 Garching bei München, Germany

1 Introduction

Learning at the level of neurons is believed to give rise to functional pathways in neuronal networks, which take part in distinguishing stimuli when performing a recognition task. For example, the primary visual cortex is organized into areas sensitive to different aspects of visual stimuli, viz., ocular dominance and orientation fields (Hubel and Wiesel 1962). Possible underlying mechanisms have been the subject of many studies (von der Malsburg 1973; Swindale 1996; Choe and Miikkulainen 1998; Elliott and Shadbolt 1999; Wenisch et al. 2005; Goodhill 2007) to reproduce and explain such synaptic self-organization (Kohonen 1982).

Synaptic plasticity describes mechanisms that take place at the "connection" site between two neurons (synapse), when the synaptic weight related to the post-synaptic response to a single pulse is strengthened (potentiation) or weakened (depression). Recent studies have established the importance of the timing of individual spikes in synaptic plasticity (Gerstner et al. 1996; Markram et al. 1997; Bi and Poo 2001), leading to the concept of spike-timing-dependent plasticity (STDP). Previous studies have shown how STDP can implement input selectivity according to the spike-time correlation structure of input spike trains for single neurons and feed-forward networks (Kempter et al. 1999; Leibold et al. 2002; Gütig et al. 2003; Burkitt et al. 2004; Meffin et al. 2006). In a companion paper (Gilson et al. 2009), we extended the theory developed by Kempter et al. (1999), Burkitt et al.

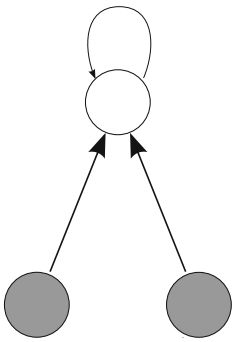


Fig. 1 Schematic illustration of the network setup studied in this paper. The neuronal network (*top circle*) has fixed recurrent connections (*thin arrows*) and is stimulated by two pools of external inputs with equal within-pool spike-time correlation (*filled bottom circles*) with plastic input weights (*thick arrows*)

(2007) to the case of a recurrent network where the input connections are subject to STDP and the recurrent connections are fixed. We showed that STDP can induce at the same time for each neuron both a homeostatic equilibrium on the input weights, in which the mean pre-synaptic input weight and the output firing rate stabilize over time, and a potentiation of some input weights depending on the input correlation structure.

In this paper, we extend that previous analysis (Gilson et al. 2009) to focus on the case where the network illustrated in Fig. 1 is stimulated by two input pools whose spike trains have similar characteristics, namely homogeneous initial input weights, identical firing rates and equal within-pool spike-time correlations. We analyze particularly the concept of symmetry breaking, which consists of the specialization of the neurons to just *one* of the input pools. It has been previously demonstrated that STDP can implement this symmetry breaking for a single neuron stimulated by two input pools having the same firing rate and spike-time correlation: STDP causes a neuron with initially homogeneous input weights to become sensitive to *only one* of the two pools (Gütig et al. 2003). In this paper we extend the study by Gütig et al. (2003) to the case of a *recurrently* connected network stimulated by two input pools as described above. We use a framework developed in a companion paper (Gilson et al. 2009), which describes how the network activity, viz. firing rates and spike-time correlations, determines the weight evolution that occurs on a slower time scale than the neuronal activation mechanisms.

After presenting STDP model (Sect. 2.1) and neuron model (Sect. 2.2) that we use, we recapitulate the theoretical framework developed by Gilson et al. (2009) to describe the evolution of the neuronal activity in a network consisting of an arbitrary number of neurons stimulated by an arbitrary number of external sources (Sect. 2.3). We then investigate the impact of fixed recurrent connections on the learning dynamics in the case of two pools of external inputs with identical characteristics (Sect. 3).

2 Modeling learning and neuronal activity

2.1 Hebbian additive STDP

Spike-timing-dependent plasticity describes the change of synaptic weight according to the precise timing of spikes. Here, we constrain our study to the contributions induced by single spikes and pairs of spikes, and use the so-called Hebbian additive STDP model (Kempter et al. 1999). We keep in mind its limitations compared to more elaborate versions of STDP involving, for example, weight-dependence (van Rossum et al. 2000; Bi and Poo 2001; Gütig et al. 2003) or triplets of spikes (Sjöström et al. 2001; Pfister and Gerstner 2006; Appleby and Elliott 2006); see Morrison et al. (2008) for a review. 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 times t^{in} and t^{out} respectively is given by the sum of three 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 constant w^{in} (resp. w^{out}) accounts for the effect of each pre-synaptic (post-synaptic) spike, which occurs at time t^{in} (t^{out}). The STDP learning window function W describes the contribution of each pair of pre- and post-synaptic spikes in terms of the difference between the spike times $t^{\text{in}} - t^{\text{out}}$ (Gerstner et al. 1996; Kempter et al. 1999). Figure 2 illustrates a typical choice of W where pre-synaptic spikes that take part in the firing of post-synaptic spikes induce a weight potentiation (Hebb 1949). 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. See Gilson et al. (2009) for more details.

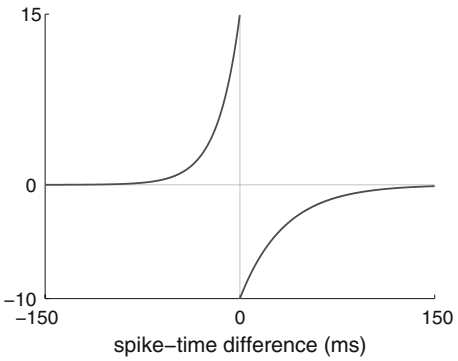


Fig. 2 Example of STDP window function W. It consists of one decaying exponential for potentiation (*left curve*) with time constant 17 ms and one for depression (*right curve*) with 34 ms. See Appendix C for details on the parameters

2.2 Poisson neuron model

In the Poisson neuron model, 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 $S_i(t)$ (Kempter et al. 1999). 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 synapses (indexed by k), whose spike-time series are denoted by $\hat{S}_k(t)$,

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

The constant ν_0 is the spontaneous firing rate (identical for all the neurons), which accounts for other pre-synaptic connections that are not considered in detail. Each pre-synaptic spike induces a variation of $\rho_i(t)$ taken care of by the post-synaptic potential (PSP), which is determined by the synaptic weights K_{ik}, the post-synaptic response kernel ϵ, and the delays $\hat{d}_{ik}$. The kernel function ϵ models the PSP due to the current injected into the post-synaptic neuron as a consequence of a single pre-synaptic spike; $\epsilon(t)$ is normalized to one, i.e., $\int \epsilon(t)\mathrm{d}t = 1$, and in order to preserve causality, we have $\epsilon(t) = 0$ for $t < 0$. A PSP also incorporates a fixed delay, denoted by $\hat{d}_{ik}$ for the kth synaptic connection to neuron i. The overall synaptic influx is the sum of the PSPs over all spike times $\hat{t}_{k,n}$ (related to the kth synapse, and indexed by n). We only consider positive weights here, i.e., excitatory synapses. See Gilson et al. (2009) for more details.

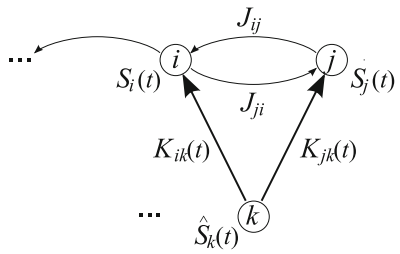
2.3 Learning equations and network activity

In a companion paper (Gilson et al. 2009), a dynamical system of equations was derived to describe the impact of STDP on the neuronal dynamics, where the variables of importance are the firing rates and the pairwise spike-time correlations (cf. Sect. 2.1), as well as the weights. In the present paper, we use this framework together with further-derived equations to study the weights dynamics in more depth. The same assumptions are used: the synaptic plasticity occurs on a much slower time scale than the neuronal activation dynamics (adiabatic hypothesis); the expectation values of the firing rates and pairwise covariances are constant in time for the external inputs, hence quasi-constant for the network neurons respectively.

2.3.1 Description of the network activity

We consider a recurrently connected network of N Poisson neurons stimulated by M Poisson spike trains (a.k.a. external inputs or sources), as shown in Fig. 3a. In addition to receiving synaptic input from some external sources, each neuron also excites other neurons via connections that may form feedback loops in the network, but without self-connections. Typically both M and N are large; n^K denotes the number of input connections and n^J the number of recurrent connections in a given network. Partially connected networks are generated by randomly assigning input-to-neuron and neuron-to-neuron connections.

a



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. 3 Presentation of the network notation. **a** Schematic representation of one of the M external inputs (*bottom circle*, indexed by $1 \le k \le N$) and two of the N network neurons (*top circles*, $1 \le i, j \le N$). The input connections have plastic weights $K_{ik}(t)$ (*thick arrows*) while the recurrent connections have fixed weights J_{ij} (*thin arrows*). The output spike trains of the input k and neuron i 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 $\boldsymbol{\nu}$; covariance coefficients $\hat{C}$ and F; and the variables related to the synaptic connections: weights K and J; delays $\hat{d}$ and d

Spikes are considered to be instantaneous events; the spike trains of the external input k ($1 \le k \le M$) and of the network neuron i ($1 \le i \le N$) are modeled using delta-functions (Dirac comb) and denoted by $\hat{S}_k(t)$ and $S_i(t)$, respectively. We define the time-averaged firing rate $\nu_i(t)$ of neuron i, using T as much larger time scale than that of the neuronal activation mechanisms but much smaller than the learning time scale related to η^{-1} (Kempter et al. 1999; Burkitt et al. 2007; Gilson et al. 2009),

$$\nu_i(t) := \frac{1}{T} \int_{t-T}^{t} \left\langle S_i(t') \right\rangle \mathrm{d}t', \qquad (3)$$

where $\langle S_i(t) \rangle$ is the instantaneous firing rate averaged over the randomness, which is self-averaging (van Hemmen 2001). Likewise for the firing rate $\hat{\nu}_k(t)$ of input k, which is constant in time.

We use the following neuron-to-input time-averaged covariance F_{ik} and covariance coefficient F_{ik}^{Ψ} for a given

kernel Ψ between neuron i and input k (Gilson et al. 2009)

$$F_{ik}(t,u) := \frac{1}{T}\int_{t-T}^{t} \left\langle S_i(t')\,\hat{S}_k(t'+u)\right\rangle \mathrm{d}t' - \frac{1}{T}\int_{t-T}^{t} \left\langle S_i(t')\right\rangle\left\langle \hat{S}_k(t'+u)\right\rangle \mathrm{d}t' , \tag{4}$$

$$F_{ik}^{\Psi}(t) := \int_{-\infty}^{+\infty} \Psi(u)\,F_{ik}(t,u-\hat{d}_{ik})\,\mathrm{d}u.$$

The terms in (4) incorporate the delay $\hat{d}_{ik}$, which accounts for the transmission time required for a spike fired by input k to reach the synaptic site $k \to i$.

Likewise, the input time-averaged covariance $\hat{C}_{kl}$ and the input covariance coefficient $\hat{C}_{kl}^{\Psi}$ between inputs k and l are defined by

$$\hat{C}_{kl}(t,u) := \frac{1}{T}\int_{t-T}^{t} \left\langle \hat{S}_k(t')\,\hat{S}_l(t'+u)\right\rangle \mathrm{d}t' - \frac{1}{T}\int_{t-T}^{t} \left\langle \hat{S}_k(t')\right\rangle\left\langle \hat{S}_l(t'+u)\right\rangle \mathrm{d}t' , \tag{5}$$

$$\hat{C}_{kl}^{\Psi}(t) := \int_{-\infty}^{+\infty} \Psi(u)\hat{C}_{kl}(t,u)\,\mathrm{d}u.$$

Note that the input covariance coefficients do not involve delays.

The covariances $\hat{C}_{kl}(t,u)$ by convention do not incorporate the "atomic" discontinuity at $u=0$ as a consequence of the autocorrelation of the stochastic processes $\hat{S}_k$ for $k=l$, namely $\langle \hat{S}_k(t)\rangle\,\delta(u)$, where δ is the Dirac delta function. See Gilson et al. (2009) for details.

In the remainder of this paper, we consider a particular network topology inspired by Kempter et al. (1999), Gütig et al. (2003), where the external inputs are divided into two independent pools, with homogeneous characteristics within each pool (viz. firing rates and spike-time correlations) and homogeneous connectivity from each input pool to each neuron group as illustrated in Fig. 1. The term 'pool' will always refer to the external inputs, the term 'group' to the neurons they input to.

2.3.2 Generation of 'delta-correlated' input spike trains

The external inputs that stimulate the groups of neurons 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 identical. Positive within-pool correlation is generated so that, for any input, a given portion of its spikes occurs at the same time as some other spikes within the pool, whereas the remainder occur at independent times (Gütig et al. 2003; Meffin et al. 2006). By 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, for $k \neq l$, pairwise covariances $\hat{C}_{kl}(t,u)$ defined in (5)

$$\hat{C}_{kl}(t,u) \simeq \hat{c}\,\hat{\nu}_0\,\delta(u), \tag{6}$$

where $0 \leq \hat{c} \leq 10^{-1}$ is the correlation strength (chosen to be small) and δ is the Dirac delta-function. Since inputs from the same pool are only correlated for $u=0$, we refer to them as 'delta-correlated' inputs. See Gilson et al. (2009, Sect. 2.3.7) for more details.

2.3.3 Analysis of the weight dynamics

In a companion paper (Gilson et al. 2009), we derived a dynamical system to analyze the steady states of the neuron firing-rates and of the weights, as well as their stability. This system of equations (Gilson et al. 2009, Eqs. 18a–c) described the evolution of the expectation value of the weights, i.e., the first order of the stochastic process. In the remainder of the present paper, we refer to this leading order as the *drift* of the dynamics, in comparison to *higher orders* of the stochastic process, some of which are related to autocorrelation effects. In the present paper, we focus on symmetry breaking, which relies upon higher-order stochastic dynamics; consequently, this phenomenon cannot be completely captured by the analysis of Gilson et al. (2009, Eqs. 18a–c). It is, however, possible to use our formalism to further analyze symmetry breaking, in a similar manner to the evaluation of the weight variance growth (Kempter et al. 1999; Burkitt et al. 2007).

We want to evaluate the second moment of the weight dynamics, so we need to consider single stochastic trajectories (realizations of the random process). In the remainder of this paper, we assume that all the input delays are identically equal to $\hat{d}$, and likewise all the recurrent delays are equal to d. The learning Eq. (1) can be rewritten as

$$\frac{\mathrm{d}K_{ik}^{\varpi}(t)}{\mathrm{d}t} = w^{\mathrm{in}}\,\hat{S}_k(t-\hat{d}) + w^{\mathrm{out}}S_i(t) + \int W(u)S_i(t)\,\hat{S}_k(t+u-\hat{d})\mathrm{d}u, \tag{7}$$

where ϖ denotes a given stochastic trajectory of the process. Time has been rescaled to remove η. Using the expression (7), we can evaluate the multidimensional matrix coefficient

$$\Upsilon_{i,k,j,l}(t,t') := \left\langle \frac{\mathrm{d}K_{ik}^{\varpi}(t)}{\mathrm{d}t}\,\frac{\mathrm{d}K_{jl}^{\varpi}(t')}{\mathrm{d}t}\right\rangle, \tag{8}$$

which is related to the second moment of the weight dynamics (cf. Appendix A.1). The ensemble average denoted by the

angular brackets in (8) is performed over all stochastic trajectories ϖ; the terms inside the brackets in (8) self-average over the randomness in the case of slow learning (van Hemmen 2001). In comparison, the previous companion paper (Gilson et al. 2009) studied the expectation value of the expression (7) over all the trajectories, which actually leads to the drift of the weight, K_{ik},

$$\dot{K}_{ik}(t) = \left\langle \frac{\mathrm{d}K_{ik}^{\varpi}(t)}{\mathrm{d}t} \right\rangle. \tag{9}$$

The term *mean* (applied to firing rates, weights, etc.) will refer to an average over the neurons, inputs, connections, etc., of the network (topological averaging), whereas *averaged* refers to time averaging, unless specified otherwise. The term *homeostatic equilibrium* will refer to the situation where the mean firing rates and mean weights 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 so that further learning may cause individual weights to change but the qualitative character of the distribution (e.g., bimodal) will remain unchanged.

We recall that it is necessary to introduce bounds on the weights in numerical simulation because of their tendency to diverge because of the competition induced by STDP both upon input and recurrent connections (Kempter et al. 1999; Gilson et al. 2009). The simulation results presented in this paper were run using the neuron and learning parameters listed in Appendix C.

3 Symmetry breaking of the distribution of input weights with fixed recurrent weights

3.1 Previous results concerning the weight drift

The analysis of the drift of the input weight dynamics for the network topology described by Fig. 1 when the recurrent weights are kept fixed has been presented in a companion paper (Gilson et al. 2009). It showed that STDP implements two types of behavior on the input weights, namely a homeostatic equilibrium (under certain conditions on the STDP parameters) and a divergence of the individual weights. When one pool has stronger within-pool correlation, the corresponding input weights will be potentiated at the expense of the weights from the other pool provided the mean input firing rates for the two pools are not too different. We assume that the two input pools have the same size and focus in the remainder of this paper on the special case where the two input pools have the same firing rate $\hat{\nu}_0$ and same correlation strength $\hat{c}_0$ as defined in (6).

The evolution of the matrix K of the input weights can be described through two vectors $K\hat{\mathbf{e}}$ and $K\hat{\mathbf{h}}$, where $\hat{\mathbf{e}}$ is a column vector with all M elements equal to one,

$$\hat{\mathbf{e}} := [1, \ldots, 1]^{\mathbf{T}}, \tag{10}$$

and $\hat{\mathbf{h}}$ is a column vector where the first $M/2$ elements are 1 and the next $M/2$ elements are -1 so that

$$\hat{\mathbf{h}} := \sum_{1 \le k \le M/2} \hat{\mathbf{x}}_k - \sum_{M/2+1 \le k \le M} \hat{\mathbf{x}}_k, \tag{11}$$

where $\hat{\mathbf{x}}_k$ is the kth M-column vector of the canonical basis of $\mathbb{R}^M$, with all M elements equal to zero except the element on the kth row, which is equal to one. The elements for index i of the column vectors $K\hat{\mathbf{e}}$ and $K\hat{\mathbf{h}}$ thus are for each neuron the lumped sum of all input weights, and the difference between the weight sums coming from each of the two pools respectively; see Gilson et al. (2009) for details.

In the case of weak correlation, the following condition ensures homeostatic equilibrium, namely the stability of the vector $K\hat{\mathbf{e}}$ of the mean input weights for all neurons (Gilson et al. 2009):

$$w^{\mathrm{out}} + \widetilde{W}\hat{\nu}_{\mathrm{av}} < 0. \tag{12}$$

When the inhomogeneities of J are neglected, we can approximate $K\hat{\mathbf{e}} \simeq n_{\mathrm{av}}^K K_{\mathrm{av}}^* \mathbf{e}$, where n_{av}^K is the mean number of input weights per neuron and K_{av}^* is the equilibrium value of the mean input weight (Gilson et al. 2009); the N-column vector $\mathbf{e}$ is defined similarly to $\hat{\mathbf{e}}$ in (10). The mean neuron firing rate ν_{av} is then also stable and its equilibrium value can be approximated by

$$\nu_{\mathrm{av}}^* \simeq -\frac{w^{\mathrm{in}}\hat{\nu}_{\mathrm{av}}}{w^{\mathrm{out}} + \widetilde{W}\hat{\nu}_{\mathrm{av}}}. \tag{13}$$

When the network is in homeostatic equilibrium, $K\hat{\mathbf{h}}$ describes the specialization of each neuron to one of the two input pools: if the ith vector element grows positively (negatively), then neuron i becomes sensitive to input pool $\hat{1}$ only (resp. $\hat{2}$). Since the input pools have identical firing rates and correlation strengths, the evolution of $K\hat{\mathbf{h}}$ is given by

$$\dot{K}\hat{\mathbf{h}} = F^W\hat{\mathbf{h}} = M\kappa\ (\mathbb{1}_N - J)^{-1}\ K\hat{\mathbf{h}}, \tag{14}$$

where F^W, as defined in (4) with $\Psi = W$, satisfies the equality $F^W = \left(\mathbb{1}_N - J\right)^{-1} K\,\hat{C}^{W*\epsilon}$ (Gilson et al. 2009, Eq. 18b). The constant κ is given by

$$\kappa = \frac{1}{2}\hat{C}_{kl}^{W*\epsilon}(0) = [W * \epsilon](0)\frac{\hat{c}_0\hat{\nu}_0}{2}, \tag{15}$$

where we have used (5) and (6) with k and l in the same input pool, $\Psi = W * \epsilon$ and $\hat{c} = \hat{c}_0$. The present case corresponds to $\gamma = \gamma' = \kappa' = 0$, in Gilson et al. (2009, Eq. 39b).

The fixed point $K(\infty)\hat{\mathbf{h}} = 0$ is unstable since $\kappa > 0$. This holds because $[W * \epsilon](0) > 0$ for any "Hebbian" choice of

STDP window function W such that $W(u) > 0$ for $u < 0$, cf. Fig. 2, which we assume throughout what follows. For initial conditions in which each neuron is already specialized to a given input, STDP should reinforce the initial specialization. However, for homogeneous initial input weights (in other words the network is unorganized), $K(0)\hat{\mathbf{h}} \simeq 0$ and the state of the dynamical system lies at the unstable fixed point so that $K\hat{\mathbf{h}}$ will grow either positively or negatively, until most input weights become either saturated or quiescent. In this case, the drift for $K\hat{\mathbf{h}}$ is initially zero according to (14) and it is not modified by the convergence towards the homeostatic equilibrium; higher-order terms may then come into play and influence the symmetry breaking. If the neurons are not recurrently connected (in a purely feed-forward network), half of them should specialize to one input pool and the other half to the other pool (Gütig et al. 2003). In the remainder of this section, we examine the dynamics of such symmetry breaking, focusing on the impact of the recurrent connections on the specialization pattern.

3.2 Impact of fixed recurrent connections

To evaluate the second moment of the stochastic evolution of the weights K, we proceed in a similar manner to Kempter et al. (1999), Burkitt et al. (2007) for the analysis of the weight variance, by evaluating $\Upsilon_{i,k,j,k}(t,t')$ as defined in (8) for indices i, j and $k = l$. For each pair of input weights K_{jk} and K_{ik} from the same external input k to two neurons i and j, a connection from neuron j to neuron i induces an extra contribution to the expectation value $\Upsilon_{i,k,j,k}(t,t')$. This contribution relates to the spike-triggering effect for each pair of spikes fired by input k and neuron j, as shown in Appendix A, which results in a *positively* correlated evolution of the weights K_{jk} and K_{ik} for positive recurrent weights. In other words, these two weights tend to vary in the same way, either potentiated or depressed. It follows that the ith and jth elements of $K\hat{\mathbf{h}}$ tend to behave in the same way at the beginning of learning, when the input weights split between the two pools. This means that neurons i and j should have similar input specialization pattern. A connection back from neuron i to neuron j reinforces this phenomenon for each pair of spikes at input k and at neuron i. The contribution is stronger when w^{in} is large, and when w^{out} and $\widetilde{W}$ have the same sign; see (27) in Appendix A.

For a randomly connected network with roughly $n^K_{\text{av}} = n^K/N$ and $n^J_{\text{av}} = n^J/N$ pre-synaptic input and recurrent connections per neuron respectively, the sum over the whole network of the terms related to the spike-triggering effect that impact upon all the weight changes is

$$\frac{n^K n^K_{\text{av}} n^J_{\text{av}}}{MN} J_{\text{av}}\, \nu_{\text{av}} \left(w^{\text{out}} + \widetilde{W}\hat{\nu}_{\text{av}}\right)^2. \tag{16}$$

Note that this expression would be multiplied by η^2 if time had not been rescaled. This is to be compared with the increase of the input variance (Kempter et al. 1999, (30) and (31)) lumped for all the n^K input weights

$$n^K \left\{(w^{\text{in}})^2\hat{\nu}_{\text{av}} + (w^{\text{out}})^2\nu_{\text{av}} + \widetilde{W^2}\hat{\nu}_{\text{av}}\nu_{\text{av}} + 2\widetilde{W}\hat{\nu}_{\text{av}}\nu_{\text{av}}\left[w^{\text{in}} + w^{\text{out}} + \widetilde{W}\left(\hat{\nu}_{\text{av}} + \nu_{\text{av}}\right)\right]\right\}, \tag{17}$$

where $\widetilde{W^2} = \int [W(u)]^2 \mathrm{d}u$. The positive correlation of the evolution of all the input weights K_{ik} coming from all the inputs k in a homogeneous pool is stronger for denser input and recurrent connectivity. The difference in magnitude related to the connectivity between the expressions in (16) and (17) is given by $n^K_{\text{av}} n^J_{\text{av}}/MN$, a fraction that practically equals one for full input and recurrent connectivity.

As a result, neurons from a recurrently connected group tend to specialize together to the same input pool, with probability 50% for each pool, as illustrated in Fig. 4a, b. This is in contrast to a network with no recurrent connections, in which each neuron specializes individually and independently to one of the two pools, as illustrated in Fig. 4c, d.

A sufficiently large correlation strength $\hat{c}_0$ is required to ensure that the diverging behavior of the input weights corresponds to a splitting between the two pools and therefore input selectivity (Gütig et al. 2003). This finding can be qualitatively reproduced in a calculation similar to Appendix A involving K_{ik} and K_{il} for one neuron i and two inputs k and l, see Appendix B.

3.3 Non-homogeneous fixed recurrent connections

Partial connectivity with low density and/or small recurrent weights weaken the group symmetry-breaking effect. This is illustrated in Fig. 5 for a network of two groups of neurons (cf. Fig. 7) with partial connectivity both for the plastic input (30%) and the fixed recurrent connections (30% within-group and 10% between-group). The specialization is weaker than that for full connectivity, cf. Fig. 4a. In addition, neuron group 1 (weight mean represented by the thick dashed trace) weakly specialized to input pool $\hat{1}$, while group 2 (thick dashed-dotted trace) specialized to pool $\hat{2}$. This relates to the fact that the neuron groups have stronger feedback within themselves than between each other and thus may evolve in an independent way. The behavior illustrated in Fig. 5 is interesting in that it contrasts with the "naturally" expected specialization of the network, where the two neuron groups select the same input pool because of positive coupling (recurrent weights) between them; in general, specialization to the same input pool is more likely to occur than to different input pools.

For stronger fixed recurrent weights, the two neuron groups tend to specialize to the same input pool, whereas

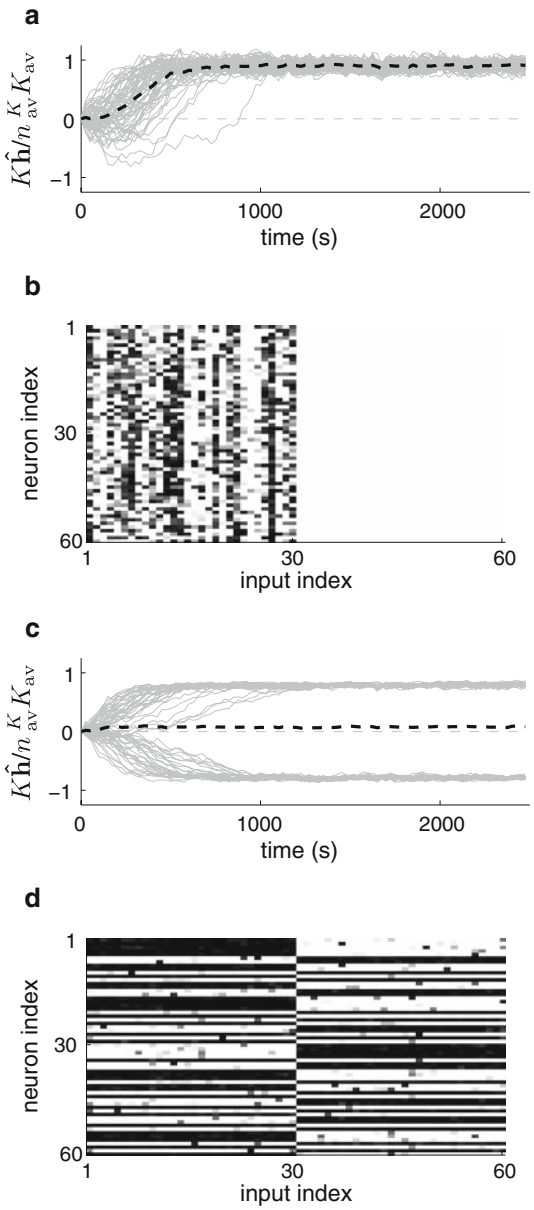


Fig. 4 Symmetry breaking of the input weights for $N = 60$ neurons and two pools of $M/2 = 30$ inputs each. **a**, **b** Full recurrent connectivity versus **c**, **d** no recurrent connections. The plots **a**, **c** show the traces of the elements of $K\hat{\mathbf{h}}$ for each neuron (*grey thin solid lines*) and the mean over all the neurons of $K\hat{\mathbf{h}}$ (*black thick dashed line*). The *grey dashed line* at zero corresponds to no specialization. The matrix graphs (**b**, **d**) show the matrix K (neuron indexed vertically; input horizontally with first pool on the left and second pool on the right) at the end of learning; darker pixels stand for potentiated weights. For full recurrent connectivity, almost all neurons (**b** 60 vs. 0) specialized to the first input pool and the mean of the $K\hat{\mathbf{h}}$ is clearly positive (**a**). In contrast, for no recurrent connections, the neurons specialized almost evenly between the two input pools (**d** 33 vs. 27) and the mean of the $K\hat{\mathbf{h}}$ is almost zero (**c**)

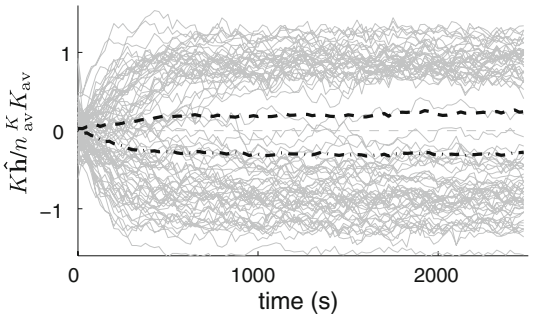


Fig. 5 Symmetry breaking of the input weights for two pools of $M/2 = 100$ correlated inputs each, and a network made of two groups of $N = 200$ neurons each. Input weights are initially random ($\pm 10\%$ of the mean value 0.03), with partial connectivity (30%). The two groups of neurons have stronger connectivity within each (30% with mean $0.015 \pm 10\%$) than between them (10% with mean $0.008 \pm 10\%$). The plot line coding is similar to Fig. 4a, c. The first group of neurons (#1–100, weight mean in *thick dashed line* with only a representative portion plotted) specialized to the first input pool while the second group (#101–200, mean in *thick dashed-dotted line*) to the second input pool

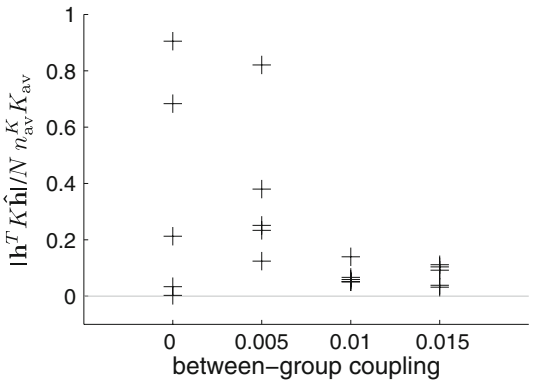


Fig. 6 Illustration of the specialization of two neuron groups as a function of the coupling between them. Each *plotted point* represents the outcome of a simulation for one network configuration. The simulated network was similar to that of Fig. 5, except for the strength of coupling between the two groups, that is, the recurrent weights between them that have a fixed mean weight (x-axis in the plot) with partial connectivity (15%). The y-axis is a measure of the difference in the specialization between the two neuron groups: high values indicate specialization to different input pools (the vector $\mathbf{h}$ is defined in the text)

for smaller recurrent weights they may specialize to different input pools, as illustrated in Fig. 6. The y-axis indicates the degree of specialization to different input pools measured by the scalar value $|\mathbf{h}^T K\hat{\mathbf{h}}|/Nn_{\text{av}}^K K_{\text{av}}$, where $\mathbf{h}$ is the N-column vector defined similar to $\hat{\mathbf{h}}$ with $N/2$ elements equal to 1 and $N/2$ equal to -1. The probability of selecting different input pools decreases when the between-group coupling increases. The recurrent connections only have a higher-order impact on the symmetry breaking of the input weights, which induces a (probabilistic) trend to jointly specialize. For more complex network architectures, the coupling related to the recurrent connections may lead to non-trivial competition between network areas.

3.4 Dependence upon neuron model, initial conditions and learning parameters

The results shown here correspond to $\widetilde{W} < 0$ and short recurrent delays (cf. Appendix C); similar results were obtained with $\widetilde{W} > 0$ and/or larger recurrent delays (e.g., 10 ms). Note that the presence of recurrent connections changes the equilibrium value of the mean input weight K^*_{av}, which has an impact on the weight saturation through the number of potentiated versus quiescent weights.

From a population-statistics point of view, the input firing rates, spike-time correlations and initial weights need not be exactly fine-tuned to ensure that symmetry breaking in one way or the other is equally probable over all neurons. In other words, when $K(\infty)\hat{\mathbf{h}}$ and $K(0)\hat{\mathbf{h}}$ are not strictly zero, the uncertainty due to the initial distribution of the input and recurrent weights, when homogeneous, still leads to an equiprobable specialization to one of the two input pools. This situation occurs in particular for partial input connectivity, where individual neurons may receive more connections from one input pool than the other. We observed that spike-triggering effects were sufficiently strong to play a role even for 30% random input connectivity and 10% spread of the initial input weights around their mean (Fig. 5).

In addition to the input connectivity and the initial distribution of the input weights, the stochastic nature of the Poisson neurons has an influence upon the weight dynamics: the intrinsic randomness of the output favors equiprobability of specialization to each of the two input pools for any input spiking history. Similar results were obtained using a deterministic version of the integrate-and-fire neuron model, which required the addition of an external source of background activity in place of the spontaneous rate ν_0 of the Poisson neuron model. For this purpose, we have used an extra input pool of uncorrelated Poisson spike trains with random and fixed input connectivity.

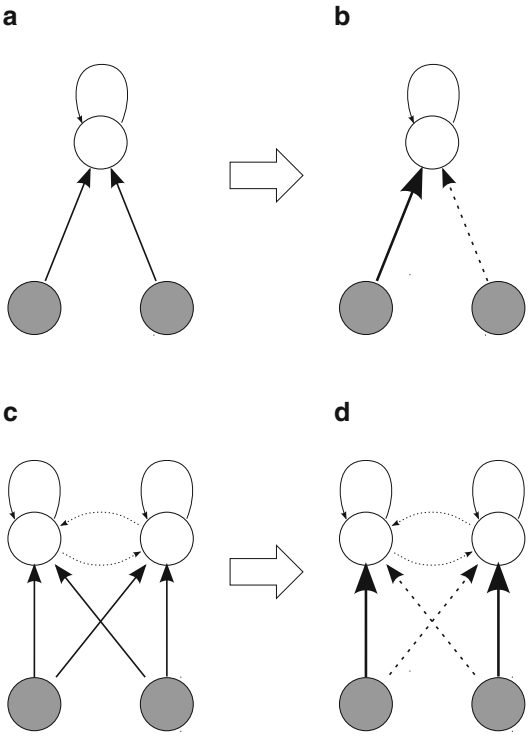


Fig. 7 Schematic representation of the input weight specialization (**a**, **c**) before and (**b**, **d**) after to the learning epoch for two network topologies (*top* vs. *bottom*). The neuron groups (*top circles*) in the network become sensitive to only one of the two correlated input pools (*bottom filled circles*) through the potentiation of some input weights (*very thick arrow*) at the expense of the other input weights that are depressed (*dashed thick arrow*). (**a** ⇒ **b**) Sufficiently strong recurrent connections (*thin arrow*) induce a group symmetry-breaking effect. **c** ⇒ **d** In the case of inhomogeneous recurrent connectivity (strong connections in *thin solid arrows* compared to weak ones in *thin dotted arrows*), this can result in some cases in neuron groups becoming specialized to distinct input pools

4 Discussion

The competition induced by STDP between the input connections from pools of spike trains with balanced firing rates and within-pool correlations results in symmetry breaking for a homogeneous initial distribution of the input weights (Gütig et al. 2003). For two input pools with balanced within-pool correlation but no between-pool correlation, sufficiently correlated inputs (but in the range of small correlations) are necessary in order to obtain an asymptotic bimodal distribution of the input weights that corresponds to a splitting between the two pools. This holds for a broad range of STDP parameters, provided they correspond to a stabilization of the firing rates; see Gilson et al. (2009) for more details on the stability conditions. The influence of non-identical but similar input firing rates and spike-time correlations has yet to be studied in more depth. This robust specialization occurs whatever the detail of the shape of the STDP learning window function W (provided it is "Hebbian", cf. Sec. 2.1), PSP kernel ϵ and homogeneous input delays.

During symmetry breaking, the non-learning connections can play a determining role, for example, in causing neurons with fixed excitatory recurrent connections to specialize in the same way, as illustrated in Fig. 7. This group effect takes place at the beginning of learning; when the neurons become sufficiently specialized, the drift takes over and reinforces the initial symmetry breaking, because of the instability of the fixed point related to the differential equation (14). "Fast" learning (as presented in this paper) implies noise in the weight dynamics but does not prevent the weight structure from emerging; a smaller learning rate η still leads to similar neuronal specialization because of the diverging weight behavior. STDP thus provides a framework for cortical self-organization, in the textbook case where learning takes place on the excitatory connections from some external

inputs whereas the remaining connections are considered fixed. Our results can be linked, for example, to the emergence of ocular-dominance areas in the primary visual cortex, when specializing to one ocular pathway (left or right eye) in the first weeks of life of new-born mammals. The two assumptions of stronger local excitatory connections than those at a longer range and of more correlation for spike trains within each ocular pathway than between the two pathways, are sufficient to qualitatively obtain the emergence of specialized recurrently connected areas sensitive to the inputs from only one eye. Other versions of STDP are expected to generate similar group specialization so long as they generate both a homeostatic equilibrium and a splitting of the weight distribution depending upon the input correlations. Higher-order effects due to the recurrent connections may combine with non-linearities in other STDP models (Gütig et al. 2003; Burkitt et al. 2004; Appleby and Elliott 2006) or specific input structures (e.g., Leibold et al. 2002) to introduce further complexity in the weight dynamics.

These results are intended to shed analytical light on previous work that used numerical simulations to show the emergence of a cortical-like organization due to STDP (Choe and Miikkulainen 1998; Wenisch et al. 2005). The present study has made minimal assumptions about the network topology and the input firing rate and correlation structures in order to explore the input specialization behavior in a recurrent network. More complex weight dynamics are likely to occur in a more detailed network topology inspired by the cortex, such as short-range excitatory and medium-range inhibitory connections in visual cortex (von der Malsburg 1973). The results presented here have some bearing on previous work on ocular dominance (von der Malsburg 1973; Swindale 1996; Elliott and Shadbolt 1999; Goodhill 2007); most of the models proposed or cited by von der Malsburg (1973) and Swindale (1996) interestingly combine the same dynamical ingredients as those shown here to be generated by STDP, namely a combination of stabilization and divergence. However, a more complete understanding of this phenomenon requires the analysis of the effect of STDP on the recurrent weights, which is the subject of subsequent papers in this series, and consequently the relationship of our work to these previous models will be discussed there.

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: Symmetry breaking within K for different neurons

This appendix details some calculations related to the study of the impact of recurrent connections on the symmetry breaking performed by STDP on input connections through the second stochastic moment of their weight dynamics.

A.1 Second moment of the stochastic evolution of K

Here, we consider $\Upsilon_{i,k,j,k}(t,t')$ defined in (8) for indices i, j and $k=l$. This coefficient relates to the relative evolution of the weights K_{ik} and K_{jk}: the sign of $\Upsilon_{i,k,j,k}(t,t')$ indicates whether K_{ik} and K_{jk} tend to evolve in the same direction or not (potentiation or depression). We only consider the simplified case of identical input firing rates $\hat{\nu}_k=\hat{\nu}_0$ and spike-time correlation ($\hat{c}_0$). From (7) we have

$$\begin{aligned}
&\frac{\mathrm{d}K_{ik}^{\varpi}(t)}{\mathrm{d}t}\,\frac{\mathrm{d}K_{jk}^{\varpi}(t')}{\mathrm{d}t}\\
&=\Big[(w^{\mathrm{in}})^2\,\hat{S}_k(t-\hat{d})\,\hat{S}_k(t'-\hat{d})\\
&\quad+(w^{\mathrm{out}})^2 S_i(t)S_j(t')\\
&\quad+w^{\mathrm{in}}w^{\mathrm{out}}\,\hat{S}_k(t-\hat{d})S_j(t')\\
&\quad+w^{\mathrm{in}}w^{\mathrm{out}}S_i(t)\,\hat{S}_k(t'-\hat{d})\\
&\quad+w^{\mathrm{in}}\int W(u)S_i(t)\,\hat{S}_k(t+u-\hat{d})\,\hat{S}_k(t'-\hat{d})\,\mathrm{d}u\\
&\quad+w^{\mathrm{in}}\int W(u)\,\hat{S}_k(t-\hat{d})\,\hat{S}_k(t'+u'-\hat{d})S_j(t')\,\mathrm{d}u'\\
&\quad+w^{\mathrm{out}}\int W(u)S_i(t)\,\hat{S}_k(t+u-\hat{d})S_j(t')\,\mathrm{d}u\\
&\quad+w^{\mathrm{out}}\int W(u')S_i(t)\,\hat{S}_k(t'+u'-\hat{d})S_j(t')\,\mathrm{d}u'\\
&\quad+\int\int W(u)W(u')S_i(t)\,\hat{S}_k(t+u-\hat{d})\\
&\qquad\hat{S}_k(t'+u'-\hat{d})S_j(t')\,\mathrm{d}u\,\mathrm{d}u'\Big].
\end{aligned}\tag{18}$$

The leading-order drift obtained when taking the expectation value of the sum of these nine terms is $\langle\frac{\mathrm{d}K_{ik}^{\varpi}(t)}{\mathrm{d}t}\rangle\langle\frac{\mathrm{d}K_{jk}^{\varpi}(t')}{\mathrm{d}t}\rangle$, while neglecting the autocorrelation effects and some probabilistic interdependence of the spike trains $\hat{S}_k$, S_i and S_j. This leading-order term is almost zero when the input mean weights are stable around their equilibrium value for all neurons, which follows from

$$\left\langle\frac{\mathrm{d}K_{ik}^{\varpi}(t)}{\mathrm{d}t}\right\rangle=\dot{K}_{ik}(t)=0\tag{19}$$

for all indices k and i. Consequently, higher orders involving autocorrelation effects of inputs and neurons may have an impact on the evolution of K_{ik} and K_{jk} when they have reached the homeostatic equilibrium. We identify different

kinds of contributions: the first-order autocorrelation terms that are independent of the network connectivity, which will not be discussed here, see Kempter et al. (1999) for details; spike-triggering effects (second-order in terms of autocorrelation) that depend on the connectivity; and further orders that will not be considered, i.e., terms that arise from recurrent synaptic paths of length two or more.

A.2 Recurrent connections and spike-triggering effect

We focus on the spike-triggering effects related to recurrent connections when taking the ensemble average of (18) for two given neurons $i \neq j$ and a given input k. First, we consider a single recurrent connection $j \to i$ with weight $J_{ij} > 0$, ignoring all other recurrent connections. Spike-triggering effects due to the autocorrelation of input k arise in the second, seventh, eighth and ninth terms of the rhs of (18).

In the second term of (18), taking the ensemble average of $S_i(t)S_j(t')$ induces an additional term $J_{ij}\,\epsilon(t-t'-d)\,\langle S_j(t')\rangle$ due to the autocorrelation of neuron j, which stems form the relationship

$$\big\langle S_i(t)S_j(t')\big\rangle = \big\langle \rho_i(t)S_j(t')\big\rangle, \tag{20}$$

where $\rho_i(t)$ involves $J_{ij}\,[\epsilon * S_j](t-d)$. This leads to the following contribution to $\Upsilon_{i,k,j,k}(t,t')$ induced by J_{ij}

$$(w^{\text{out}})^2\,J_{ij}\,\epsilon(t-t'-d)\,\big\langle S_j(t')\big\rangle. \tag{21}$$

For each spike fired by neuron j at time t', there is a non-zero contribution given by (21) for all times $t \geq t'+d$ such that $\epsilon(t-t'-d) \neq 0$.

The seventh term of (18) gives

$$\begin{aligned} w^{\text{out}}J_{ij}\int W(u)\,\epsilon(t-t'-d) &\\ \times\big\langle S_j(t')\,\hat{S}_k(t+u-\hat{d})\big\rangle\mathrm{d}u &\\ \simeq w^{\text{out}}\,J_{ij}\,\epsilon(t-t'-d)\big\langle S_j(t')\big\rangle &\\ \times\int W(u)\big\langle\hat{S}_k(t+u-\hat{d})\big\rangle\,\mathrm{d}u, & \end{aligned} \tag{22}$$

where $S_j(t')$ and $\hat{S}_k(t+u-\hat{d})$ are taken to be independent, which is equivalent to considering only the leading-order in terms of the autocorrelation of neuron j. Likewise, the eighth term of (18) gives

$$\begin{aligned} w^{\text{out}}J_{ij}\int W(u')\,\epsilon(t-t'-d) &\\ \times\big\langle S_j(t')\,\hat{S}_k(t'+u'-\hat{d})\big\rangle\mathrm{d}u' &\\ \simeq w^{\text{out}}\,J_{ij}\,\epsilon(t-t'-d)\big\langle S_j(t')\big\rangle &\\ \times\int W(u')\big\langle\hat{S}_k(t'+u'-\hat{d})\big\rangle\,\mathrm{d}u'. & \end{aligned} \tag{23}$$

Finally, the ninth term in (18) gives

$$\begin{aligned} J_{ij}\int\int W(u)W(u')\,\epsilon(t-t'-d) &\\ \times\big\langle S_j(t')\,\hat{S}_k(t+u-\hat{d})\,\hat{S}_k(t'+u'-\hat{d})\big\rangle\mathrm{d}u\,\mathrm{d}u' &\\ \simeq J_{ij}\,\epsilon(t-t'-d)\big\langle S_j(t')\big\rangle\int\int W(u)W(u') &\\ \times\big\langle\hat{S}_k(t+u-\hat{d})\big\rangle\big\langle\hat{S}_k(t'+u'-\hat{d})\big\rangle\mathrm{d}u\,\mathrm{d}u'. & \end{aligned} \tag{24}$$

Summing the four terms in (21), (22), (23) and (24), we obtain the total contribution to $\Upsilon_{i,k,j,k}(t,t')$ induced by the single weight J_{ij}

$$\begin{aligned} J_{ij}\,\epsilon(t-t'-d)\,\big\langle S_j(t')\big\rangle &\\ \left[w^{\text{out}}+\int W(u)\big\langle\hat{S}_k(t+u-\hat{d})\big\rangle\mathrm{d}u\right]^2, & \end{aligned} \tag{25}$$

which is positive since the instantaneous firing rate $\langle S_j(t')\rangle$, ϵ and the recurrent weight J_{ij} are positive.

This additional contribution implies that $\Upsilon_{i,k,j,k}(t,t')$ is more positive in the presence of the recurrent connection $j \to i$. This induces a more positively correlated evolution of K_{ik} and K_{jk}, which means that they tend to evolve in the same direction: either they both increase or decrease.

Since weights vary slowly compared to the time scale of the neuronal activation mechanisms related to ϵ, d and $\hat{d}$, we integrated (25) over time to obtain the time-averaged effect. In the case of homogeneous inputs, this leads to

$$J_{ij}\,\nu_{\text{av}}\,\left(w^{\text{out}}+\widetilde{W}\hat{\nu}_{\text{av}}\right)^2, \tag{26}$$

since the integral of ϵ is normalized to one. Using the approximation of the equilibrium value of ν_{av} in (13), the expression (26) becomes

$$-\,J_{ij}\hat{\nu}_{\text{av}}\,w^{\text{in}}\left(w^{\text{out}}+\widetilde{W}\hat{\nu}_{\text{av}}\right) \;>\; 0. \tag{27}$$

Recall that $(w^{\text{out}}+\widetilde{W}\hat{\nu}_{\text{av}}) < 0$ is required for homeostatic stability.

Note that, because J has no self-connections, the diagonal terms J_{ii} do not contribute to the variance of the input weights, which is related to $\Upsilon_{i,k,i,k}(t,t')$.

A.3 Arbitrary homogeneous connectivity

We now consider the situation when each input and recurrent connection have the probability n^K/NM and $n^J/N(N-1)$ resp. of existing (recall that n^K and n^J are the number of input and recurrent connections resp.). We average (26) over the whole network for all triplets (k,i,j) to obtain the time-averaged contribution to $\sum_{k\to i}\sum_{k\to j}\Upsilon_{i,k,j,k}(t,t')$ due to

all recurrent connections

$$MN(N-1)\left(\frac{n^K}{NM}\right)^2 \frac{n^J}{N(N-1)} J_{\text{av}}\,\nu_{\text{av}}\left(w^{\text{out}}+\widetilde{W}\hat{\nu}_{\text{av}}\right)^2 \simeq \frac{n^K n^K_{\text{av}} n^J_{\text{av}}}{MN} J_{\text{av}}\,\nu_{\text{av}}\left(w^{\text{out}}+\widetilde{W}\hat{\nu}_{\text{av}}\right)^2, \quad (28)$$

where $n^K_{\text{av}} = n^K/N$ and $n^J_{\text{av}} = n^J/N$ are the mean numbers per neuron of pre-synaptic external input connections and pre-synaptic recurrent connections, respectively.

Appendix B: Symmetry breaking through competition between input weights

We consider the equivalent of (18) for $\frac{\mathrm{d}K^{\varpi}_{ik}(t)}{\mathrm{d}t}\frac{\mathrm{d}K^{\varpi}_{il}(t')}{\mathrm{d}t}$ with two external inputs $k \neq l$ and recurrently connected neuron i. When k and l come from the same correlated input pool with correlation strength $\hat{c}$, an additional contribution for $t = t'$ to $\Upsilon_{i,k,i,l}(t,t')$ defined in (8) arises from the autocorrelation of inputs $k \neq l$, namely

$$\hat{c}\hat{\nu}_{\text{av}}\left(w^{\text{in}}+\widetilde{W}\nu_{\text{av}}\right)^2. \quad (29)$$

This contribution multiplied by the number of external input connections n^K is to be compared with the evaluation of the increase of the external input weight variance in (17), which "generates" the symmetry breaking. In order for the symmetry breaking to occur between different external input pools and not within the pools, it is necessary that the correlation strength $\hat{c}$ be sufficiently large that the expression in (29) is comparable with that in (17), as shown by Gütig et al. (2003).

Table 1 List of simulation parameters

Time step	10^{-4} s
Simulation duration	2.5×10^3 s
Input Poisson spike trains	
Firing rates	$\hat{\nu}_{\text{av}} = 30$ Hz
Correlation strength	$\hat{c}_{\text{av}} = 0 - 0.1$
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	$d = 0.4$ ms
Spread of recurrent delays	± 0.2 ms
Mean of input delays	$\hat{d} = 7$ ms
Spread of input delays	± 1 ms
STDP	
Learning parameter	$\eta = 10^{-5}$
Pre-synaptic rate-based coefficient	$w^{\text{in}} = 4$
Post-synaptic rate-based coefficient	$w^{\text{out}} = -0.5$
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 C: 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

$$W(u) = \begin{cases} c_P \exp(u/\tau_P) & \text{for } u < 0 \\ -c_D \exp(-u/\tau_D) & \text{for } u > 0. \end{cases} \quad (30)$$

The PSP kernel ϵ is defined by

$$\epsilon(t) = \begin{cases} \frac{\exp(t/\tau_B)-\exp(t/\tau_A)}{\tau_B-\tau_A} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0. \end{cases} \quad (31)$$

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 two orders of magnitude below (10^{-2}) 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

Appleby PA, Elliott T (2006) Stable competitive dynamics emerge from multispike interactions in a stochastic model of spike-timing-dependent plasticity. Neural Comput 18(10):2414–2464

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

Choe Y, Miikkulainen R (1998) Self-organization and segmentation in a laterally connected orientation map of spiking neurons. Neurocomputing 21(1-3):139–157

Elliott T, Shadbolt NR (1999) A neurotrophic model of the development of the retinogeniculocortical pathway induced by spontaneous retinal waves. J Neurosci 19(18):7951–7970

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 (2009) Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. I. Input selectivity—strengthening correlated input pathways. doi:10.1007/s00422-009-0319-4

Goodhill GJ (2007) Contributions of theoretical modeling to the understanding of neural map development. Neuron 56(2):301–311

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

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

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

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in cats visual cortex. J Physiol (Lond) 160(1):106

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

Leibold C, Kempter R, van Hemmen JL (2002) How spiking neurons give rise to a temporal-feature map: from synaptic plasticity to axonal selection. Phys Rev E 65(5):051915

von der Malsburg C (1973) Self-organization of orientation sensitive cells in striate cortex. Kybernetik 14(2):85–100

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

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

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

Pfister JP, Gerstner W (2006) Triplets of spikes in a model of spike timing-dependent plasticity Source. J Neurosci 26(38):9673–9682

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

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

Swindale NV (2006) The development of topography in the visual cortex: A review of models. Network Comput Neural 7(2):161–247

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