博士論文

Synaptic Dynamics and Learning:

How biological mechanisms of plasticity provide efficient learning schemes for neural computation

(シナプスのダイナミクスと学習：
いかにして可塑性の生物学的メカニズムは、神経情報処理を
可能とする効率的な学習則を実現するか。)

平谷 直輝

Abstract

Learning is the key feature of mammalian brain. It is widely believed that change in synaptic connection between neurons is the essential substrate for learning. Experimental studies in last two decades further revealed rules and constraints on synaptic plasticity. Correspondingly, many theoretical studies were conducted on synaptic plasticity. However, many of these studies were only concerning on its dynamic properties, and did not provide functional implication. In addition, most studies were limited to single spine or single cell levels due to numerical and analytical complexity arise from interaction of synaptic dynamics and neural dynamics, and little is known about properties of synaptic dynamics in neural circuits. On the other hand, many attempts were also made from machine learning perspective, but in these studies, biological constraints were often taken for granted, in particular, little is explained on how functional neural circuits are self-organized in the absence of absolute teaching signals, or explicit objective functions.

This thesis is an interim report of an attempt to bridge the gap between two kind of studies. The thesis consists of four independent works related to the question above. Each work capture one or two aspects of complicated synaptic plasticity, and performs both analysis based on dynamic systems theory, and functional investigation based on information theory or machine learning study. Four works were arranged in the order of spatial scale of phenomenon considered in the study.

The first work is focused on a single synapse and a dendritic hotspot consist of several number of synapses. In the work, I studied on how nearby synapses on a dendritic tree interact with each other. Especially, I investigated the functional role of recently discovered heterosynaptic spike-timine-dependent plasticity (h-STDP).

In the second work, the main focus is still a single synapse, but here I studied long-term dynamics of a dendritic spine. In the long timescale, elimination and creation of spines is expected to play crucial role in addition to synaptic weight plasticity, because such spine turnover is known to be active even in the cortex of adult mammalian. Thus, in this study, I investigate how elimination and creation of spines helps learning and computation in collaboration with synaptic weight plasticity.

On the third work, I shifted my focus to neural circuits. Although, the actual neural circuits in the brain are highly complicated, there are number of basic circuit motifs. Feedback-type circuit is one of such motifs, and indeed observed in many neural systems. In the study, I investigated how feedback-type neural circuits can perform learning with spike-timing-dependent plasticity (STDP). In particular, I revealed how propagation of spike correlation in a feedback-type circuit drives STDP learning.

The last work is about recurrent neural circuits. Although there are many theoretical studies on synaptic learning in recurrent circuits, most of them are about input-driven learning, and little is known about modulation of learned memory traces by spontaneous activity. By considering simple neural models, I studied how cell assemblies are selectively retained or merged by STDP through spontaneous activity. Especially, my study revealed possible functional roles of short-term plasticity for

the modulation of memory traces.

Detailed results of those studies are explained in each chapter, but key findings of the thesis are

- Calcium-based synaptic plasticity model can replicate various results of heterosynaptic spike-timing-dependent plasticity by adding current-based heterosynaptic terms (Chapter 2).
- By considering h-STDP, critical period plasticity of binocular matching is explained by GABA-maturation (Chapter 2).
- To perform inference by a feedforward neural network with limited connections, under certain conditions, it is better to encode information by synaptic connections not by synaptic weights, because signal variability is reduced in the former case (Chapter 3).
- Under the presence of random noise, spike correlation should be as precise as possible to perform correlation-based learning. However, in the presence of cross-talk noise, non-precise spike-correlation is beneficial for learning (Chapter 4).
- In feedback-type neural circuits, STDP-based learning mimics Bayesian independent component analysis, because membrane potential dynamics approximates likelihood functions of hidden sources (Chapter 4).
- Alternation of cell assemblies, which is observed in the hippocampus of rodents, possibly supports cell assembly retention by inducing activity-dependent long-term potentiation(LTP). In addition, presynaptic release probability should be non-zero small value in order to achieve such alternation-based retention (Chapter 5).
- Selective retention and merging of memory traces are possibly supported by dynamic modulation of cell assemblies during awake-quiet or sleep states (Chapter 5).

As all these studies are purely theoretical, their impacts on the understanding of the brain are limited. Still, my study provides several novel interpretations for previously observed phenomena, as well as several experimentally testable predictions. Therefore, I believe these works extend our knowledge on brain science in a tiny but significant portion.

Acknowledgement

I thank to Dr. Tomoki Fukai for supervising the PhD study, Dr. Matthieu Gilson for mentoring projects on synaptic plasticity, Dr. Jun-nosuke Teramae for overseeing projects on neural dynamics, and Dr. Masato Okada for general mentoring. I would also like to thank to Dr. Kaoru Inokuchi for helpful advises on the memory modulation study, and Dr. Haruo Kasai for insightful comments on the model of heterosynaptic plasticity. In addition, I would like to acknowledge Drs. Kensuke Arai, Toshitake Asabuki, Tom Close, Peter Dayan, Anthony Decostanzo, Tatsuya Haga, Takashi Handa, David Higgins, Grace Huckins, Jun Igarashi, Ryo Karakida, Florence Kleberg, Tomoki Kurikawa, Máté Lengyel, Yoshinori Nakanishi-Ohno, Amanda Reilly, Thomas Sharp, Takuma Tanaka, Sina Tootoonian, Jochen Triesch, Yasuhiro Tsubo, and Xiao-Jing Wang for helpful discussion.

Abbreviations

AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic Acid
EPSP	Excitatory Post Synaptic Potential
GABA	γ-Aminobutyric Acid
h-STDP	Heterosynaptic Spike-Timing-Dependent Plasticity
i-STDP	Inhibitory Spike-Timing-Dependent Plasticity
ICA	Independent Component Analysis
IPSP	Inhibitory Post Synaptic Potential
LIF	Leaky Integrate-and-Fire (model)
LTD	Long-Term Depresstion
LTP	Long-Term Potentiation
NMDA	N-Methyl-D-aspartic Acid
PCA	Principal Component Analysis
STD	Short-Term Depression
STDP	Spike-Timing-Dependent Plasticity
VDCC	Voltage-Dependent Calcium Channel

Copyright Statement

Hiratani and Fukai, 2014 Hiratani N, Fukai T. Interplay between short- and long-term plasticity in cell-assembly formation. PloS One. 2014;9: e101535. doi:10.1371/journal.pone.0101535

Hiratani and Fukai, 2015a Hiratani, N., Fukai, T., 2015. Mixed Signal Learning by Spike Correlation Propagation in Feedback Inhibitory Circuits. PLoS Comput Biol 11, e1004227. doi:10.1371/journal.pcbi.1004227

Hiratani and Fukai, 2015b Hiratani N, Fukai T. Hebbian Wiring Plasticity Generates Efficient Network Structures for Robust Inference with Synaptic Weight Plasticity. bioRxiv. 2015; 024406. doi:10.1101/024406

Contents

Chapter 1

Background

(Reply of the senses to Intellect): 'Miserable Mind, you get your evidence from us, and do you try to overthrow us? The overthrow will be your downfall'.

— Dêmocritus of Abdêra, *Ancilla to the Pre-Socratic Philosophers, 68:125*

Synaptic Dynamics and Learning

The central theme of this thesis is synaptic dynamics and learning, but relationship between synaptic plasticity and learning is not trivial. Synapses change their structure in response to synaptic inputs, neuromodulators, neuronal activity, or even spontaneously. In general, we can understand those changes as learning if such changes help neural circuits to perform better computation or information processing, and enhance the adaptability and survivability of the animal. For example, synaptic degeneration in Alzheimer's deceases does not satisfy this criteria because the degeneration typically degrades information storage capacity of the circuits [234]. On the other hand, enhanced spine elimination during developmental period is learning because an animal usually acquire better sensory information processing and motor control skills through developmental change in neural circuits [105].

The brain can perform object recognition, decision making, sensory-motor control, and many other functions that require appropriate computational procedures, and arguably these different computations require different types of learning. For a well-defined computation, in principle, we can predict how neural circuits should be organized, and ideally we can evaluate optimality of neural activity or synaptic weights organization. However, except for simple systems, such prediction or evaluation is mostly impractical because of various difficulty arise from complexity and inscrutability of the brain. Still, by assuming that the brain is operated in a near-optimal regime, we can relate synaptic dynamics and learning. This means that if some characteristic synaptic dynamics is observed in various areas of many species in wide range of conditions, the dynamics should be related to learning. Following this

principle, I first briefly list up typical behavior of synapses, then from the next chapter, I investigate their functions.

Synaptic plasticity mechanisms

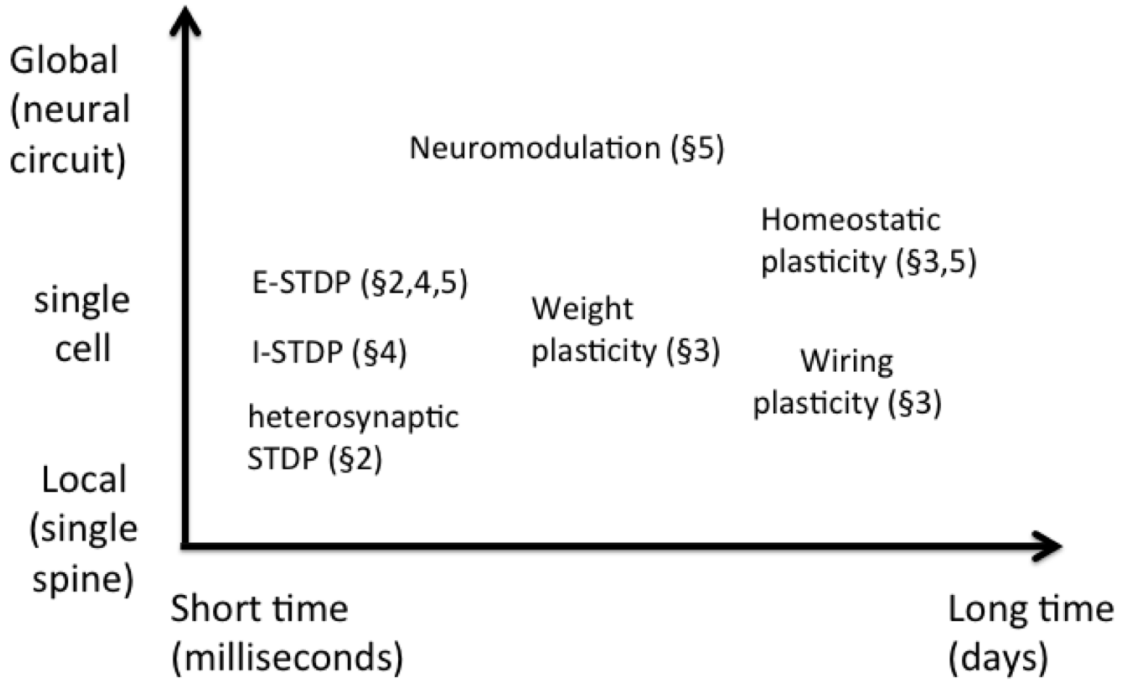


Figure 1.1. Plasticities in various temporal and spatial scales. Numbers written besides the mechanism are the section in the thesis at which the plasticity mechanism was considered.

Plasticity is the fundamental mechanism of learning in the brain. Thus, expectedly there are various different mechanisms that cause changes in neural circuits. Figure 1.1 summarizes the synaptic plasticity mechanisms I employed in this thesis. Here, X-axis of the figure represents the main timescale of plasticity mechanism and the y-axis represents the main spatial scale. Below, I explain the main plasticity mechanisms.

Spike-Timing Dependent Plasticity

Change in EPSP size does not only depend on firing rates of presynaptic and postsynaptic neurons, but also influenced by relative timing of spikes at presynaptic and postsynaptic neurons. This spike-timing-dependent form of synaptic plasticity is called STDP. Although, synaptic weight change by STDP depends on the membrane potential of the postsynaptic neuron [43], firing rates of the postsynaptic and presynaptic neurons [209], number of AMPA receptor on the postsynaptic spine [20] [225], relative timings of successive spikes [187] [84], neuromodulation [253], relative position of the synapse on the dendrite [138] [210], and the timing of inputs at neighboring synapses [94] [176], still in the simplest

form, synaptic weight change can be approximated as below,

$$\Delta w(\Delta t) = \begin{cases} A_p e^{-\Delta t} \text{ if } \Delta t > 0 \\ -A_d e^{\Delta t} \text{ if } \Delta t < 0 \end{cases} \tag{1.1}$$

where Δt is the relative spike timing between the postsynaptic spike and the presynaptic spike (i.e. if the post-neuron fires before the pre-neuron fires, Δt becomes positive in this case). Note that this simple form of formalization does not reproduce any of dependence listed above, including the firing rates dependence. Despite lack of firing rate dependence, this form of STDP is often called as Hebbian [213], because the rule is sensitive to causal relationship between presynaptic and postsynaptic activities.

Inhibitory spike-timing-dependent plasticity

Although, most studies on STDP are focused on excitatory synaptic connections from excitatory neurons to other excitatory neurons, partly because excitatory neurons are believed to play primary role in cortical information processing, recent experimental results suggest that other types of synaptic connections also show STDP-type plasticity [231]. For instance, Woodin and colleagues revealed that GABAergic synapses on excitatory neurons show coincidence-detection type STDP, but interestingly, their results suggest that synaptic plasticity is not realized by the potentiation of synapse itself, but by change in cotransporter activity [243]. It is also known that glutaminergic synapses on inhibitory neurons show STDP [146] [66], in particular, STDP at excitatory-to-inhibitory connections is suggested to play a critical role in critical period plasticity [248]. Theoretical studies suggest that inhibitory STDP supports retention of the detailed excitatory-to-inhibitory balance [230]. Especially, recent studies indicate that inhibitory STDP play a critical role in stabilization of recurrent circuits [142] [252] that is difficult to achieve when only excitatory-to-excitatory connections are plastic [166].

Spinogenesis, Wiring plasticity

Most synaptic connections are projected to the dendritic tree of the postsynaptic neuron. In particular, a majority of excitatory connections are projected to dendritic spines protruded from the dendritic tree. These spines are known to change their sizes in response to long-term potentiation (LTP) or depression (LTD), and also eliminated or created depending on neural activity or even spontaneously [105] [116]. Creation and elimination of spine, which is often called spinogenesis, is most active in the developmental period, but even in the adult cortex, spinogenesis is frequently induced [105]. In case of rodents, previous studies suggest that the spine turn over rate is up to 15% per day in the sensory cortex [104], and even 5 % per day in the motor cortex [255]. In addition, recent experimental results suggest that spine turnover is tightly correlated with task acquisition [245] [244]. For instance, Xu and colleagues revealed that in motor learning task, performance after learning is positively correlated with the amount of spines

created during the training period and survived until the test [245]. Therefore, spinogenesis is possibly important in synaptic learning, but their functions are not yet well characterized.

Heterosynaptic plasticity

In cortical circuits, synapses are projected to dendritic tree, and it has long been known that synapses on the dendritic tree interact with each other in their plastic changes [85] [174], yet these heterosynaptic plasticity mechanisms were known to work on timescale of minutes to hours. For instance, in hippocampal synapses, by inducing strong LTP at one excitatory synapse, at nearby excitatory spines, thresholds for LTP decreases several minutes after the original LTP due to spreading of Ras activity [93]. Recently results further suggest that heterosynaptic plasticity is also caused by spike correlation between nearby synapses in milliseconds timescale [182] [94]. For example, in Schaffer-collateral synapses, by inducing GABA uncaging right before pre and postsynaptic stimulation, time window of STDP at the stimulated excitatory synapse changes due to heterosynaptic effect from the inhibitory input [94]. In the next chapter, I consider the function of these spike-timing-dependent heterosynaptic plasticity.

Homeostatic plasticity

In addition to activity dependent plasticity mechanisms, synapses are also modified through homeostatic mechanism [224]. For instance, when the activity of neurons stays high for a certain period, synapses are down regulated to reduce the postsynaptic firing rate. These homeostatic changes typically occur at the timescale of days. Note that, many activity dependent plasticity mechanisms have intrinsic homeostatic effects. For example, synaptic weight dependence of STDP prevents divergence of synaptic weights (see equation (4.9) for details).

Neuromodulation

In addition to local plasticity mechanisms listed above, there exits global regulation mechanisms through neuro-modulators. In particular, cortical synapses are known to change their STDP rules in response to neuromodulation [207] [34]. For instance, under the presence of dopamine, the STDP window of glutamate synapses turns nearly symmetric in rat hippocampus [253]. We discuss their functional merits in section 5.

Chapter 2

GABA-driven Synaptic Organization by Heterosynaptic Spike-Timing-Dependent Plasticity

Balance between excitatory and inhibitory inputs is a key feature of cortical dynamics. Such balance is arguably preserved in dendritic branches, yet its underlying mechanism and functional roles are still unknown. Here, by introducing spike-timing dependent heterosynaptic plasticity, I show that the detailed balance on dendritic branch is robustly achieved, as a result of GABA-driven local synaptic clustering. A neuron with the local balance can optimally perform abstract change detection task, due to functional specialization at each branch. I further demonstrate that heterosynaptic plasticity explains critical period plasticity of binocular matching. My study provides a theoretical basis for functional investigation of heterosynaptic plasticity.

Introduction

Activity dependent synaptic plasticity is essential for learning. Especially, spike time difference between presynaptic and postsynaptic neurons is a critical factor for synaptic learning [20] [32]. Recent experimental results further revealed that the relative spike timings among neighboring synapses on a dendritic branch have significant influence on changes in synaptic efficiency of these synapses [94] [182] [176]. Especially, the timing of GABAergic input exerts a great impact on synaptic plasticity at nearby Glutamatergic synapses. Similar phenomenon were also observed in biophysical simulations [46] [14]. This heterosynaptic form of spike-timing-dependent plasticity (h-STDP) is potentially important for synaptic organization on dendritic tree, and resultant dendritic computation [163] [25] [117]. However, the functional role of h-STDP remains elusive, partly due to lack of simple analytical model.

In the understanding of homosynaptic STDP, simple mathematical formulation of plasticity has been

playing important roles [73] [213] [230]. Following these studies, I constructed a mathematical model of h-STDP based on calcium-based synaptic plasticity models [208] [86], and then considered potential functional merits of the plasticity. The model reproduces the several effects of h-STDP observed in the hippocampal CA1 area and the striatum of rodents, and provides analytical insights for the underlying mechanism. The model further indicates that h-STDP causes the detailed balance between excitatory and inhibitory inputs on a dendritic branch, because long-term depression (LTD) at excitatory synapses is shunted by correlated inhibitory inputs on neighboring dendrite. This result suggests that not only the number and the total current of excitatory/inhibitory synapses are balanced at a branch [143] [242], but temporal input structure is also balanced as observed in the soma [58] [67]. Moreover, by considering supervised learning on a two-layered single cell model, I show that such detailed balance is beneficial for change detection. The model also reconciles with critical period plasticity of binocular matching observed in V1 of mice [235] [236], and provides a candidate explanation on how GABA-maturation modulates the orientation selectivity of excitatory neurons.

Results

Calcium-based synaptic plasticity model with current-based heterosynaptic interaction explains h-STDP.

I constructed a model of a dendritic spine as shown in Fig. 2.1A (see *Model* A_1 in Methods for details). In the model, the membrane potential of the spine is modulated by influx/outflux from AMPA/NMDA receptors (x^A and $g_N(u)x^N$ in Fig. 2.1A), back-propagation (x^{BP}), and heterosynaptic currents from nearby excitatory/inhibitory synapses (x^E and x^I). Calcium concentration in the spine is controlled through NMDA receptors and voltage-dependent calcium channels (VDCC) [99]. Because, both NMDA and VDCC are voltage-dependent [147], the calcium level in the spine is indirectly controlled by pre, post, and heterosynaptic activities (Fig. 2.1B top and middle panels). For synaptic plasticity, I used calcium-based plasticity model, in which LTP/LTD are initiated if the Ca^{2+} level is above LTP/LTD thresholds (orange and cyan lines in Fig. 2.1B middle). This plasticity model is known to well capture homosynaptic STDP [208] [86]. I introduced an intermediate variable $y(t)$ to capture non-graded nature of synaptic weight change [185]. Thus, change in Ca^{2+} level is first embodied in the intermediate $y(t)$ (Fig. 2.1B bottom), then reflected to the synaptic weight $w(t)$ upon accumulation.

I first consider the effect of inhibitory input to synaptic plasticity at nearby excitatory spines. A recent experimental result revealed that, in medium spiny neuron, a synaptic connection from a cortical excitatory neuron typically shows anti-Hebbian type STDP under pairwise stimulation protocol, but if GABA-A receptor is blocked, STDP time window flips to Hebbian [182] (points in Fig. 2.2A). The proposed model can explain this phenomenon in the following way. Let us first consider the case when the presynaptic excitatory input arrives before the postsynaptic spike. If the GABAergic

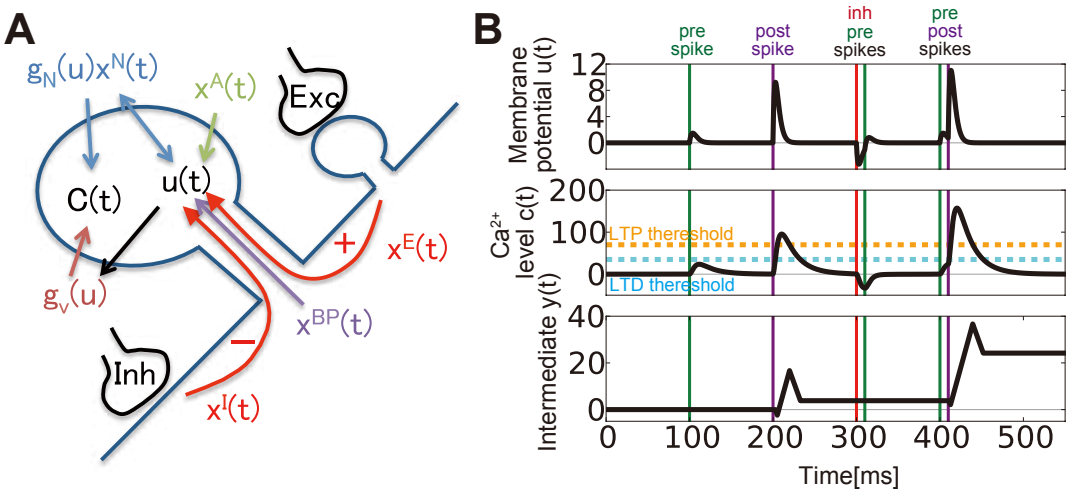


Figure 2.1. Schematic figure of the model of heterosynaptic spike-timing-dependent plasticity (h-STDP). (**A**) Two variables in the spine $u(t)$ and $c(t)$ represent the normalized membrane potential and Ca^{2+} concentration respectively. Presynaptic input modulates $u(t)$ through AMPA (x^A) and NMDA ($g_N(u)x^N$) receptors. In addition, $u(t)$ is changed by back-propagation (x^{BP}), and heterosynaptic current caused by excitatory (x^E) and inhibitory (x^I) inputs. Calcium level $c(t)$ is modulated by influx/outflux through NMDA ($g_N(u)x^N$) and VDCC ($g_V(u)$). $c(t)$ is consequently controlled by $u(t)$ because both NMDA and VDCC are voltage-dependent. (**B**) An example of dynamics of the membrane potential variable $u(t)$ (top), Ca^{2+} concentration $c(t)$ (middle), and the intermediate variable $y(t)$ (bottom).

input is blocked, because of membrane depolarization at the excitatory spine due to presynaptic and postsynaptic spikes, calcium concentration rises up above the LTP threshold (red line in Fig 2.2B upper-right), hence induces LTP after repetitive stimulation (red line in Fig 2.2B lower-right). On the other hand, if the GABAergic input arrives coincidentally with the presynaptic input, depolarization at the excitatory spine is attenuated by negative current influx though the inhibitory synapse. As a result, calcium concentration cannot go up enough to cause LTP although it is still enough to cause LTD, thus eventually LTD is induced (black lines in Fig 2.2B right). Similarly when the postsynaptic spike arrives to the spine before the presynaptic spike does, without any GABAergic input, the presynaptic spike causes slow decay in the level of calcium concentration that may induce LTD (red lines in Fig 2.2B left). On the contrary, if the GABAergic input is provided simultaneously with the presynaptic input, slow decay in the calcium concentration is blocked because the inhibitory input causes hyperpolarization of the membrane potential at the excitatory spine. As a result, LTP is more likely achieved (black lines in Fig. 2.2B left). Therefore, when a GABAergic input is provided in coincidence with the presynaptic spike, the STDP time window changes sign in both pre-post and post-pre regimes (lines in Fig. 2.2A).

GABAergic effect on excitatory synaptic plasticity is also observed in CA1 [94]. In this case, post-pre stimulation does not induce LTD unless GABA uncaging is conducted near the excitatory spine right before the postsynaptic spike arrives at the spine, whereas LTP is induced by pre-post stimulation regardless of GABA uncaging (blue and cyan points in Fig. 2.2C). The proposed model can also replicate this result. In pre-post stimulation, due to positive feedback through NMDA receptor, membrane potential of the spine shows strong depolarization even if inhibitory current is delivered through GABA

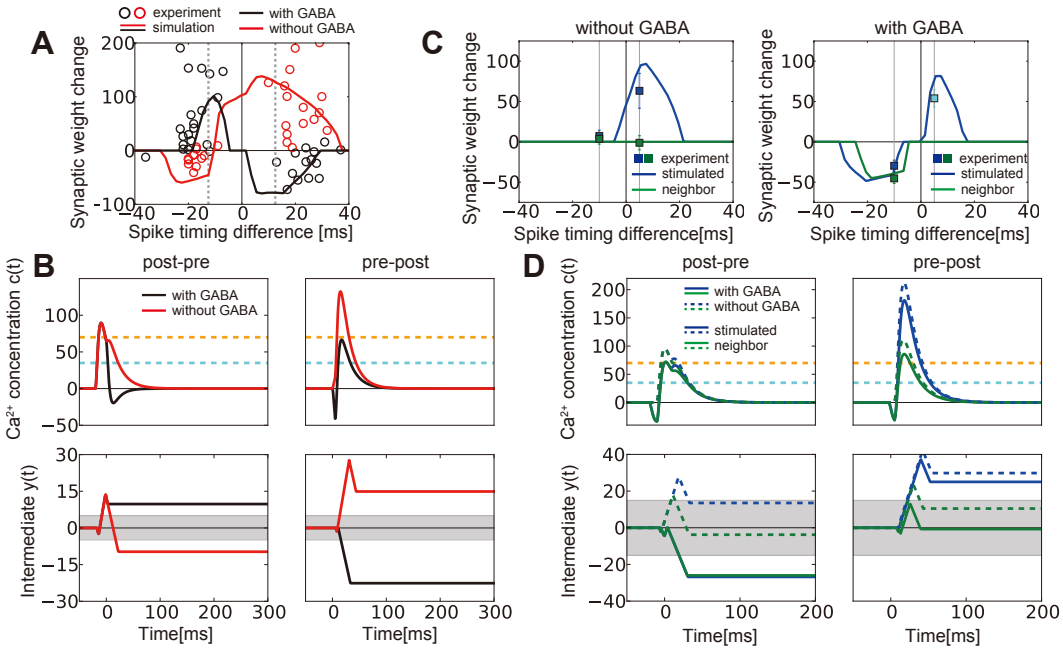


Figure 2.2. The model reproduces spike-timing-dependent heterosynaptic effects. (**A**) Spike timing window with/without a di-synaptic GABAergic input. Lines are simulation data, and points are experimental data taken from Paille et al., 2013 [182]. Vertical lines represent the timing differences at which Fig. B is calculated ($\Delta t = \pm 12.5$ ms). (**B**) Dynamics of calcium concentration $c(t)$ (top) and the intermediate variable $y(t)$ (bottom) at the stimulated spine. Gray areas in the bottom figures represent regions satisfying $y(t) < y_{th}/K_{rep}$, in which the change in the intermediate is not reflected into synaptic weight, where K_{rep} represents the number of paired stimulation given in the simulation for fig. **A,C**. (**C**) Synaptic weight change with/without GABAergic inputs right before pre/post stimulation. Data points were taken from Hayama et al., 2013 [94]. The cyan point is a result from muscimol application, not GABA uncaging. (**D**) Dynamics of $c(t)$ and $y(t)$ at the stimulated spine (blue lines) and a neighboring spine (green lines).

uncaging (blue lines in Fig. 2.2D upper-right). Thus, LTP is caused after repetitive stimulation (blue lines in Fig. 2.2D lower-right). On the other hand, in post-pre protocol, in the absence of GABAergic input, LTP/LTD effects tend to cancel each other, whereas LTD becomes dominant under GABA condition (blue lines in Fig. 2.2D left).

In addition to inhibitory-to-excitatory effect, excitatory-to-excitatory effect is also observed in case of CA1 [94]. If GABA uncaging is performed right before postsynaptic firing, LTD is also observed in neighboring excitatory spines. This E-to-E heterosynaptic effect is not observed for LTP or in the absence of GABAergic input (green points in Fig 2.2C). In my framework, excitatory current influx from a nearby synapse causes mild potentiation of calcium concentration in cooperation with inhibitory current influx, hence eventually induces LTD (green lines in Fig 2.2D left). Note that for this E-to-E effect, interaction at latter stage of synaptic plasticity may also play a dominant role [94].

Phase transitions underlying h-STDP.

Though, in the previous section, I introduced a complicated model to achieve correspondence with the biological process and get insight into the underlying mechanism, not all components of the model above are necessary to reproduce effects of h-STDP. Here, I provide a simple analytically tractable model to investigate the robustness of the proposed mechanism.

To this end, I shrink the model to the one in which calcium level at spine is directly modulated by pre-, post-, and heterosynaptic activity as below,

$$\begin{aligned}\frac{dC_i(t)}{dt} &= -\frac{C_i(t)}{\tau_C} + C_{pre}X_i(t-d_a) + C_{post}\left[1+g_C(C_i(t-\Delta t))\right]X_{post}(t-d_d) \\ &\quad -C_I\sum_{j\in\Omega_i^I} X_j^I(t-d_I) + C_E\sum_{j\in\Omega_i^E} X_j^E(t-d_E). \end{aligned} \tag{2.1}$$

Here, $C_i(t)$ represents Ca^{2+} concentration at spine i, X_i and X_{post} represent presynaptic and postsynaptic spikes respectively, and Ω_i^I and Ω_i^E are sets of neighboring inhibitory and excitatory synapses (see *Model B* in Methods for the details of the model). In addition, d_a, d_d are axonal delay and dendritic delay, and d_E, d_I are delays in heterosynaptic interaction. Despite simplicity, the model can reproduce heterosynaptic effects observed in striatum and CA1 neurons, though quantitative correspondence is hard to achieve in this case (Fig. 2.3A and B respectively). The model further provides analytical insights for the phenomena.

Let us first consider how the inhibitory effect parameter C_I controls I-to-E heterosynaptic effect observed in the CA1 experiment. If we characterize STDP time windows by the total number of local minimum/maximum, the parameter space can be divided into several different phases (Fig. 2.3C). If LTP threshold θ_p satisfies $C_{pre} < \theta_p < C_{post}$, Hebbian type STDP time window appears when the strength of heterosynaptic inhibitory effect C_I satisfies $(C_{post}-\theta_p)e^{\delta_I/\tau_C} < C_I < C_{pre}e^{\delta_I/\tau_C}$ (Orange phase in the middle of Fig. 2.3C). Here I defined δ_I as the spike timing difference between inhibitory spike

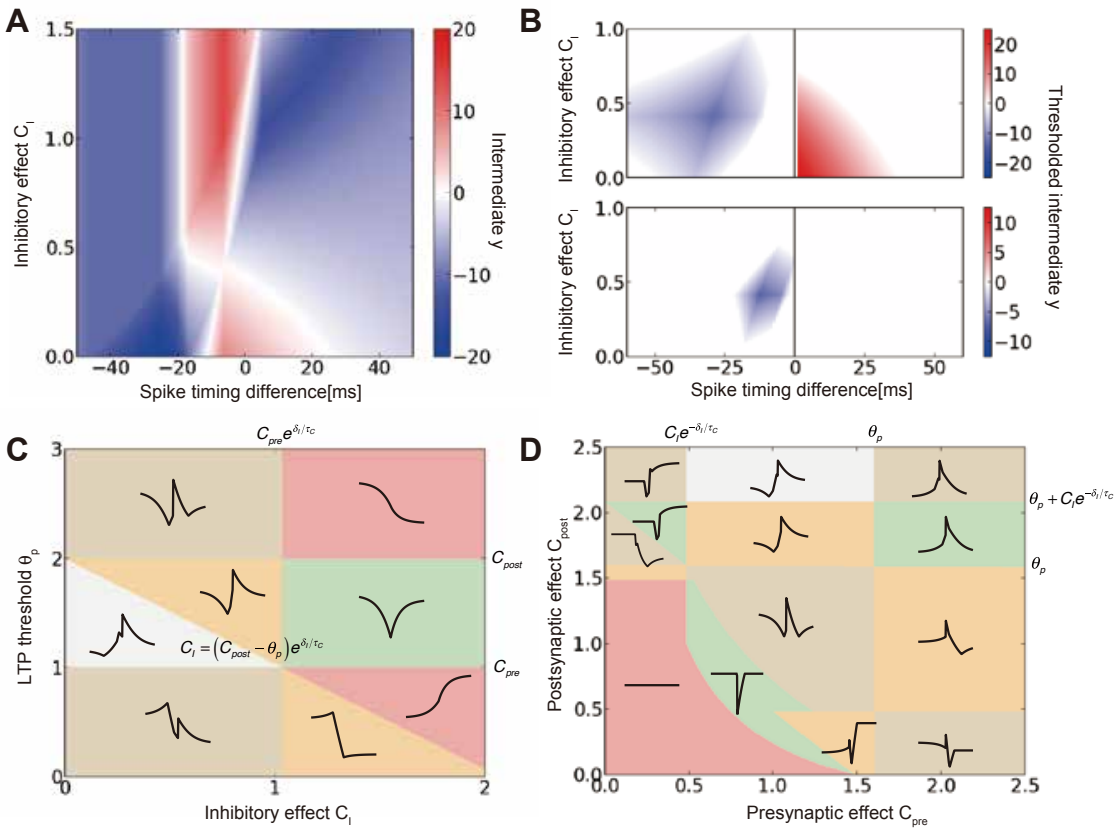


Figure 2.3. Heterosynaptic STDP can be understood as phase transitions on STDP time window in the analytical model (**A, B**) STDP windows at various strength of heterosynaptic inhibitory effect. Fig. **A** corresponds to the striatum experiment, while Fig. **B** reproduces the CA1 experiment. Note that values in Fig. **B** were calculated by $\tilde{y} = \mathrm{sgn}(y) \cdot [y - 15]_+$ to reflect the effect of thresholding. (**C**) Phase diagram of STDP time window calculated for inhibitory effect and LTP threshold. Colors show the number of local minimum/maximum. (**D**) Phase diagram calculated for pre and postsynaptic effect parameters at a fixed inhibitory effect ($C_I = 0.5$).

and presynaptic (postsynaptic) spikes in pre-post (post-pre) stimulation protocols. If C_I is larger than $C_{pre}\exp(\delta_I/\tau_C)$, strong inhibitory effect causes LTD even in pre-post regime (green phase in Fig. 2.3C), whereas LTD in post-pre regime is suppressed when C_I is smaller than $(C_{post}-\theta_p)\exp(\delta_I/\tau_C)$ (gray phase in Fig. 2.3C). This analysis confirms that, to observe heterosynaptic LTD, the heterosynaptic spike timing difference δ_I should be smaller than the timescale of Ca^{2+} dynamics τ_C [94], because $\delta_I < \tau_C \log\frac{C_I}{C_{post}-\theta_p}$ is necessary for a significant heterosynaptic LTD, and typically C_I is smaller than C_{post} and θ_p. In addition, heterosynaptic suppression of pre-post LTP is very unlikely to happen, because $C_I > C_{pre}$ is necessary even if $\delta_I = 0$, but heterosynaptic effect on Ca^{2+} dynamics in the spine is expected to be smaller than the homosynaptic effect (i.e. $C_I < C_{pre}$).

The model also provides analytical insight for E-to-E interaction. When the postsynaptic effect parameter Cpost satisfies $\theta_p < C_{post} < \theta_p + C_I e^{-\delta_I/\tau_C}$, and the presynaptic effect parameter C_{pre} fits into $C_I e^{-\delta_I/\tau_C} < C_{pre} < \theta_p$, STDP time window shows Hebbian-type timing dependency (orange phase in Fig. 2.3D). On the other hand, if C_{pre} is smaller than $C_I e^{-\delta_I/\tau_C}$ while satisfying $\theta_p + C_I e^{-\delta_I} - C_{post} < C_{pre}$, then STDP curve becomes LTD dominant (green phase in Fig. 2.3D). In E-to-E interaction, neighboring synapses receive small heterosynaptic calcium transient C_E, instead of presynaptic input C_{pre}. As a result, even if C_{pre} is large enough to cause Hebbian plasticity, C_E is typically smaller than C_{pre}, thus only LTD is observed in neighboring synapses as in experiments [94] [174]. These analytical results revealed that heterosynaptic effects are observable if parameters of calcium dynamics belong to a certain phase in the parameter space, thus h-STDP is robustly reproducible in my framework.

h-STDP induces local functional E/I balance at dendritic hotspots

Results so far suggest that the proposed model gives a good approximation of h-STDP. I next considered how this h-STDP rule shapes synaptic organization on the dendrite of a simulated neuron to investigate its possible functions. To this end, I first consider a model of a dendritic hotspot [113] that receives 10 excitatory inputs and 1 inhibitory input (Fig. 2.4A). Excitatory inputs are organized in 5 pairs, and each pair of excitatory synapses receives correlated inputs (Fig. 2.4B; see *Model* A_2 in Methods for details). In addition, the inhibitory input is correlated with one excitatory pair (Blue ones in Fig. 2.4A). Here, I assumed that postsynaptic activity follows a Poisson process, because influence of a single hotspot to the soma is usually negligible. In addition, I neglected the effect of morphology and hypothesized that delay between all spines within the hotspot is the same. In this configuration, surprisingly, excitatory synapses correlated with the inhibitory one are potentiated while other synapses experience minor de-potentiation (Fig. 2.4C). This potentiation is only observable when inhibitory activity is tightly correlated with excitatory activities, and slightly larger when inhibitory spike precedes excitatory spikes compared to the opposite case (Fig. 2.4D). In addition, heterosynaptic inhibitory effect γ_G needs to be relatively small in order to observe correlated potentiation (red area in Fig. 2.4E).

Otherwise, inhibitory input causes strong hyperpolarization on nearby synapses, resulting depression at correlated excitatory synapses instead of potentiation (blue area in Fig. 2.4E). These results indicate that h-STDP induces local functional E/I balance by potentiating excitatory synapses correlated with inhibitory synapses.

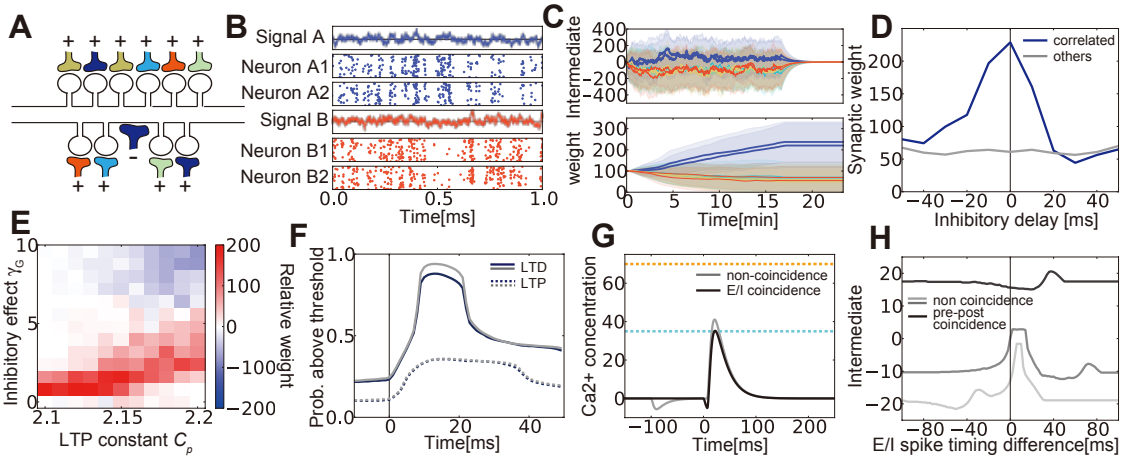


Figure 2.4. Emergence of local functional Excitatory/Inhibitory balance by heterosynaptic STDP (**A**) Schematic figure of a dendritic hotspot model. (**B**) Examples of correlated spike inputs. (**C**) Changes in intermediate y (top) and weight w (bottom) by h-STDP. (**D**) Synaptic weight change at the excitatory synapses correlated with the inhibitory inputs, at various inhibitory delays. (**E**) Relative weight changes w_R calculated at various parameters. I defined w_R by $w_R \equiv \langle w_i^E \rangle_{i \in corr} - \langle w_i^E \rangle_{i \in un-corr}$, where "corr" represents a set of excitatory synapses correlated with the inhibitory synapse, and "un-corr" stands for uncorrelated ones. (**F**) Probability of LTP/LTD occurrence calculated from simulation. (**G, H**) Results in single-spike simulations. E/I coincidence prevents LTD effect due to pre-spike (**G**), without affecting LTP effect due to pre-post coincidence (**H**). In fig. **G**, inhibitory spikes were provided at $t = 0$ in the black line, $t = -100$ms in the gray line, and the presynaptic spike was given at $t = 0$ in both lines. Similarly, in fig. **H**, postsynaptic spikes were provided at $t = -50$ (light-gray), 0 (black), $+50$ ms (dark-gray), and the presynaptic spike was given at $t = 0$ in all lines.

To reveal underlying mechanism of this E/I balance generation, from simulation data, I calculated the probability of calcium level being above the LTD/LTP thresholds after a presynaptic spike. LTP probability shows the same trajectory after a presynaptic spike on average, regardless of whether presynaptic activity is correlated with inhibitory input or not (dotted lines in Fig. 2.4F). On the other hand, LTD probability is lower for spine correlated with inhibitory inputs (solid lines in Fig. 2.4F), although the probability goes up after the presynaptic spike in both cases. This asymmetry between LTP and LTD can be understood in the following way; LTD is mainly caused when the presynaptic neuron spikes and the postsynaptic neuron stays silent both in the experiment [151] and in the model (gray line in Fig. 2.4G). However, if inhibitory input is provided at nearby dendrite in coincidence, calcium boost caused by excitatory presynaptic input is attenuated by heterosynaptic inhibitory effect (black line in Fig. 2.4G). As a result, LTD is shunted by correlated inhibitory inputs. On the other hand, LTP is mainly caused by coincidence between pre and postsynaptic neurons, which induces large increase in calcium level compared to attenuation by heterosynaptic inhibitory effect. Thus, inhibitory activity at nearby site does not prevent LTP at correlated excitatory synapses (Fig. 2.4H). Therefore, correlated spines experiences less depression, as a result, tend to be potentiated as a net sum.

To check the generality of the local E/I balance, I extended the model to a two-layered single cell by modeling each branch with one dendritic hotspot (Fig. 2.5A; see *Model* A_3 in Methods for details), and investigated the dendritic organization by h-STDP. Even in this case, when inhibitory inputs show diverse selectivity, each dendritic hotspot shapes its synaptic organization based on the selectivity of the inhibitory input (Fig. 2.5B). These result further imply that correlation-based clustering of excitatory synapses observed in previous experiments [122] [217] are possibly caused by common inhibitory inputs instead of direct interaction among excitatory spines.

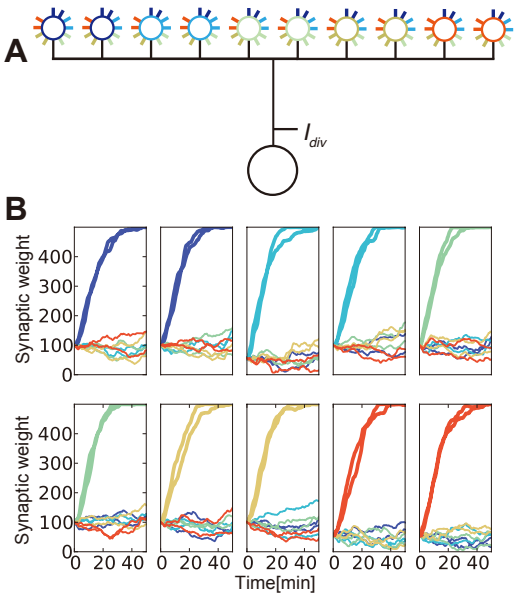


Figure 2.5. Local functional E/I balance in a two-layered single cell model (**A**) Schematics of the single cell model. In the model, each branch receives 10 excitatory inputs and 1 inhibitory input. Colors of the circles correspond to the feature-selectivity of inhibitory inputs. For instance, in the far left branch, the inhibitory neuron is correlated with excitatory synapses represented with blue lines in fig. **B**. (**B**) Synaptic weight change at each branch.

Local functional E/I balance enables optimal change detection from noisy stimuli

In the previous section, I demonstrated that h-STDP induces local E/I balance on a dendritic branch. I next investigate functional advantages of this local balance by considering what kind of teaching signal induces local E/I balance in supervised learning. In this way, we can perform rigid comparison with all the other possible synaptic organization.

To this end, based on the two-layered single neuron model in the previous section, I constructed a model of an excitatory neuron in primary sensory cortex (see *Model C* in Methods for details). The postsynaptic neuron has K numbers of dendritic branch each receives N_b^E excitatory inputs and one inhibitory input. To mimic the sensory stimuli, we introduced one external variable θ, which shows a random walk and occasionally jumps to a distant value (Fig 2.6A top). For instance, in case of primary visual cortex, variable θ would be the direction of moving bar stimuli. For input neurons, I assumed that

both excitatory and inhibitory input neurons show feature selectivity, and follow rate-modulated Poisson processes based on the external variable (Fig. 2.6A bottom). I additionally assumed that inhibitory response is broader and slightly delayed as often seen in the sensory cortex [148] [67]. I performed supervised learning on the somatic membrane potential by minimizing the error between the desired potential and the actual potential, which is calculated as the nonlinear sum of activities in dendritic branches. For the objectives of supervised signals, I considered two cases. One is change detection task in which the neuron should be depolarized if the external variable shows a rapid change within past 20 milliseconds, and otherwise should be hyperpolarized (Fig. 2.6D left). The other task is excitability maximization, in which the neuron should be in the depolarized state regardless of the external variable (Fig. 2.6D right).

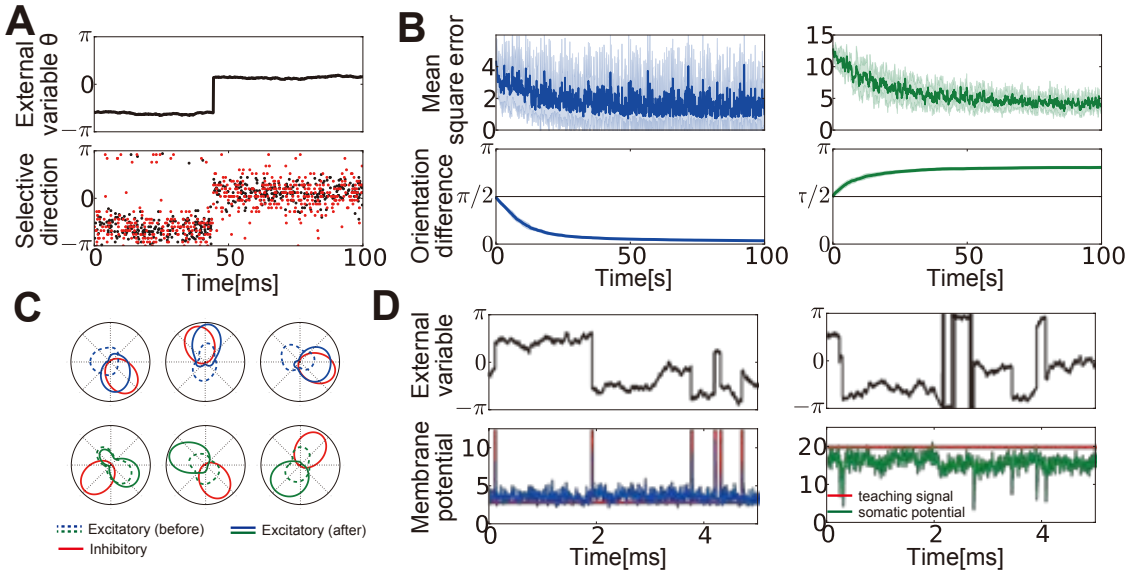


Figure 2.6. Supervised learning on a two-layered single cell model with two different teaching signals (**A**) An example of dynamics of the external variable $\theta(t)$ (top), and the spike responses of excitatory (black dots) and inhibitory (red dots) neurons. (**B**) Mean square error (top) and the orientation difference between excitatory population and inhibitory neuron (bottom) on a dendritic branch. (**C**) Polar representation of orientation selectivity of excitatory population and inhibitory neuron. (**D**) Examples of dynamics after learning. Red lines represent teaching signals, and blue/green lines are somatic potential.

By error back-propagating algorithm [197], we can modify weights of excitatory synapses on dendritic branches for minimizing the error. Indeed, for both teaching signals, mean square errors decrease as the learning progresses (Fig. 2.6B top). Interestingly, orientation difference between the inhibitory input and the excitatory population weights at each dendritic branch converged to the opposite values in two tasks (Fig. 2.6B bottom). In case of the change detection task, the orientation difference decreases so that excitatory and inhibitory inputs show the similar orientation selectivity (Fig. 2.6C top). On the other hand, in the excitability maximization task, the excitatory selectivity develops to the opposite direction with the inhibitory selectivity (Fig. 2.6C bottom). These results indicate that local functional E/I balance is not suitable for excitability maximization, but highly beneficial for the change detection. In visual [28] and auditory [54] cortices, many excitatory neurons are known to sensitive for change

in the external environment. My result predicts that in these neurons, inhibitory input show similar selectivity with excitatory inputs on nearby spines. Note that, in the model, each dendritic branch is specialized for detecting a change for one direction (Fig. 2.6C top), as a result, somatic potential can response to change in arbitrary directions, whereas in previous point-neuron models, typically an output neuron can barely response to one specific change [125] [230].

h-STDP explains critical period plasticity of binocular matching

Results so far indicate h-STDP induces GABA-driven circuit formation. To confirm that these results are applicable for the developmental plasticity, I next consider a model of critical period plasticity in binocular matching [235] [236]. In mice, two weeks after the eye opening, typically binocular neurons in V1 still have different orientation selectivity for inputs from two eyes. Nevertheless, after a month, selective orientation for both eyes get closer, and almost coincides with each other eventually [235]. Importantly, this phenomenon is not likely to be explained by simple Hebbian plasticity, because in that case, binocular matching should be initiated in first two weeks upon eye opening.

I modeled this process with a two-layered single cell introduced in Fig.5 (see *Model* A_4 in Methods for details). Inputs were modeled as rate modulated Poisson processes driven by a circular variable θ, as in the previous section. I assumed that (i) inputs from ipsi- and contralateral eyes already have some weak orientation selectivity at the eye opening [235] [61], (ii) Inhibitory cells are driven by both ipsi- and contralateral eyes [248] [130], (iii) Average selectivity of inhibitory inputs comes in between the selectivity for ipsilateral inputs and contralateral inputs. The last assumption is not supported from experimental evidence, but if inhibition is provided from neighboring interneurons, these inhibitory neurons are likely to be driven by similar feedforward excitatory inputs to those drive the output neuron. Here, I consider direction selectivity instead of orientation selectivity for simplicity, but the same argument holds for the latter.

In the simulation, I first run the process without inhibition then introduced GABAergic inputs after a while (red lines in Fig. 2.7A-C represent the starting point of inhibitory inputs). After the introduction, mean excitatory input direction in each branch converged to the direction of the local inhibition (Fig. 2.7A top), though synaptic weight development was biased toward the global direction selectivity (Fig. 2.7D; here, the bias is toward the right side). Thus, even if we consider all the synapses, the difference between ipsi- and contralateral selectivity became smaller (Fig. 2.7A middle). As a result, binocular selectivity became stronger (Fig. 2.7A bottom), and the response for monocular inputs matches with each other (Fig. 2.7E). When I deprived inputs from the contralateral eye right after the introduction of inhibition, binocular matching was blocked (Fig. 2.7B), while the matching was not disrupted when the deprivation was performed before the introduction of GABAergic input (Fig. 2.7C). These results indicate that GABA-maturation and resultant h-STDP can be a part of the underlying mechanism for critical period plasticity in binocular matching.

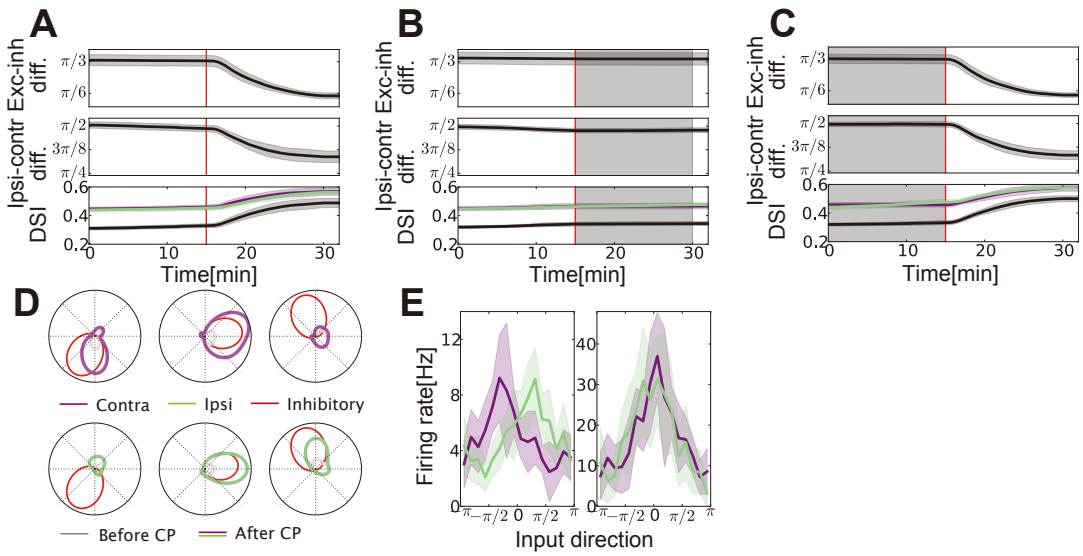


Figure 2.7. heterosynaptic plasticity can trigger binocular matching (**A**-**C**) (top): Difference between mean excitatory direction selectivity and inhibitory direction selectivity in a branch. (middle): Difference between mean ipsi-driven excitatory direction selectivity and mean contra-driven excitatory direction selectivity over all synapses on the neuron. (bottom): Direction selectivity index calculated for contralateral inputs (purple), ipsilateral inputs (light-green), and binocular input (black). In all figures, red vertical lines represent the timing for introduction of inhibitory inputs. In shadowed areas of Fig. **B**, **C**, to mimic monocular deprivation, contra-driven inputs were replaced with rate-fixed Poisson inputs. (**D**) Examples of direction selectivity of a branch before (gray lines) and after (purple/light-green lines) the learning. Red lines represent the selectivity of the inhibitory input. (**E**) Firing responses of the neuron before (left) and after (right) the learning.

Discussion

In this study, I first showed that a calcium-based plasticity model robustly captures several different characteristics of plasticity-related interaction between neighboring synapses in millisecond timescale by introducing current-based heterosynaptic interaction terms (Fig. 2.2,2.3). Based on this proposed model, I next investigated the possible functions of h-STDP. My study revealed that correlated E/I synaptic inputs on the same hotspot causes the local function E/I balance (Fig. 2.4,2.5), which is beneficial for change detection (Fig. 2.6). Furthermore, I found that h-STDP can induces binocular matching upon GABA maturation, and support accurate input estimation (Fig. 2.7).

Experimental predictions

My study provides three experimental testable predictions: First, the results in Figure 4 indicate that LTD at an excitatory synapse is cancelled out by coincident inhibitory inputs to the nearby dendrite. Thus, LTD by low frequency stimuli [151] can be attenuated by coincident GABA uncaging around the stimulated spine. Note that this result would not contradict with GABA-driven heterosynaptic LTD observed in paired stimulation, because in that experiment, the excitatory spine was presumably overexcited for inducing LTD in the absence of GABA [94]. Indeed, coincident GABAergic inputs may induce hetersosynaptic LTD by combining with presynaptic stimulation at moderately high frequency that itself does not cause LTD [22].

Secondly, my results provide a hypothesis on synaptic organization on dendritic tree. It is known that excitatory synaptic inputs to a dendritic hotspot often show correlated activities [122] [217]. My results indicate that an inhibitory input should also be correlated to excitatory inputs projected to the nearby dendrite (Fig. 2.4,2.5), especially on a dendritic tree of an excitatory neuron that is sensitive to change in the external environment (Fig. 2.6). Also the model explains why feature selectivity of these spines typically shows weak similarity despite their correlation [113] [38]. Suppose a synaptic cluster is shaped by a common inhibitory heterosynaptic effect, not by excitatory-to-excitatory interaction, variability within the cluster tends to be large, because inhibitory neurons in sensory cortex typically have wider feature selectivity than excitatory neurons [148] [165]. In addition, it should also be noted that, E-to-E heterosynaptic LTP is typically induced as a meta-plasticity in a timescale of minutes [93], which itself is not sufficient to create a correlation-based synaptic cluster.

The third implication of the model is about binocular matching. My model indicates that GABA-maturation plays a critical role in binocular matching, but the phenomenon can also be explained by Hebbian plasticity plus some kind of meta-plasticity. If binocular matching is purely induced by Hebbian plasticity not through heterosynaptic mechanism, selective orientation after the matching depends solely on the initial selectivity for monocular inputs. Especially when the contralateral input is larger than the ipsilateral input, resultant selectivity should nearly coincides with the original contralateral selectivity. On the other hand, if the proposed mechanism takes part in the development, the consequent selectivity should also be influenced by mean selectivity of input inhibitory neurons. Thus, long-term imaging of monocular selectivity at binocular neurons in V1 would reveal whether a covariance-based rule is sufficient enough to explain the phenomena, or some other mechanisms including the proposed one also play a role in the shift. In addition, it is known that precocious GABA maturation disrupts binocular matching [236]. My model suggests that the disruption is possibly related to the violation of the third assumption in the model, which is correlation of mean inhibitory inputs to both ipsi- and lateral selectivity of the postsynaptic neuron.

Carrier of heterosynaptic interaction

Heterosynaptic plasticity has been observed in various spatial and temporal scales, and arguably underlying molecular mechanisms are different from one to one [174]. In case of milliseconds-order interaction, single-atomic ions are strong candidates, because poly-atomic ions such as IP3 are too big to move rapidly from spine to spine [202]. If change in Ca^{2+} concentration at an un-stimulated spine is crucial for synaptic plasticity, Ca^{2+} influx/outflux from either intra or extracellular sources are necessary for induction of heterosynaptic plasticity. Because inhibitory synaptic inputs often change the Ca^{2+} concentration in the dendritic branch locally [168], intracellular modulation of calcium is plausible, but at the same time, extracellular Ca^{2+} influx/outflux through NMDA and VDCC should also be driven strongly by heterosynaptic activity, because often inhibitory inputs modulate membrane voltage

locally [75]. In addition, most of intracellular calcium-ions exist within calcium-buffer [99], and they may also important for induction of synaptic plasticity. In my model, both current-based interaction (*Model A*) and calcium-based interaction (*Model B*) replicate the experimental results (Fig. 2.2 and 2.3). Nevertheless, my study implies that intracellular supply may not be sufficient, because, according to my analytical study, the heterosynaptic Ca^{2+} change typically needs to be comparable with the homosynaptic change in order to cause significant heterosynaptic plasticity (Fig. 2.3C, D).

Note that heterosynaptic interaction does not need to work in milliseconds order to interfere with STDP. For instance, E-to-E heterosynaptic LTD can be initiated by spreading of LTD-related molecules, not by neural dynamics-related components [94]. In addition, Paille et al. proposed that the shift in STDP time window observed in their experiment is possibly explained by change in the ratio between calcium influx through NMDA and the influx through VDCC [182].

Inhibitory cell types

Somatostatin positive (SOM^+) inhibitory neurons are the major candidate for heterosynaptic STDP, because they are typically projected to the dendrite, their IPSP curves is shorter than the timescales of NMDA or Ca^{2+} dynamics [153], and they often show strong feature selectivity compared to other inhibitory neuron types [148]. However, the model does not exclude parvalbumin positive (PV^+) inhibitory neurons, because they usually have projections to proximal dendrites, and they are typically fast spiking [153]. In particular, h-STDP through PV^+ cell may play important roles in critical period plasticity [218].

In addition, many inhibitory synapses are projected to dendritic spine [39], though I mainly considered inhibitory synapses on dendritic shaft in this study. My result implies that heterosynaptic effect by these synapses would be more specific and possibly strong.

Related theoretical studies

Previous biophysical simulation studies revealed that synaptic plasticity at excitatory synapse critically depends on inhibitory inputs at nearby dendrite [46] [14], but these studies did not reveal general rules nor functional roles of the heterosynaptic plasticity. On the other hands, network modeling studies found that heterosynaptic plasticity provides a homeostatic mechanism [40] [252], but in these models, heterosynaptic plasticity were modeled as a global homeostatic plasticity without any branch specificity, and the advantage over other homeostatic mechanisms was unclear. In my work, by considering intermediate abstraction with analytical but biologically plausible models, I proposed candidate mechanisms for experimental results that have not modeled before, and potential functions of h-STDP in neural circuit formation.

Methods

Model A_1: Calcium-based STDP model with current-based heterosynaptic interaction

Let us first consider membrane dynamics of a dendritic spine. Membrane potential of a spine is mainly driven by presynaptic inputs through AMPA/NMDA receptors, backpropagation of postsynaptic spike, leaky currents, and current influx/outflux caused by excitatory/inhibitory synaptic inputs at nearby synapses. Hence, I modeled membrane dynamics of spine i with the following differential equation:

$$\frac{du_i(t)}{dt} = -\frac{u_i(t)}{\tau_m} + \gamma_A x_i^A(t) + \gamma_N g_N(u_i) x_i^N(t) + \gamma_{BP} x_i^{BP}(t) - \gamma_I \sum_{j \in \Omega_i^I} x_j^I(t - d_I) + \gamma_E \sum_{j \in \Omega_i^E} x_j^E(t - d_E), \tag{2.2}$$

where u_i is the membrane potential of the spine, and τ_m is the membrane time constant. Here, conductance changes were approximated by current changes. The resting potential was renormalized to zero for simplicity. In next terms, x_i^A and x_i^N are glutamate concentration on AMPA/NMDA receptors respectively. The function $g_N(u_i) = \alpha_N u_i + \beta_N$ represents voltage dependence of current influx through NMDA receptors. This positive feedback is enhanced when additional current is provided through back-propagation. As a result, the model reproduces large depolarization caused by coincident spike between presynaptic and postsynaptic neurons. Although AMPA receptor also shows voltage dependence, here I neglected the dependence, as the relative change is small around the resting potential [147]. x_i^{BP} is the effect of backpropagation from soma, and the last two term of the equation represents heterosynaptic current, which is given as the sum of inhibitory/excitatory currents x_j^I, x_j^E at nearby synapses. I defined sets of nearby inhibitory/excitatory synapses as Ω_i^I and Ω_i^E respectively, and their delays were denoted as d_I and d_E. Each input x_i^Q (Q = A,N,BP,I,E) is given as convoluted spikes: $\frac{dx_i^Q(t)}{dt} = -\frac{x_i^Q(t)}{\tau_Q} + \sum_{s^k} \delta\left(t - s^k\right)$, where s_k represents the spike timing of the k-th spike. In the simulation, although convolution is calculated at the heterosynaptic synapse, this does not influence results because exponential decay is linear.

I next consider calcium influx to a spine through NMDA receptors and VDCC. For a given membrane potential u_i, calcium concentration at spine i can be written as

$$\frac{dc_i}{dt} = -\frac{c_i}{\tau_C} + g_N(u_i) x_i^N(t) + g_V(u_i), \tag{2.3}$$

where $g_V(u_i) = \alpha_V u_i$ represents calcium influx through VDCC, and $g_N(u_i) x_i^N(t)$ is the influx from NMDA.

Calcium concentration at spine is the major indicator of synaptic plasticity, and many results indicate that high Ca^{2+} concentration on a spine typically induces LTP, while low concentration often causes LTD [147]. Previous modeling studies revealed calcium-based synaptic plasticity model constructed on that principle well replicate various homosynaptic STDP time window observed in *in vitro* experiments

[208] [86]. Hence, here I employed their framework for plasticity model. I additionally introduced an intermediate variable to reflect all-or-none nature of synaptic weight change [185]. This variable is expected to correspond with concentration of plasticity related enzymes such as CaMKII or PP1. Similar results are expected for stochastic bi-stable attractor model [87] [86]. In the proposed model the intermediate y_i and synaptic weight w_i follow

$$\begin{aligned} \frac{dy_i(t)}{dt} &= -\frac{y_i(t)}{\tau_y} + C_p[c_i - \theta_p]_+ - C_d[c_i - \theta_d]_+, \\ \frac{dw_i(t)}{dt} &= B_p[y_i - y_{th}]_+ \left[x_i^H - h_{th}\right]_+ - B_d[-(y_i + y_{th})]_+. \end{aligned} \tag{2.4}$$

$[X]_+$ is a sign function which returns 1 if $X \geq 0$, returns 0 otherwise. x_i^H is a gating term introduced for preventing pathological behavior. Without this term, modeled synapse can show potentiation without any presynaptic input if calcium concentration is high enough, though such phenomena are not reported yet. Note that this term is a constraint for plausibility, hence not necessary for reproduction of experimental results. In addition, as observed in recent experiments [71], in the model, back-propagation is not necessary for LTP, if presynaptic inputs are given when the membrane potential at the spine is well depolarized.

In the simulation, I set common parameters as $\tau_C = 18.0$, $\tau_M = 3.0$, $\tau_N = 15.0$, $\tau_A = 3.0$, $\tau_{BP} = 3.0$, $\tau_I = 3.0$, $\tau_E = 6.0$, $\tau_Y = 50,000$, $\tau_H = 50,000$, $d_I = 0.0$, $\alpha_N = 1.0$, $\alpha_V = 2.0$, $\gamma_A = 1.0$, $\theta_p = 70$, $\theta_d = 35$, $C_p = 2.3$, $C_d = 1.0$, $B_p = 0.001$, $B_d = 0.0005$, $h_{th} = 0.01$. In the model of STDP at striatum, in addition, I used $\beta_N = 1.0$, $\gamma_N = 0.0$, $\gamma_{BP} = 8.0$, $\gamma_I = 5.0$, $y_{th} = 250$, while for the model of Schaffer collateral synapses, I used $d_E = 1.0$, $\beta_N = 0.0$, $\gamma_N = 0.2$, $\gamma_{BP} = 8.5$, $\gamma_I = 3.0$, $\gamma_E = 1.0$, $y_{th} = 750$. In the parameter search, decay time constants were chosen from biologically reasonable ranges [123], α_N, γ_A, C_d, B_d were fixed at unitary values, and other parameters were manually tuned. Synaptic weight variables $\{w\}$ were bounded to $0 < w < 500$, and initialized at $w = 100$. All other variables were initialized at zero in the simulation. Paired stimulation was given every 1 second for 100 seconds, and synaptic weight changes were calculated from the values 400 seconds after the end of stimulation. In the cortico-striatal synapse model, the inhibitory spike was presented at the same timing with the presynaptic spike, and for Schaffer collateral synapses, inhibitory spikes were given 10 milliseconds before pre (post) spikes in pre-post (post-pre) stimulation protocols. In calculation of intermediate variable $y(t)$ in Fig. 2.2B,D, I ignored the effect of exponential term, because of the difference in timescale. I subtracted 7.5 milliseconds of axonal delay from the timing of presynaptic stimulation in the calculation of spike timing difference.

Model A_2: Models of a dendritic hotspot

Dendritic hotspot model was constructed based on the Schaffer collateral synapse model described above. For simplicity, I hypothesized that heterosynaptic effect by inhibitory spike arrives at excitatory

spines at the same time, and I disregarded E-to-E interaction by setting $gamma_E = 0.0E$. Correlated spikes were generated using a hidden variable as in previous studies [230] [102]. I generated five dynamic hidden variables, and updated them at each time step by $s_\mu(t+\Delta t) = (\zeta - \frac{1}{2})(1-\alpha_s) + s_\mu(t)\alpha_s$, where $\alpha_S = \exp[-\Delta t/\tau_S]$, $\tau_S = 10$ms, and ζ is a random variable uniformly chosen from [0,1). In the simulation, the time step was set at $\Delta t = 0.1$ms. Activities of presynaptic neurons were generated by rate-modulated Poisson process with $r_i^E(t) = [r_X^E + r_s^E s_\mu(t)]_+$ for excitatory neuron i modulated by the hidden variable μ. Similarly, the presynaptic inhibitory neuron was described by a Poisson-model with $r^I(t) = [r_X^I + r_S^I s_0(t)]_+$. Activity of the postsynaptic neuron was given as a Poisson-model with a fixed rate r_{post}. I set parameters $\{r_x^E, r_s^E, r_{post}\}$ in a way that all pre and postsynaptic excitatory neurons show the same firing rate, to avoid the effect of firing-rate difference on synaptic plasticity.

For parameters, I used $\gamma_I = 0.5$, $\beta_N = 1.0$, $\gamma_{BP} = 8.0$, $C_p = 2.1$, $y_{th} = 250$ and other parameters were kept at the same value with the original Schaffer collateral model. Except for Fig. 2.4D, the delay of inhibitory spike was set as zero. Presynaptic activities were given by $r_X^E = 1.0$, $r_S^E = 500.0$, $r_X^I = 2.0$, $r_S^I = 1000.0$, and postsynaptic firing rate was set as $r_{post} = 5.0$.

Model A_3: a two-layered single cell model Previous studies suggest that complicated dendritic computation

can be approximated by a two-layered single cell model [189] [145]. Thus, I constructed a single cell model by assuming that each hotspot works as a subunit of a two-layered model. I defined the mean potential of a dendritic subunit k by $u_b^k(t) \equiv \sum_{i=1}^{N_b^E} w_i^k u_i^k(t)/(w_o^E N_b^E)$, and calculated the somatic membrane potential by $u_{soma}(t) \equiv -\gamma_S x^{BP}(t) + \sum_k g_b(u_b^k(t))$. Here, I subtracted the back propagation term to reproduce the effect of refractory period. Postsynaptic spikes were given as a rate-modulated Poisson model with the rate $u_{soma}(t)/I_{dv}(t)$. $I_{dv}(t)$ is the divisive inhibition term introduced to keep the output firing rate at r_{post}. By using the mean somatic potential $\frac{d\bar{u}_{soma}(t)}{dt} = -\frac{1}{\tau_V}(\bar{u}_{soma} - u_{soma}(t))$, $I_{dv}(t)$ was calculated as $I_{dv}(t) \equiv \bar{u}_{soma}(t)/r_{post}$. In the simulation, I used $\gamma_S = 10.0$, $C_p = 2.0$, $\tau_V = 1$ seconds, and other parameters were kept at the same values with the hotspot model.

Model A_4: A model of binocular matching

For the model of critical period plasticity of binocular matching, I used the two-layered single cell model introduced in the previous section (*Model* A_3). The neuron has $K = 100$ dendritic branches, each receives $N_b^E = 20$ excitatory inputs and 1 inhibitory input. At each branch, half of excitatory inputs are from the contralateral eye, and the other half are from the ipsilateral eye. Each excitatory input neuron have direction selectivity characterized with $\theta_{k,i}^E$, and shows rate-modulated Poisson firing with

$$r_{k,i}(t) = r_x^E \exp\left[\beta_E \cos(\theta(t) - \theta_{k,i}^E)\right] / I_0(\beta_E),$$

where where $I_0(\beta)$ is the modified Bessel function of order 0 (input configure is basically the same with *Model C*). Similarly, firing rate of an inhibitory neuron is given as $r_k^I(t) = r_x^I \exp\left[\beta_I \cos(\theta(t) - \theta_k^I)\right] / I_0(\beta_I)$. For each input neuron, mean direction selectivity $\{\theta_{k,i}^E, \theta_k^I\}$ were randomly chosen from a von Mises distribution $\exp\left[\beta_S \cos(\theta - \theta_Q)\right] / 2\pi I_0(\beta_S)$, where Q=$\{$contra, ipsi, inh$\}$. In the simulation, I used $\theta_{contra} = -\pi/4$, $\theta_{ipsi} = \pi/4$, $\theta_{inh} = 0$. Direction of visual stimulus $\theta(t)$ changes randomly with $\theta(t + \Delta t) = \theta(t) + \sigma_{sr}\zeta_G$ where ζ_G is a Gaussian random variable, and Δt is the time step of the simulation. To mimic monocular deprivation, in the shadowed area of Fig. 2.7B,C, I replaced contra-driven input neuron activity with a Poisson spike with constant firing rate r_{md}^E. In addition, in Fig. 2.7B, to simulate the lack of contra-driven inputs to inhibitory neurons, I replaced inhibitory activity with $r_k^I(t) = r_{md}^I + (r_x^I/2) \exp\left[\beta_I \cos(\theta(t) - \theta_{k,md}^I)\right] / I_0(\beta_I)$, where $\theta_{k,md}^I$ was sampled from $\exp\left[\beta_S \cos(\theta(t) - \theta_{ipsi})\right] / 2\pi I_0(\beta_S)$.

To evaluate the development of binocular matching, I introduced three order parameters. First, the difference between mean excitatory direction selectivity and inhibitory selectivity at a branch k was evaluated by $\theta_{b,k}^d = \left|\arg\left(\sum_i w_{k,i}^E e^{i(\theta_{k,i}^E - \theta_k^I)}\right)\right|$. Similarly, the global direction selectivity difference between inputs from the ipsi- and contralateral eyes were defined by

$$\theta_G^d = \hat{d}\left[\arg\left(\sum_{k=1}^{K}\sum_{i\in ipsi} w_{k,i}^E e^{i\theta_{k,i}^E}\right), \arg\left(\sum_{k=1}^{K}\sum_{i\in contrai} w_{k,i}^E e^{i\theta_{k,i}^E}\right)\right],$$

where the function $\hat{d}[\theta_1, \theta_2]$ calculates the phase difference between two angles. Finally, direction selectivity index DSI for binocular input was calculated by

$$DSI = \left|\frac{\sum_{k=1}^{K}\sum_{i=1}^{N_b^E} w_{k,i} e^{i\theta_{k,i}^E}}{\sum_{k=1}^{K}\sum_{i=1}^{N_b^E} w_{k,i}}\right|.$$

For the calculation of the monocular direction selectivity index, at each branch k, I took sum over $N_b^E/2$ excitatory inputs corresponding to the each eye instead of all N_b^E inputs. In Fig. 2.7E, I measured direction selectivity by providing monocular inputs for 100 seconds. The sensory stimulus $\theta(t)$ was randomly sampled every 100 milliseconds.

In the simulation, mostly I used the same parameters with the *model* A_3. In addition, I set $\gamma_I = 4.0$, $C_p = 1.85$, $y_{th} = 75.0$. Inputs parameters were set at $\beta_E = 4.0$, $\beta_I = 2.0$, $\beta_S = 1.0$, $\theta_{contra} = -\pi/4$, $\theta_{ipsi} = \pi/4$, $\theta_{inh} = 0$, $r_X^E = 5.0$, $r_X^I = 10.0$, $r_{md}^E = 1.0$, $r_{md}^I = 1.0$, $\sigma_{sr} = 0.1\sqrt{\Delta t}$.

Model B: A reduced analytical model of a spine

If we shrink equations for membrane potential and calcium concentration into one, the reduced equation would be written as,

$$\frac{dC_i(t)}{dt} = -\frac{C_i(t)}{\tau_C} + C_{pre}X_i(t - d_a) + C_{post}\left[1 + g_C(C_i(t - \Delta t))\right] X_{post}(t - d_d)$$

$$-C_I \sum_{j\in\Omega_i^I} X_j^I(t-d_I) + C_E \sum_{j\in\Omega_i^E} X_j^E(t-d_E),$$

where $g_c(X) = \eta[X]_+$ captures the nonlinear effect caused by pre-post coincidence. g_c was calculated from the value of C_i at $t = t - \Delta t$ to avoid pathological divergence caused by the delta function. In the simulation, I simply used value of C_i one time step before. Here, all input X_i, X_{post}, X_j^I, X_j^E are given as point processes. For the intermediate y, I used the same equation as before. Note that above equation is basically same with the one in [86] except for the nonlinear term $g_c(C)$ and the heterosynaptic terms.

Let us consider weight dynamics of an excitatory synapse that has only one inhibitory synapse in its neighbor. For analytical tractability, I consider the case when presynaptic, postsynaptic, and inhibitory neurons fire only one spikes at $t = t_{pre}, t_{post}, t_I$. In case of the CA1 experiment, because GABA uncaging was always performed before pre and postsynaptic spike, the timing of inhibitory spike is given as $t_I = \min(t_{pre}, t_{post}) - \delta_I$ for $\delta_I > 0$. Note that spike timings are counted at the excitatory spine, so the actual timings are $t'_{pre} = t_{pre} - d_{axon}$, $t'_{post} = t_{post} - d_{dendrite}$, $t'_I = t_I I - d_I - d_{inhaxon}$. In this case, the change in intermediate variable of the excitatory synapse is given as

$$\Delta y = \begin{cases} G_1\left(C_1, t_{pre} - t_{post}\right) + G_2\left(C_{pre} + C_1 e^{-(t_{pre}-t_{post})/\tau_C}\right) (\text{if} t_{post} < t_{pre}) \\ G_1\left(C_2, t_{post} - t_{pre}\right) + G_2\left(C_{post}\left[1 + g_C\left(C_2 e^{-(t_{post}-t_{pre})/\tau_C}\right)\right] + C_2 e^{-(t_{post}-t_{pre})/\tau_C}\right) (\text{otherwise}) \end{cases}$$

where,

$$\begin{aligned} C_1 &\equiv C_{post} - C_I e^{-(t_{post}-t_I)/\tau_C},\ C_2 \equiv C_{pre} - C_I e^{-(t_{pre}-t_I)/\tau_C} \\ G_1(C, \Delta t) &\equiv B_p[C-\theta_p]_+ \left(\left[\tau_C \log\tfrac{C}{\theta_p} - \Delta t\right]_+ \Delta t + \left[\Delta t - \tau_C \log\tfrac{C}{\theta_p}\right]_+ \tau_C \log\tfrac{C}{\theta_p} \right) \\ &\quad - B_d[C-\theta_d]_+ \left(\left[\tau_C \log\tfrac{C}{\theta_d} - \Delta t\right]_+ \Delta t + \left[\Delta t - \tau_C \log\tfrac{C}{\theta_d}\right]_+ \tau_C \log\tfrac{C}{\theta_d} \right), \\ G_2(C) &\equiv B_p[C-\theta_p]_+ \tau_C \log\tfrac{C}{\theta_p} - B_d[C-\theta_d]_+ \tau_C \log\tfrac{C}{\theta_d}. \end{aligned}$$

Similarly, in case of the striatum experiment, by setting $\eta = 0$, the change in the intermediate variable is given as

$$\Delta y = \begin{cases} G_1\left(C_{post}, t_{pre} - t_{post}\right) + G_1\left(C_3, t_I - t_{pre}\right) + G_2\left(-C_I + C_3 e^{-(t_I - t_{pre})/\tau_C}\right) (\text{if} t_{post} < t_{pre} < t_I) \\ G_1\left(C_{pre}, t_I - t_{pre}\right) + G_1\left(C_4, t_{post} - t_I\right) + G_2\left(C_{post} + C_4 e^{-(t_{post} - t_I)/\tau_C}\right) (\text{if} t_{pre} < t_I < t_{post}) \\ G_1\left(C_{pre}, t_{post} - t_{pre}\right) + G_1\left(C_5, t_I - t_{post}\right) + G_2\left(-C_I + C_5 e^{-(t_I - t_{post})/\tau_C}\right) (\text{if} t_{pre} < t_{post} < t_I), \end{cases}$$

where $C_3 \equiv C_{pre} + C_{post} e^{-(t_{pre}-t_{post})/\tau_C}$, $C_4 \equiv -C_I + C_{pre} e^{-(t_I - t_{pre})/\tau_C}$, and $C_5 \equiv C_{post} + C_{pre} e^{-(t_{post}-t_{pre})/\tau_C}$.

In the simulation, parameters were set at $\tau_C = 30$ms, $C_{post} = 2.0$, $\theta_p = 1.6$, $\theta_d = 1.0$. Additionaly, in the model of a Schaffer collateral synapse, I used $\delta_I = 1.0$, $C_{pre} = 1.0$, $C_E = 0.30$, $\eta = 2.0$, and

for the model of a cortico-striatal synapse, I employed $\delta_I = 5.0$, $C_{pre} = 0.75$, $C_E = 0.0$, $\eta = 0.0$.

Model C: Supervised learning with a two-layered single cell model

I constructed the model by considering a two-layered single cell model with $K = 100$ dendritic branches, each receives $N_b^E = 20$ excitatory inputs and 1 inhibitory input. Input neuron activity depends on an external variable θ, defined on a ring as $\theta \in [-pi, pi)$. The variable θ follows a random walk process plus occasional jump as:

$$\theta(t+\Delta t) = \begin{cases} \theta(t) + \left(\frac{1}{2} + \zeta_U\right)\pi & \text{(with prob. } \Delta t/\tau_\theta) \\ \theta(t) + \sigma_\theta \zeta_G & \text{(otherwise)}, \end{cases}$$

where ζ_U is a random variable uniformly sampled from [0,1), and ζ_G is a Gaussian random variable. As in *model* A_4, the response of excitatory input neurons follow a rate-modulated Poisson process with rate $r_{k,j}(t)$, which is defined from von Mises distribution as

$$r_{k,j}(t) = r_x^E \exp\left[\beta_E \cos(\theta(t) - \theta_{k,i}^E)\right] / I_0(\beta_E),$$

where $I_0(\beta)$ is the modified Bessel function of order 0. The firing rate of an inhibitory input neuron is given as $r_k^I(t) = r_x^I \exp\left[\beta_I \cos(\theta(t-\delta_\theta) - \theta_k^I)\right] / I_0(\beta_I)$, where δ_θ is the delay in inhibitory response ($\delta_\theta = 5.0$ms in the simulation). Membrane dynamics of the output neuron was defined as

$$\begin{aligned} v_k(t) &= \sum_{i \in \Omega_k^E} w_{k,i}^E \int_0^\infty d\tau \cdot \varepsilon_E(\tau)\, s_{k,i}(t-\tau) - \sum_{j \in \Omega_k^I} w_o^I \int_0^\infty d\tau \cdot \varepsilon_I(\tau)\, s_{k,j}(t-\tau) \\ v_{soma}(t) &= \sum\nolimits_k g(v_k(t)),\ g(v) = \left(1 + \exp\left[-\beta_g(v - \alpha_g)\right]\right)^{-1} \end{aligned}$$

where EPSP/IPSP curves were given by double exponential kernels $\epsilon_E(t) = \left(e^{-t/\tau_a^E} - e^{-t/\tau_b^E}\right)/(\tau_a^E - \tau_b^E)$, and $\epsilon_I(t) = \left(e^{-t/\tau_a^I} - e^{-t/\tau_b^I}\right)/(\tau_a^I - \tau_b^I)$ for $t > 0$. In this formulation, for a given target somatic potential v_{teach}, supervised learning by error back-propagation [197] is exactly calculable. By considering stochastic gradient descendent on the squared error $(v_{soma}(t) - v_{teach}(t))$, the learning rule is derived as

$$\Delta w_{k,i}^E \propto \left[v_{teach}(t) - v_{soma}(t)\right] \cdot g(v_k(t)) \cdot \left[1 - g(v_k(t))\right] \cdot \int_0^\infty d\tau \cdot \varepsilon_E(\tau)\, s_{k,i}(t-\tau). \tag{2.5}$$

Note that the rule is written as a function of the somatic potential, the local dendritic potential, and presynaptic activity. Error back-propagation learning is often criticized as biologically implausible, because the learning rule is non-local in the conventional implementation [180], but here we can avoid implausibility by considering branch-dependent plasticity.

For the change detection task, the teaching signal was defined as

$$U_{teach}(t) = \begin{cases} U_H & (\text{if } t - t_j < t_{cd}) \\ U_L & (\text{otherwise}) \end{cases}$$

where t_j represents the timing of the most recent jump in the external variable θ, and t_{cd} is the desired duration of the response. In the excitability maximization task, $U_{teach}(t)$ was fixed at U_H regardless of the external variable.

In the simulation, parameters were set at $\tau_a^E = 1.0$, $\tau_b^E = 5.0$, $\tau_a^I = 1.0$, $\tau_b^I = 10.0$, $\tau_\theta = 1000.0$ ms, $\sigma_\theta = 0.03\sqrt{\Delta t}$, $\beta_E = 4.0$, $\beta_I = 4.0$, $r_x^E = 50.0$, $r_x^I = 200.0$, $\alpha_g = 4.0$, $\beta_g = 1.0$, $w_o^I = 10.0$, and initial value of excitatory weights were set at $w_{k,i}^E = 2.0$. Stimulus selectivity of input neurons $\{\theta_{k,i}^E\}$ and $\{\theta_k^I\}$ were randomly selected from $[-\pi, \pi)$. For supervised signals, I used $U_H = 20.0$, $U_L = 2.0$. In the change detection task, the learning rate was set at $\eta = 0.001$, and for excitability maximization, I used $\eta = 0.0002$. The particular learning rate parameters were chosen to achieve error reduction in a similar timescale at two tasks, and the convergence was robust against parameter choice.

Chapter 3

Wiring Plasticity Generates Efficient Network Structure for Synaptic Plasticity

In the adult mammalian cortex, a small fraction of spines are created and eliminated every day, and the resultant synaptic connection structure is highly nonrandom, even in local circuits. However, it remains unknown whether a particular synaptic connection structure is functionally advantageous in local circuits, and why creation and elimination of synaptic connections is necessary in addition to rich synaptic weight plasticity. To answer these questions, I studied an inference task model through theoretical and numerical analyses. I demonstrate that a robustly beneficial network structure naturally emerges by combining Hebbian-type synaptic weight plasticity and wiring plasticity. Especially in a sparsely connected network, wiring plasticity achieves reliable computation by enabling efficient information transmission. Furthermore, the proposed rule reproduces experimental observed correlation between spine dynamics and task performance.

Introduction

The amplitude of excitatory and inhibitory postsynaptic potentials (EPSPs and IPSPs), often referred to as synaptic weight, is considered a fundamental variable in neural computation [23] [49]. In the mammalian cortex, excitatory synapses often show large variations in EPSP amplitudes [214] [108] [31], and the amplitude of a synapse can be stable over trials [135] and time [247], enabling rich information capacity compared with that at binary synapses [26] [100]. In addition, synaptic weight shows a wide variety of plasticity which depend primarily on the activity of presynaptic and postsynaptic neurons [32] [64]. Correspondingly, previous theoretical results suggest that under appropriate synaptic plasticity, a randomly connected network is computationally sufficient for various tasks [149] [72].

On the other hand, it is also known that synaptic wiring plasticity and the resultant synaptic connection structure are crucial for computation in the brain [42] [105]. Elimination and creation of dendritic spines are active even in the brain of adult mammalians. In rodents, the spine turnover rate is up to 15% per day in sensory cortex [104] and 5% per day in motor cortex [255]. Recent studies further revealed that spine dynamics are tightly correlated with the performance of motor-related tasks [245] [244]. Previous modeling studies suggest that wiring plasticity helps memory storage [188] [215] [126]. However, in those studies, EPSP amplitude was often assumed to be a binary variable, and wiring plasticity was performed in a heuristic manner. Thus it remains unknown what should be encoded by synaptic connection structure when synaptic weights have a rich capacity for representation, and how such a connection structure can be achieved through a local spine elimination and creation mechanism, which is arguably noisy and stochastic [116].

To answer these questions, I constructed a theoretical model of an inference task. I first studied how sparse connectivity affects the performance of the network by analytic consideration and information theoretic evaluations. Then, I investigated how synaptic weights and connectivity should be organized to perform robust inference, especially under the presence of variability in the input structure. Based on these insights, I proposed a local unsupervised rule for wiring and synaptic weight plasticity. In addition, I demonstrated that connection structure and synaptic weight learn different components under a dynamic environment, enabling robust computation. Lastly, I investigated whether the model is consistent with various experimental results on spine dynamics.

Results

Connection structure reduces signal variability in sparsely connected networks

What should be represented by synaptic connections and their weights, and how are those representations acquired? To explore the answers to these questions, I studied a hidden variable estimation task (Fig. 3.1A), which appears in various stages of neural information processing [17] [144]. In the task, at every time t, one hidden state is sampled with equal probability from p number of external states $s^t = \{0, 1, ..., p-1\}$. Neurons in the input layer show independent stochastic responses $r^t_{X,j}\ \ N(\theta_{j\mu}, \sigma_X)$ due to various noises (Fig. 3.1B middle), where $\theta_{j\mu}$ is the average firing rate of neuron j to the stimulus μ, and σ_X is the constant noise amplitude. Although, I used Gaussian noise for analytical purposes, the following argument is applicable for any stochastic response that follows a general exponential family, including Poisson firing (Supplementary Fig. 1). Neurons in the output layer estimate the hidden variable from input neuron activity and represent the variable with population firing $\{r_{Y,i}\}$. This task is computationally difficult because most input neurons have mixed selectivity for several hidden inputs, and the responses of the input neurons are highly stochastic (Fig. 3.1C). Let

me assume that the dynamics of output neurons are written as follows:

$$r^t_{Y,i} = r^o_Y \exp\left[\sum_{j=1}^{M} c_{ij}\left(w_{ij} r^t_{X,j} - h_w\right) - I^t_{inh}\right], \quad I^t_{inh} = \log\left[\sum_{i=1}^{N} \exp\left(\sum_{j=1}^{M} c_{ij}\left[w_{ij} r^t_{X,j} - h_w\right]\right)\right] \tag{3.1}$$

where c_{ij} (= 0 or 1) represents connectivity from input neuron j to output neuron i, w_{ij} is its synaptic weight (EPSP size), and h_w is the threshold. M and N are population sizes of the input and output layers, respectively. In the model, all feedforward connections are excitatory, and the inhibitory input is provided as the global inhibition I^t_{inh}.

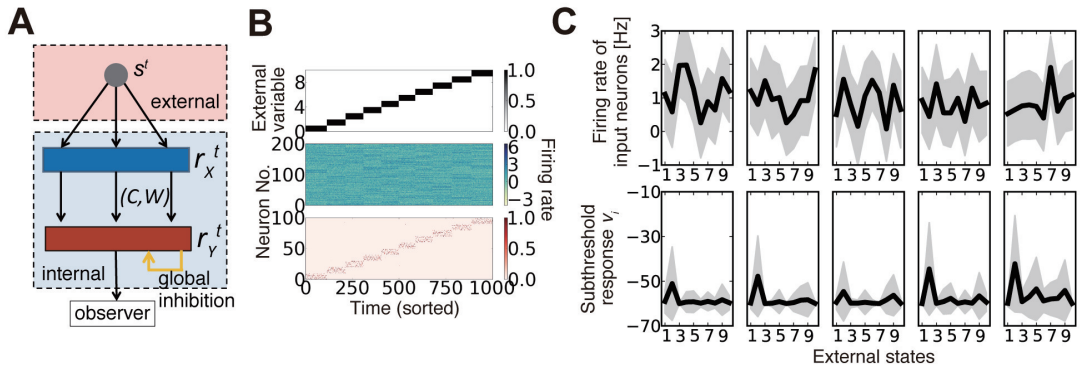


Figure 3.1. Description of the model. (**A**) Schematic diagram of the model. (**B**) An example of model behavior calculated at $\rho = 0.16$, when the synaptic connection is organized using the weight-coding scheme. The top panel represents the external variable, which takes an integer 0 to 9 in the simulation. The middle panel is the response of input neurons, and the bottom panel shows the activity of output neurons. In the simulation, each external state was randomly presented, but here the trials are sorted in ascending order. (**C**) Examples of neural activity in a simulation. Graphs on the top row represent the average firing rates of five randomly sampled input neurons for given external states (black lines) and their standard deviation (gray shadows). The bottom graphs are subthreshold responses of output neurons that represent the external state $s = 1$. Because the boundary condition for the membrane parameter $v_i \equiv \sum_j c_{ij}(w_{ij} r^t_{X,j} - h_w)$ was introduced as $v_i > \max_{i'}\{v_{i'} - v_d\}$, v_i is typically bounded at $-v_d$. Note that v_i is the unnormalized log-likelihood, and the units on the y-axis are arbitrary.

If the feedforward connection is all-to-all (i.e., $c_{ij} = 1$ for all i, j pairs), by setting the weights as $w_{ij} = q_{j\mu} \equiv \theta_{j\mu}/\sigma^2_X$ for output neuron i that represents external state μ, the network gives an optimal inference from the given firing rate vector r^t_X, because the value $q_{j\mu}$ represents how much evidence the firing rate of neuron j provides for a particular external state μ. (For details, see Methods 1.1). However, if the connectivity between the two layers is sparse, as in most regions of the brain [190], optimal inference is generally unattainable because each output neuron can obtain a limited set of information from the input layer. How should one choose connection structure and synaptic weights in such a case? Intuitively, we could expect that if we randomly eliminate connections while keeping the synaptic weights of output neuron i that represents external state μ as $w_{ij} \propto q_{j\mu}$ (below, I call it as weight coding), the network still works at a near-optimal accuracy. On the other hand, even if the synaptic weight is a constant value, if the connection probability is kept at $\rho_{ij} \propto q_{j\mu}$ (i.e. connectivity coding; see Methods 1.2 for details of coding strategies), the network is expected to achieve near-optimal performance. Figure 3.2A describes the connection matrices between input/output layers in

two strategies. In the weight coding, if we sort input neurons with their preferred external states, the diagonal components of the connection matrix show high synaptic weights, whereas in the connectivity coding, the diagonal components show dense connection (Fig. 3.2A). Both of realizations asymptotically converge to optimal solution when the number of neurons in the middle layer is sufficiently large, though in a finite network, not strictly optimal under given constraints. In addition, both of them are obtainable through biologically plausible local Hebbian learning rules as I demonstrate in subsequent sections.

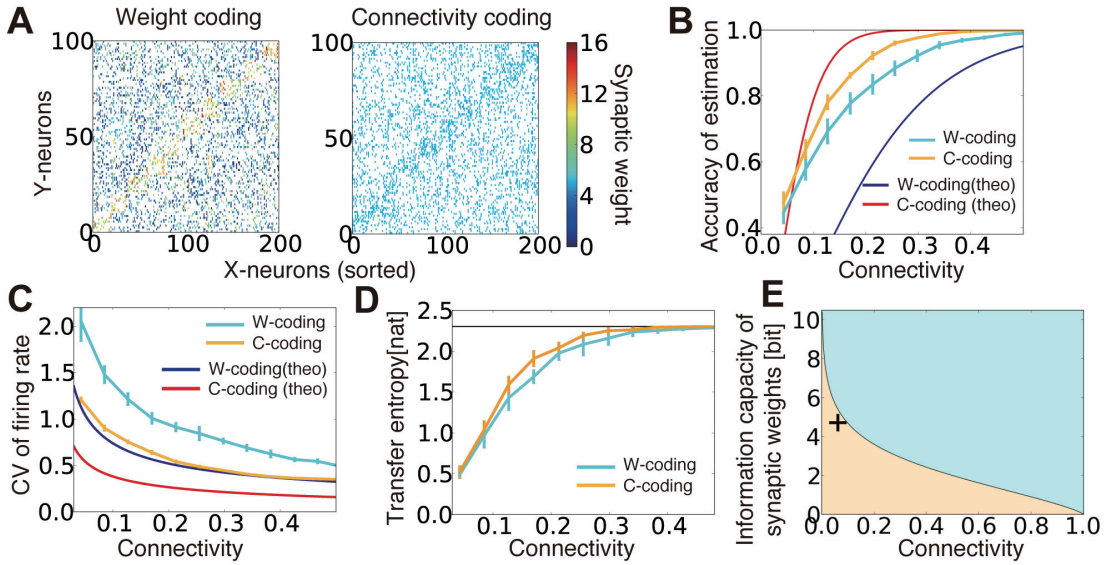


Figure 3.2. Performance comparison between connectivity coding and weight coding. (**A**) Examples of synaptic weight matrices in weight coding (W-coding) and connectivity coding (C-coding) schemes. X-neurons were sorted by their selectivity for external states. (**B**) Comparison of the performance between connectivity coding and weight coding schemes at various sparseness of connectivity. Orange and cyan lines are simulation results. The error bars represent standard deviation over 10 independent simulations. In the following panels, error bars are trial variability over 10 simulations. Red and blue lines are analytical results. (**C**) Analytically evaluated coefficient of variation (CV) of output firing rate and corresponding simulation results. For simulation results, the variance was evaluated over whole output neurons from their firing rates for their selective external states. (**D**) Estimated maximum transfer entropy for two coding strategies. Black horizontal line is the maximal information $\log_e p$. (**E**) Relative information capacity of connection structure versus synaptic weight is shown at various values of synaptic connectivity. In the orange (cyan) area, the synaptic connectivity has higher (lower) information capacity than the synaptic weights. Plus symbol represents the data point obtained from CA3-to-CA1 connections.

I evaluated the accuracy of the external state estimation using a bootstrap method (Methods 3.2) for both coding strategies. Under intermediate connectivity, both strategies showed reasonably good performance (as in Fig. 3.1B bottom). Intriguingly, in sparsely connected networks, the connectivity coding outperformed the weight coding, despite its binary representation (Fig. 3.2B cyan/orange lines). The analytical results confirmed this tendency (Fig. 3.2B red/blue lines; see Methods 2.1 for the details) and indicated that the firing rates of output neurons selective for the given external state show less variability in connectivity coding than in the weight coding, enabling more reliable information transmission (Fig. 3.2C). To further understand this phenomenon, I evaluated the maximum transfer entropy of the feed forward connections: $T_E = \langle H(s^t) - H(s^t|r_X^t, C)\rangle_t$. Because of limited connectivity, each output neuron obtains information only from the connected input neurons. Thus, the transfer

entropy was typically lower under sparse than under dense connections in both strategies (Fig. 3.2D). However, in the connectivity coding scheme, because each output neuron can get information from relevant input neurons, the transfer entropy became relatively large compared to the weight coding (orange line in Fig. 3.2D). Therefore, analyses from both statistical and information theory-based perspectives confirm the advantage of connectivity coding over the weight coding in the sparse regions.

The result above can also be extended to arbitrary feedforward network as below. For a feedforward network of M times N neurons with connection probability ρ, information capacity of connections is given as $I_C(\rho) \equiv \log_{MN} C_{\rho MN}$, where H represents the entropy function . Similarly, for a given connections between two layers, information capacity of synaptic weights is written as $H(\rho) \equiv -\rho \log \rho - (1-\rho)\log(1-\rho)$, where b is the number of distinctive synaptic states [227]. Therefore, when the connection probability ? satisfies , synaptic connections and weights have the same information capacities. This means that, as depicted in Figure 3.2E, in a sparsely connected network, synaptic connections tend to have larger relative information capacity, compared to a dense network with the same b. This result is consistent with the model above, because stochastic firing of presynaptic neuron can be translated as synaptic noise. Furthermore, in the CA3-to-CA1 connection of mice, connection probability is estimated to be around 6% [204], and information capacity of synaptic weight is around 4.7 bits [15], thus the connection structure should also play an active role in neural coding in the real brain (data point in Fig. 3.2E).

Dual coding by synaptic weights and connections enables robust inference

In the section above, I demonstrated that a random connection structure highly degrades information transmission in a sparse regime to the degree that weight coding with random connection fell behind connectivity coding with a fixed weight. Therefore, in a sparse regime, it is necessary to integrate representations by synaptic weights and connections, but how should we achieve such a representation? Theoretically speaking, we should choose a connection structure that minimizes the loss of information due to sparse connectivity. This can be achieved by minimizing the KL-divergence between the distribution of the external states estimated from the all-to-all network, and the distribution estimated from a given connection structure (i.e. $\underset{\|C\|_0=\rho MN}{\arg\min} \langle D_{KL}\left[p(s^t|r_X, C_{all})||p(s^t|r_X, C)\right]\rangle_{r_X}$, see Methods 2.2 for details). However, this calculation requires combinatorial optimization, and local approximation is generally difficult [57], thus expectedly the brain employs some heuristic alternatives. Experimental results indicate that synaptic connections and weights are often representing similar features. For example, the EPSP size of a connection in a clustered network is typically larger than the average EPSP size [135] [184], and a similar property is suggested to hold for interlayer connections [250] [191]. Therefore, we could expect that by simply combining the weight coding and connectivity coding in the previous section, low performance at the sparse regime can be avoided. On the other hand, in the previous modeling studies, synaptic rewiring and resultant connection structure were often gen-

erated by cut-off algorithm in which a synapse is eliminated if the weight is smaller than the given criteria [35] [170]. Thus, let us next compare the representation by combining the weight coding and connectivity coding (I call it as the dual coding below), with the cut-off coding strategy.

Figure 3.3A describes the synaptic weight distributions in the two strategies, as well as in random connection (see Methods 1.3 for details of the implementation). When connectivity coding and weight coding are combined (i.e. in the dual coding), connection probability becomes larger in proportion to its synaptic weight (Fig. 3.3A middle), and the resultant distribution exhibits a broad distribution as observed in the experiments [214] [108], whereas in the cut-off strategy, the weight distribution is concentrated at a non-zero value (Fig. 3.3A right). Intuitively, the cut-off strategy seems more selective and beneficial for inference. Indeed, in the original task, the cut-off strategy enabled near-optimal performance, though the dual coding also improved the performance compared to a randomly connected network(Fig. 3.3C). However, under the presence of variability in the input layer, cut-off strategy is no longer advantageous. For instance, let me consider the case when noise amplitude σ_X is not constant but pre-neuron dependent. If the firing rate variability of input neuron j is given by $\sigma_{X,j} \equiv \sigma_X \exp\left(2\zeta_j \log \sigma_r\right)/\sigma_r$, where ζ_j is a random variable uniformly sampled from [0, 1), and σ_r is the degree of variability, in an all-to-all network, optimal inference is still achieved by setting synaptic weights as $w_{ij} = q_{j\mu} \equiv \theta_{j\mu}/\sigma^2_{X,j}$. On the contrary, in the sparse region, the performance is disrupted especially in the cut-off strategy, so that the dual coding outperformed the cut-off strategy (Fig. 3.3D).

To further illustrate this phenomenon, let us next consider a case when a quarter of input neurons show a constant high response for all of the external states as $\tilde{\theta}_{j\mu} = \theta_{const}$ and the rest of input neurons show high response for randomly selected half of external states (i.e. $\Pr\left[\tilde{\theta}_{j\mu} = \theta_{high}\right] = \Pr\left[\tilde{\theta}_{j\mu} = \theta_{low}\right] = \frac{1}{2}$), where $\theta_{low} < \theta_{high} < \theta_{const}$, and $\theta_{j\mu} = \tilde{\theta}_{j\mu}/Z_\mu$ with the normalization factor $Z_\mu = r^o_X \Big/ \sqrt{\sum_{j=1}^{M} \tilde{\theta}_{j\mu} \Big/ M}$. Even in this case, $w_{ij} = q_{j\mu} \equiv \theta_{j\mu}/\sigma^2_X$ is the optimal synaptic weights configuration in the all-to-all network, but if we create a sparse network with cut-off algorithm, the performance drops dramatically at certain connectivity, whereas in the dual coding, the accuracy is kept at some high levels even in the sparse connectivity (Fig. 3.3E).

To get insights on why the dual coding is more robust against variability in the input layer, for three input configurations described above, I calculated the relationship between synaptic weight w_{ij} and the information gained by a single synaptic connection ΔI_{ij}. Here, I defined the information gain ΔI_{ij} by the mean reduction in the KL divergence $D_{KL}\left[p(s^t|r_X, C_{all})||p(s^t|r_X, C)\right]$, achieved by adding one synaptic connection c_{ij} to a randomly connected network C (see Method 2.2 for details). In the original model, ΔI_{ij} has nearly a linear relationship with the synaptic weight w_{ij} (gray points in Fig. 3.3B), thus by simply removing the connections with small synaptic weights, a near-optimal connection structure was acquired (Fig. 3.3C). On the other hand, when the input layer is not homogeneous, large synapses tend to have negative (black circles in Fig. 3.3B) or zero (black points in Fig. 3.3B) gains, as

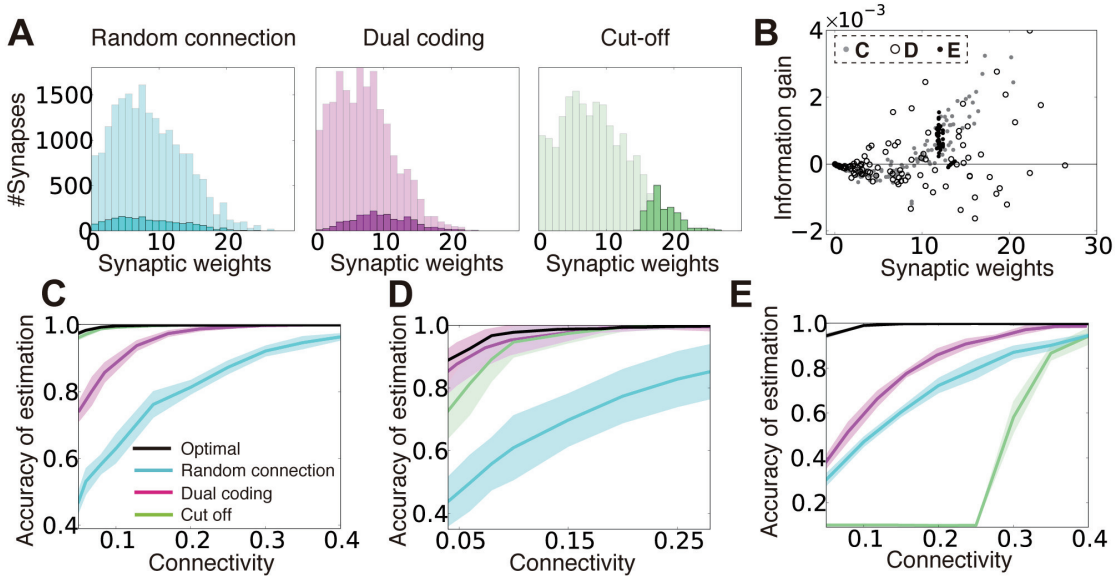


Figure 3.3. Dual coding yields robust information representation compared to fixed random connections and cut-off strategy. (**A**) Synaptic weight distributions in random connection (left), dual coding (middle), and cut-off (right) strategies. Light colors represent possible connections (i.e. distributions of synaptic weights under all-to-all connections), while dark colors show the actual connections. Connection probability was set at $\rho = 0.1$. (**B**) Relationships between the synaptic weight and the information gain per connection for three input configurations described in panels **C**-**E**. The open black circles were calculated with $\sigma_r = 2.0$ instead of $\sigma_r = 4.0$ for illustration purpose. (**C**-**E**) Comparisons of performance among different connection structure organizations. Note that black lines represent lower bounds for the optimal performance, but not the exact optimal solutions. In panel **D**, the means and standard deviations were calculated over 100 simulation trials instead of 10 due to intrinsic variability.

---

a result, the linear relationship between the weight and the information gain was lost. Thus, in these cases, the dual coding is less likely to be disrupted by non-beneficial connections.

Although my consideration here is limited to a specific realization of synaptic weights, in general, it is difficult to represent the information gain by locally acquired synaptic weight, so we could expect that the cut-off strategy is not the optimal connectivity organization in many cases.

Local Hebbian learning of the dual coding

The argument in the previous section suggest that, by combining the weight coding and connectivity coding, the network can robustly perform inference especially in sparsely connected regions. However, in the previous sections, a specific connection and weight structure were given a priori, although structures in local neural circuits are expected to be obtained with local weight plasticity and wiring plasticity. Thus, I next investigate whether dual coding can be achieved through a local unsupervised synaptic plasticity rule.

Let us first consider learning of synaptic weights. In order to achieve the weight coding, synaptic weight w_{ij} should converge to $w_{ij} = q_{j\mu}/\sigma_X^2\bar{\rho} = \langle r_{X,j}^t r_{Y,i}^t / \left(\sigma_X^2 \bar{\rho} r_{Y,i}^t\right)\rangle$ when output neuron i represents external state μ, and $\bar{\rho}$ represents the mean connectivity of the network. Thus, synaptic weight change

$\Delta w_{ij} = w_{ij}^{t+1} - w_{ij}^{t}$ is given as:

$$\Delta w_{ij} = (\eta_X/\gamma)\left(r_{Y,i}^t\left[r_{X,j}^t - \sigma_X^2\bar{\rho}w_{ij}\right] + b_h\left[r_Y^o/N - r_{Y,i}^t\right]\right). \tag{3.2}$$

The second term is the homeostatic term heuristically added to constrain the average firing rates of output neurons [224]. Note that the first term corresponds to stochastic gradient descending on $D_{KL}\left[p^*(r_X^t)||p(r_X^t|C,W)\right]$, because the weight coding approximates the optimal representation by synaptic weights [171](see Methods 1.4 for details). I performed this unsupervised synaptic weight learning on a randomly connected network. When the connectivity is sufficiently dense, the network successfully acquired a suitable representation (Fig. 3.4A). Especially under a sufficient level of homeostatic plasticity (Fig. 3.4B), the average firing rate showed a narrow unimodal distribution (Fig. 3.4C top), and most of the output neurons acquired selectivity for one of external states (Fig. 3.4C bottom).

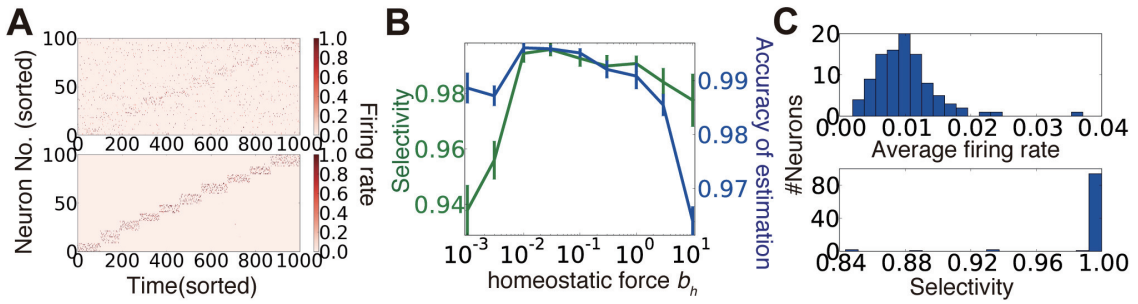


Figure 3.4. Synaptic weight learning on random connection structures. (**A**) An example of output neuron activity before (top) and after (bottom) synaptic weight learning calculated at connectivity $\rho = 0.4$. (**B**) Selectivity of output neurons and accuracy of estimation at various strengths of homeostatic plasticity at $\rho = 0.4$. Selectivity was defined as $\sum_{s^t=\mu} r_{Y,i}^t / \sum_t r_{Y,i}^t$ for $i \in \Omega_\mu$. (**C**) Histogram of average firing rates of output neurons (top), and selectivity of each neuron calculated for the simulation depicted in panel **A**.

I next investigated the learning of connection structures by wiring plasticity. Unlike synaptic weight plasticity, it is not yet well understood how we can achieve functional connection structure with local wiring plasticity. In particular, rapid rewiring may disrupt the network structure, and possibly worsen the performance [35]. Thus, let us first consider a simple rewiring rule, and discuss the biological correspondence later. Here, I introduced a variable ρ_{ij}, for each combination (i,j) of presynaptic neuron j and postsynaptic neuron i, which represents the connection probability. If we randomly create a synaptic connection between neuron (i,j) with probability ρ_{ij}/τ_c and eliminate it with probability $(1-\rho_{ij})/\tau_c$, on average there is a connection between neuron (i,j) with probability ρ_{ij}, when the maximum number of synaptic connections is bounded by 1. In this way, the total number of synaptic connections is kept constant on average, without any global regulation mechanism.

From a similar argument done for synaptic weights, the learning rule for connection probability ρ_{ij} is derived as:

$$\Delta\rho_{ij} = \eta_\rho r_{Y,i}^t\left[r_{X,j}^t - \sigma_X^2\rho_{ij}w_o\right], \tag{3.3}$$

where w_o is the expected mean synaptic weight (Methods 1.5). Under this rule, the connection probabilities converge to the connectivity coding. Moreover, although this rule does not maximize the transfer entropy of the connections, direction of learning is on average close to the direction of the stochastic gradient on transfer entropy. Therefore, the above rule does not reduce the transfer entropy of the connection on average (see Methods 1.6).

Figure 3.5A shows the typical behavior of ρ_{ij} and w_{ij} under combination of this wiring rule (equation (3)) and the weight plasticity rule described in equation (2) (I call this combination as the dual Hebbian rule because both equations (2) and (3) have Hebbian forms). When the connection probability is low, connections between two neurons are rare, and, even when a spine is created due to probabilistic creation, the spine is rapidly eliminated (Fig. 3.5A top). In the moderate connection probability, spine creation is more frequent, and the created spine survives longer (Fig. 3.5A middle). When the connection probability is high enough, there is almost always a connection between two neurons, and the synaptic weight of the connection is large because synaptic weight dynamics also follow a similar Hebbian rule (Fig. 3.5A bottom).

I implemented the dual Hebbian rule in my model and compared the performance of the model with that of synaptic weight plasticity on a fixed random synaptic connection structure. Because spine creation and elimination are naturally balanced in the proposed rule (Fig. 3.5B top), the total number of synaptic connections was nearly unchanged throughout the learning process (Fig. 3.5B bottom). As expected, the dual Hebbian rule yielded better performance (Fig. 3.5C,D) and higher estimated transfer entropy than the corresponding weight plasticity only model (Fig. 3.5E). This improvement was particularly significant when the frequency of rewiring was in an intermediate range (Fig. 3.5F). When rewiring was too slow, the model showed essentially the same behavior as that in the weight plasticity only model, whereas excessively frequent probabilistic rewiring disturbed the connection structure. Although a direct comparison with experimental results is difficult, the optimal rewiring timescale occurred within hours to days, under the assumption that firing rate dynamics (equation (1)) are updated every 10 to 100 ms. Initially, both connectivity and weights were random (Fig. 3.5G left), but after the learning process, the diagonal components of the weight matrix developed relatively larger synaptic weights, and, at the same time, denser connectivity than the off-diagonal components (Fig. 3.5G right). Thus, through dual Hebbian learning, the network can indeed acquire a connection structure that enables efficient information transmission between two layers; as a result, the performance improves when the connectivity is moderately sparse (Fig. 3.5D, E). Although the performance was slightly worse than that of a fully-connected network, synaptic transmission consumes a large amount of energy [206], and synaptic connection is a major source of noise [62]. Therefore, it is beneficial for the brain to achieve a similar level of performance using a network with fewer connections.

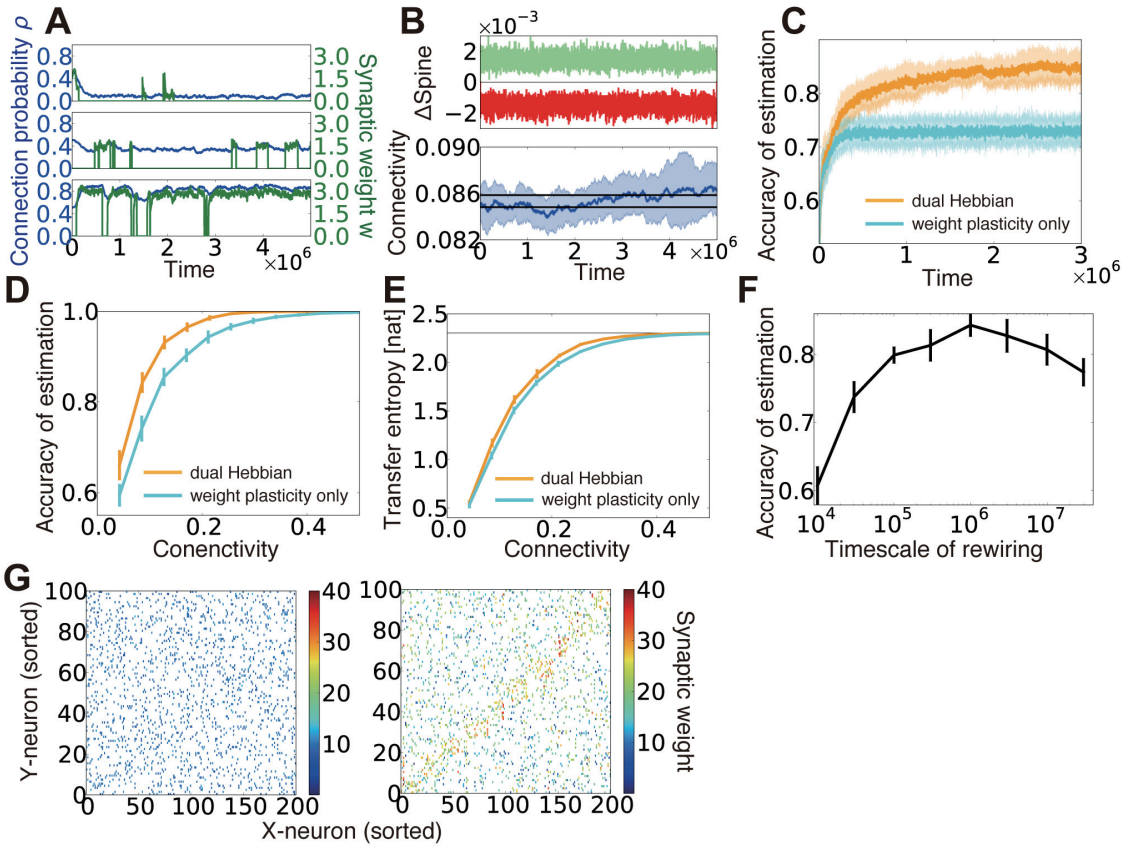


Figure 3.5. Dual Hebbian learning for synaptic weights and connections. (**A**) Examples of spine creation and elimination. In all three panels, green lines show synaptic weights, and blue lines are connection probability. When there is not a synaptic connection between two neurons, the synaptic weight becomes zero, but the connection probability can take a non-zero value. Simulation was calculated at $\rho = 0.48$, $\eta_\rho = 0.001$, and $\tau_c = 10^5$. (**B**) Change in connectivity due to synaptic elimination and creation. Number of spines eliminated (red) and created (green) per unit time was balanced (top). As a result, connectivity did not appreciably change due to rewiring (bottom). Black lines in the bottom graph are the mean connectivity at $\gamma = 0.1$ and $\gamma = 0.101$ in the model without rewiring. (**C**) Accuracy of estimation for the model with/without wiring plasticity. For the dual Hebbian model, the sparseness parameter was set as $\gamma = 0.1$, whereas $\gamma = 0.101$ was used for the weight plasticity model to perform comparisons at the same connectivity (see panel B). (**D, E**) Comparison of the performance (**D**) and the maximum estimated transfer entropy (**E**) after learning between the dual Hebbian model and the model implemented with synaptic plasticity only at various degrees of connectivity. Horizontal line in panel **E** represents the total information $\log_e p$. (**F**) Accuracy of estimation with various timescales for rewiring τ_c. Note that the simulation was performed only for 5×10^6 time steps, and the performance did not converge for the model with a longer timescale. (**G**) Synaptic weight matrices before (left) and after (right) learning. Both X-neurons (input neuron) and Y-neurons (output neurons) were sorted based on their preferred external states.

Connection structure can acquire constant components of stimuli and enable rapid learning

I have shown that the dual coding by synaptic weights and connections robustly helps computation in a sparsely connected network, and the desirable weight and connectivity structures are naturally acquired through the dual Hebbian rule. Although I was primary focused on sparse regions, the rule potentially provides some beneficial effects even in densely connected networks. To consider this issue, I extended the previous static external model to a dynamic one, in which at every interval T_2, response probabilities of input neurons partly change. If we define the constant component as θ_{const} and the variable component as θ_{var}, then the total model becomes $\theta_{j\mu} = \frac{1}{Z}\left[\kappa_m \theta_{j\mu}^{const} + (1-\kappa_m)\theta_{j\mu}^{var}\right]$, where the normalization term is given as $\frac{1}{MZ^2}\sum_{j=1}^{M}\left[\kappa_m \theta_{j\mu}^{const} + (1-\kappa_m)\,\theta_{j\mu}^{\mathrm{var}}\right]^2 = \left(r_X^o\right)^2$ (Fig. 3.6A). In this case, when the learning was performed only with synaptic weights based on fixed random connections, although the performance rapidly improved, every time a part of the model changed, the performance dropped dramatically and only gradually returned to a higher level (cyan line in Fig. 3.6B). By contrast, under the dual Hebbian learning rule, the performance immediately after the model shift (i.e., the performance at the trough of the oscillation) gradually increased, and convergence became faster (Fig. 3.6B,C), although the total connectivity stayed nearly the same (Fig. 3.6D). After learning, the synaptic connection structure showed a higher correlation with the constant component than with the variable component (Fig. 3.6E; see Methods 3.3). By contrast, at every session, synaptic weight structure learned the variable component better than it learned the constant component (Fig. 3.6F). The timescale for synaptic rewiring needed to be long enough to be comparable with the timescale of the external variability T_2 to capture the constant component. Otherwise, connectivity was also strongly modulated by the variable component of the external model (Fig. 3.6G). After sufficient learning, the synaptic weight w and the corresponding connection probability ρ roughly followed a linear relationship (Fig. 3.6H). Remarkably, some synapses developed connection probability $\rho = 1$, meaning that these synapses were almost permanently stable because the elimination probability $(1-\rho)/\tau_c$ became nearly zero.

Approximated dual Hebbian learning rule reconciles with experimentally observed spine dynamics

My results up to this point have revealed functional advantages of dual Hebbian learning. In this last section, I investigated the correspondence between the experimentally observed spine dynamics and the proposed rule. To this end, I first studied whether a realistic spine dynamics rule approximates the proposed rule, and then examined if the rule explains the experimentally known relationship between synaptic rewiring and motor learning [245] [244].

Previous experimental results suggest that a small spine is more likely to be eliminated [247] [116], and spine size often increases or decreases in response to LTP or LTD respectively, with a certain

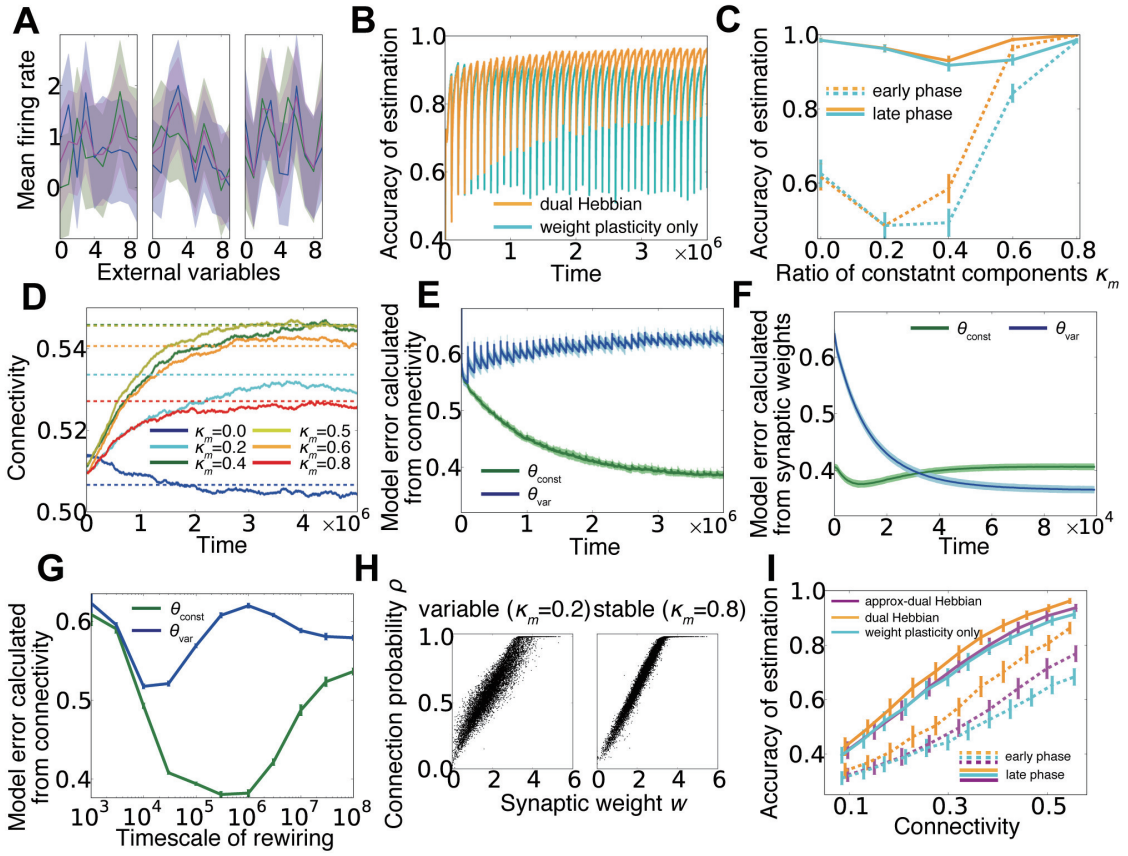


Figure 3.6. Dual learning under a dynamic environment. (**A**) Examples of input neuron responses. Blue lines represent the constant components θ_{const}, green lines show the variable components θ_{var}, and magenta lines are the total external models θ calculated from the normalized sum. (**B**) Learning curves for the model with or without wiring plasticity, when the variable components change every 10^5 time steps. (**C**) Accuracy of estimation for various ratios of constant components. Early phase performance was calculated from the activity within 10,000 steps after the variable component shift, and the late phase performance was calculated from the activity within 10,000 steps before the shift. As in panel **B**, orange lines represent the dual Hebbian model, and cyan lines are for the model with weight plasticity only. (**D**) Trajectories of connectivity change. Connectivity tends to increase slightly during learning. Dotted lines are mean connectivity at (κ_m, γ) = (0.0, 0.595), (0.2, 0.625), (0.4, 0.64), (0.5, 0.64), (0.6, 0.635), and (0.8, 0.620). In panel **C**, these parameters were used for the synaptic plasticity only model, whereas γ was fixed at $\gamma = 0.6$ for the dual Hebbian model. (**E,F**) Model error calculated from connectivity (**E**) and synaptic weights (**F**). Note that the timescale of panel F is the duration in which the variable component is constant, not the entire simulation (i.e. the scale of x-axis is 10^4 not 10^6). (**G**) Model error calculated from connectivity for various rewiring timescales τ_c. For a large τ_c, the learning process does not converge during the simulation. (**H**) Relationship between synaptic weight w and connection probability ρ at the end of learning. When the external model is stable, w and ρ have a more linear relationship than that for the variable case. (**I**) Comparison of performances among the model without wiring plasticity (cyan), the dual Hebbian model (orange), the approximated model (magenta).

delay [160] [240]. In addition, though spine creation is to some extent influenced by postsynaptic activity [127] [246], the creation is expected to be more or less a random process [105]. Thus, changes in the connection probability can be described as

$$\rho_{ij}^{t} = \begin{cases} \rho_{ij}^{t-1} + \eta_\rho \left[\gamma^2 w_{ij} - \rho_{ij}^{t-1}\right] \ (\mathrm{if} c_{ij} = 1) \\ \gamma^2 w_o \ (\mathrm{if} c_{ij} = 0) \, . \end{cases} \tag{3.4}$$

By combining this rule and the Hebbian weight plasticity described in equation (2), the dynamics of connection probability well replicated the experimentally observed spine dynamics [247] [116] (Fig. 3.7A-C). Moreover, the rule outperformed the synaptic weight only model in the inference task, although the rule performed poorly compared to the dual Hebbian rule due to the lack of activity dependence in spine creation (magenta line in Fig. 3.6I). This result suggests that plasticity rule by equations (2) and (4) well approximates the dual Hebbian rule (equations (2)+(3)). This is because, even if the changes in the connection probability are given as a function of synaptic weight as in equation (4), as long as the weight plasticity rule follows equation (2), wiring plasticity indirectly shows a Hebbian dependency for pre- and postsynaptic activities as in the original dual Hebbian rule (equation (3)). As a result, the approximated rule gives a good approximation of the original dual Hebbian rule.

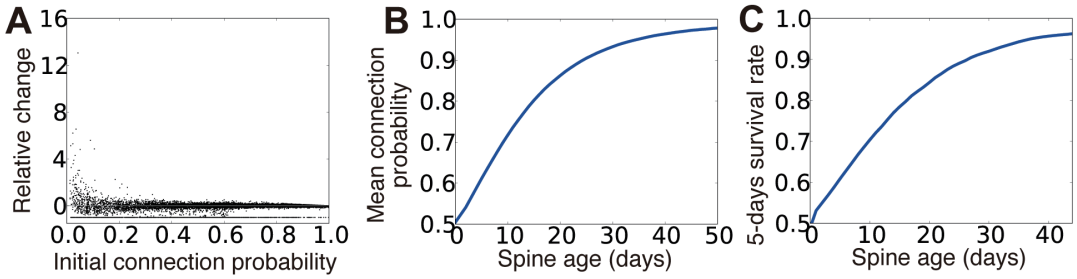


Figure 3.7. Spine dynamics of the approximated dual Hebbian model. (**A**) Relative change of connection probability in 10^5 time steps. If the initial connection probability is low, the relative change after 10^5 time steps has a tendency to be positive, whereas spines with a high connection probability are more likely to show negative changes. The line at the bottom represents eliminated spines (i.e., relative change = -1). (**B,C**) Relationships between spine age and the mean connection probability (**B**) and the 5-days survival rate (**C**). Consistent with the experimental results, survival rate is positively correlated with spine age. 5-days survival rate was calculated by regarding 10^5 time steps as one day.

I next applied this approximated learning rule to motor learning tasks. The primary motor cortex has to adequately read-out motor commands based on inputs from pre-motor regions [199] [216]. In addition, the connection from layer $2/3$ to layer 5 is considered to be a major pathway in motor learning [156]. Thus I hypothesized that the input and output layers of my model can represent layers $2/3$ and 5 in the motor cortex. I first studied the influence of training on spine survival [244] (Fig. 3.8A). To compare with experimental results, below I regarded 10^5 time steps as one day, and described the training and control phases as two independent external models θ_{ctrl} and θ_{train}. In both training and control cases, newly created spines were less stable than pre-existing spines (solid lines vs. dotted

lines in Fig. 3.8B), because older spines tended to have a larger connection probability (Fig. 3.7B). In addition, continuous training turned pre-existed spines less stable and new spines more stable than their respective counterparts in the control case (red lines vs. lime lines in Fig. 3.8B). The 5-day survival rate of a spine was higher for spines created within a couple of days from the beginning of training compared with spines in the control case, whereas the survival rate converged to the control level after several days of training (Fig. 3.8C). I next considered the relationship between spine dynamics and task performance [245]. For this purpose, I compared task performance at the beginning of the test period among simulations with various training lengths (Fig. 3.8D). Here, I assumed that spine elimination was enhanced during continuous training, as is observed in experiments [245] [244]. The performance was positively correlated with both the survival rate at day 7 of new spines formed during the first 2 days, and the elimination rate of existing spines (left and right panels of Fig. 3.8E). By contrast, the performance was independent from the total ratio of newly formed spines from day 0 to 6 (middle panel of Fig. 3.8E). These results demonstrate that complex spine dynamics are well described by the approximated dual Hebbian rule, suggesting that the brain uses a dual learning mechanism.

Discussion

In this study, I first analyzed how random connection structures impair performance in sparsely connected networks by analyzing the change in signal variability and the transfer entropy in the weight coding and the connectivity coding strategies (Fig. 3.2). Subsequently, I showed that connection structures created by the cut-off strategy are not beneficial under the presence of input variability, due to lack of positive correlation between the information gain and weight of synaptic connections (Fig. 3.3). Based on these insights, I proposed that the dual coding by weight and connectivity structures as a robust representation strategy, then demonstrated that the dual coding is naturally achieved through dual Hebbian learning by synaptic weight plasticity and wiring plasticity (Fig. 3.4, 3.5). I also revealed that, even in a densely connected network in which synaptic weight plasticity is sufficient in terms of performance, by encoding the time-invariant components with synaptic connection structure, the network can achieve rapid learning and robust performance (Fig. 3.6). Even if spine creation is random, the proposed framework still works effectively, and the approximated model with random spine creation is indeed sufficient to reproduce various experimental results (Fig. 3.7, 3.8).

Model evaluation

Spine dynamics depend on the age of the animal [104], the brain region [255], and many molecules play crucial roles [116] [33], making it difficult for any theoretical models to fully capture the complexity. Nevertheless, my simple mathematical model replicated many key features [247] [245] [244] [116]. For instance, small spines often show enlargement, while large spines are more likely to show shrinkage (Fig. 3.7A). Older spines tend to have a large connection probability, which is proportional to spine

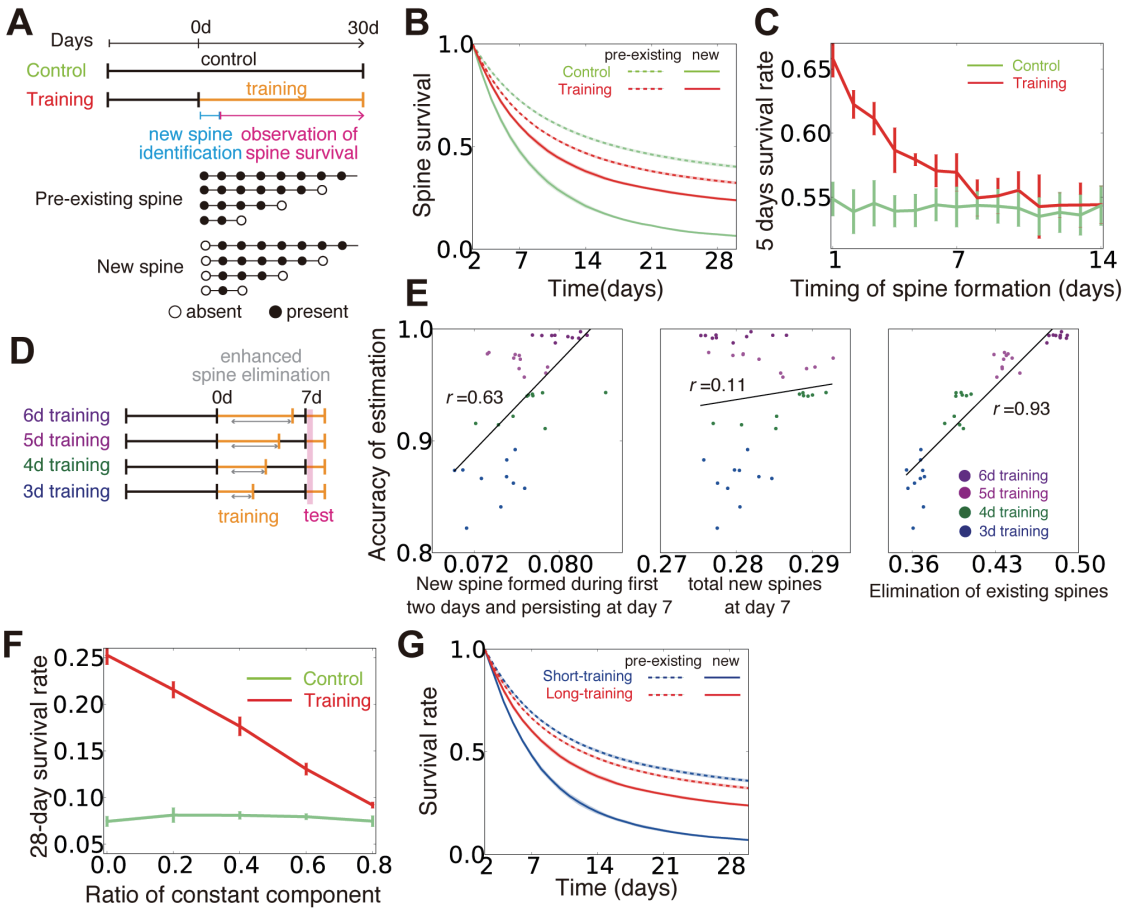


Figure 3.8. Influence of training on spine dynamics. (**A**) Schematic diagrams of the simulation protocols for panels **B,C**, and **F,G**, and examples of spine dynamics for pre-existing spines and new spines. (**B**) Spine survival rates for control and training simulations. Dotted lines represent survival rates of pre-existing spines (spines created before day 0 and existing on day 2), and solid lines are new spines created between day 0 and day 2. (**C**) The 5-day survival rate of spines created at different stages of learning. (**D,E**) Relationships between creation and elimination of spines and task performance. Performance was calculated from the activity within 2,000-7,000 time steps after the beginning of the test phase. In the simulation, the synaptic elimination was increased fivefold from day 1 to the end of training. (**F**) Effect of similarity between the control condition and training on the new spine survival rate. The value of κ_m was changed as in Figure 3.6C to alter the similarity between the two conditions. Note that $\kappa_m = 0$ in panels **A**-**E**, and **G**. (**G**) Spine survival rates for short-training (2 d) and long-training (30 d) simulations. Pre-existing and new spines were defined as in panels **A,B**.

size (Fig. 3.7B), and they are more stable (Fig. 3.7C). In addition, training enhances the stability of newly created spines, whereas it degrades the stability of older spines (Fig. 3.8B).

Experimental prediction

In the developmental stage, both axon guidance [169] and dendritic extension [159] show Hebbian-type activity dependence, but in the adult cortex, both axons and dendrites seldom change their structures [105]. Thus, although recent experimental results suggest some activity dependence for spine creation [127] [246], it is still unclear to what extent spine creation depends on the activity of presynaptic and postsynaptic neurons. My model indicates that in terms of performance, spine creation should fully depend on both presynaptic and postsynaptic activity (Fig. 3.6I). However, I also showed that it is possible to replicate a wide range of experimental results on spine dynamics without activity-dependent spine creation (Fig. 3.8).

Furthermore, whether or not spine survival rate increases through training is controversial [245] [244]. My model predicts that the stability of new spines highly depends on the similarity between the new task and control behavior (Fig. 3.8F). When the similarity is low, new spines created in the new task are expected to be more stable than those created in the control case, because the synaptic connection structure would need to be reorganized. By contrast, when the similarity is high, the stability of the new spines would be comparable to that of the control. In addition, my model replicates the effect of varying training duration on spine stability [245]. When training was rapidly terminated, newly formed spines became less stable than those undergoing continuous training (Fig. 3.8G).

Related studies

Previous theoretical studies revealed candidate rules for spine creation and elimination [50] [254] [63], yet their functional benefits were not fully clarified in those studies. Some modeling studies considered the functional implications of synaptic rewiring [188] or optimality in regard to benefit and wiring cost [37], but the functional significance of synaptic plasticity and the variability of EPSP size were not considered in those models. In comparison, my study revealed functional roles of wiring plasticity that cooperates with synaptic weight plasticity and obeys local unsupervised rewiring rules. In addition, I extended the previous results on single-spine information storage and synaptic noise [227] into a network, and provided a comparison with experimental results (Fig. 3.2E).

Previous studies on associative memory models found the cut-off coding as the optimal strategy for maximizing the information capacity per synapse [35] [126]. My results suggest that the above result is the outcome of the tight positive correlation between the information gain and synaptic weight in associative memory systems, and not generally applicable to other paradigms (Fig. 3.3BC). In addition, although cut-off strategy did not yield biologically plausible synaptic weight distributions in my task setting (Fig. 3.3A right), in perceptron-based models, this unrealistic situation can be avoided by

tuning the threshold of neural dynamics [26] [198]. Especially, cut-off strategy may provide a good approximation for developmental wiring plasticity [129], though the algorithm is not fully consistent with wiring plasticity in the adult animals.

Finally, my model provides a biologically plausible interpretation for multi-timescale learning processes. It was previously shown that learning with two synaptic variables on different timescales is beneficial under a dynamically changing environment [70]. In my model, both fast and slow variables played important roles, whereas in previous studies, only one variable was usually more effective than others, depending on the task context.

Methods

1. Model

1.1 Model dynamics

I first define the model and the learning rule for general exponential family, and derive equations for two examples (Gaussian and Poisson). In the task, at every time t, one hidden state s^t is sampled from prior distribution $p(s)$. Neurons in the input layer show stochastic response $r^t_{X,j}$ that follows probabilistic distribution $f(r_{X,j}|s^t)$:

$$f(r_{X,j}|\mu) \equiv \exp\left[h(\theta_{j\mu})g(r_{X,j}) - A(\theta_{j\mu}) + B(r_{X,j})\right]. \tag{3.5}$$

From these input neuron activities, neurons in output layer estimate the hidden variables. Here I assume maximum likelihood estimation for decision making unit, as the external state is a discrete variable. In this framework, in order to detect the hidden signal, firing rate of neuron i should be proportional to posterior

$$r^t_{Y,i} \propto \Pr\left[s^t = \sigma_i | r^t_X\right]. \tag{3.6}$$

where σ_i represents the index of the hidden variable preferred by output neuron i [17] [144]. Note that $\{r_{X,j}\}$ represent firing rates of input neurons, whereas $\{r_{Y,i}\}$ represent the rates of output neurons. Due to Bayes rule, estimation of s^t is given by,

$$\begin{aligned}\log p(s^t = \mu | r^t_X) &= \sum_{j=1}^{M} \log p(r^t_{X,j}|s^t = \mu) + \log p(s^t = \mu) - \log p(r^t_X) \\ &= \sum_{j=1}^{M} \left[q_{\mu j} g(r^t_{X,j}) - \alpha(q_{\mu j}) + B(r^t_{X,j})\right] + \log p(s^t = \mu) - \log p(r^t_X),\end{aligned} \tag{3.7}$$

where $q_{j\mu} \equiv h(\theta_{j\mu})$, $\alpha(q_{j\mu}) \equiv A\left(h^{-1}(q_{j\mu})\right)$. If I assume the uniformity of hidden states as $\log p(s^t = \mu)$: const and $\frac{1}{M}\sum_{j=1}^{M}\alpha(q_{j\mu}) = \alpha_o$, the equation above becomes

$$\log p(s^t = \mu | r_X^t) = \sum_{j=1}^{M}\left[q_{\mu j} g(r_{X,j}^t) + B(r_{X,j}^t)\right] - \log p(r_X^t) + \text{const.}$$

To achieve neural implementation of this inference problem, let us consider a neural dynamics in which the firing rates of output neurons follow,

$$r_{Y,i}^t = r_Y^o \exp\left[\sum_{j=1}^{M} c_{ij}\left(w_{ij} g(r_{X,j}^t) - h_w\right) - I_{inh}^t\right], \tag{3.8}$$

where,

$$I_{inh}^t \equiv \log\left[\sum_{i=1}^{N}\exp\left(\sum_{j=1}^{M} c_{ij}\left[w_{ij} g(r_{X,j}^t) - h_w\right]\right)\right],$$

and h_w is the threshold. If connection is all-to-all, $w_{ij} = q_{j\mu}$ gives optimal inference, because

$$\frac{r_{Y,i}^t}{r_Y^o} = \frac{\exp\left[\sum_j q_{j\mu} g(r_{X,j}^t)\right]}{\sum_\nu \exp\left[\sum_j q_{j\nu} g(r_{X,j}^t)\right]} = p(s^t = \mu | r_X^t) \tag{3.9}$$

Note that h_w is not necessary to achieve optimal inference, however, under a sparse connection, h_w is important for reducing the effect of connection variability. In this formalization, even in non-all-to-all network, if the sparseness of connectivity stays in reasonable range, near-optimal inference can be performed for arbitrary feedforward connectivity by adjusting synaptic weight to $w_{ij} = w_{\mu j} \equiv q_{j\mu}/\rho_{\mu j}$ where $\rho_{\mu j} = \frac{1}{|\Omega_\mu|}\sum_{i\in\Omega_\mu} c_{ij}$.

1.2. Weight coding and connectivity coding

Let us first consider the case when the connection probability is constant (i.e. $\rho_{ij} = \rho$). By substituting $\rho_{ij} = \rho$ into the above equations, c and w are given with $\Pr[c_{ij} = 1] = \rho$ and $w_{ij} = w_{\mu j} = q_{j\mu}/\rho$, where the mean connectivity is given as $\rho = \gamma\bar{q}$, and $barq$ is the average of the normalized mean response $q_{j\mu}$ (i.e., $\bar{q} = \frac{1}{Mp}\sum_j\sum_\mu q_{j\mu}$). Parameter γ is introduced to control the sparseness of connections, and here I assumed that neuron i represents the external state $\mu = \text{floor}(\frac{p\times i}{N})$(i.e., if $\frac{\mu N}{p} < i \leq \frac{(\mu+1)N}{p}$, output neuron i represents the state μ). Under this configuration, the representation is solely achieved by the synaptic weights, thus I call this coding strategy as the weight coding.

On the other hand, if the synaptic weight is kept at a constant value, the representation is realized by synaptic connection structure (i.e. connectivity coding). In this case, the model is given by $\Pr[c_{ij} = 1] = \rho_{\mu j}$ and $w_{ij} = w_{\mu j} = 1/\gamma$, where $\rho_{\mu j} = \min(\gamma q_{j\mu}, 1)$.

1.3 Dual coding and cut-off coding

By combining the weight coding and connectivity coding described above, the dual coding is given as $w_{ij} = w_{\mu j} = q_{j\mu}/\rho$, $\Pr[c_{ij} = 1] = \rho_{\mu j}$, $\rho_{\mu j} = \min(\gamma q_{j\mu}, 1)$, where ρ was defined by $\rho = \gamma\bar{q}$, $\bar{q} = \frac{1}{Mp}\sum_j \sum_\mu q_{j\mu}$, as in the weight coding. For the cut-off coding strategy, the synaptic weight was chosen as $w_{ij} = w_{\mu j} = q_{j\mu}/\rho_o$ where ρ_o is the mean connection probability. Based on these synaptic weights, for each output neuron, I selected $M\rho_o$ largest synaptic connections, and eliminated all other connections. Thus, connection matrix C was given as $c_{ij} = \left[\sum_{j'} [w_{ij} \leq w_{ij'}]_+ \leq M\rho_o\right]_+$, where $[\text{true}]_+ = 1$, $[\text{false}]_+ = 0$. When multiple connections have the same weight, I randomly selected the connections so that the total number of inbound connections becomes $M\rho_o$. Finally, in the random connection strategy, synaptic weights and connections were determined as $w_{ij} = w_{\mu j} = q_{j\mu}/\rho_o$, $\Pr[c_{ij} = 1] = \rho_o$.

1.4 Synaptic weight learning

To perform maximum likelihood estimation from output neuron activity, synaptic weight matrix between input neurons and output neurons should provide a reverse model of input neuron activity. If the reverse model is faithful, KL-divergence between the true input and the estimated distributions would be minimized [48] [171]. Therefore, synaptic weights learning can be performed by $\text{argmin}_W D_{KL}[p^*(r^t_X)||p(r^t_X|C,W)]$. Likelihood $p(r^t_X|C,W)$ is approximated as

$$\begin{aligned} p(r^t_X|C,W) &\propto \sum_\mu p(r^t_X|s^t = \mu, C, W)p(s^t = \mu|C,W) \\ &= \sum_\mu p(s^t = \mu|C,W)\exp\left[\sum_j \left(h(\theta^{C,W}_{j,\mu})g(r^t_{X,j}) - A(\theta^{C,W}_{j,\mu}) + B(r^t_{X,j})\right)\right] \\ &\simeq \sum_\mu p(s^t = \mu)\exp\left[\sum_j \left(q^{C,W}_{j\mu} g(r^t_{X,j}) - \alpha(q^{C,W}_{j\mu}) + B(r^t_{X,j})\right)\right]. \end{aligned} \tag{3.10}$$

in the second line is the average response estimated from connectivity matrix C, and weight matrix W. In the last equation, $q^{C,W}_{j\mu}$ is substituted for $h(\theta^{C,W}_{j,\mu})$. If we approximate the estimated parameter $q^{C,W}_{j\mu}$ with $q^{C,W}_{j\mu} \simeq \rho_o w_{ij}$ by using the average connectivity ρ_o, a synaptic weight plasticity rule is given by stochastic gradient descending as

$$\begin{aligned} \Delta w_{ij} &\propto \frac{\partial \log p(r^t_X|C,W)}{\partial w_{ij}} \\ &= p(s^t = \mu|r^t_X, C, W)\rho_o\left(g(r^t_{X,j}) - \alpha'(\rho_o w_{ij})\right) \\ &\simeq r^t_{Y,i}\rho_o\left(g(r^t_{X,j}) - \alpha'(\rho_o w_{ij})\right) \end{aligned} \tag{3.11}$$

Especially, in a Gaussian model, the synaptic weight converges to the weight coding as $w_{ij} = \left\langle r^t_{Y,i} r^t_{X,j} / \left(\sigma^2_X \rho_o r^t_{Y,i}\right)\right\rangle = q_{j\mu}/\rho_o$, where μ is the external state that output neuron i learned to represent (i.e. $i \in \Omega_\mu$).

As I was considering population representation, in which the total number of output neuron is larger than the total number of external states (i.e. $p < N$), there is a redundancy in representation. Thus,

to make use of most of population, homeostatic constraint is necessary. For homeostatic plasticity, I set a constraint on the output firing rate. By combining two terms, synaptic weight plasticity rule is given as

$$\Delta w_{ij} = \frac{\eta_X}{\gamma}\left(r^t_{Y,i}\left[g(r^t_{X,j}) - \alpha'\left(\rho_o w_{ij}\right)\right] + b_h\left[r^o_Y/N - r^t_{Y,i}\right]\right). \tag{3.12}$$

By changing the strength of homeostatic plasticity b_h, the network changes its behavior. The learning rate is divided by γ, because the mean of w is proportional to $1/\gamma$. Although, this learning rule is unsupervised, each output neuron naturally selects an external state in self-organisation manner.

1.5 Synaptic connection learning

Wiring plasticity of synaptic connection can be given in a similar manner. As shown in Figure 3.3, if the synaptic connection structure of network is correlated with the external model, the learning performance typically gets better. Therefore, by considering $\mathrm{argmin}_\rho\, D_{KL}\left[p^*(r^t_X)||p(r^t_X|\rho, W)\right]$, the update rule of connection probability is given as

$$\Delta\rho_{ij} \propto r^t_{Y,i} w_o\left[g(r^t_{X,j}) - \alpha'(\rho_{ij} w_o)\right]. \tag{3.13}$$

Here, I approximated w_{ij} with its average value w_o. In this implementation, if synaptic weight is also plastic, convergence of KL-divergence is no longer guaranteed, yet as shown in Figure 3.3, redundant representation robustly provides a good heuristic solution.

Let us next consider the implementation of the rewiring process with local spine elimination and creation based on the connection probability ρ_{ij}. To keep the detailed balance of connection probability, creation probability $c_p(\rho)$ and elimination probability $e_p(\rho)$ need to satisfy

$$(1-\rho)c_p(\rho) = \rho e_p(\rho).$$

The simplest functions that satisfy above equation is $c_p(\rho) \equiv \rho/\tau_c$, $e_p \equiv (1-\rho)/\tau_c$. In the simulation, I implemented this rule by changing c_{ij} from 1 to 0 with probability $(1-\rho)/\tau_c$ for every connection with $c_{ij} = 1$, and shift c_{ij} from 0 to 1 with probability ρ/τ_c for non-existing connection ($c_{ij} = 0$) at every time step.

1.6 Dual Hebbian rule and estimated transfer entropy

The results in the main texts suggest that non-random synaptic connection structure can be beneficial either when that increases estimated transfer entropy or is correlated with the structure of the external model. To derive dual Hebbian rule, I used the latter property, yet in the simulation, estimated transfer entropy also increased by the dual Hebbian rule. Here, I consider relationship of two objective functions.

Estimation of the external state from the sampled inputs is approximated as

$$\langle p(s^t=\mu)|\{c_{ij}r^t_{X,j}\}\rangle_{i\in\Omega_\mu} \simeq \frac{1}{|\Omega_\mu|}\sum_{i\in\Omega_\mu}\frac{p(s^t=\mu)\exp\left(\sum_j \rho_{ij}\left[q_{\mu j}g(r^t_{X,j})-\alpha(q_{\mu j})+B(r^t_{X,j})\right]\right)}{\sum_\nu p(s^t=\nu)\exp\left(\sum_j c_{ij}\left[q_{\nu j}g(r^t_{X,j})-\alpha(q_{\nu j})+B(r^t_{X,j})\right]\right)} \tag{3.14}$$

Therefore, by considering stochastic gradient descending, an update rule of ρ_{ij} is given as

$$\Delta\rho_{ij} \propto \left(1+\log r^t_{Y,i}/r^o_Y\right) r^t_{Y,i}\left[g(r^t_{X,j})-\alpha(q_{\mu j})/q_{\mu j}+B(r^t_{X,j})/q_{\mu j}\right] \tag{3.15}$$

If I compare this equation with the equation for dual Hebbian rule (equation (13)), both of them are monotonically increasing function of $r^t_{Y,i}$ and have the same dependence on $g(r^t_{X,j})$ although normalization terms are different. Thus, under an adequate normalization, the inner product of change direction is on average positive. Therefore, although dual Hebbian learning rule does not maximize the estimated maximum transfer entropy, the rule rarely diminishes it.

1.7 Gaussian model

I constructed mean response probabilities $\{\theta_{j\mu}\}^{\mu=1,\ldots,p}_{j=1,\ldots,M}$ by following 2 steps. First, non-normalized response probabilities $\{\tilde{\theta}_{j\mu}\}^{\mu=1,\ldots,p}_{j=1,\ldots,M}$ were chosen from a truncated normal distribution $N(\mu_M,\sigma_M)$ defined on $[0,\infty)$. Second, I defined $\{\theta_{j\mu}\}^{\mu=1,\ldots,p}_{j=1,\ldots,M}$ by $\theta_{j\mu}=\tilde{\theta}_{j\mu}/Z_\mu$, where $Z_\mu = r^o_X \Big/ \sqrt{\sum_{j=1}^M \tilde{\theta}_{j\mu}\Big/M}$. When the noise follows a Gaussian distribution, the response functions in equation (5) are given as

$$h(\theta)=\frac{\theta}{\sigma_x^2},\ g(r)=r,\ A(\theta)=\frac{\theta^2}{2\sigma_x^2}+\log(\sqrt{2\pi}\sigma_x),\ B(r)=-\frac{r^2}{2\sigma_x^2}. \tag{3.16}$$

Because $h^{-1}(q)=\sigma_X^2 q$, $\alpha(q)$ is given as $\alpha(q)\equiv A\left(h^{-1}(q)\right)=\sigma_x^2q^2/2+\log(\sqrt{2\pi}\sigma_x)$. By substituting above values into the original equations, the neural dynamics is given as

$$r^t_{Y,i}=r^o_Y\exp\left[\sum_{j=1}^M c_{ij}\left(w_{ij}r^t_{X,j}-w_o\right)-I^t_{inh}\right]. \tag{3.17}$$

Similarly, dual Hebbian rule becomes

$$\Delta w_{ij} = \frac{\eta_X}{\gamma}\left(r^t_{Y,i}\left[r^t_{X,j}-\sigma_X^2\rho_o w_{ij}\right]+b_h\left[r^o_Y/N-r^t_{Y,i}\right]\right) \tag{3.18}$$

$$\Delta\rho_{ij} = \eta_\rho r^t_{Y,i}\left(r^t_{X,j}-\sigma_x^2\rho_{ij}w_o\right). \tag{3.19}$$

1.8 Poisson model

For Poisson model, I defined mean response probabilities $\{\theta_{j\mu}\}^{\mu=1,\ldots,p}_{j=1,\ldots,M}$ from a log-normal distribution instead of a normal distribution. Non-normalized values were sampled from a truncated log-normal distribution $\log N(\mu^p_M,\sigma^p_M)$ defined on $(l^p_{\min},l^p_{\max})$. Normalization was performed as $\theta_{j\mu}=\tilde{\theta}_{j\mu}/Z_\mu$ for $\{\tilde{\theta}_{j\mu}\}^{\mu=1,\ldots,p}_{j=1,\ldots,M}$, where $Z_\mu = r^o_X M\Big/\sum_j\theta_{j\mu}$. Because the noise follows a Poisson distribution $p(r|\theta)=$

$\exp[-q + r\log q - \log r!]$, the response functions are given as

$$h(\theta) = \log\theta,\ g(r) = r,\ A(\theta) = \theta,\ B(r) = -\log r!\ . \tag{3.20}$$

As a result, $\alpha(q)$ is defined as $\alpha(q) \equiv A\left(h^{-1}(q)\right) = e^q$. By substituting them to the original equations, the neural dynamics also follows equation (17). If connection is all-to-all, by setting $w_{ij} = \log\theta_{j\mu}/\theta_o$ for $i \in \Omega_\mu$, optimal inference is achievable. Here, I normalized θ_j by θ_o, which is defined as $\theta_o = \frac{1}{2}\min_{j,\mu}\theta_{\mu j}$, in order to keep synaptic weights in non-negative values.

Learning rules for synaptic weight and connection are given as

$$\Delta w_{ij} = \frac{\eta_x}{\gamma}\left(r^t_{Y,i}\left[r^t_{X,j} - \theta_{min}\exp[\rho_o w_{ij}]\right] + b_h\left[r^o_Y/N - r^t_{Y,i}\right]\right), \tag{3.21}$$

$$\Delta\rho_{ij} = \eta_\rho r^t_{Y,i}\left(r^t_{X,j} - \theta_{min}\exp(\rho_{ij}w_o)\right). \tag{3.22}$$

Note that the first term of the synaptic weight learning rule coincides with a previously proposed optimal learning rule for spiking neurons [171] [90]. In calculation of model error, error was calculated as $d = \sqrt{\frac{1}{pM}\sum_\mu\sum_j\left(\tilde{q}_{j\mu} - q*_{j\mu}\right)^2}$, where estimated parameter $\{\tilde{q}_{j\mu}\}$ was given by $\tilde{q}_{j\mu} = \frac{\langle q^*_{j\mu}\rangle\bar{q}_{j\mu}}{\sum_q\sum_j\bar{q}_{j\mu}/pM}$. Here, $\langle q^*_{j\mu}\rangle$ represents the mean of true $\{q_{j\mu}\}$, and non-normalized estimator $\bar{q}_{j\mu}$ was calculated as $\bar{q}_{j\mu} = \frac{1}{\langle c_{ij}\rangle|\Omega_\mu|}\sum_{i\in\Omega_\mu} c_{ij}w_{ij}$. In Figure S1D, estimation from connectivity was calculated from $\bar{q}^C_{j\mu} = \frac{1}{\langle c_{ij}\rangle|\Omega_\mu|}\sum_{i\in\Omega_\mu} c_{ij}$, and similarly, estimation from weights was calculated by $\bar{q}^W_{j\mu} = \frac{1}{|\Omega_\mu|\sum_{i\in\Omega_\mu}c_{ij}}\sum_{i\in\Omega_\mu} c_{ij}w_{ij}$. For parameters, I used $\mu^p_M = 0.0$, $\sigma^p_M = 1.0$, $l^p_{min} = 0.2$, $l^p_{max} = 20.0$, $w_o = 1/\gamma$, $r^o_X = 0.3$, and for other parameters, I used same values with the Gaussian model.

2 Analytical evaluations

2.1 Evaluation of performances in weight coding and connectivity coding

In Gaussian model, we can analytically evaluate the performance in two coding schemes. As the dynamics of output neurons follows $r_{Y,i} = r^o_Y\exp\left[\sum_j c_{ij}(w_{ij}r^t_{X,j} - w_o) - I^t_{inh}\right]$, membrane potential variable u_i, which is defined as

$$u_i \equiv \sum\nolimits_j c_{ij}(w_{ij}r^t_{X,j} - w_o), \tag{3.23}$$

determines firing rates of each neuron. Because $\{\theta_{j\mu}\}$ is normalized with $\sum_{j=1}^M\theta^2_{j\mu}/M = (r^o_X)^2$, mean and variance of $\{\theta_{j\mu}\}$ are given as

$$\mu_\theta = \frac{\mu_M r^o_X}{\sqrt{\mu^2_M + \sigma^2_x}}, \sigma^2_\theta = \frac{\left(\sigma_M r^o_X\right)^2}{\mu^2_M + \sigma^2_M}, \tag{3.24}$$

where μ_M and σ_M are the mean and variance of the original non-normalized truncated Gaussian distribution $\{\tilde{\theta}_{j\mu}\}$. Because both $r_{X,j}$ and $\{\theta_{j\mu}\}$ approximately follow Gaussian distribution, u_i is expected to follow Gaussian. Therefore, by evaluating its mean and variance, we can characterize the

distribution of u_i for a given external state [11].

Let us first consider the distribution of u_i in the weight coding. In weight coding scheme, w_{ij} and c_{ij} are defined as

$$w_{ij} = \theta_{j\mu}/\rho\sigma_x^2,\ \Pr[c_{ij}=1] = \rho \tag{3.25}$$

where $\rho = \gamma\mu_\theta/\sigma_X^2$. By setting $w_o = \mu_\theta^2/(\rho\sigma_X^2)$, the mean membrane potential of output neuron i selective for given signal (i.e. $i \in \Sigma_\mu$ for $s^t = \mu$) is calculated as,

$$\langle u_i \rangle = \left\langle \sum\nolimits_j \left(\theta_{j\mu}^2 - \langle \theta_{j\mu} \rangle^2 \right) \Big/ \sigma_x^2 \right\rangle = M\sigma_\theta^2/\sigma_x^2.$$

Similarly, the variance of ui is given as

$$\begin{aligned}\langle (u_i - \langle u_i \rangle)^2 \rangle &= \left\langle \left(\tfrac{1}{\rho\sigma_X} \sum_j c_{ij}\theta_{j\mu}\zeta_j + \tfrac{1}{\rho\sigma_X^2} \sum_j (c_{ij}-\rho)\left(\theta_{j\mu}^2 - \mu_\theta^2\right) + \tfrac{1}{\sigma_X^2} \sum_j \left(\theta_{j\mu}^2 - \left[\mu_\theta^2 + \sigma_\theta^2\right]\right) \right)^2 \right\rangle \\ &= \tfrac{M}{\rho\sigma_X^2}\left(\mu_\theta^2 + \sigma_\theta^2\right) + \tfrac{M\sigma_\theta^2}{\rho\sigma_X^4}\left[2\left(2\mu_\theta^2 + \sigma_\theta^2\right) + (1-\rho)\,\sigma_\theta^2\right]\end{aligned} \tag{3.26}$$

where ζ_i is a Gaussian random variable. On the other hand, if output neuron i is not selective for the presented stimuli (if $s^t \neq \mu$ and $i \in \Sigma_\mu$), w_{ij} and $r_{X,j}$ are independent. Thus, the mean and the variance of u_i are given as,

$$\langle u_i \rangle = 0,\ \langle (u_i - \langle u_i \rangle)^2 \rangle = \frac{M}{\rho\sigma_x^2}(\mu_\theta^2 + \sigma_\theta^2) + \frac{M\sigma_\theta^2}{\rho\sigma_x^4}\left(2\mu_\theta^2 + \sigma_\theta^2\right)$$

In addition to that, due to feedforward connection, output neurons show noise correlation. For two output neurons i and l selective for different states (i.e. $i \in \Omega_\mu$ and $l \neq \Omega_\mu$), the covariance between u_i and u_l satisfies

$$\langle (u_i - \langle u_i \rangle)(u_l - \langle u_l \rangle) \rangle = \left\langle \rho^2 \sum\nolimits_j w_{ij}w_{lj}(r_{X,j} - \theta_{j\mu})^2 \right\rangle = M\mu_\theta^2/\sigma_x^2$$

Therefore, approximately (u_i, u_l) follows a multivariable Gaussian distributions

$$\begin{pmatrix} u_i \\ u_l \end{pmatrix} = N\left(\begin{pmatrix} \frac{M\sigma_\theta^2}{\sigma_x^2} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{M\left(\mu_\theta^2+\sigma_\theta^2\right)}{\rho\sigma_X^2} + \frac{M\sigma_\theta^2\left[2\left(2\mu_\theta^2+\sigma_\theta^2\right)+(1-\rho)\sigma_\theta^2\right]}{\rho\sigma_X^4} & \frac{M\mu_\theta^2}{\sigma_x^2} \\ \frac{M\mu_\theta^2}{\sigma_x^2} & \frac{M(\mu_\theta^2+\sigma_\theta^2)}{\rho\sigma_x^2} + \frac{M\sigma_\theta^2\left(2\mu_\theta^2+\sigma_\theta^2\right)}{\rho\sigma_x^4} \end{pmatrix} \right). \tag{3.27}$$

In maximum likelihood estimation, the estimation fails if a non-selective output neuron shows higher firing rate than the selective neuron. When there are two output neurons, probability for such an event is calculated as

$$\epsilon_w = \Pr\left[\sum\nolimits_j c_{lj}(w_{lj}r_{X,j}^t - w_o) > \sum\nolimits_j c_{ij}(w_{ij}r_{X,j}^t - w_o) | s^t = \mu, i \in \Omega_\mu, l \notin \Omega_\mu \right].$$

In the simulation, there are $p-1$ distractors per one selective output neuron. Thus, approximately,

accuracy of estimation was evaluated by $(1-\epsilon_w)^{p-1}$. In Figure 3.2B, I numerically calculated this value for the analytical estimation.

Similarly, in connectivity coding, w_{ij} and c_{ij} are given as

$$w_{ij} = 1/\gamma,\ \Pr[c_{ij}=1] = \rho_{ij},\ \rho_{ij} = \gamma\theta_{j\mu}/\sigma_x^2.$$

By setting $w_o = \mu_\theta/\gamma$, from a similar calculation done above, the mean and the variance of (u_i, u_l) are derived as

$$\begin{pmatrix} u_i \\ u_l \end{pmatrix} = N\left(\begin{pmatrix} \frac{M\sigma_\theta^2}{\sigma_x^2} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{M\mu_\theta}{\gamma} + \frac{M\sigma_\theta^2\left[\mu_\theta\sigma_x^2 - \gamma\sigma_\theta^2\right]}{\gamma\sigma_x^4} & \frac{M\mu_\theta^2}{\sigma_x^2} + \frac{M\mu_\theta^2\sigma_\theta^2}{\sigma_x^4} \\ \frac{M\mu_\theta^2}{\sigma_x^2} + \frac{M\mu_\theta^2\sigma_\theta^2}{\sigma_x^4} & \frac{M\mu_\theta}{\gamma} + \frac{M\mu_\theta\sigma_\theta^2}{\gamma\sigma_x^2} \end{pmatrix}\right). \tag{3.28}$$

If we compare the two coding schemes, means are the same for two coding schemes, and as γ satisfies $\gamma = \sigma_X^2\rho/\mu_\theta$ variance of non-selective output neuron are similar. The main difference is the second term of signal variance. In the weight coding, signal variance is proportional to $1/\gamma$, on the other hands, in the connectivity coding, the second term of signal variance is negative, and does not depend on the connectivity. As a result, in the adequately sparse regime, firing rate variability of selective output neuron becomes smaller in connectivity coding, and the estimation accuracy is better. In the sparse limit, the first term of variance becomes dominant and both schemes do not work well, consequently, the advantage for connectivity coding disappears. Coefficient of variation calculated for signal terms is indeed smaller in connectivity coding scheme (blue and red lines in Fig 2C), and the same tendency is observed in simulation (cyan and orange lines in Fig 2C).

2.2 Optimality of connectivity

To evaluate optimality of a given connection matrix C, I calculated the posterior probability of the external states estimated from C and r_X, and compared then to that from the fully connected network C_{all}. Below, I denote the mean KL-divergence $\langle D_{KL}\left[p\left(s^t|r_X, C_{all}\right)||p\left(s^t|r_X, C\right)\right]\rangle_{r_X}$ as $I(C_{all}, C)$ for readability. When the true external state is $s^t = \nu$, firing rates of input neurons are given by $r_{X,j}^t\ N(\theta_{j\nu}, \sigma_X)$, hence this $I(C_{all}, C)$ is approximately evaluated as

$$\begin{aligned} I\left(C_{all}, C\right) &\approx \tfrac{1}{p}\sum_\nu \left\langle D_{KL}\left[p\left(s^t|r_{X|\nu}, C_{all}\right)||p\left(s^t|r_{X|\nu}, C\right)\right]\right\rangle_{r_X} \\ &\approx \tfrac{1}{p}\sum_\nu D_{KL}\left[\left\langle p\left(s^t|\left\{\theta_{j\nu} + \sigma_X\zeta_j\right\}, C_{all}\right)\right\rangle_{\{\zeta_j\}}||\left\langle p\left(s^t|\left\{\theta_{j\nu} + \sigma_X\zeta_j\right\}, C\right)\right\rangle_{\{\zeta_j\}}\right] \end{aligned}$$

where $\{\zeta_i\}$ are Gaussian random variables, and C_{all} represents the all-to-all connection matrix. By taking integral over Gaussian variables, the posterior probability is evaluated as

$$\left\langle p\left(s^t = \mu|\left\{\theta_{j\nu} + \sigma_X\zeta_j\right\}, C\right)\right\rangle_{\{\zeta_j\}} \cong \frac{1}{|\Omega_\mu|}\sum_{i\in\Omega_\mu}\frac{\exp\left(\phi_{\mu\nu}^{i,C} + \frac{1}{2}\psi_\mu^{i,C}\right)}{\sum_{\mu'}\exp\left(\phi_{\mu'\nu}^{i,C} + \frac{1}{2}\psi_{\mu'}^{i,C}\right)} \equiv p_\nu\left(s^t = \mu|C\right),$$

where

$$\phi_{\mu\nu}^{i,C} \equiv \sum\nolimits_j c_{ij}\left(2\theta_{\mu j}\theta_{\nu j} - \theta_{\mu j}^2\right)/\left(2\sigma_X^2\right),\ \psi_\mu^{i,C} \equiv \sum\nolimits_j c_{ij}(\theta_{\mu j}/\sigma_X)^2.$$

Thus, the KL-divergence between estimations by two connection structures C_{all} and C is approximated as:

$$I\left(C_{all}, C\right) \approx \frac{1}{p}\sum_\nu \sum_\mu p_\nu\left(s^t = \mu | C_{all}\right) \log \frac{p_\nu\left(s^t = \mu | C_{all}\right)}{p_\nu\left(s^t = \mu | C\right)} \tag{3.29}$$

In the black lines in Figures 3.3C-E, I maximized the approximated KL-divergence $I(C_{all}, C)$ with a hill-climbing method from various initial conditions, thus the lines may not be the exact optimal, but rather lower bounds of the optimal performance. Information gain by a connection c_{ij} was evaluated by

$$\Delta I_{ij} \equiv \left\langle I\left(C_{all}, C\right) - I\left(C_{all}, C + \eta_{ij}\right)\right\rangle_C, \tag{3.30}$$

where η_{ij} is a $N \times M$ matrix in which only (i, j) element takes 1, and all other elements are 0. In Figure 3.3B, I took average over 1000 random connection structures with connection probability $\rho = 0.1$.

3 Model settings

3.1 Details of simulation

In the simulation, the external variable s^t was chosen from 10 discrete variables ($p = 10$) with equal probability ($\Pr[s^t = q] = 1/p$, for all q). The mean response probability $\theta_{j\mu}$ was given first by randomly chosen parameters $\{\tilde{\theta}_{j\mu}\}_{j=1,\ldots,M}^{\mu=0,\ldots,p-1}$ from the truncated normal distribution $N(\mu_M, \sigma_M)$ in $[0, \infty)$, and then normalized using $\theta_{j\mu} = \tilde{\theta}_{j\mu}/Z_\mu$, where $Z_\mu = r_X^o \Big/ \sqrt{\sum_{j=1}^M \tilde{\theta}_{j\mu} \Big/ M}$. Mean weight w_o was defined as $w_o = r_X^o/\gamma$. The normalization factor h_w was defined as $h_w = \bar{q}/\gamma$ in Figures 3.1?2 and 3.4-5, where $\bar{q} = \frac{1}{Mp}\sum_j \sum_\mu \theta_{j\mu}/\sigma_X^2$, and as $h_w = r_X^o/\gamma$ in Figures 3.6?7, as the mean of θ depends on κ_m. In Figure 3.3, I used $h_w = \bar{q}/\gamma$ for the dual coding, and $h_w = \bar{q}/\rho_o$ for the rest. Average connectivity $\bar{\rho}$ was calculated from the initial connection matrix of each simulation. In the calculation of the dynamics, for the membrane parameter $v_i \equiv \sum_j c_{ij}\left(w_{ij} r_{X,j}^t - h_w\right)$, a boundary condition $v_i > \max_l\{v_l - v_d\}$ was introduced for numerical convenience, where $v_d = -60$. In addition, synaptic weight w_{ij} was bounded to a non-negative value ($w_{ij} > 0$), and the connection probability was defined as $\rho \in [0, 1]$. For simulations with synaptic weight learning, initial weights were defined as $w_{ij} = \left(1 + \sigma_w^{init}\right)/\gamma$, where $\sigma_w^{init} = 0.1$, and ζ is a Gaussian random variable. Similarly, in the simulation with structural plasticity, the initial condition for the synaptic connection matrix was defined as $\Pr[c_{ij} = 1] = \gamma\langle\theta_{j\mu}\rangle/\sigma_X^2$. In both the dual Hebbian rule and the approximated dual Hebbian rule, the synaptic weight of a newly created spine was given as $w_{ij} = \left(1 + \sigma_w^{init}\zeta\right) w_o$, for a random Gaussian variable $\zeta \leftarrow N(0, 1)$. In Figure 3.8, simulations were initiated at -20 days (i.e., 2×10^6 steps before stimulus onset) to ensure convergence for the control condition. For model parameters, $\mu_M = 1.0$, $\sigma_M = 1.0$, $\sigma_X = 1.0$, $M = 200$, $N = 100$, $r_X^o = 1.0$, and $r_Y^o = 1.0$ were used, and for learning-related

parameters, $\eta_X = 0.01$, $b_h = 0.1$, $\eta_\rho = 0.001$, $\tau_c = 10^6$, $T_2 = 10^5$, and $\kappa_m = 0.5$ were used. In Figures 3.7 and 3.8, $\eta_\rho = 0.0001$, $\tau_c = 3 \times 10^5$, and $\gamma = 0.6$ were used, unless otherwise stated.

3.2 Accuracy of estimation

The accuracy was measured with the bootstrap method. By using data from $t - T_o <= t' < t$, the selectivity of output neurons was first decided. Ω_μ was defined as a set of output neurons that represents external state μ. Neuron i belongs to set Ω_μ if i satisfies

$$\mu = \arg\max_{\mu'} \frac{\sum_{t'=t-T_o}^{t} [s^t = \mu']_+ r_{Y,i}^t}{\sum_{t'=t-T_o}^{t} [s^t = \mu']_+},$$

where operator $[X]_+$ returns 1 if X is true; otherwise, it returns 0. By using this selectivity, based on data from $t <= t' < t + T_o$, the accuracy was estimated as

$$\frac{1}{T_o} \sum_{t'=t}^{t+T_o-1} \left[\frac{1}{|\Omega_{s^{t'}}|} \sum_{i \in \Omega_{s^{t'}}} r_{Y,i}^{t'} > \max_{\mu \neq s^{t'}} \frac{1}{|\Omega_\mu|} \sum_{i \in \Omega_\mu} r_{Y,i}^{t'} \right]_{tof}.$$

In the simulation, $T_o = 10^3$ was used because this value is sufficiently slow compared with weight change but sufficiently long to suppress variability.

3.3. Model error

Using the same procedure, model error was estimated as

$$d = \sqrt{\frac{1}{pM} \sum_{\mu=1}^{p} \sum_{j=1}^{M} \left(\tilde{\theta}_{j\mu} - \theta_{j\mu} \right)^2},$$

where $\tilde{\theta}_{j\mu}$ represents the estimated parameter. $\tilde{\theta}_{j\mu}$ was estimated by

$$\bar{\theta}_{j\mu} = \frac{1}{\langle c_{ij} \rangle |\Omega_\mu|} \sum_{i \in \Omega_\mu} c_{ij} w_{ij}, \ \tilde{\theta}_{j\mu} = r_o^X \bar{\theta}_{j\mu} \Big/ \sqrt{\tfrac{1}{M} \textstyle\sum_{j=1}^{M} \bar{\theta}_{j\mu}^2}.$$

In Figure 3.6E, the estimation of the internal model from connectivity was calculated by

$$\bar{\theta}_{j\mu}^C = \frac{1}{\langle c_{ij} \rangle |\Omega_\mu|} \sum_{i \in \Omega_\mu} c_{ij}.$$

Similarly, the estimation from the synaptic weight in Figure 3.6F was performed with

$$\bar{\theta}_{j\mu}^W = \frac{1}{|\Omega_\mu|} \sum_{i \in \Omega_\mu} c_{ij} w_{ij} \Big/ \sum_{i \in \Omega_\mu} c_{ij}.$$

3.4 Transfer entropy

Entropy reduction caused by partial information on input firing rates was evaluated by transfer entropy:

$$T_E = \left\langle H\left(s^t\right) - H\left(s^t | r_X^t, C\right)\right\rangle_t,$$

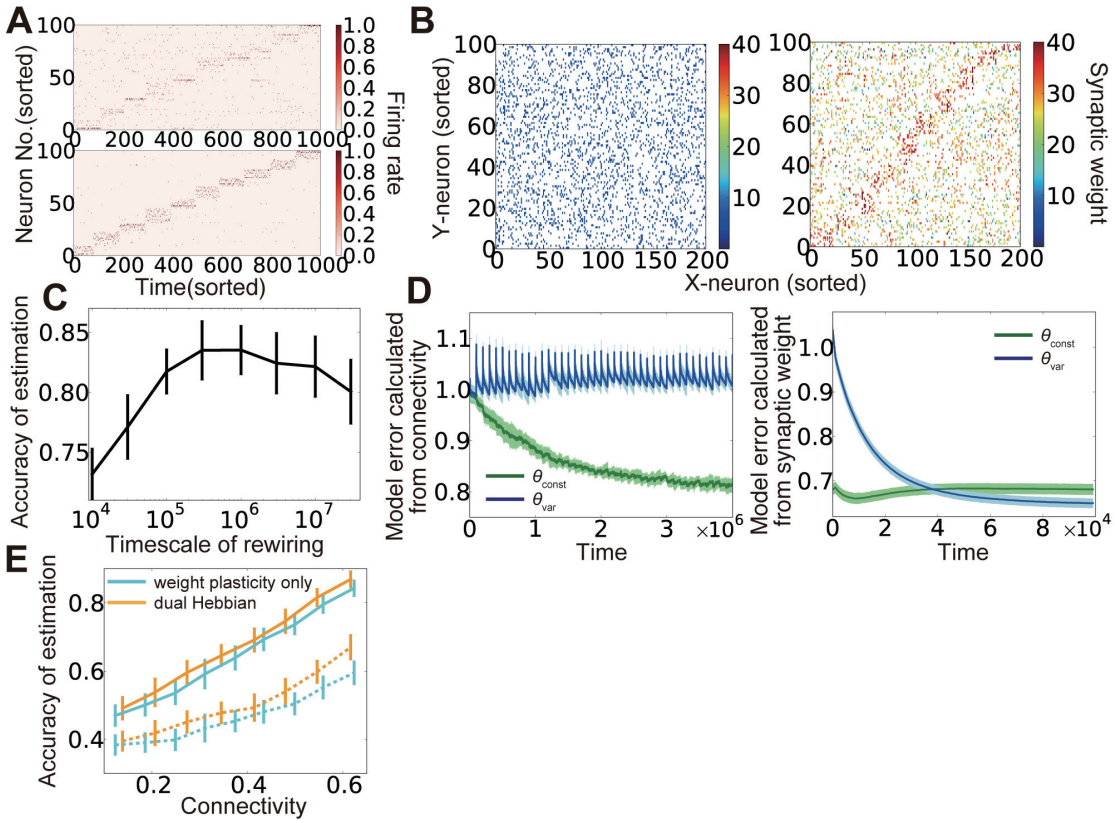
where

$$\begin{aligned} H\left(s^t | r_X^t, C\right) &= -\sum_{\mu=1}^{p} p\left(s^t = s_\mu | r_X^t, C\right) \log p\left(s^t = s_\mu | r_X^t, C\right) \\ &\cong -\sum_{\mu=1}^{p} \left\langle p\left(s^t = s_\mu | \left\{c_{ij} r_{X,j}^t\right\}\right)\right\rangle_{i\in\Omega_\mu} \log \left\langle p\left(s^t = s_\mu | \left\{c_{ij} r_{X,j}^t\right\}\right)\right\rangle_{i\in\Omega_\mu}, \end{aligned}$$

$$\begin{aligned} \left\langle p\left(s^t = s_\mu | \left\{c_{ij} r_{X,j}^t\right\}\right)\right\rangle_{i\in\Omega_\mu} &\cong \frac{1}{|\Omega_\mu|} \sum_{i\in\Omega_\mu} p\left(s^t = s_\mu\right) \prod_{c_{ij}=1} p\left(r_{X,j}^t | s^t = s_\mu\right) \\ &= \frac{1}{|\Omega_\mu|} \sum_{i\in\Omega_\mu} \frac{p\left(s^t=s_\mu\right) \exp\left(\sum_{j=1}^{M} c_{ij}\left[q_{\mu j} g\left(r_{X,j}^t\right) - \alpha(q_{\mu j}) + B\left(r_{X,j}^t\right)\right]\right)}{\sum_\nu p(s^t=s_\nu) \exp\left(\sum_{j=1}^{M} c_{ij}\left[q_{\nu j} g\left(r_{X,j}^t\right) - \alpha(q_{\nu j}) + B\left(r_{X,j}^t\right)\right]\right)}. \end{aligned}$$

Output group Ω_μ was determined as described above. Here, the true model was used instead of the estimated model to evaluate the maximum transfer entropy achieved by the network.

Supplementary Figures



Supplementary Figure 1. Results in Poisson model. (**A**) An example of output neuron activity before (top) and after (bottom) synaptic weight learning at connectivity $\rho = 0.25$. (**B**) Synaptic weight matrices before (left) and after (right) learning. Both X-neurons and Y-neurons were sorted based on their preferred external states. (**C**) Accuracy of estimation at various timescale of rewiring τ_c. (**D**) Model error calculated from connectivity (left) and synaptic weights (right). (**E**) Comparison of performance among the model without wiring plasticity (cyan), and dual Hebbian model(orange). Corresponding results in the Gaussian model are described in Fig. 3.4A, Fig. 3.5F, Fig. 3.5G, Fig. 3.6EF, Fig. 3.6I respectively.

Chapter 4

Mixed Signal Learning by Spike Correlation Propagation in Feedback Inhibitory Circuits

Introduction

Neurons receive inputs from a large number of other neurons encoding a variety of information about various signals. Despite the diversity and variability of input spike trains, neurons can learn and represent specific information during developmental processes and according to specific task requirements. Spike-timing-dependent plasticity (STDP) [152] [20] is a candidate mechanism of neural learning. Extensive studies have revealed the type of information that a single neuron can learn through STDP [73] [213] [88] [136] [82]; however, the type of information that a population of neurons interacting with each other learns through STDP has not yet been determined. Understanding this extension from a single neuron to a population of neurons is crucial because a single neuron learns and represents only a limited amount of information that may be transmitted to it from thousands of inputs.

Among neural interactions, lateral inhibition is a basic interaction widely observed in various regions, such as the olfactory bulb [9], visual cortex [134], somatosensory cortex [3], and entorhinal cortex [45]. Previous theoretical results showed that neural circuits with lateral inhibition enhance signal detection [5] [239] and improve learning performance [164] [68] [16]. Several simulation studies further revealed that neurons acquire receptive field [238] [203] [120] or spike patterns [158] through STDP by introducing lateral inhibition; yet, those studies were limited to simplified cases for which a large population of independent neurons was suggested to be sufficient [88] [157] [43]. Therefore, it remains unclear whether lateral inhibition plays a crucial role in STDP learning; in particular, the spike level effects of lateral inhibition remain elusive. Moreover, recent experimental results suggest that animals learn and

discriminate mixed olfactory signals [241] [173] [195] or auditory signals masked by noise [162] [194], but it is still unknown how feedback interactions contribute to such learning.

Here, by considering a simple feedback network model of spiking neurons, I investigated the algorithm inherent to STDP in neural circuits containing feedback. I analyzed the propagation of spike correlations through inhibitory circuits, and revealed how such secondary correlations influence STDP learning at both feedforward and feedback connections. I discovered that the timescale of spike correlation preferable for learning depends on whether the noise is independent from any signal (random noise) or generated from the mixing of signals (cross-talk noise). I also found that excitatory and inhibitory STDP cooperatively shapes lateral circuit structure, making it suitable for signal detection. I further found a possible link between stochastic membrane dynamics and sampling process, which is necessary for neural approximation of learning algorithm of Bayesian independent component analysis (ICA). I applied my findings by demonstrating that STDP implements a spike-based solution in neural circuits for the cocktail party problem [162] [41] [95].

Results

Model

I constructed a network model with three feedforward layers as shown in Fig. 4.1A (see *Neural dynamics* in Methods for details). The external source layer represents the external environment or neural activity at sensory systems. The external layer also provides common inputs to the input layer and induces correlations in the neurons in the input layer. The input layer shows rate-modulated Poisson firing based on events at the external layer and external noise, which is approximated with the constant firing rate r_i^o. Subsequently, spike activity at the input layer projects to the output layer, which also receives inhibitory feedback from the lateral layer. Neurons in the lateral layers are excited by inputs from the output layer. I assumed that all neurons in the input layer and the output layer are excitatory, whereas lateral-layer neurons are assumed to be inhibitory. Although excitatory lateral interactions also exist in the sensory cortex, they are typically sparse [103] and weak [3] compared with inhibitory interactions; thus I concentrated on the latter. For the analytical treatment, the neurons in the output and lateral layers were modeled with a linear Poisson model. I first studied synaptic plasticity at the feedforward connections (connections from the input layer to the output layer), while fixing lateral connections (i.e., connections from the output layer to the lateral layer and connections from the lateral layer to the output layer). For STDP, I used pairwise log-STDP (Fig. 4.1B) [81], which replicates the experimentally observed long-tailed synaptic weight distribution [214] [31].

I considered the case for information encoded in the correlated activity of input neurons [233] [131], and fixed the average firing rate of all input neurons at the constant value ν_o^X (See Table 1 and 2 for the list of variables and parameters). If the firing rate of input neuron i is given as

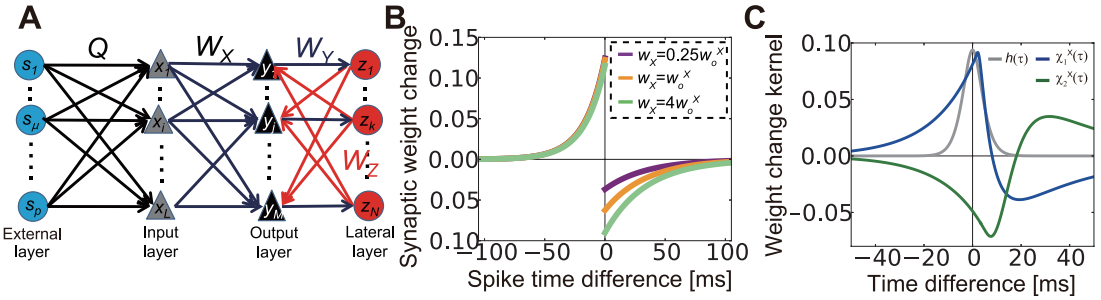


Figure 4.1. Description of the model. (**A**) Schematic figure of the model. (**B**) Spike-time dependent synaptic weight change in log- spike-timing-dependent plasticity (STDP). (**C**) Normalized temporal cross-correlogram of input neurons receiving common sources (gray line), and kernel functions of plasticity propagated by feedforward correlation (blue line) and feedback correlation (green line).

$r_i^o + \sum_{\mu=1}^{p} q_{i\mu} \int_0^\infty \phi(t') s_\mu(t-t')dt'$, for external event $s_\mu(t)$ and the response probability of the neuron $q_{i\mu}$, then common inputs from the external layer induce a temporal correlation proportional to

$$h(\tau;\theta_t) \equiv \int_{\max(\tau,0)}^{\infty} dt' \phi(t')\phi(t'-\tau), \tag{4.1}$$

where $\phi(t)$ is a response kernel (see equation (14) and (24) in Methods for details). If we use $\phi(t) = t^2 e^{-t/\theta_t}/2\theta_t^3$, where θ_t is the parameter that controls the timescale of spike correlations, then $h(\tau;\theta_t) = \frac{1}{16\theta_t^3}(\tau^2 + 3\theta_t|\tau| + 3\theta_t^2)e^{-|\tau|/\theta_t}$ (gray line in Fig. 4.1C). For the kernel function, I used the gamma distribution with shape parameter $k_g = 3$ in order to reproduce broad spike correlations typically observed in cortical neurons [132] [12]. Synaptic weight dynamics by STDP is written as

$$\frac{dw_{ji}^X}{dt} = x_i(t - d_{ji}^{Xa}) \int_0^\infty F_d(w_{ji}^X, s) y_j(t - s - d_{ji}^{Xd}) ds + y_j(t - d_{ji}^{Xd}) \int_0^\infty F_p(w_{ji}^X, s) x_i(t - s - d_{ji}^{Xa}) ds$$

for $F_d(w_{ij}^X, s) = f_d(w_{ij}^X)e^{-s/\tau_d}$, $F_p(w_{ij}^X, s) = f_p(w_{ij}^X)e^{-s/\tau_p}$, where $f_d(w)$ and $f_p(w)$ are synaptic weight dependence of LTD/LTP (long-term depression/potentiation), respectively. By taking the average of above equation over time and ensemble (see *Average synaptic weight velocity* in Methods for details), the weight change of the feedforward connection W_X can be approximated as

$$\dot{W}_X \approx W_X \left(g_1^X E - g_2^X W_Z W_Y\right) C^t, \tag{4.2}$$

where g_1^X and g_2^X are scalar coefficients, C is the correlation matrix, and E is the identity matrix (see equations (25)-(30) for derivation). The first term describes the synaptic weight change directly caused by an input spike correlation and can be rewritten into the convolution of the temporal correlation and correlation kernel function χ_1^X as

$$\begin{aligned} g_1^X &\equiv G_1^X(w_o^X),\ G_1^X(w) \equiv \int_{-\infty}^{\infty} \chi_1^X(\tau;w)h(\tau)d\tau, \\ \chi_1^X(\tau;w) &= \int_{-\tau+2d_{Xd}}^{\infty} ds F(w,s)\epsilon_X(\tau + s - 2d_{Xd}), \end{aligned} \tag{4.3}$$

where $F(w,s) = F_d(w,-s)$ if $s < 0$, else $F(w,s) = F_p(w,s)$, and ϵ_X is the EPSP curve of input neurons (see equation (15) and (31) in the Methods). By the deconvolution of $G_1^X(w)$, we can separate the effect of the intrinsic network property χ_1^X and that of the input correlation $h(\tau)$ for STDP-based learning. Due to causality, LTP/LTD balance, and dendritic delay, $\chi_1^X(\tau;w)$ typically becomes LTP-dominant around $\tau = 0$ (blue line in Fig. 4.1C; I set $w = w_o^X$), so that g_1^X takes positive values, which enables coincidence-based learning [213] [88] [76]. The second term of equation (2), which is of particular interest in this model, describes how the input correlation influences STDP learning at feedforward connections through lateral inhibition:

$$\begin{aligned} g_2^X &\equiv G_2^X(w_o^X),\ G_2^X(w) \equiv \int_{-\infty}^{\infty} \chi_2^X(\tau;w)h(\tau)d\tau \\ \chi_2^X(\tau;w) &= \int_{-\tau+D}^{\infty} dsF(w,s) \int_0^{\tau+s-D} dr\epsilon_Z(r) \int_0^{\tau+s-r-D} dq\epsilon_Y(q)\epsilon_X(\tau+s-r-q-D) \end{aligned} \quad (4.4)$$

where $D = 2d_{Xd} + d_Y + d_Z$, and ϵ_Y and ϵ_Z are EPSP/IPSP curves of output/inhibitory neurons, respectively. This term primarily causes LTD as the sign flips through lateral inhibition ($-\chi_2^X(\tau;w)$; shown as the green line in Fig. 4.1C). Previous simulation studies showed lateral inhibition has critical effects on excitatory STDP learning [238] [203] [120]; however, it has not yet been well studied how a secondary correlation generated through the lateral circuits influences STDP at feedforward connections, and it is still largely unknown how lateral inhibition functions with various stimuli in different neural circuits. For example, the correlation kernel of the feedback term exhibits a delay as the signal propagates through the inhibitory circuit; yet, we do not know how much delay is permitted for effective learning or if realistic synaptic delays satisfy such a condition. Furthermore, it is also unknown what information a circuit can learn if there are several mixed signals with different amplitudes for which symmetry-breaking learning [88] [77] is not valid. Therefore, using theoretical analysis and simulation, I first investigated the properties of the inhibitory kernel $-\chi_2^X(\tau;w)$ in STDP learning.

Lateral inhibition enhances minor source detection by STDP

In equation (2), if lateral inhibition is negligible (i.e., $g_2^X/g_1^X = 0$), all output neurons acquire the principal component of the response probability matrix Q, and the other information is neglected [82] [177] [4]. On the other hand, if lateral inhibition is effective, different output neurons may acquire various components of the external structure. I first examined that point in a simple network model with two independent external sources (Fig. 4.2A). In the model, each external source drives an independent subgroup of input neurons (I defined those input neurons as A-neurons and B-neurons), which project excitatory inputs to all of the output neurons. Here, I assume that source A drives input neurons with a higher probability than source B ($q_A = 0.6$, $q_B = 0.5$), so that input neurons projected by source A show higher correlations ($c_A = 0.36$) than those receiving the output of source B ($c_B = 0.25$). In the

matrix form,

$$Q = \begin{pmatrix} q_A & 0 \\ 0 & q_B \\ 0 & 0 \end{pmatrix}, \; C = \begin{pmatrix} c_A & 0 & 0 \\ 0 & c_B & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

The third row in Q represents response probabilities of background neurons in the input layer (gray triangles in Fig. 4.2A; note that $C = QQ^t$). I refer to this as the minor source detection task below. Here, for lateral connections, I assumed that both excitatory-to-inhibitory (E-to-I) and inhibitory-to-excitatory (I-to-E) connections are well organized such that inhibition only works mutually between two output neuron groups (Fig. 4.2A; blue lines are E-to-I and red lines are I-to-E connections. See also equation (30) in Methods). The origin of these structured lateral connections is discussed later. When the network is excited by inputs from external sources, excitatory postsynaptic potential (EPSP) sizes of feedforward connections WX change according to STDP rules. Initially, in all output neurons, synaptic weights from A-neurons (blue triangles in Fig. 4.2A) become larger because A-neurons are more strongly correlated with one another than B-neurons are. However, as learning proceeds, one of the output neuron groups becomes selective for the minor source B (Fig. 4.2B). After 30 min, the network successfully learns both sources. If we focus on the peristimulus time histogram (PSTH) for the average membrane potential of output neurons aligned to external events, both neuron groups initially show weak responses to both correlation events, and yet the depolarization is relatively higher for source A than for source B (Fig. 4.2C left). After 10 min of learning, both neuron groups show relatively stronger initial responses for source A, but group 1 shows a hyperpolarization soon after the initial response (Fig. 4.2C middle). As a result, synaptic weights from A-neurons to group 1 become weaker, and group 1 neurons eventually become selective for the minor source B (Fig. 4.2C right). The mean cross-correlation (see *cross-correlation* in Methods for details) between the external sources and the population activity of output neurons is maximized when the delay is approximately 10-15 ms (Fig. 4.2E). If we fix the delay at 14 ms, then the cross-correlation gradually increases as the network learns both sources (Fig. 4.2D). The same argument holds if mutual information is used for performance evaluation (green lines in Figures 4.2D, 4.2E). Interestingly, the network better detects the minor source when it is learned with a highly correlated source compared with when it is learned with another minor source (Fig. 4.2F), because a highly correlated opponent source causes strong lateral inhibition on the output neurons, which enhances minor source learning. Similar results are also obtained for conductance-based leaky integrate-and-fire (LIF) neurons (Supplementary Figure 1).

Lateral inhibition should be strong, fast, and sharp

To investigate how and when the network can acquire multiple sources represented by correlated inputs, I further analyzed the model above (see *Mean-field approximation of a two-source model* in Methods for details). Because both output excitatory neurons and lateral inhibitory neurons are bundled into

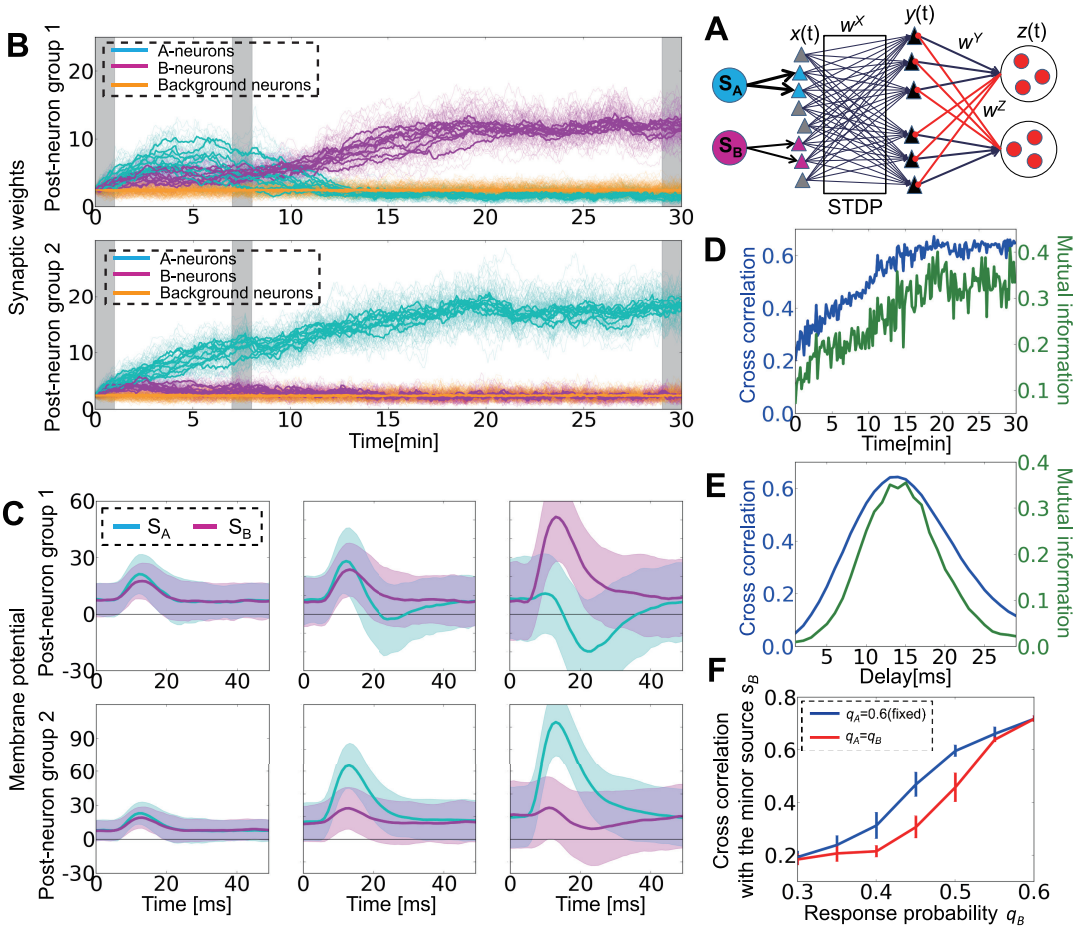


Figure 4.2. Lateral inhibition enables minor source detection by spike-timing-dependent plasticity (STDP) through membrane hyperpolarization. (**A**) Schematic figure of the simplified model. S_A and S_B (on the left side) are the sources that project to subsets of input neurons (colored triangles). Gray triangles are background neurons, black triangles (on the right) are output neurons, and red circles are inhibitory neurons. (**B**) Development of synaptic weights. Thick lines are mean synaptic weights from A-neurons (blue), B-neurons (red), and Background-neurons (orange) to each output neuron. Thin lines are traces of individual synaptic weights. Gray bar shows the timing at which **figure C** is calculated. (**C**) Peristimulus time histograms (PSTHs) of membrane potentials averaged within output neuron groups. $T = 0$ indicates the timing of events at external layers. The three figures are calculated from the data at t = 0-1 min, 7-8 min, and 29-30 min. (**D**) Development of mean cross-correlation and mutual information between external sources and population activity of output neurons for the simulation depicted in panels **B** and **C**. (**E**) Delay dependence of mean cross-correlation and mutual information. Both values were calculated from five simulations. (**F**) Cross-correlation between the output group that detected the minor source and the minor source activity for various response probabilities q_B with a fixed $q_A(= 0.6)$. When none of output groups detected the minor source, the larger value calculated for the two output groups was used. Throughout the study, error bars represent standard deviation calculated from five simulations, unless otherwise indicated.

groups, in the mean-field approximation, we can approximate M excitatory populations and N inhibitory populations into two representative output neurons and two inhibitory neurons. Similarly, input neurons can be bundled into three groups (A-neurons, B-neurons, and Background-neurons). In addition, I assumed that the synaptic connections from Background-neurons to output neurons are fixed because they showed little weight change in the simulation (orange lines in Fig. 4.2B). In this approximation, by inserting equation (32) into equation (29), the mean synaptic weight changes of feedforward connections follow

$$\begin{aligned}\frac{dw^X_{\mu\nu}}{dt} \simeq & \sum_{\nu'}^{L/L_a} L_a w^X_{\mu\nu'} \nu_o^S G_1^X(w^X_{\mu\nu}) \sum_{\rho} q_{\nu\rho} q_{\nu'\rho} - N_a w_Z M_a w_Y \sum_{\nu'=1}^{L/L_a} L_a w^X_{\bar{\mu}\nu'} \nu_o^S G_2^X(w^X_{\mu\nu}) \sum_{\rho} q_{\nu\rho} q_{\nu'\rho} \\ & + \bar{F}(w^x_{\mu\nu}) \left[\left(\nu_o^X\right)^2 \sum_{\nu'=1}^{L/L_a} L_a w^X_{\mu\nu'} - (\nu_o^X)^2 N_a w_Z M_a w_Y \sum_{\nu'=1}^{L/L_a} L_a w^X_{\bar{\mu}\nu'} + \left(N_a w_Z\right)^2 M_a w_Y \nu_o^X \nu_1^Z \right] \end{aligned} \quad (4.5)$$

where $\mu = 1, 2$ and $\bar{\mu} = 2, 1$ ($\mu \neq \bar{\mu}$), and $\nu = A, B$. The first two terms are correlation-based learning, and the last term is the homeostatic effect intrinsic to STDP [88]. G_1^X and G_2^X are coefficients determined by synaptic delays, EPSP/IPSP (Inhibitory postsynaptic potential) shapes, and correlation structure, as shown in equations (3) and (4). By solving the self-consistency condition (equation (34) in Methods), the firing rates of inhibitory neurons are approximated as

$$\begin{aligned}\nu_1^Z &= \frac{M_a w_Y \nu_o^X}{1 - \left(M_a w_Y N_a w_Z\right)^2} \left[\left(L_a w_{1A} + L_a w_{1B} + 2 L_a w_o^X\right) - \left(M_a w_Y N_a w_Z\right) \left(L_a w_{2A} + L_a w_{2B} + 2 L_a w_o^X\right) \right] \\ \nu_2^Z &= \frac{M_a w_Y \nu_o^X}{1 - \left(M_a w_Y N_a w_Z\right)^2} \left[\left(L_a w_{2A} + L_a w_{2B} + 2 L_a w_o^X\right) - \left(M_a w_Y N_a w_Z\right) \left(L_a w_{1A} + L_a w_{1B} + 2 L_a w_o^X\right) \right] \end{aligned} \quad (4.6)$$

I estimated the nullclines by calculating the lines that satisfy $\dot{w}\left(w_{1A}, w_{1B}, w^*_{2A}(w_{1A}, w_{1B}), w^*_{2B}(w_{1A}, w_{1B})\right) = 0$ for $\mu = A$ or B. As a result, I found that when the mutual inhibition is weak ($w_I = 10$), the system has only one stable point at which w_{1A} is larger than w_{1B} (Fig. 4.3A left). At this point, w_{2A} is also larger than w_{2B} ($w_{2A} = 9.64$, $w_{2B} = 3.60$; not shown in the figure), which means that both output neuron groups are specialized for the major source A (I call this state a winner-take-all state or T-state); however, if the inhibition is moderately strong ($w_I = 21.5$), two new stable fixed points and two unstable fixed points appear in the system (Fig. 4.3A middle). In the stable point on the left, neuron group 1 picks up source B while neuron group 2 picks up source A ($w_{2A} = 12.52$, $w_{2B} = 2.87$). On the right-hand side, neuron group 1 selects source A while neuron group 2 selects source B (I denote those two states as winners-share-all states or S-states below). At the stable point in the middle, both groups detect source A ($w_{1A} = w_{2A} = 9.47$, $w_{1B} = w_{2B} = 3.61$). Note that because of the mutual inhibition, the synaptic weight from A-neuron is smaller when both groups learn A than it is when only group 1 learns A. For strong inhibition ($w_I = 40.0$), the stable point in the middle disappears, and the system is stable only when two neuron groups detect different sources (Fig. 4.3A right). Simulation results confirm this analysis because strong inhibition indeed causes a winner-share-all state in which multiple neuron groups survive in competition [68], whereas the network tends to show a winner-take-all learning

when the inhibition is weak (Fig. 4.3B). I measured the degree of winner-share-all/winner-take-all states by defining the specialization index w_{SI} as

$$w'_{SI} = (w_{1A} - w_{1B})(w_{2B} - w_{2A}),\ w_{SI} = w'_{SI}/\sqrt{|w'_{SI}|} \tag{4.7}$$

If $w'_{SI} = 0$, I set $w_{SI} = 0$. If two output groups are specialized for different sources, w_{SI} becomes positive, whereas if two groups are specialized for the same source, w_{SI} becomes negative. When the synaptic delay in the lateral connections is small, only S-states are stable, whereas at longer delays, both S-states and T-states are stable. In the simulation, the network typically grows toward the latter state in the bistable strategy (Fig. 4.3C). Moreover, if we change the shape of the IPSP curve while keeping $\tau_B^Z = 5\tau_A^Z$, for steep IPSP curves (i.e., both τ_A^Z and τ_B^Z are small), only the S-states are stable, whereas T-states also become stable for slower IPSPs (Fig. 4.3D). Therefore, both analytical and simulation studies indicate that lateral inhibition should be strong, fast and sharp to detect higher correlation structure. Moreover, lateral inhibition does not need to be pathologically strong because the I/E balance of $N_a w_Z / L w_o^X \simeq 20\%$ is sufficient to cause multistability.

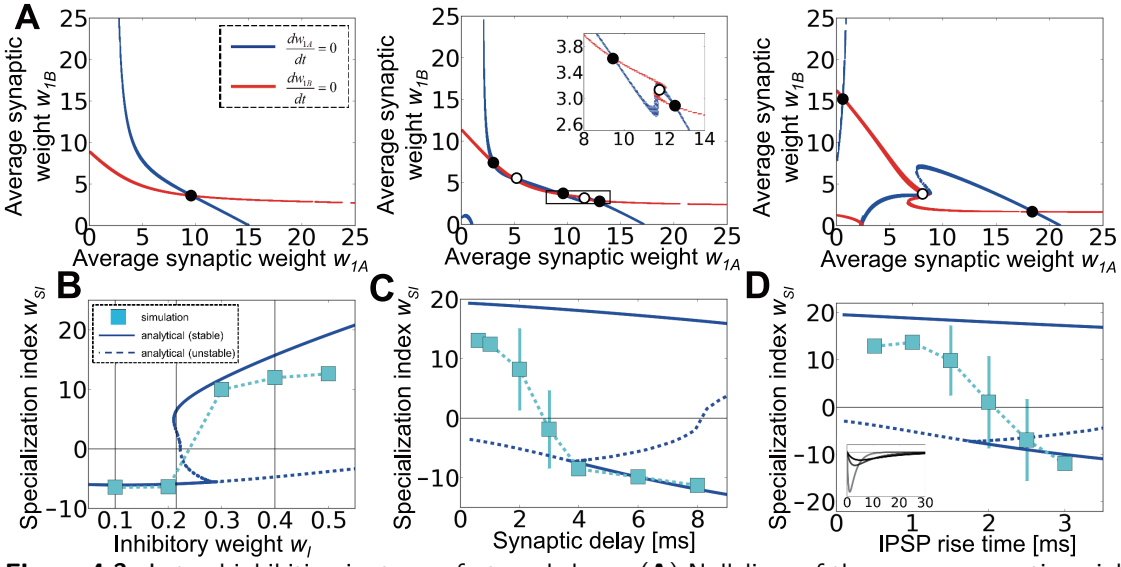


Figure 4.3. Lateral inhibition is strong, fast, and sharp. (**A**) Nullclines of the average synaptic weight changes at different inhibitory amplitudes $w_Z = 0.1, 0.215, 0.4$. The inset in the middle graph is a magnified view of boxed area. (**B**) Specialization indices wSI for various inhibitory weights. Positive w_{SI} indicates the winner-share-all state, whereas negative w_{SI} indicates the winner-take-all state. Blue lines are analytical estimations and cyan squares are the results of simulations. Vertical lines correspond to the values at which the nullclines in **Fig. 4.A** are calculated. (**C**) The same graphs for various synaptic delays. The average synaptic delay of both lateral excitatory $(d_{min}^Y + d_{max}^Y)/2$ and inhibitory $(d_{min}^Z + d_{max}^Z)/2$ connections was changed, while the variability was kept at $d_{max}^Y - d_{min}^Y = d_{max}^Z - d_{min}^Z = 1.0$ ms. (**D**) IPSP rise time dependence. The inset shows IPSP curves at $\{\tau_A^Z, \tau_B^Z\} = \{0.5, 2.5\}$ (gray line), $\{1.5, 7.5\}$ (dark gray line), and $\{2.5, 12.5\}$ (black line).

Optimal correlation timescale changes depend on the noise source

In the previous section, I revealed the effects of network properties for a fixed input correlation structure; however, actual neurons show various timescales for correlations depending on the brain region [42] [12]

and characteristics of the stimuli [150] [124], and it is largely unknown how different timescales influence correlation-driven learning. Therefore, I next considered the effect of correlation timescales, especially on noise tolerance. In my current model, input neurons respond to external sources with input kernel $\phi(t) = t^2 e^{-t/\theta_t}/2\theta_t^3$ (Fig. 4.4A left), and so the correlation between input neuron i and l is given as

$$C_{il}(s) = \nu_o^S \sum_{\mu=1}^{p} q_{i\mu} q_{l\mu} h(s)$$

By changing the parameter θ_t, I studied the effect of the correlation timescale on learning. The correlation is precise when θ_t is small, whereas it becomes broad at large values of θ_t (Fig. 4.4A right, Fig. 4.4B). Because STDP causes homeostatic plasticity that does not depend on a correlation, as shown in the third term of equation (5), in a more precise approximation, equation (2) should be written as

$$\dot{W}_X \approx W_X \left(g_1^X E - g_2^X W_Z W_Y \right) C^t + \langle \text{homeostatic term} \rangle. \tag{4.8}$$

I first calculated g_1^X and g_2^X at various θ_t. Both g_1^X and g_2^X become smaller for a larger θ_t, but decreases in g_2^X are slower than those in g_1^X, and, as a result, $\kappa = g_2^X/g_1^X$ becomes larger for a longer correlation timescale (Fig. 4.4C). This means that a longer temporal correlation is more suitable for the detection of multi-components. This is indeed confirmed in the simulation (Fig. 4.4D). When $\theta_t = 0.5$ and the minor component is slightly weaker than the major one ($c_A = 0.36$, $c_B = 0.25$), the minor component is no longer detectable. On the other hand, at $\theta_t = 2.0$, the minor component is detectable even if the strength of the induced correlation is less than half ($c_A = 0.36$, $c_B = 0.16$). At $\theta_t = 4.0$, g_1^X becomes smaller so that even the major signal is not fully detectable.

Similar results hold for crosstalk noise. In the model above, the noise is provided through the spontaneous Poisson firing of input neurons as random noise (Fig. 4.4E top, black dots are spikes caused by random noise). In reality, however, there would be crosstalk noise among input spike trains caused by the interference of external sources. I implemented this crosstalk noise by introducing non-diagonal components in the response probability matrix as

$$Q = \begin{pmatrix} q_S & q_N \\ q_N & q_S \\ 0 & 0 \end{pmatrix},$$

where q_S is the response probability to the preferred signal and q_N is that to the non-preferred signal (Fig. 4.4E bottom). I refer to this as the noisy source detection task below. To make a clear comparison, in the simulation of random noise, I kept $q_N = 0$ and changed the spontaneous firing rate of the input neurons (r_i^o) to modify the noise intensity, whereas in simulation of crosstalk noise I removed random noise (i.e., $r_i^o = 0$) and changed q_N. For random noise, a smaller θ_t enables better learning because a large g_1^X competes with the homeostatic force (Fig. 4.4F). By contrast, for crosstalk noise, the

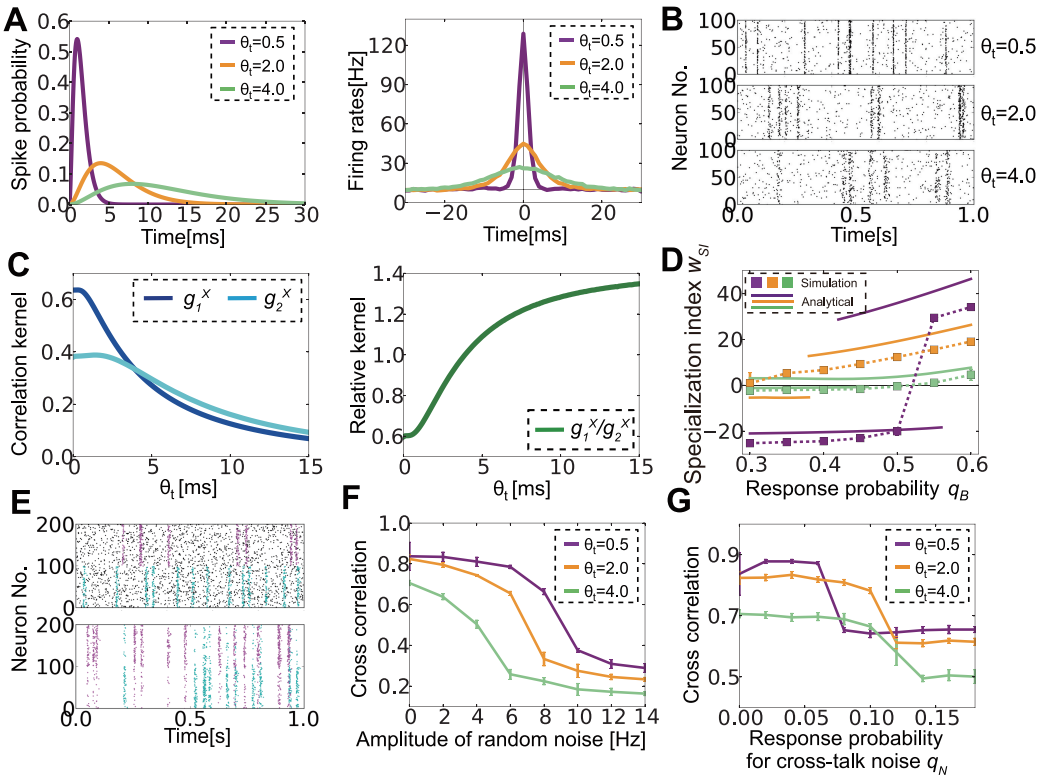


Figure 4.4. Optimal correlation timescale changes depending on noise characteristics. (**A**) Response kernels of input neurons to external events (left) and cross-correlation among input neurons responding to the same source calculated from simulated data (right) for three different correlation timescale parameters θ_t. (**B**) Raster plots of input neurons for various θ_t. Only 100 correlated neurons are plotted although there are 400 input neurons in total. (**C**) Analytically calculated correlation kernels g_1^X, g_2^X (left), and their ratio g_1^X/g_2^X. (**D**) Specialization index w_{SI} for various response probabilities q_B while fixing $q_A = 0.6$. Lines represent w_R at analytically estimated stable points, and dotted squares represent simulation results. (**E**) Raster plots of two types of noise. The upper panel shows random noise, whereas the lower panel depicts crosstalk noise. In both panels, the first 100 neurons respond primarily to the cyan source, and the next 100 neurons respond to the purple source. For random noise, the noise (black dots) is independent from the signals, whereas the crosstalk noise (purple dots in the lower half, cyan dots in the upper half) is correlated with the signal for the other population. (**F, G**) The effects of random noise (**F**) and crosstalk noise (**G**) at various correlation timescales.

performance is better at $\theta_t = 2.0$ than at $\theta_t = 0.5$ because strong lateral inhibition suppresses crosstalk noise (Fig. 4.4G). Although for small noise regimens, the network performs better at $\theta_t = 0.5$ than at $\theta_t = 2.0$, but the difference is almost negligible. Therefore, to cope with crosstalk noise, the spike correlation needs to be broad, whereas a narrow spike correlation is better for random noise. I note that qualitatively the same arguments as above also hold for the exponential kernel (Supplementary Figure 3D,E). However, the ratio of two coefficients (i.e., $\kappa_e = g^X_{e2}/g^X_{e1}$) is typically smaller for this kernel than for the kernel I used throughout this study (Supplementary Figure 3B,C vs. Fig. 4.4D) because lateral inhibition is less effective due to highly peaked spike correlation (Supplementary Figure 3A).

Excitatory and inhibitory STDP cooperatively shape structured lateral connections

To this point, I have considered a network already clustered into two assemblies that inhibit one another (as in Fig. 4.5A left). This means that the network somehow knows a priori that the number of external sources is two; however, in reality, a randomly connected network should also learn such information. To test this idea, I introduced STDP-type synaptic plasticity in lateral excitatory connections and feedback inhibitory connections and investigated how different STDP rules cause different structures in the circuit.

I first checked whether structured lateral connections were helpful for learning. For comparison, I also considered a model with random lateral connections in which all output neurons and inhibitory neurons are randomly connected with probability 0.5 (Fig. 4.5A middle). When lateral connections are random, mean-field equations are modified as

$$\begin{aligned}
\frac{dw^X_{\mu\nu}}{dt} &\simeq \sum_{\nu'=1}^{L/L_a} w^X_{\mu\nu'}\nu^S_o G^X_1(w^X_{\mu\nu})\sum_\rho q_{\nu\rho}q_{\nu'\rho} - N_a w_Z M_a w_Y \sum_{\mu'=1}^{p}\sum_{\nu'=1}^{L/L_a} L_a w^X_{\mu'\nu'}\nu^S_o G^X_2(w^X_{\mu\nu})\sum_\rho q_{\nu\rho}q_{\nu'\rho} \\
&\quad + \bar{F}(w^W_{\mu\nu})\left[(\nu^X_o)^2\sum_{\nu'=1}^{L/L_a} L_a w^X_{\mu\nu'} - (\nu^X_o)^2 N_a w_Z M_a w_Y \sum_{\mu'=1}^{p}\sum_{\nu'=1}^{L/L_a} L_a w^X_{\mu'\nu'} + (N_a w_z)^2 M_a w_Y \nu^X_o \nu^Z_{tot}\right], \\
\nu^z_{tot} &\equiv \frac{2M_a w_Y \nu^X_o (L_a w_{1A} + L_a w_{1B} + 2L_a w^x_o)}{1 + 2M_a w_y N_a w_Z}.
\end{aligned} \tag{4.9}$$

I separated lateral connections into two groups as in the previous case, but this approximation is legitimate only when two input sources are symmetrical (i.e., $q_A = q_B$). In other cases, neurons are often organized into two groups with different population sizes. In such cases, for evaluating performance, I measured average weights from source A on the output neurons receiving stronger inputs from A-neurons than from B-neurons or Background-neurons. For randomly connected lateral inhibition, learning performance dropped significantly in noisy source detection (Fig. 4.5B) and in minor source detection (Fig. 4.5C); thus clustered connectivity is indeed advantageous for learning.

I next investigated whether such structure can be learned using STDP rules. I first introduced Hebbian STDP for both E-to-I and I-to-E connections. With these learning rules, the lateral connections

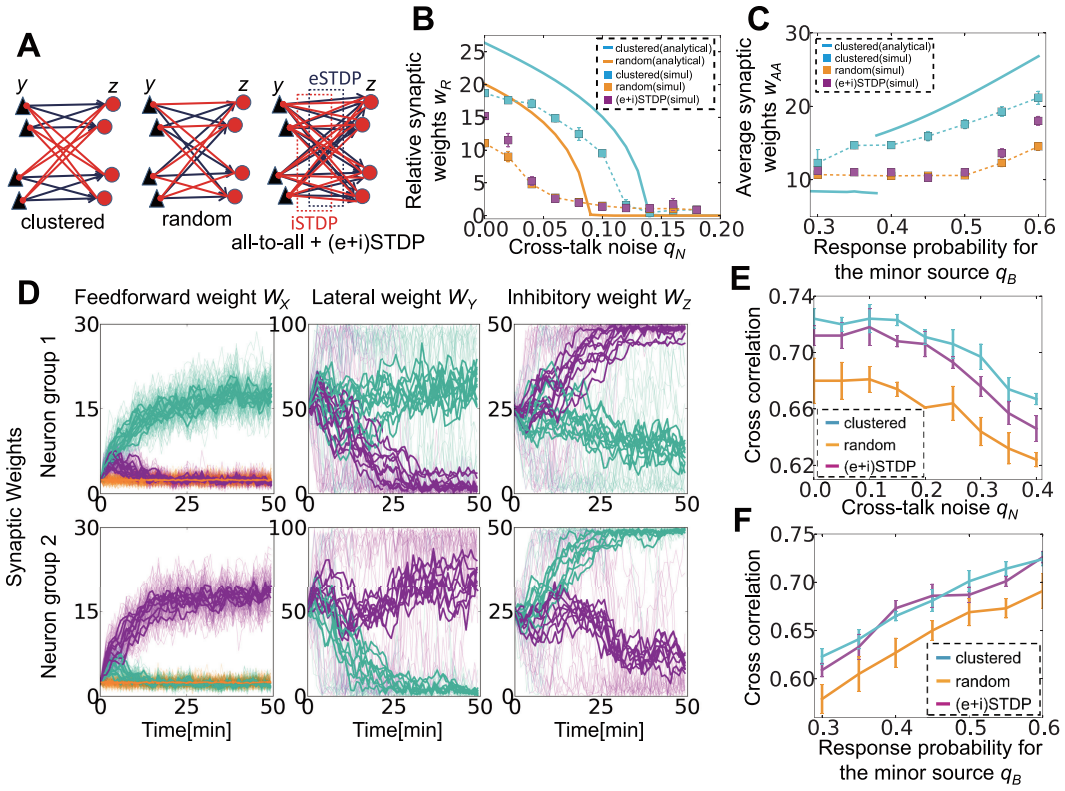


Figure 4.5. Lateral connection structuring by excitatory and inhibitory spike-timing-dependent plasticity (STDP). (**A**) Schematic figures of connections between the output layer and the lateral layer. In the simulation, each layer consists of 20 neurons. (**B**) The effect of crosstalk noise on different lateral structures. Analytical results are shown as bold lines, and the results from simulations are shown as dotted lines. (**C**) Minor source detection with different lateral structures. Because the specialization index is not well defined for a network with random lateral connections, the average synaptic weights from source A to those output neurons that prefer source A were measured instead. (**D**) Synaptic weight development at three connections. In the left and right columns, panels show synaptic weights of excitatory/inhibitory synapses projected to the neuron group 1 (top) and group 2 (bottom). In the middle graph, panels correspond to excitatory synapses projected from the neuron group 1 (top) and group 2 (bottom). In all panels, thin lines indicate the development of individual synapses, thick lines represent average weights onto output neurons, and colors indicate A-neurons (blue), B-neurons (red), and Background-neurons (orange). (**E, F**) Performance of the network with different lateral structures in noisy signal detection (**E**) and minor signal detection (**F**). Here (and only here), a pre-learned network is used to investigate responses for various inputs.

successfully learn a mutual inhibition structure (Fig. 4.5D); however, this learning is achievable only when the learning of a hidden external structure is possible from the random lateral connections (magenta lines in Fig. 4.5B, C; note that orange points are hidden by magenta points because they show similar behaviors in noisy cases), which means either when crosstalk noise is low or two sources have similar amplitudes. Nevertheless, once a structure is obtained in easy settings ($q_N = 0$ or $q_A = q_B$), that network outperforms the network with random lateral connections in both noisy source detection (Fig. 4.5E) and minor source detection (Fig. 4.5F). In Fig. 4.5E, I evaluated the performance of noisy source detection by first conducting STDP learning at $q_N = 0$, and then I terminated STDP and performed simulations at the various noise levels q_N. Similarly, in the minor source detection task depicted in Fig. 4.5F, I first performed STDP learning with $q_A = q_B = 0.6$, and then evaluated the performance for a smaller q_B. STDP can also generate similar lateral connection structures when the total number of input sources is larger than two (Supplementary Figures S2A, S2B). Therefore, STDP at lateral connections helps signal detection by efficiently organizing the connection structure.

I next studied the analytical conditions for learning of the clustered structure (see *Analytic approach for STDP in lateral and inhibitory connections* in Methods for details). The synaptic weight dynamics of lateral excitatory and inhibitory connections are approximately given as

$$\begin{aligned}\dot{W}_Y &\approx g_1^Y W_Y W_X C^t W_X^t,\ g_1^Y \equiv \int_{-\infty}^{\infty} ds F^Y(s) \int D_r^X \int D_u^Y \int D_{r'}^X h(u + r' - s - r) \\ \dot{W}_Z &\approx g_1^Z W_X C W_X^t W_Y^t,\ g_1^Z \equiv \int_{-\infty}^{\infty} ds F^Z(s) \int D_r^X \int D_u^Y \int D_{r'}^X h(r - s - u - r' - d_z - d_x)\end{aligned} \tag{4.10}$$

Both equations represent indirect effects of the input correlation propagated into the lateral circuit. From a linear analysis, we can expect that when g_1^Y is positive, E-to-I connections tend to be feature selective (see equation (35) in Methods). Each inhibitory neuron receives stronger inputs from one of the output neuron groups and, as a result, shows a higher firing rate for the corresponding external signal. On the other hand, if g_1^Z is positive, I-to-E connections are organized in reciprocal form, where one of the reciprocal connections is enhanced and the other is suppressed (see equation (36) in Methods). We can evaluate feature selectivity of inhibitory neurons by

$$\varphi^Y = \frac{1}{N} \sum_{k=1}^{N} \left(\frac{1}{|\Omega_A^Y|} \sum_{j \in \Omega_A^Y} w_{kj}^Y - \frac{1}{|\Omega_B^Y|} \sum_{j \in \Omega_B^Y} w_{kj}^Y \right) \cdot \left(\frac{1}{M} \sum_{j=1}^{M} w_{kj}^Y \right)^{-1} \tag{4.11}$$

where Ω_A^Y and Ω_B^Y are the sets of excitatory neurons responding preferentially to sources A and B, respectively. Indeed, when the LTD time window is narrow, analytically calculated g_1^Y tends to take negative values (the green line in Fig. 4.6A), and E-to-I connections organized in the simulation are not feature selective (the blue points in Fig. 4.6A). By contrast, for a long LTD time window (i.e., when LTD is weakly spike-timing dependent), g_1^Y tends to take positive values, and E-to-I connections become clustered. In the simulation, W_Z is also plastic, but as shown in equation (10), the effect of

W_Z on the plasticity of W_Y is negligible in first-order approximations.

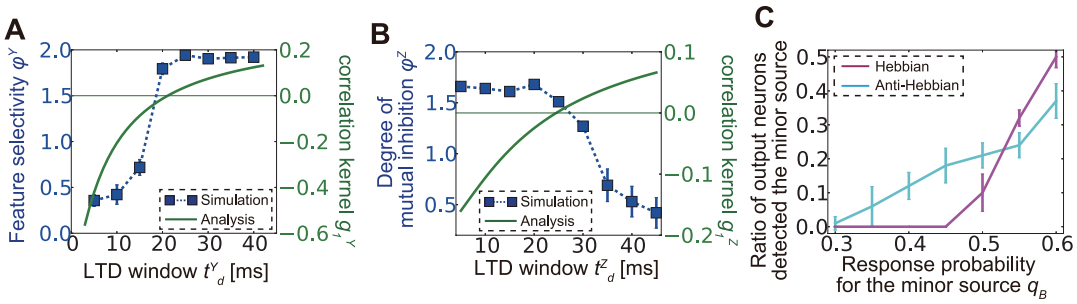


Figure 4.6. Correlation propagation shapes lateral connection structure (**A**) Comparison between feature selectivity (blue dots) calculated from simulation results and analytically calculated correlation kernel function g_1^Y (green line) for lateral excitatory connections. Thin green horizontal line represents$g_1^Y = 0$. (**B**) Comparison between the degree of mutual inhibition (blue dots) calculated from the simulation and analytically calculated correlation kernel g_1^Z (green line) for lateral inhibitory connections. Negative g_1^Z is correlated with a high degree of mutual inhibition, as expected (see Methods). (**C**) Ratio of output neurons tuned for the minor source in a minor source detection task under Hebbian and anti-Hebbian inhibitory spike-timing-dependent plasticity.

Similarly, for I-to-E connections, I measure the degree of mutual inhibition (non-reciprocity) with

$$\varphi^Z = \frac{1}{N}\sum_{k=1}^{N}\left|\frac{w_{kj}^Y}{\sum_{j=1}^{M} w_{kj}^Y} - \frac{w_{jk}^Z}{\sum_{j=1}^{M} w_{jk}^Z}\right| \tag{4.12}$$

When LTD is strongly spike-timing dependent, g_1^Z is negative and φ^Z calculated from the simulation data tends to be large (Fig. 4.6B), which means that inhibitory connections are organized such that the inhibition functions as mutual inhibition between excitatory neuron groups. Note that the organized neuronal wiring patterns are not a pure product of the pre-post causality of STDP but the effect of spike correlations propagating through lateral inhibitory circuits. If the structural plasticity is merely caused by the pre-post causality, both φ^Y and φ^Z can decrease with increases in the inhibitory population while maintaining the total synaptic weights because the causal effect becomes weaker as each synaptic weight becomes smaller [119]; however, in my simulations, the values of both quantities generally increased for larger inhibitory populations (Supplementary Figure S2C).

Hebbian inhibitory STDP at lateral connections is not always beneficial for learning. For example, in minor source detection, if I use Hebbian inhibitory STDP, a slightly minor source is not detectable, whereas for anti-Hebbian STDP, a small number of neurons still detect the minor source because reciprocal connections from strong-source responsive inhibitory neurons to strong-source responsive output neurons inhibit synaptic weight development for the stronger source (Fig. 4.6C).

Neural Bayesian ICA and blind source separation

My results to this point have revealed that correlation-based STDP learning combined with lateral inhibition can successfully detect signals from mixed inputs masked by noises. To confirm this mechanism is indeed effective in realistic tasks, I applied the above method to blind source separation. I first examined

the condition in which the network could capture external sources. I extended the previous network to include four independent sources mixed at the input layer (Fig. 4.7A). In the present application, I used structured lateral connections because learning for clustered structures is difficult with noisy stimuli, as shown in the preceding section. The response probability matrix Q and correlation matrix C are given as

$$Q = \begin{pmatrix} q_S & q_N & 0 & q_N \\ q_N & q_S & q_N & 0 \\ 0 & q_N & q_S & q_N \\ q_N & 0 & q_N & q_S \end{pmatrix},\ C = \begin{pmatrix} q_S^2 + 2q_N^2 & 2q_S q_N & 2q_N^2 & 2q_S q_N \\ 2q_S q_N & q_S^2 + 2q_N^2 & 2q_S q_N & 2q_N^2 \\ 2q_N^2 & 2q_S q_N & q_S^2 + 2q_N^2 & 2q_S q_N \\ 2q_S q_N & 2q_N^2 & 2q_S q_N & q_S^2 + 2q_N^2 \end{pmatrix}.$$

Therefore, the principal components of matrix Q (i.e., eigenvectors of C) are {1, 1, 1, 1,}, {-1, 0, 1, 0}, {0, -1, 0, 1}, {-1, 1, -1, 1}. Because the first-order approximation of synaptic weight dynamics follows $\dot{W}_X \approx g_1^X W_X C^t$, we may expect that synaptic weight vectors converge to the eigenvectors of the principal components; however, this was not the case in my simulations, even if we took into account the non-negativity of synaptic weights (see Fig. 4.7B, where I renormalized the principal vectors to [0, 1]). Instead, each weight vector converged to a column of the response probability matrix Q (Fig. 4.7B, the left panel is the projection to the first two dimensions, and the right panel is the projection to the other two dimensions). This result implies that the network can extract independent sources, rather than principal components, from multiple intermixed inputs.

I next evaluated the performance of hidden external source detection, especially its tolerance against crosstalk noise. To this end, I compared the performance of the model with that of the Bayesian ICA algorithm, in which independence of external sources is treated as a prior [193] [128]. In the algorithm, the learned mixing matrix may converge to its Bayesian optimal value estimated from a stream of inputs. Although we cannot directly argue the optimality of cross-correlations, if the mixing matrix is accurately estimated, external activity is also well inferred, and thus we can use the mean cross-correlation as a measure for the optimality of learning. In terms of discretized input activity X, the external source activity S and prior information I, we can express the conditional probability of the estimated response probability matrix $\tilde{Q}$ as $P[\tilde{Q}|X,I] = \frac{P[\tilde{Q}|I}{P[X|I]} \int P[X|S,\tilde{Q},I]P[S|I]dS$ (see *Bayesian ICA* in Methods for details). This means that even if no prior information is given for $\tilde{Q}$ itself (i.e. $P[\tilde{Q}|I] = \text{const.}$), posterior $P[\tilde{Q}|X,I]$ still depends on a prior given for S. If we introduce a prior that each external source follows an independent Bernoulli Process (i.e. $P[S|I] = \prod_{k=1}^{T/\Delta t} \prod_{i=1}^{L} (r_s \Delta t)^{s_\mu^k} (1 - r_s \Delta t)^{1 - s_\mu^k}$), then the stochastic gradient descendent of posterior function is given as,

$$\frac{\partial}{\partial \tilde{q}_{i\mu}} \log P[\tilde{Q}|X,I] = \frac{1}{Z_p} \sum_{k=1}^{T/\Delta t} \int P[S,X|\tilde{Q},I] \frac{2x_i^k - 1}{x_i^k p_i^k/(1-p_i^k) + (1-x_i^k)} \frac{\sum_{k'=0}^{\infty} \phi_{k'} s_\mu^{k-k'}}{1 - \tilde{q}_{i\mu} \sum_{k'=0}^{\infty} \phi_{k'}^{k-k'}} dS,$$

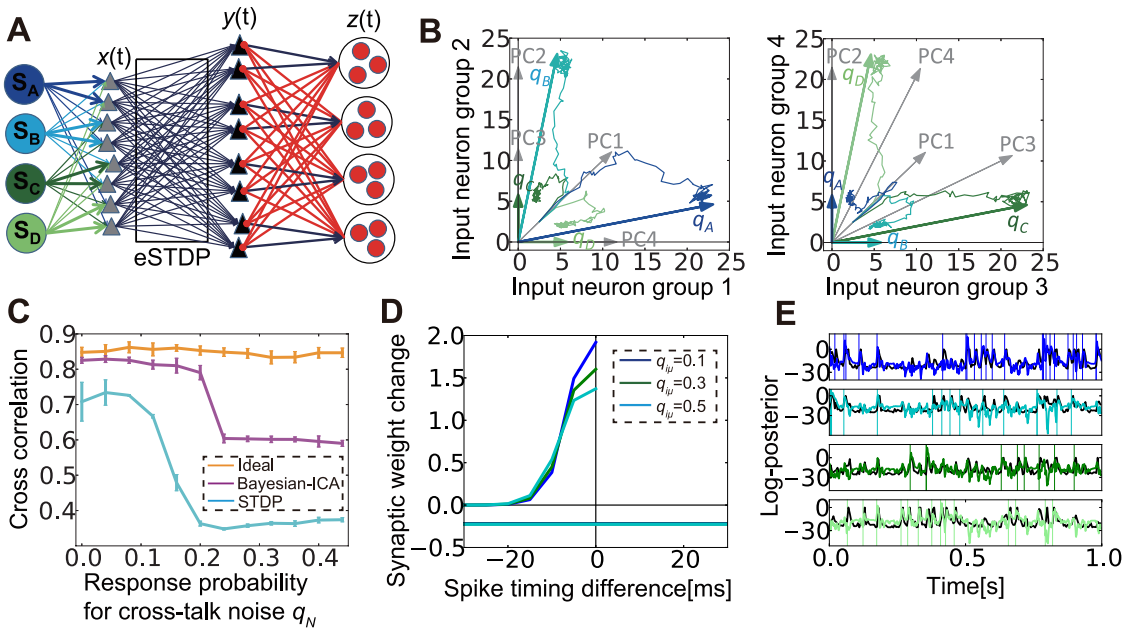


Figure 4.7. With lateral inhibition, spike-timing-dependent plasticity (STDP) mimics Bayesian independent component analysis (ICA). (**A**) Schematic figure of the model with four sources. (**B**) Synaptic weight development in input neuron space. Arrows q_A to q_D are response probability vectors of the four sources, and PC1 to PC4 are normalized principal components of the correlation matrix C. Lines represent traces of average synaptic weight from each input group to the output groups that learned corresponding sources during the learning process. (**C**) Comparison of performance among the ideal observer, Bayesian ICA learning, and STDP learning. (**D**) LTP/LTD time window of Bayesian ICA learning. (**E**) Behaviors of log-membrane potential (color lines) in the STDP model, and estimated log-posterior (black lines) in the Bayesian ICA algorithm for the same stimuli. Vertical lines represent timings of external events. Log-membrane potentials are normalized to align the mean and the variance to the corresponding log-posteriors.

where

$$p_i^k = 1 - (1 - r_i^o \Delta t) \prod_{\mu=1}^{p} \left[1 - \tilde{q}_{i\mu} \sum_{k'=0}^{\infty} \phi_{k'} s_\mu^{k-k'} \right], \ \phi_k = \frac{1}{2\theta_t^3} \left[(k+1/2)\Delta t \right]^2 \exp\left[-(k+1/2)\Delta t/\theta_t \right]$$

I approximated this Bayesian ICA algorithm by a sequential sampling source activity instead of calculating the integral over all possible combinations in the estimation of the log-posterior of the response probability matrix Q. In this approximation, the learning rule of the estimated response probability matrix $\tilde{Q}$ obeys

$$\begin{aligned} \Delta \tilde{q}_{i\mu}^k &\propto \frac{2x_i^k - 1}{x_i^k p_i^k(Y^{1:k-1})/(1 - p_i^k(Y^{1:k-1})) + (1 - x_i^k)} \times \frac{\sum_{k'=0}^{\infty} \phi_{k'} y_\mu^{k-k'}}{1 - \tilde{q}_{i\mu}^k \sum_{k'=0}^{\infty} \phi_{k'} y_\mu^{k-k'}} \\ p_i^k(Y^{1:k-1}) &= 1 - (1 - r_i^o \Delta t) \prod_{\mu=1}^{p} \left[1 - \tilde{q}_{i\mu}^k \sum_{k'=0}^{\infty} \phi_{k'} y_\mu^{k-k'} \right], \end{aligned} \tag{4.13}$$

where Y is the sampled sequence, and $p_i^k(Y^{1:k-1})$ is the sample based approximation of p_i^k in the previous equation. This rule has spike-timing and weight dependence similar to those seen in STDP (Fig. 4.7D). Although the performance of STDP is much worse than the ideal case (when the true Q is given), this performance is similar to that for the sample-based learning algorithm discussed above (Fig. 4.7C). Therefore, the network detects independent sources if crosstalk noise is not large. I further studied the response of the models for the same inputs and found that the logarithm of the average membrane potential $u_\mu^E = \frac{1}{|\Omega_\mu|} \sum_{j \in \Omega_\mu} u_j^E$ well approximates the log-posterior estimated in Bayesian ICA, even in the absence of a stimulus (Fig. 4.7E). This result suggests that in the STDP model, expected external states are naturally sampled through membrane dynamics that are generated through the interplay of feedforward and feedback inputs.

I finally performed the blind separation task using the same network as shown in Fig. 4.7A. I created "sensory" inputs by mixing four artificially created auditory sequences (Fig. 4.8A). In the auditory cortex, various frequency components of a sound, particularly high-frequency components, are represented by specific neurons typically organized in a tonotopic map structure [205], whereas low-frequency components are expected to be perceived as a change in sound pressure. Furthermore, populations of neurons in the primary auditory cortex are known to synchronize the relative timing of their spikes during auditory stimuli and provide correlated spike inputs for higher cortical areas in which the auditory scene is fully analyzed and perceived [52] [10]. I modeled these features by assuming that input neurons have a preferred frequency $\{f_i\}$ defined as

$$f_i = \exp\left[\frac{i}{L} (\log f_{max} - \log f_{min}) + \log f_{min} \right],$$

and auditory inputs are provided as time-dependent response probabilities, which follow $q_i(t) = q_o \sum_q a_l^q(t) a_h^q(f_i)$, where $a_h^q(f)$ is the spectrum of auditory source q (left panel of Fig. 4.8C), and $a_l^q(t)$ is the temporal change of the sound pressure (black lines in Fig. 4.8B). In this representation, each sound source

is represented by correlated spikes of neural populations (right panel of Fig. 4.8C). Even if signals have overlapping frequency components $\{a_h^q(f)\}_q$, blind separation is possible as long as $\{a_l^q(t)\}_q$ are independent and have sharp rising profiles sufficient to cause spike correlations. After learning, four output neuron groups successfully detected changes in the sound pressure of the four original auditory signals (colored lines in Fig. 4.8B) by correctly identifying the input neurons that encoded the signals. Therefore, STDP rules implemented in a feedforward neural network with lateral inhibition serve as a spike-based solution to the blind source separation or cocktail party effect problem.

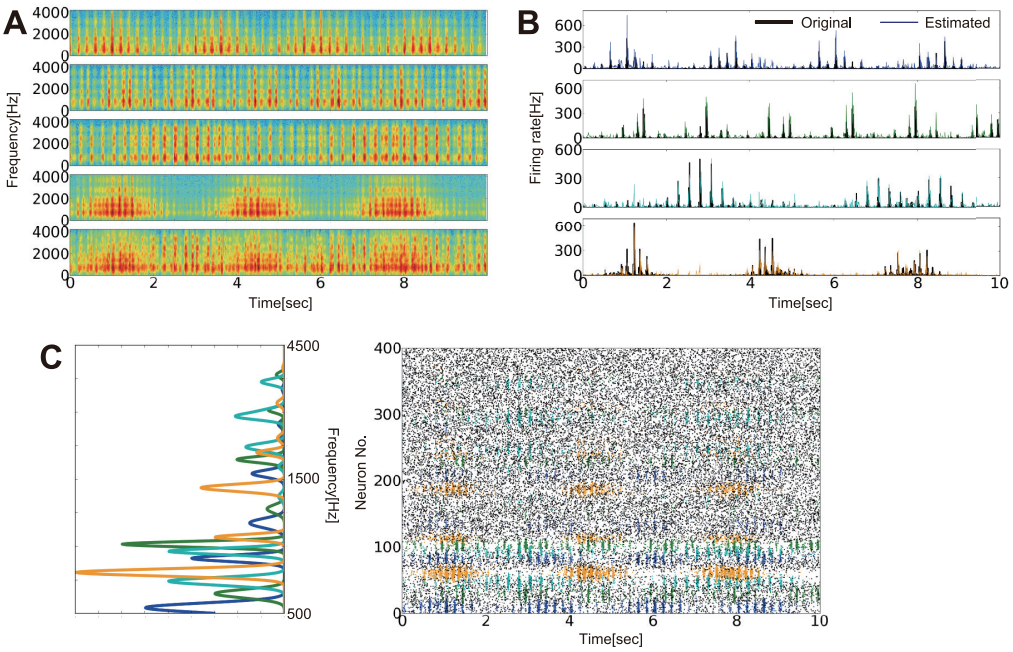


Figure 4.8. Blind source separation by spike-timing-dependent plasticity (STDP). (**A**) Four original auditory signals (from the top to the fourth set of signals) and one mixed signal (bottom). (**B**) Amplitudes of original signals (black lines) and those estimated from output firing rates (colored lines). (**C**) Spectra of auditory sources $a_h^q(f)$ (left). Raster plots of input neuron activity. Colors were probabilistically assigned based on expected sources. All figures were calculated from the 30'00"-30'10" portion of the auditory signals and simulation.

Discussion

By analytically investigating the propagation of input correlations through feedback circuits, I revealed how lateral inhibition influenced plasticity at feedforward connections. I showed that a population of neurons could learn multiple signals with different strengths or mixed levels. In addition, I found that to perform learning from signals corrupted with random noise, the timescale of the input correlations needed to be in the range of milliseconds, whereas the timescale was broader for crosstalk noise, which may explain why the spike correlation of cortical neurons often exhibits a large jitter (approximately 10 ms) [12] [132]. I also investigated the functional roles of STDP at lateral excitatory and inhibitory connections to demonstrate that Hebbian STDP shaped the lateral structure to improve signal detection performance. The results also suggested that anti-Hebbian plasticity was helpful for learning from minor

sources and implied that different STDP rules at lateral connections induced different algorithms at feedforward connections. Furthermore, I derived an STDP-like online learning rule by considering an approximation of Bayesian ICA with sequence sampling. This result suggested that lateral inhibition adjusted the membrane potentials of postsynaptic neurons so that their spiking processes accurately performed sequence sampling. I also demonstrated that this mechanism was applicable to blind source separation of auditory signals.

Noise characteristics and correlation timescales

Simultaneously recorded neurons in close proximity often show correlated spiking, yet the precision of these correlations varies across brain regions. Neurons in the lateral geniculate nucleus show strong spike correlations [42] [47], while correlations in V1 [132] [221] or higher visual areas [12] are less precise. My results indicate the interesting possibility that these differences may reflect the different characteristics of the noise with which the various cortical areas need to contend. At an early stage of sensory processing, the major noise component may be environmentally produced background noise from various sources; thus precise spike correlation is beneficial at this stage for noise reduction during signal detection and learning (Fig. 4.4G). By contrast, in higher sensory cortices, crosstalk noise accumulated through signal propagation in circuits may form the primary noise source, so less precise spike correlation is preferable (Fig. 4.4H). It would be intriguing to examine whether lower and higher cortical areas similarly change the strength of spike correlations for other sensory modalities.

STDP in E-to-I and I-to-E connections

It is known that both glutaminergic synapses on inhibitory neurons [146] [66] and GABAergic synapses on excitatory neurons [243] [89] show STDP, and it is also known that STDP at E-to-I connections plays an important role in developmental plasticity [248]; however, detailed properties of these plasticities are still largely disputable [133] [231] and, reportedly, highly dependent on inhibitory cell type [172], neuromodulator [107], and region [133]. I showed that in a feedback circuit, Hebbian inhibitory STDP preferred winner-take-all while anti-Hebbian inhibitory STDP tended to cause winner-share-all (see Fukai and Tanaka 1997 for winner-share-all) at excitatory neurons (Fig. 4.6D). This result indicates that different inhibitory STDP imposes different functions for excitatory STDP, which suggests that a neural circuit may select optimal inhibitory STDP for a specific purpose or strategy of learning, and this may differ across regions and be modified by neuromodulators. A recent study showed that inhibitory plasticity even directly influences the plasticity at excitatory synapses of the postsynaptic neuron [237]. In such cases, algorithm selection would play a more important role than it did for the standard STDP implemented in my model.

Recently, inhibitory neurons in the rodent hippocampus CA1 were shown to display context-dependent activity rate changes during a spatial learning task, in association with the activity rate changes in ex-

citatory cells [59]. In addition, the authors suggested the candidate mechanism for this change in activity is STDP at E-to-I synapses. My results examining E-to-I STDP confirmed this configuration of inhibitory cells modulated by plasticity at feedforward excitatory connections (Figures 4.5D, S2A, S2B). In my model, although inhibitory neurons are not directly projected from input sources, as excitatory neurons learn a specific input source (Fig. 4.5D, left panel), inhibitory neurons acquire feature selectivity through Hebbian STDP at synaptic connections from those excitatory neurons (Fig. 4.5D, middle panel). Furthermore, my results indicate an important function of these feature-selective inhibitory neurons. Once an adequate circuit structure is learned and inhibitory connections are organized into a feature-selective pattern, even if the input to the network becomes noisy or faint, the network can still robustly detect signals (Figures 4.5E, 4.5F). This robustness would be useful for spatial learning, as contextual information is often uncertain.

STDP and Bayesian ICA

Results above indicated that STDP in a lateral inhibition circuit mimicked Bayesian ICA [193] [128]. First, output neurons were able to detect hidden external sources, without capturing principal components (Fig. 4.7B). Previous results suggest that for a single output neuron, an additional homeostatic competition mechanism is necessary to detect an independent component [82] [43]. In addition, when information is coded by firing rate, homeostatic plasticity is critically important, because STDP itself does not mimic Bienenstock-Cooper-Munro learning [203]. However in my model, information was encoded by correlation, and mutual inhibition naturally induced intercellular competition so that intracellular competition through homeostatic plasticity was unnecessary. Moreover, my analytical results suggested the reason that independent sources are detected. To perform a principal components analysis using neural units, the synaptic weight change needs to follow

$$\dot{W}_X = W_X C - LT[W_X C W_X^t] W_X,$$

where LT[] means lower triangle matrix [201] [178]. This LT transformation protects principal components caused by the lateral modification from higher order components; however in my model, because all output neurons receive the same number of inhibitory inputs (equation 2), all neurons are decorrelated with one another and develop into independent components.

Recently, it was shown that STDP can perform Bayesian optimal learning [171] [90]. In the model used by those authors, the synaptic weight matrix is treated as a hyper parameter and estimated by considering the maximum likelihood estimation of input spike trains. By contrast, in the Bayesian ICA framework, the mixing matrix (corresponding to synaptic weight matrix) is treated as a probabilistic variable. Using this framework, we needed to calculate an integral over all possible source activities in the past to derive stochastic gradient descendent; however, as shown in Fig. 4.7C, the stochastic learning was well performed by employing an approximation with sequential sampling. Moreover, I

naturally derived an adequate LTP time window from the response kernel of input neurons to external events (Fig. 4.7D). I also found that STDP self-organized a lateral circuit structure that performed better than a random global inhibition (Figures 4.5E, 4.5F). Mathematically, to perform sampling from a probabilistic distribution, we first needed to calculate the occurrence probability of each state; however, in a neural model, membrane potentials of output neurons approximately represent the occurrence probability through membrane dynamics. In machine learning methods, integration over possible source activities is often approximated using Markov chain Monte Carlo (MCMC) sampling [167]. Interestingly, a recent study showed that a recurrent network performed MCMC sampling [27] [222], suggesting that my network may perform a more accurate sampling in the presence of recurrent excitatory connections.

Suboptimality of STDP

Previous theoretical results suggest that STDP can modulate synaptic weights in a way that optimizes information transmission between pre- and postsynaptic neurons [222] [97]. In the Bayesian ICA framework, blind source separation can be formulized as an optimization problem, but, in this case, the problem itself is ill-defined because optimality does not guarantee the true solution. In addition, local minima are often unavoidable for online learning rules. Nevertheless, the problems faced by the brain are often ill-defined, and suboptimality is inevitable [18]. Because I performed both nonlinear dynamics-based and machine learning-based analyses, I can offer some insights regarding the origins of local minima in stochastic gradient descendent learning. In the initial state, synaptic weights are typically homogeneously distributed, and this state is often locally stable. As a result, the homogeneous stable point is more likely to be selected in learning (Figures 4.2C, 4.2D) than the non-homogenous, more desirable, points; however, introducing additional noise may change this situation. Indeed, in Figures 4.4B and 4.7C, the performance of the model was improved by adding a small amount of noise to input activities, although the improvement was not significant; however, because a large amount of noise is harmful for computations and stable learning, the benefit of noise addition is highly limited, and the brain may recruit other mechanisms for near optimal learning.

Neural mechanism of blind source separation

Humans and nonhuman animals can detect a specific auditory sequence from a mixed, noisy auditory stimulus, a phenomenon often called the cocktail party effect. The mechanism underlying the cocktail party effect remains elusive [162] [41] [95], although several solutions have been proposed [232] [8]. An effective solution for this problem is ICA [44] [19] [6], and the neural implementation of the algorithm has been studied by several authors [164] [203] [114] [83]. My study extended these results through a rigorous analytical treatment on biologically plausible STDP learning of spiking neurons, and my analyses enabled us to discover interesting functions of correlation coding. Moreover, by explicitly modeling inhibitory neurons, I found that STDP at E-to-I and I-to-E connections cooperatively organized

a lateral structure suitable for blind source separation. In addition, I successfully extended a previous model for the formation of static visual receptive fields [203] [120] to a more dynamic model in an auditory blind source separation task. In realistic auditory scene analysis, the frequency spectrum of acoustic signals is first analyzed in the cochlea, where each frequency component is the mixture of sound components from independent sources. Components belonging to the same source may be separated and integrated by downstream auditory neurons for the perception of the original signal. These frequency components can be considered a mixed signal in the ICA problem [211]; thus even if signals are mixed in frequency space, if the amplitudes of the signals are temporally independent, blind separation is still achievable. In the neural implementation of the problem, if two frequencies are commonly activated in the same signal, neurons representing those frequencies show spike correlation under the presence of the signal; thus the learning process is naturally achieved by STDP learning. These results indicate an active role of spike correlation and STDP in efficient biological learning.

Methods

Model

Neural dynamics Based on the previous study [82], I constructed a network model with one external layer and three layers of neurons (Fig. 4.1A). The first layer is the external layer that corresponds to external stimuli or the sensory system's response to these stimuli. For simplicity, I approximated the activity of external sources using a Poisson process with the constant rate ν_o^S. If I define the Poisson process with rate r as $\hat{\sigma}(r)$, the activity of the external source μ at time t is written as $s_\mu(t) = \hat{\sigma}(\nu_o^S)$ (see Table 1 for the list of variables). Neurons in the input layer fire spikes in response to activity in the external layer. By assuming a rate-modulated Poisson process, the spiking activity of the input neuron i follows

$$x_i(t) = \hat{\sigma}\left[r_i^o + \sum_{\mu=1}^{p} q_{i\mu} \int_0^\infty \phi(t') s_\mu(t-t') dt'\right], \tag{4.14}$$

where r_i^o is the instantaneous firing rate defined with $r_i^o = \nu_o^x - \sum_{\mu=1}^p q_{i\mu}\nu_o^S$, $q_{i\mu}$ is the response probability for the hidden external source μ, and $\phi(t) = t^2 e^{-t/\theta_t}/2\theta_t^3$ is the response kernel for each external event. In most theoretical studies, cross-correlations give an exponential decay or a delta function [88] [76], but here I used a response kernel that produces broader correlations (Fig. 4.4A right), because the actual correlations observed in the cortex are usually not sharply peaked [12] [132]. For instance, for the exponential kernel $\phi_e(t) = e^{-t/\theta_t}/\theta_t$, correlations show a peaked distribution even if the timescale parameter θ_t is several milliseconds (Supplementary Figure 3A). Because of the common inputs from the external layer, input neurons show highly correlated activity, which enables population coding of the hidden structure. Although here I explicitly assumed the presence of the external layer, these analytical results can also be applied for arbitrary realization of a spatiotemporal correlation.

Output neurons are modeled with the Poisson neuron model [88] [76] [119] in which the membrane

potential of neuron j at time t is described as

$$u_j^E(t) = \sum_{j=1}^{M} w_{ji}^X \int_0^\infty \epsilon_X(r) x_i(t - r - d_{ji}^X) dr - \sum_{k=1}^{N} w_{jk}^Z \int_0^\infty \epsilon_Z(r) z_k(t - r - d_{jk}^Z) dr, \tag{4.15}$$

where w_{ji}^X and w_{jk}^Z are the EPSPs/IPSPs of input currents from input neuron x_i and lateral neuron z_k, respectively, convolution functions are defined as $\epsilon_X(r) = \frac{e^{-r/\tau_A^X} - e^{-r/\tau_B^X}}{\tau_A^X - \tau_B^X}$ and $\epsilon_Z(r) = \frac{e^{-r/\tau_A^Z} - e^{-r/\tau_B^Z}}{\tau_A^Z - \tau_B^Z}$, and synaptic delays in the feedforward excitatory and feedback inhibitory connections are d_{ij}^X and d_{jk}^Z. For feedforward excitatory connections, the synaptic delay d_{ij}^X is given by the sum of the axonal delay d_{ij}^a and dendritic delay d_{ij}^d, whereas for inhibitory connections, I assume for simplicity that the delay is purely axonal. The response of the output neuron follows $y_j(t) = \hat{\sigma}\left[g_E(u_j^E)\right]$. Similarly, inhibitory neurons in the lateral layer show Poisson firing based on the membrane potential $\{u_k^I\}_{k=1,\ldots,N}$ which is defined as

$$u_k^I(t) = \sum_{j}^{M} w_{kj}^Y \int_0^\infty \epsilon_Y(r) y_j(t - r - d_{kj}^Y) dr, \tag{4.16}$$

for EPSPs of a lateral connection w_{kj}^Y, convolution function $\epsilon_Y(r) = \frac{e^{-r/\tau_A^Y} - e^{-r/\tau_B^Y}}{\tau_A^Y - \tau_B^Y}$, and synaptic delay of the lateral connection d_{kj}^Y. The synaptic delay of the excitatory lateral connection is also approximated as the axonal delay. The spiking activity of the inhibitory neurons is given with $z_k(t) = \hat{\sigma}\left[g_I(u_k^I)\right]$. For analytical tractability, I use a linear response curve $g_E(u) = u$ and $g_I(u) = u$.

Synaptic Plasticity For most of this study, I focused on synaptic plasticity in the feedforward connection W_X, with fixed lateral synaptic weights W_Y and W_Z. When the timing of the spikes at the cell bodies of pre- and postsynaptic neurons is t_{pre} and t_{post}, spike timings at the synaptic sites are $t_{pre}^s = t_{pre} + d_{ji}^a$ and $t_{post}^s = t_{post} + d_{ji}^d$ with axonal and dendritic delays of d_{ji}^a and d_{ji}^d. For every pair of t_{pre}^s and t_{post}^s, synaptic weight change is given with

$$\Delta w_{ji}^X = \begin{cases} \eta^X f_p(w_{ji}^X) \exp\left[-(t_{post}^s - t_{pre}^s)/\tau_p\right] & (\text{if } \; t_{post}^s > t_{pre}^s) \\ \eta^X f_d(w_{ji}^X) \exp\left[-(t_{pre}^s - t_{post}^s)/\tau_d\right] & (\text{if } \; t_{post}^s < t_{pre}^s) \end{cases}. \tag{4.17}$$

For the synaptic weight dependence of STDP, I considered a pairwise log-STDP [81] in which LTP/LTD follows

$$f_p(w) = C_p(1 + \sigma_{stdp}\xi)e^{-w/(\beta w_o)}, \; f_d(w) = -C_d(1 + \sigma_{stdp}\xi)\frac{\log(1 + \alpha w/w_o)}{\log(1 + \alpha)} \tag{4.18}$$

where ξ is a Gaussian random variable. The log-weight dependence well replicates experimentally observed synaptic weight distributions [214] [31] and is suggested to have an important function in memory modulation [101]. Analytical treatment below is applicable to other types of synaptic weight dependence, yet in the additive STDP (i.e. $f_p(w) = C_p$ and $f_d(w) = C_d$), the mean-field equation typically does not have any stable fixed point. In addition, under the multiplicative STDP in which

LTD has a linear rather than a logarithmic dependence on synaptic weight, strong correlation is often necessary to induce salient LTP [81]. The coefficients $C_p = 1$ and $C_d = C_p\tau_p/\tau_d$ are chosen so that total LTP and LTD are balanced around the referential synaptic weight.

The STDP at E-to-I connections and I-to-E connections is similarly defined. For simplicity, I assume that synaptic delays are solely axonal (i.e., $d^Y_{k,j} = d^{Y,a}_{k,j}$,$d^Z_{j,k} = d^{Z,a}_{j,k}$), and the change in synaptic weight does not depend on the synaptic weight. To maintain the balance between LTP and LTD, coefficients are chosen as $C^Y_p = 1$, $C^Y_d = \gamma^Y C^Y_p \tau^Y_p/\tau^Y_d$, $\eta^Y = 0.3\eta w^Y_o/w^X_o$. Similarly, for I-to-E connections, $C^Z_p = 1$, $C^Z_d = \gamma^Z C^Z_p \tau^Z_p/\tau^Z_d$, $\eta^Z = 0.3\eta w^Z_o/w^X_o$. I also modify constant (initial) synaptic weights to $w^Y_o = 50.0$ and $w^Z_o = 25.0$, and bounded synaptic weights with $w^Y_{max} = 100.0$ and $w^Z_{max} = 50.0$. In this normalization, the total lateral inhibition takes the same value as that in the non-plastic model at the initial state. Time windows are defined as $\tau^Y_p = \tau^Y_d = \tau^Z_p = \tau^Z_d = 20.0$ ms.

In Fig. 4.6C, anti-Hebbian STDP was calculated by

$$\Delta w^Q = \begin{cases} -\eta^Q \exp\left[-(t^s_{post} - t^s_{pre})/\tau^Q_d\right] & (\text{if } \ t^s_{post} > t^s_{pre}) \\ \eta^Q\gamma^Q(\tau^Q_d/\tau^Q_p)\exp\left[-(t^s_{pre} - t^s_{post})/\tau^Q_p\right] & (\text{if } \ t^s_{post} < t^s_{pre}) \end{cases}$$

for Q = Y or Z. Similarly, the correlation detector type of STDP in Supplementary Figure 2 was defined as

$$\Delta w^Q = \begin{cases} \eta^Q\left(\exp\left[-(t^s_{post} - t^s_{pre})/\tau^Q_p\right] - (\tau^Q_p/\tau^Q_d)\exp\left[-(t^s_{post} - t^s_{pre})/\tau^Q_d\right]\right) & (\text{if } \ t^s_{post} > t^s_{pre}) \\ \eta^Q\gamma^Q\left(\exp\left[-(t^s_{pre} - t^s_{post})/\tau^Q_p\right] - (\tau^Q_p/\tau^Q_d)\exp\left[-(t^s_{pre} - t^s_{post})/\tau^Q_d\right]\right) & (\text{if } \ t^s_{pre} > t^s_{post}) \end{cases}$$

The anti-correlation detector was calculated by changing the sign of above equations.

Leaky Integrate-and-Fire (LIF) Model In the main text, I performed all simulations with a linear Poisson model for analytical purposes, although I also confirmed those results with a conductance-based LIF model (Supplementary Figure 1). In the LIF model, the membrane potentials of excitatory neurons follow

$$\begin{aligned} \frac{dv^E_j}{dt} &= -\frac{1}{\tau^E_m}\left(v^E_j - V_L\right) - g^{EE}_j\left(v^E_j - V_E\right) - g^{EI}_j\left(v^E_j - V_I\right), \\ \frac{dg^{EE}_j}{dt} &= -\frac{g^{EE}_j}{\tau^{EE}_s} + \sum_{i=1}^{L} w^X_{ji}\sum_{t^s_i}\delta(t - t^s_i), \text{ and } \frac{dg^{EI}_j}{dt} = -\frac{g^{EI}_j}{\tau^{EI}_s} + \sum_{k=1}^{N} w^{Zd}_{jk}\sum_{t^s_k}\delta(t - t^s_k). \end{aligned}$$

where g^{EE}_j and g^{EI}_j are excitatory and inhibitory conductances, respectively, and t^s_i and t^s_k are the spike timings of input neuron i and lateral neuron k. Similarly, for inhibitory neurons in the lateral layer,

$$\frac{dv^I_k}{dt} = -\frac{1}{\tau^I_m}\left(v^I_k - V_L\right) - g^{IE}_k\left(v^I_k - V_E\right) - g^{II}_k\left(v^I_k - V_I\right),$$

$$\frac{dg_k^{IE}}{dt} = -\frac{g_k^{IE}}{\tau_s^{IE}} + \sum_{i=1}^{L} w_{kj}^{Y} \sum_{t_j^s} \delta(t - t_j^s), \text{ and } \frac{dg_k^{II}}{dt} = -\frac{g_k^{II}}{\tau_s^{II}} + w_{II} \sum_{t_r^s} \delta(t - t_r^s).$$

In addition to the excitatory inputs from the output layer, I added random inhibitory inputs as Poisson processes with a fixed firing rate r_o^{II} for inhibitory neurons. A neuron fires if the membrane potential exceeds the threshold V_{th}, and immediately goes into a refractory period in which the membrane potential stays at V_{ref} for 1 ms after spiking. Plasticity was implemented for w_{ji}^X in the same manner as that for the Poisson model. Parameters were chosen as $V_L = -70.0$, $V_E = 0.0$, $V_I = -80.0$, $V_{ref} = -60.0$, $V_{th} = -50.0$ mV, $t_m^E = 20.0$, $t_m^I = 10.0$, $t_s^{EE} = 5.0$, $t_s^{EI} = 2.5$, $t_s^{IE} = 4.0$,$t_s^{II} = 5.0$ ms, $w_o^X = 0.001$, $w_o^I = 0.008$, $w_o^L = 1.0$, $r_o^{II} = 1000.0$ Hz, $w_o^{II} = 0.005$, $C_d = 1.8 C_p \tau_p^X / \tau_d^X$, and $\alpha = 50.0$. All other parameters were the same as those used in the Poisson model (Table 2).

In the LIF model, synaptic weights develop in a manner similar to that for the linear Poisson model, although change occurs more rapidly (Figures 4.1B, S1A). Both cross-correlation and mutual information behave as they do in the Poisson model, but the performance is slightly better, possibly because the dynamics are deterministic (Figures 4.1D, 4.1E, S1B, S1C); however, membrane potentials show different responses for correlation events (Figure S1D) because output neurons are constantly in high-conductance states, so that correlation events immediately cause spikes. As a result, membrane potentials drop to the V_{ref}, and the average potential goes down. Interestingly, after neuron groups detect different signals, a preferred signal initially causes hyperpolarization due to firing, but, subsequently, a non-preferred signal causes hyperpolarization due to lateral inhibition (Fig. 4.1D right). The PSTH of firing shows that the behavior of the membrane potential in the Poisson model is similar (Figures 4.1C and S1E). This is natural, because in the linear Poisson model, the firing rate has linear relationship with the membrane potential, whereas in LIF model relationship between the average membrane potential and firing rate is highly non-linear.

Bayesian ICA If discretized with Δt, the time series of the external source activity is written as $S = \{s_{\mu k}\}_{\mu=1,\ldots,p}^{k=1,\ldots,T/\Delta t}$, and input activity becomes $X = \{x_{ik}\}_{i=1,\ldots,L}^{k=1,\ldots,T/\Delta t}$. Therefore, for prior information I, the joint probability of sources S and the estimated response probability matrix Q is

$$P\left[S, \tilde{Q} | X, I\right] = P\left[X | S, \tilde{Q}, I\right] P\left[S, \tilde{Q} | I\right] / P\left[X | I\right]$$

Therefore, by considering marginal probability,

$$P[\tilde{Q}|X, I] = \frac{P[\tilde{Q}|I]}{P[X|I]} \int P[X|S, \tilde{Q}, I] P[S|I] dS. \tag{4.19}$$

By considering maximum likelihood estimation for a given prior $P[S|I]$, Q can be optimally estimated [193] [128]. In my problem setting, by assuming that external signals are independent, and input

neurons respond to signals with a Bernoulli process,

$$P[X|S,\tilde{Q},I]=\prod_{k=1}^{T/\Delta t}\prod_{i=1}^{L}\left[x_i^k p_i^k+(1-x_i^k)(1-p_i^k)\right],\ P[S|I]=\prod_{k=1}^{T/\Delta t}\prod_{i=1}^{L}(r_s\Delta t)^{s_\mu^k}(1-r_s\Delta t)^{1-s_\mu^k},$$

where

$$p_i^k=1-(1-r_i^o\Delta t)\prod_{\mu=1}^{p}\left[1-\tilde{q}_{i\mu}\sum_{k'=0}^{\infty}\phi_{k'}s_\mu^{k-k'}\right],\ \phi_k=\frac{1}{2\theta_t^3}\left[(k+1/2)\Delta t\right]^2\exp\left[-(k+1/2)\Delta t/\theta_t\right].$$

Therefore, log-likelihood becomes

$$\log P[\tilde{Q}|X,I]=\log\left(\int dS\prod_{k=1}^{T/\Delta t}\left[\prod_{i=1}^{L}\left(x_i^k p_i^k+(1-x_i^k)(1-p_i^k)\right)\times\prod_{\mu=1}^{p}(r_s\Delta t)^{s_\mu^k}(1-r_s\Delta t)^{1-s_\mu^k}\right]\right). \tag{4.20}$$

By taking gradient descendent,

$$\frac{\partial}{\partial\tilde{q}_{i\mu}}\log P[\tilde{Q}|X,I]=\frac{1}{Z_p}\sum_{k=1}^{T/\Delta t}\int P[S,X|\tilde{Q},I]\frac{2x_i^k-1}{x_i^k p_i^k/(1-p_i^k)+(1-x_i^k)}\frac{\sum_{k'=0}^{\infty}\phi_{k'}s_\mu^{k-k'}}{1-\tilde{q}_{i\mu}\sum_{k'=0}^{\infty}\phi_{k'}s_\mu^{k-k'}}dS.$$

Therefore, we need to calculate the integral over all possible combinations of sources in the past to obtain stochastic gradient descendent; however, such a calculation is computationally difficult and incompatible with neural computation. Instead, I used sequential sampling of $Y=\{y_{\mu k}\}_{\mu=1,\ldots,p}^{k=1,\ldots,T/\Delta t}$, which is randomly sampled from

$$\begin{aligned}P[y^k=s^k]&\propto P[s^k,x^k|Y^{1:k-1},\tilde{Q},I]\\&=\prod_{i=1}^{L}\left(x_i^k p_i^k(s^k,Y^{1:k-1})+(1-x_i^k)\left(1-p_i^k(s^k,Y^{1:k-1})\right)\right)\\&\quad\times\prod_{\mu=1}^{p}(r_s\Delta t)^{y_\mu^k}(1-r_s\Delta t)^{1-y_\mu^k},\end{aligned} \tag{4.21}$$

where

$$p_i^k(y^k,Y^{1:k-1})=1-(1-r_i^o\Delta t)\prod_{\mu=1}^{p}\left[1-\tilde{q}_{i\mu}\sum_{k'=0}^{\infty}\phi_{k'}y_\mu^{k-k'}\right].$$

Note in the above equations, x^k is given as a fixed value and not a random variable. Under this sample-based approximation, the stochastic gradient descendent follows

$$\Delta q_{i\mu}^k\propto\frac{2x_i^k-1}{x_i^k p_i^k(y^k,Y^{1:k-1})/(1-p(y^k,Y^{1:k-1}))+(1-x_i^k)}\times\frac{\sum_{k'=0}^{\infty}\phi_{k'}y_\mu^{k-k'}}{1-\tilde{q}_{i\mu}\sum_{k'=0}^{\infty}\phi_{k'}y_\mu^{k-k'}} \tag{4.22}$$

For Fig. 4.7C, I discretized the activity of hidden sources and input neurons with 5 ms bins, and performed learning with a learning rate $\eta^{SGD}=0.001$. Cross-correlation was evaluated using the sample sequence Y. For the ideal case, we performed sequential sampling from the true response probability Q.

If $y_\mu^{k-k'}=1$ and $y_\mu^{k-k''}=0$ for all other nearby $k''(\neq k')$, and if $q_{i\nu}=0$ for all $\nu(\neq\mu)$, then LTP at the connection $q_{i\mu}$ caused by an output spike $y_\mu^{k-k'}=1$ for $x_i^k=1$ is written as

$$\Delta q_{i\mu}^{k,k',LTP}=\frac{\left(1-\left[r_o^X-r_o^S\tilde{q}_{i\mu}\right]\right)\left(1-\tilde{q}_{i\mu}\phi_{k'}\right)}{1-\left(1-\left[r_o^X-r_o^S\tilde{q}_{i\mu}\right]\right)\left(1-\tilde{q}_{i\mu}\phi_{k'}\right)}\times\frac{\phi_{k'}}{1-\tilde{q}_{i\mu}\phi_{k'}}. \tag{4.23}$$

In the absence of the input spike ($x_i^k=0$), an output spike $y_\mu^{k-k'}=1$ causes LTD in total $\Delta q_{i\mu}^{LTD}=-\sum_{k'=0}^{\infty}\frac{\phi_{k'}}{1-\tilde{q}_{i\mu}\phi_{k'}}$. Therefore, this learning rule has weight dependence and temporal dependence similar to those in STDP. In Fig. 4.7D, I plotted $\Delta q_{i\mu}^{k,k',LTP}$ and $\Delta q_{i\mu}^{LTD}$ for different $\tilde{q}_{i\mu}$ ($\tilde{q}_{i\mu}=0.1,0.3,0.5$).

Blind source separation In the blind source separation task, I created the original source by calculating high-frequency and low-frequency components separately. First, the spectrum of the signal q at a high frequency was defined as

$$a_h^q(f)=\sum_i\sum_k\frac{a_{h,i}^q b_{h,k}^q}{\sqrt{2\pi}\sigma_{hf,k}}\exp\left[-(f-kf_{h,i}^q)^2/(2\sigma_{hf,k}^2)\right],$$

where $f_{h,i}^q$ is a characteristic frequency of signal q, and $kf_{h,i}^q$ are the harmonics of that frequency. The standard deviation was defined as $\sigma_{hf,k}=k\sigma_{hf}^o$ for $\sigma_{hf}^o=20$Hz. Low-frequency components were directly given as exponential oscillations as below.

$$a_l^q(t)=\frac{1}{Z_l}\exp\left[\beta_l\sum_i a_{l,i}^q\cos\left(2\pi f_{l,i}^q(t-\delta_{l,i}^q)\right)\right],$$

$f_{l,i}^q$ is a characteristic frequency, and $\delta_{l,i}^q$ is the delay. By combining these two components, the amplitude of a mixed sound is given as

$$a(t)=\sum_q a_l^q(t)\sum_i a_h^q(f_i)\cos\left(2\pi f_i(t-\delta_f^q)\right).$$

Summation over frequency f is performed using 400 representative values that correspond to the tuned frequency of each input neuron:

$$f_i=\exp\left[\frac{i}{L}(\log f_{max}-\log f_{min})+\log f_{min}\right].$$

In neural implementation, input neurons were stimulated with the response probability $q_i(t)=q_o\sum_q a_l^q(t)a_h^q(f_i)$ where $q_o=0.05$.

In the simulated example, for high-frequency components, I defined $f_{h,i}^q=\{\{523.3,784.0\},\{587.4,880.0\},\{650.0,830.6\},\{698.5,932.4\}\}$, $a_{h,i}^q=\{\{0.6,0.4\},\{0.3,0.7\},\{0.5,0.5\},\{0.9,0.3\}\}$, $b_{h,k}^q=\{\{1.0,0.5,0.2,0.1\},\{1.0,0.5,0.3,0.2\},\{1.0,0.1,1.0,0.8\},\{1.0,0.8,0.1,0.1\}\}$, and $\sigma_{h,f}^o=20$ Hz. Each column represents four different sources. Similarly for low-frequency components, I used $f_{l,i}^q=\{\{0.4,5.0,10.0,40.0,88.0\},\{0.6,6.0,8.0,42.0,86.0\},\{0.2,4.0,7.5,44.0,84.0\},\{0.3,6.0,7.0,46.0,82.0\}\}$, $a_{l,i}^q=\{\{0.3,0.4,0.2,0.5,0.5\},$

{0.25,0.5,0.2,0.5,0.5}, {0.24,0.3,0.4,0.5,0.5}, {0.61,0.2,0.2,0.5,0.5}}, $\delta^q_{l,i} = \{\{1.0,0.25,0.65,0.17,0.01\}$, {3.0,0.12,0.32,0.13,0.02}, {7.8,0.55,0.40,0.11,0.03}, {4.5,0.22,0.71,0.07,0.05}}, $\beta_I = 5.0$, and $Z_l = 27.24$. I chose $f_{min} = 500$ Hz, $f_{min} = 4,500$ Hz, and δ^q_f was randomly selected from 0 to $1/f_{min}$. Fig. 4.8A was generated by performing Fourier transformations with 25 ms sliding bins at every 2.5 ms.

Details of the simulation Simulations were calculated using the Runge-Kutta method, with a 0.05 ms time step. Initial synaptic weights were randomly chosen with $w^Q_{ij} = w^Q_o(1+\sigma^{init}_W\zeta)$ for Q = X, Y, Z and a random Gaussian variable ζ. Similarly, synaptic delays were decided as $d^Q_{ij} = d^Q_{min} + (d^Q_{max} - d^Q_{min})\xi$ for a random variable ξ uniformly chosen from [0, 1].

Analytical consideration of synaptic weight dynamics

Correlation among input neurons Because input neurons receive common inputs from external sources, I define cross-correlation among input neurons as $C_{il}(s) \equiv \langle x_i(t)x_l(t-s)\rangle - \langle x_i(t)\rangle\langle x_l(t)\rangle$, and cross-correlation among input neurons satisfies

$$\begin{aligned} C_{il}(s) &= \left\langle \hat{\sigma}\left[r^o_i + \sum_{\mu=1}^{p} q_{i\mu}\int_0^\infty \phi(t')s_\mu(t-t')dt'\right] \times \hat{\sigma}\left[r^o_l + \sum_{\mu=1}^{p} q_{l\mu}\int_0^\infty \phi(t'')s_\mu(t-s-t'')dt''\right]\right\rangle - (\nu^X_o)^2 \\ &\cong \nu^S_o \sum_{\mu=1}^{p} q_{i\mu}q_{l\mu}\int_0^\infty dt'\int_0^\infty dt''\phi(t')\phi(t'')\delta(t'-t''-s) \\ &= \nu^S_o \sum_{\mu=1}^{p} q_{i\mu}q_{l\mu}\int_{\max(0,s)}^\infty dt'\phi(t')\phi(t'-s). \end{aligned} \tag{4.24}$$

When $\phi(t) = t^2e^{-t/\theta_t}/2\theta_t^3$, $C_{il}(s)$ becomes

$$C_{il}(s) = \nu^S_o \sum_{\mu=1}^{p} q_{i\mu}q_{l\mu}\frac{1}{16{\theta_t}^3}\left(s^2 + 3\theta_t|s| + 3{\theta_t}^2\right)e^{-|s|/\theta_t} = \nu^S_o \sum_{\mu=1}^{p} q_{i\mu}q_{l\mu}h(s),$$

where $h(s) \equiv \frac{1}{16\theta_t^3}\left(s^2 + 3\theta_t|s| + 3\theta_t^2\right)e^{-|s|/\theta_t}$.

Average synaptic weight velocity The synaptic weight dynamics defined above can be rewritten as

$$\frac{dw^X_{ji}}{dt} = x_i(t-d^{Xa}_{ji})\int_0^\infty F_d(w^X_{ji},s)y_j(t-s-d^{Xd}_{ji})ds + y_j(t-d^{Xd}_{ji})\int_0^\infty F_p(w^X_{ji},s)x_i(t-s-d^{Xa}_{ji})ds, \tag{4.25}$$

for $F_d(w^X_{ij},s) = f_d(w^X_{ij})e^{-s/\tau_d}$, $F_p(w^X_{ij},s) = f_p(w^X_{ij})e^{-s/\tau_p}$. By taking an average over a short period of time and also using a stochastic Poisson process, synaptic weight change follows

$$\begin{aligned} \left\langle \frac{dw^X_{ji}}{dt}\right\rangle &= \left\langle x_i(t-d^{Xa}_{ji})\int_0^\infty F_d(w^X_{ji},s)y_j(t-s-d^{Xd}_{ji})ds\right\rangle + \left\langle y_j(t-d^{Xd}_{ji})\int_0^\infty F_p(w^X_{ji},s)x_i(t-s-d^{Xa}_{ji})ds\right\rangle \\ &= \left\langle x_i(t-d^{Xa}_{ji})\int_{-\infty}^0 F_d(w^X_{ji},-s')y_j(t+s'-d^{Xd}_{ji})ds'\right\rangle + \left\langle y_j(t-d^{Xd}_{ji})\int_0^\infty F_p(w^X_{ji},s)x_i(t-s-d^{Xa}_{ji})ds\right\rangle \\ &= \left\langle \int_{-\infty}^0 F_d(w^X_{ji},-s')x_i(t'-s'-d^{Xa}_{ji})y_j(t'-d^{Xd}_{ji})ds'\right\rangle + \left\langle \int_0^\infty F_p(w^X_{ji},s)x_i(t-s-d^{Xa}_{ji})y_j(t-d^{Xd}_{ji})ds\right\rangle \end{aligned}$$

$$\cong \int_{-\infty}^{\infty} F(w_{ji}^X, s) \left\langle x_i(t-s-d_{ji}^{Xa}) y_j(t-d_{ji}^{Xd}) \right\rangle ds,$$

where

$$F(w,s) \equiv \begin{cases} F_p(w,s) & (\text{if } s \geq 0) \\ F_d(w,-s) & (\text{if } s < 0). \end{cases}$$

Therefore, by calculating the cross-correlation between pre-spikes x_i and post-spikes y_j, synaptic weight dynamics can be analytically estimated. Because the spike probability linearly depends on the membrane potential in my model, cross-correlation follows

$$\begin{aligned} \left\langle x_i(t-s-d_{ji}^{Xa}) y_j(t-d_{ji}^{Xd}) \right\rangle &\cong \left\langle x_i(t-s-d_{ji}^{Xa}) u_j^E(t-d_{ji}^{Xd}) \right\rangle \\ &\cong \sum_{l=1}^{L} w_{jl}^X \int_0^{\infty} dr \varepsilon_X(r) \left\langle x_i(t-s-d_{ji}^{Xa}) x_l(t-d_{ji}^{Xd}-r-d_{ji}^X) \right\rangle \\ &\quad - \sum_{k=1}^{N} w_{jk}^Z \int_0^{\infty} dr \varepsilon_Z(r) \left\langle x_i(t-s-d_{ji}^{Xa}) z_k(t-d_{ji}^{Xd}-r-d_{jk}^Z) \right\rangle. \end{aligned}$$

Since I define cross-correlation among input neurons as

$$C_{il}(s) \equiv \langle x_i(t) x_l(t-s) \rangle - \langle x_i(t) \rangle \langle x_l(t) \rangle,$$

the first term is written as

$$\sum_{l=1}^{L} w_{jl}^X \int_0^{\infty} dr \varepsilon_X(r) \left\langle x_i(t-s-d_{ji}^{Xa}) x_l(t-d_{ji}^{Xd}-r-d_{ji}^X) \right\rangle \cong \sum_{l=1}^{L} w_{jl}^X \left[\left(\nu_o^X\right)^2 + \int_0^{\infty} dr \varepsilon_X(r) C_{il}(r-s+2d_{Xd}) \right]. \tag{4.26}$$

This result is consistent with that in previous studies [88] [76] [119]. The analysis can be extended to the cross-correlation between an input neuron and a lateral inhibitory neuron as

$$\begin{aligned} &\left\langle x_i(t-s-d_{ji}^{Xa}) z_k(t-d_{ji}^{Xd}-r-d_{jk}^Z) \right\rangle \cong \left\langle x_i(t-s-d_{ji}^{Xa}) u_k^I(t-d_{ji}^{Xd}-r-d_{jk}^Z) \right\rangle \\ &\cong \sum_{m=1}^{M} w_{km}^Y \int_0^{\infty} dq \varepsilon_Y(q) \left\langle x_i(t-s-d_{ji}^{Xa}) y_m(t-d_{ji}^{Xd}-r-d_{jk}^Z-q-d_{km}^Y) \right\rangle \\ &\cong \sum_{m=1}^{M} w_{km}^Y \sum_{l=1}^{L} w_{ml}^X \int_0^{\infty} dq \varepsilon_Y(q) \int_0^{\infty} dr' \varepsilon_X(r') \left\langle x_i(t-s-d_{ji}^{Xa}) x_l(t-d_{ji}^{Xd}-r-d_{jk}^Z-q-d_{km}^Y-r'-d_{ml}^X) \right\rangle \\ &- \sum_{m=1}^{M} w_{km}^Y \sum_{n=1}^{N} w_{mn}^Z \int_0^{\infty} dq \varepsilon_Y(q) \int_0^{\infty} dr' \varepsilon_Z(r') \left\langle x_i(t-s-d_{ji}^{Xa}) z_n(t-d_{ji}^{Xd}-r-d_{jk}^Z-q-d_{km}^Y-r'-d_{mn}^Z) \right\rangle \\ &\cong \sum_{m=1}^{M} w_{km}^Y \sum_{l=1}^{L} w_{ml}^X \left[\left(\nu_o^X\right)^2 + \int_0^{\infty} dq \varepsilon_Y(q) \int_0^{\infty} dr' \varepsilon_X(r') C_{il}(r+q+r'-s+2d_{Xd}+d_Z+d_Y) \right] \\ &\quad - \sum_{m=1}^{M} w_{km}^Y \sum_{n=1}^{N} w_{mn}^Z \nu_o \nu_n^Z. \end{aligned} \tag{4.27}$$

Theoretically, expansion over a lateral connection should be performed infinite times to obtain the exact solution, but at each expansion, the delay caused by synaptic delay $d_Z + d_Y$ and EPSP/IPSP rise times is accumulated so that the effect on correlation rapidly becomes small, especially when the original input cross-correlation $C(t)$ is narrow; however, even if $C(t)$ is broad, the effect for learning is bounded

by the STDP time window. Therefore, higher order terms practically influence weight dynamics only through firing rates, so that by applying the approximation

$$\int_0^\infty dq\varepsilon_Y(q)\int_0^\infty dr'\varepsilon_Z(r')\left\langle x_i(t-s-d_{ji}^{Xa})z_n(t-d_{ji}^{Xd}-r-d_{jk}^Z-q-d_{km}^Y-r'-d_{mn}^Z)\right\rangle \cong \nu_o^X\nu_n^Z,$$

the last term can be obtained. In general, ν_n^Z is not analytically calculable, but by considering the balanced condition, it can be estimated. Therefore, the second term is given as

$$\begin{aligned}&\sum_{k=1}^N w_{jk}^Z\int_0^\infty dr\varepsilon_Z(r)\left\langle x_i(t-s-d_{ji}^{Xa})z_k(t-d_{ji}^{Xd}-r-d_{jk}^Z)\right\rangle\\ &\cong \sum_{k=1}^N w_{jk}^Z\sum_{m=1}^M w_{km}^Y\sum_{l=1}^L w_{ml}^X\left[(\nu_o^X)^2+\int_0^\infty dr\varepsilon_Z(r)\int_0^\infty dq\varepsilon_Y(q)\int_0^\infty dr'\varepsilon_X(r')C_{il}(r+q+r'-s+2d_{Xd}+d_Z+d_Y)\right]\\ &-\sum_{k=1}^N w_{jk}^Z\sum_{m=1}^M w_{km}^Y\sum_{n=1}^N w_{mn}^Z\nu_o^X\nu_n^Z.\end{aligned}$$

Therefore, if I denote

$$\begin{aligned}\Gamma_{il}^{X1}(w_{ji}^X) &\equiv \int_{-\infty}^\infty dsF(w_{ji}^X,s)\int_0^\infty dr\varepsilon_X(r)C_{il},\\ \Gamma_{il}^{X2}(w_{ji}^X) &\equiv \int_{-\infty}^\infty dsF(w_{ji}^X,s)\int_0^\infty dr\varepsilon_Z(r)\int_0^\infty dq\varepsilon_Y(q)\int_0^\infty dr'\varepsilon_X(r')C_{il}(r+q+r'-s+2d_{Xd}+d_Z+d_Y),\\ \bar{F}(w_{ji}^X) &\equiv \int_{-\infty}^\infty F(w_{ji}^X,s)ds,\end{aligned} \tag{4.28}$$

average synaptic weight dynamics satisfies

$$\begin{aligned}\left\langle\tfrac{dw_{ji}^X}{dt}\right\rangle &\cong \sum_{l=1}^L w_{jl}^X\Gamma_{il}^{X1}(w_{ji}^X)-\sum_{k=1}^N w_{jk}^Z\sum_{m=1}^M w_{km}^Y\sum_{l=1}^L w_{ml}^X\Gamma_{il}^{X2}(w_{ji}^X)\\ &+\bar{F}(w_{ji}^X)\left[\sum_{l=1}^L w_{jl}^X(\nu_o^X)^2-\sum_{k=1}^N w_{jk}^Z\sum_{m=1}^M w_{km}^Y\sum_{l=1}^L w_{ml}^X(\nu_o^X)^2+\sum_{k=1}^N w_{jk}^Z\sum_{m=1}^M w_{km}^Y\sum_{n=1}^N w_{mn}^X\nu_o^X\nu_n^Z\right].\end{aligned} \tag{4.29}$$

The first two terms are Hebbian terms that depend on correlation by Γ^{X1} and Γ^{X2}, whereas the remainders are homeostatic terms. In all terms, synaptic weight dependence is primarily caused by w_{ji}^X and not by other synapses. By inserting the explicit representation of correlation into the equation above, Γ^{X1} and Γ^{X2} can be rewritten as

$$\begin{aligned}\Gamma_{il}^{X1}(w_{ji}^X) &= \nu_o^S G_1^X(w_{ji}^X)\sum_{\mu=1}^p q_{i\mu}q_{l\mu},\ \Gamma_{il}^{X2}(w_{ji}^X)=\nu_o^S G_2^X(w_{ji}^X)\sum_{\mu=1}^p q_{i\mu}q_{l\mu},\\ G_1^X(w_{ji}^X) &\equiv \int_{-\infty}^\infty dsF(w_{ji}^X,s)\int_0^\infty dr\varepsilon_X(r)\int_{\max(0,r-s+2d_{Xd})}^\infty dt'\phi(t')\phi(t'-(r-s+2d_{Xd})),\\ G_2^X(w_{ji}^X) &\equiv \int_{-\infty}^\infty dsF(w_{ji}^X,s)\int_0^\infty dr\varepsilon_Z(r)\int_0^\infty dq\varepsilon_Y(q)\int_0^\infty dr'\varepsilon_X(r')\\ &\quad\times\int_{\max(0,t'')}^\infty dt'\phi(t')\phi(t'-(r+q+r'-s+2d_{Xd}+d_Z+d_Y)),\end{aligned} \tag{4.30}$$

where $t''=r+q+r'-s+2d_{Xd}+d_Z+d_Y$. Note that G_1^X and G_2^X do not depend on any indexes of the neurons, except for synaptic weight dependency, and so the two values are considered basic constants

that decide how correlation shapes learning.

If I ignore the homeostatic term, then the synaptic weight dynamic is written in the matrix form as $\dot{W}_X \approx (W_X C^t).G_1^X - (W_X W_Z W_Y C^t).G_2^X$, where the dot product is defined as $(A.B)_{ij} = A_{ij}B_{ij}$. Especially if I approximate G_1^X and G_2^X with $g_1^X \equiv G_1^X(w_o^X)$ and $g_2^X \equiv G_2^X(w_o^X)$ (or if weight dependence is negligible as in additive-STDP), $\dot{W}_X \approx W_X \left(g_1^X E - g_2^X W_Z W_Y\right) C^t$.

The correlation kernel χ_1^X was derived from

$$\begin{aligned} G_1^X(w_{ji}^X) &= \int_{-\infty}^{\infty} dsF(w_{ji}^X, s) \int_0^{\infty} dr\varepsilon_X(r) \int_{\max(0,r-s+2d_{Xd})}^{\infty} dt' \int_{-\infty}^{\infty} d\tau \\ &\quad \times\phi(t')\phi(t'-(r-s+2d_{Xd}))\delta\left(\tau-(r-s+2d_{Xd})\right) \\ &= \int_{-\infty}^{\infty} d\tau \int_{-\tau+2d_{Xd}}^{\infty} dsF\left(w_{ji}^X, s\right)\varepsilon_X\left(s-2d_{Xd}\right) \int_{\max(0,\tau)}^{\infty} dt'\phi(t')\phi(t'-\tau) \\ &= \int_{-\infty}^{\infty} \chi_1^X(\tau; w_{ji}^X)h(\tau)d\tau \end{aligned} \tag{4.31}$$

where $\chi_1^X(\tau; w) = \int_{-\tau+2d_{Xd}}^{\infty} dsF(w,s)\,\varepsilon_X(\tau+s-2d_{Xd})$, and $h(\tau;\theta_t) \equiv \int_{\max(\tau,0)}^{\infty} dt'\phi(t')\phi(t'-\tau)$. The second correlation kernel χ_2^X was calculated in a similar way.

Mean-field approximation of a two-source model If the correlation structure $C(s)$ is simply organized, further analytical consideration is possible. In the two-source model shown in Fig. 4.2A, lateral connections are structured non-reciprocally, and EPSP/IPSP sizes are constants. The synaptic weight matrices are written as

$$W_{km}^Y = \begin{cases} w_Y & (\text{ if } \lfloor k/N_a \rfloor = \lfloor m/M_a \rfloor) \\ 0 & (\text{ otherwise }) \end{cases}, \; W_{jk}^Z = \begin{cases} w_Z & (\text{ if } \lfloor j/M_a \rfloor \neq \lfloor k/N_a \rfloor) \\ 0 & (\text{ otherwise }). \end{cases} \tag{4.32}$$

Therefore, the original L x M differential equations can be reduced into 2 x 2 equations of representative neurons as

$$\begin{aligned} \frac{dw_{\mu\nu}^X}{dt} &\cong \sum_{\nu'=1}^{L/L_a} L_a w_{\mu\nu'}^X \nu_o^S G^X(w_{\nu\nu'}^X) \sum_{\rho} q_{\nu\rho}q_{\nu'\rho} - N_a w_Z M_a w_Y \sum_{\nu'=1}^{L/L_a} L_a w_{\bar{\mu}\nu'}^X \nu_o^S G^Y(w_{\nu\nu'}^X) \sum_{\rho} q_{\nu\rho}q_{\nu'\rho} \\ &\quad + \bar{F}(w_{\mu\nu}^X)\left[(\nu_o^X)^2 \sum_{\nu'=1}^{L/L_a} L_a w_{\mu\nu'}^X - (\nu_o^X)^2 N_a w_Z M_a w_Y \sum_{\nu'=1}^{L/L_a} L_a w_{\bar{\mu}\nu'}^X + (N_a w_Z)^2 M_a w_Y \nu_o^X \nu_{\bar{\mu}}^Z\right] \end{aligned} \tag{4.33}$$

The firing rates of inhibitory neurons can be approximated as

$$\nu_\mu^Z \cong \frac{1}{N_a} \sum_{k\in\Omega_\mu^Z} u_k^I \cong M_a w_Y \nu_\mu^Y \cong M_a w_Y \left(\left(L_a w_{\mu A} + L_a w_{\mu B} + 2L_a w_o^X\right)\nu_o^X - N_a w_z \nu_{\bar{\mu}}^Z\right). \tag{4.34}$$

Therefore, by solving the simultaneous equations for ν_1^Z and ν_2^Z,

$$\nu_1^Z = \frac{M_a w_Y \nu_o^X}{1-(M_a w_Y N_a w_Z)^2}\left[\left(L_a w_{1A} + L_a w_{1B} + 2L_a w_o^X\right) - (M_a w_Y N_a w_Z)\left(L_a w_{2A} + L_a w_{2B} + 2L_a w_o^X\right)\right],$$

$$\nu_2^Z = \frac{M_a w_Y \nu_o^X}{1-\left(M_a w_Y N_a w_Z\right)^2}\left[\left(L_a w_{2A} + L_a w_{2B} + 2L_a w_o^X\right) - \left(M_a w_Y N_a w_Z\right)\left(L_a w_{1A} + L_a w_{1B} + 2L_a w_o^X\right)\right].$$

This analytical approach is applicable only when the synaptic weight change is sufficiently slow relative to the neural dynamics. Also, because I ignored the variance in the synaptic weights, numerically the accuracy is limited.

Analytic approach for STDP in lateral and inhibitory connections Using a similar calculation as above, synaptic weight development of the lateral connections is given as

$$\cdot W_Y \approx g_1^Y W_Y W_X C^t W_X^t - g_2^Y W_Y W_X C^t W_X^t W_Y^t W_z^t - g_3^Y W_Y W_Z W_Y W_X C^t W_X^t, \tag{4.35}$$

where

$$g_1^Y \equiv \int_{-\infty}^{\infty} ds F^Y(s) \int D_r^X \int D_u^Y \int D_{r'}^X h(u + r' - s - r)$$

$$g_2^Y \equiv \int_{-\infty}^{\infty} ds F^Y(s) \int D_r^X \int D_u^Y \int D_q^Z \int D_{u'}^Y \int D_{r'}^X h(u' + r' - s - q - u - r - d_Y - d_Z)$$

$$g_3^Y \equiv \int_{-\infty}^{\infty} ds F^Y(s) \int D_r^X \int D_u^Y \int D_q^Z \int D_{u'}^Y \int D_{r'}^X h(u + q + u' + r' + d_Y + d_Z - s - r),$$

where $\int D_r^X \equiv \int_0^\infty dr \epsilon_x(r)$. The meaning of these equations is made clear by summarizing the correlation propagation in the diagrams (Figure S2D i-iii). In the diagram, blue wavy lines represent intrinsic correlation, and arrows are synaptic connections. To estimate how a blue correlation influences STDP at a red arrow, we need to determine all the major trajectories in which the correlation reaches pre- and postsynaptic neurons. In the linear Poisson framework, for a given trajectory, the propagation of a correlation is calculated by simply using integrals as above. From this diagram, we can safely assume that g_2^Y and g_3^Y are negligibly smaller than g_1^Y, because trajectories (ii) and (iii) are secondary correlations and also contain synaptic delays. In this approximation, I additionally assume that

$$C = \begin{pmatrix} c_s & 0 \\ 0 & c_s \end{pmatrix},\ W_X = \begin{pmatrix} w_s^X & w_w^X \\ w_w^X & w_s^X \end{pmatrix}.$$

Then,

$$\frac{d}{dt}\begin{pmatrix} w_{11}^Y \\ w_{12}^Y \\ w_{21}^Y \\ w_{22}^Y \end{pmatrix} \approx \begin{pmatrix} A^L & B^L & 0 & 0 \\ B^L & A^L & 0 & 0 \\ 0 & 0 & A^L & B^L \\ 0 & 0 & B^L & A^L \end{pmatrix}\begin{pmatrix} w_{11}^Y \\ w_{12}^Y \\ w_{21}^Y \\ w_{22}^Y \end{pmatrix},$$

$$A^L \equiv c_S g_1^Y \left(\left(w_s^X\right)^2 + \left(w_w^X\right)^2\right),\ B^L \equiv 2c_s g_1^Y w_s^X w_w^X.$$

Therefore, $\left(w_{11}^Y, w_{12}^Y, w_{21}^Y, w_{22}^Y\right) \propto (+1,-1,-1,+1)$ is a eigenvector of the transition matrix, and the eigenvalue is $c_s g_1^Y \left(w_s^X - w_w^X\right)^2$. Because the eigenvector develops by $\exp\left[c_s g_1^Y \left(w_s^X - w_w^X\right)^2 t\right]$, when

g_1^Y is positive, the E-to-I connections are more likely to be structured in a way that the inhibitory neurons become feature selective. On the other hand, if that value is negative, such structure may not be obtained. Note that (1, -1, -1, 1) is not the principal eigenvector in this simple analysis, because the eigensystem of the matrix is $\{ \{A^L + B^L, A^L + B^L, A^L - B^L, A^L - B^L\}; \{1,1,0,0\}, \{0,0,1,1\}, \{1,-1,0,0\}, \{0,0,1,-1\}\}$. Similarly, for inhibitory connections

$$\dot{W}_Z \approx g_1^Z W_X C W_X^t W_Y^t - g_2^Z W_Z W_Y W_X C W_X^t W_Y^t$$

$$\begin{aligned} g_1^Z &\equiv \int_{-\infty}^{\infty} F^Z(s) \int D_r^X \int D_u^Y \int D_{r'}^X h(r - s - u - r' - d_Z - d_Y) \\ g_2^Z &\equiv \int_{-\infty}^{\infty} F^Z(s) \int D_q^Z \int D_u^Y \int D_r^X \int D_{u'}^Y \int D_{r'}^X h(r + u + q - s - u' - r') \end{aligned} . \quad (4.36)$$

I approximated with only two terms because the third term is negligible (Figure S2D iv-vi). If we assume $W_Y = \begin{pmatrix} w_d^Y & w_r^Y \\ w_r^Y & w_d^Y \end{pmatrix}$, and $g_2^Z = 0$, then the synaptic weight change follows $\Delta w_{11}^Z - \Delta w_{12}^Z = \Delta w_{22}^Z - \Delta w_{21}^Z = c_S g_1^Z \left(w_s^X - w_s^X\right)^2 \left(w_d^Y - w_r^Y\right)$. This means that if g_1^Z is positive, reciprocal connections are enhanced (or inhibitory connections to the neurons coding a similar feature are enhanced), whereas for negative g_1^Z, inhibitory connections develop non-reciprocally (i.e., lateral connections function as mutual inhibition between output excitatory neuron groups).

I have restricted my consideration to Hebbian STDP, but the properties of STDP on E-to-I and I-to-E connections are still debatable [133] [231]. Although it is difficult to study all combinations of STDPs, we can still provide analytical insights by investigating the behaviors of g_1^Y and g_1^Z. Supplementary Fig. 4.2E shows the behaviors of four different types of STDP. This indicates that the anti-correlation detector type of E-to-I STDP [146] tends to cause non-feature-selective lateral connections. In addition, under the anti-coincidence detector type of I-to-E STDP [243], mutual inhibition structures would be preferred; however, the implication of my analytical method is limited, and further study will be necessary to fully understand the functions of the various types of STDP.

Evaluation of the performance

Cross-correlation I evaluated the performance by measuring the mean cross-correlation between the external sources and population activity of the output neurons. For time bin $\Delta t = 10$ ms, the activity of source μ is defined as $s_\mu^k = \frac{1}{\Delta t} \int_{k\Delta t}^{(k+1)\Delta t} s_\mu(t) dt$, and, similarly, the population activity of the output neuron group ν is $y_\nu^k(\tau_D) = \sum_{j \in \Omega_\nu^Y} \frac{1}{\Delta t} \int_{k\Delta t}^{(k+1)\Delta t} y_j(t + \tau_D) dt$, where Ω_ν^Y is a set of output neurons coding a source ν. For these, cross-correlation is defined as

$$c_{\mu\nu}(\tau_D) \equiv \frac{1}{\sigma_\mu^s \sigma_\nu^y} \sum_{k=1}^{T_c/\Delta t} (s_\mu^k - \bar{s}_\mu)(y_\nu^k - \bar{y}_\nu),$$

where $\bar{s}_\mu \equiv \frac{1}{T_c}\int_{T_o}^{T_o+T_c} s_\mu(t)dt$, $\bar{y}_\nu \equiv \frac{1}{T_c}\int_{T_o}^{T_o+T_c} y_\nu(t)dt$, $\sigma_\mu^s \equiv \sqrt{\sum_{k=1}^{T_c/\Delta t}\left(s_\mu^k - \bar{s}_\mu\right)^2}$, and $\sigma_\nu^y \equiv \sqrt{\sum_{k=1}^{T_c/\Delta t}\left(y_\nu^k - \bar{y}_\nu\right)^2}$. I used $T_c = 10$ ms for the analysis. Correspondence between sources and output groups are arbitrary, and so the learned correlation should be given as $c(\tau_D) \equiv \max_\psi \frac{1}{p}\sum_{\mu=1}^{p} c_{\mu\psi(\mu)}(\tau_D)$ for all the $p!$ number of combinations with function between sources and output groups. For example, when p = 2, $c(\tau_D) = \max\{\frac{1}{2}\left[c_{A1}(\tau_D) + c_{B2}(\tau_D)\right], \frac{1}{2}\left[c_{A2}(\tau_D) + c_{B1}(\tau_D)\right]\}$. Although, in reality, supervised or reinforcement learning is necessary to perform this readout, for simplicity I did not implement readout neurons explicitly. In Fig. 4.2F, I plotted $\max_\nu c_{B\nu}(\tau_D)$ for the minor source B.

For the models with randomly connected lateral inhibition and (e+i) STDP, I defined output neuron j as belonging to Ω_μ^Y if

$$\frac{1}{\left|\Omega_\mu^X\right|}\sum_{i\in\Omega_\mu^X} w_{j\mu}^X > \alpha_{th}\max_{\nu\neq\mu}\left\{\frac{1}{\left|\Omega_\nu^X\right|}\sum_{i\in\Omega_\nu^X} w_{j\nu}^X\right\}$$

for $\alpha_{th} = 1.5$, and the cross-correlation was calculated based on Ω_μ^Y.

Mutual information Based on the discretized hidden external source/output neuron activity s_μ^k, y_ν^k, I defined the binary variables

$$\hat{s}_\mu^k \equiv \begin{cases} 1 & (\text{ if } s_\mu^k > \bar{s}_\mu^k + \sigma_\mu^s) \\ 0 & (\text{otherwise}) \end{cases}, \quad \hat{y}_\nu^k \equiv \begin{cases} 1 & (\text{ if } y_\nu^k > \bar{y}_\nu^k + \sigma_\nu^y) \\ 0 & (\text{otherwise}) \end{cases}.$$

Based on these variables, the states at time k can be defined as $\hat{s}^k \equiv (\hat{s}_1^k, ..., \hat{s}_p^k)$, $\hat{y}^k \equiv (\hat{y}_1^k, ..., \hat{y}_p^k)$. Therefore, the probability that the external state takes one particular state is $p_s(\hat{s} = \hat{s}') \equiv \frac{1}{T_c/\Delta t}\sum_{k=1}^{T_c/\Delta t}\left[\hat{s}^k = \hat{s}'\right]_{tof}$, where $[X]_{tof}$ takes 1 if X is true, otherwise it takes 0, for the statement X. Therefore, mutual information can be defined as

$$MI \equiv \sum_{\hat{s}'}\sum_{\hat{y}'} p_{sy}(\hat{s} = \hat{s}', \hat{y} = \hat{y}')\log_2\left(\frac{p_{sy}(\hat{s} = \hat{s}', \hat{y} = \hat{y}')}{p_s(\hat{s} = \hat{s}')p_y(\hat{y} = \hat{y}')}\right).$$

Tables

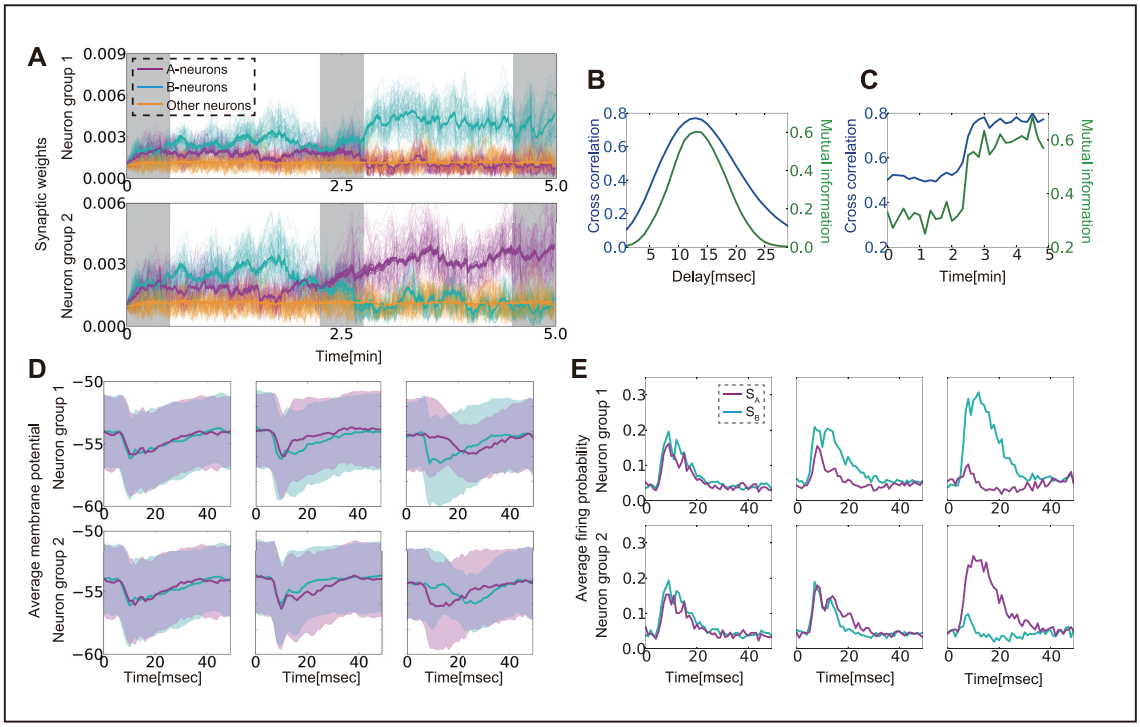
Table 1. Definition of variables

$s_\mu(t)$	The activity of external source μ	$s_\mu(t) = \hat{\sigma}[\nu_o^S]$
$x_i(t)$	The spiking activity of input neuron i	equation (14)
$u_j^E(t)$	Membrane potential of output neuron j	equation (15)
$y_j(t)$	The spiking activity of output neuron j	$y_j(t) = \hat{\sigma}[u_j^E(t)]$
$u_k^I(t)$	Membrane potential of inhibitory neuron k	equation (16)
$z_k(t)$	The spiking activity of inhibitory neuron k	$z_k(t) = \hat{\sigma}[u_k^I(t)]$
w_{ji}^X	The synaptic weight of a feed-forward excitatory connection from j to i	equation (17)
$q_{i\mu}$	Response probability of input neuron i to external source μ	equation (14)
C_{il}	Non-normalized correlation between input neuron i and l	$C_{il} = \sum_\mu q_{i\mu} q_{l\mu}$
$C_{il}(s)$	Cross correlation between input neuron i and l	equation (24)
$G_1^X(w)$, $G_2^X(w)$	Coefficients of correlation-based synaptic weight change	equation (30)
χ_1^X, χ_2^X	The correlation kernel functions	equation (3),(4)

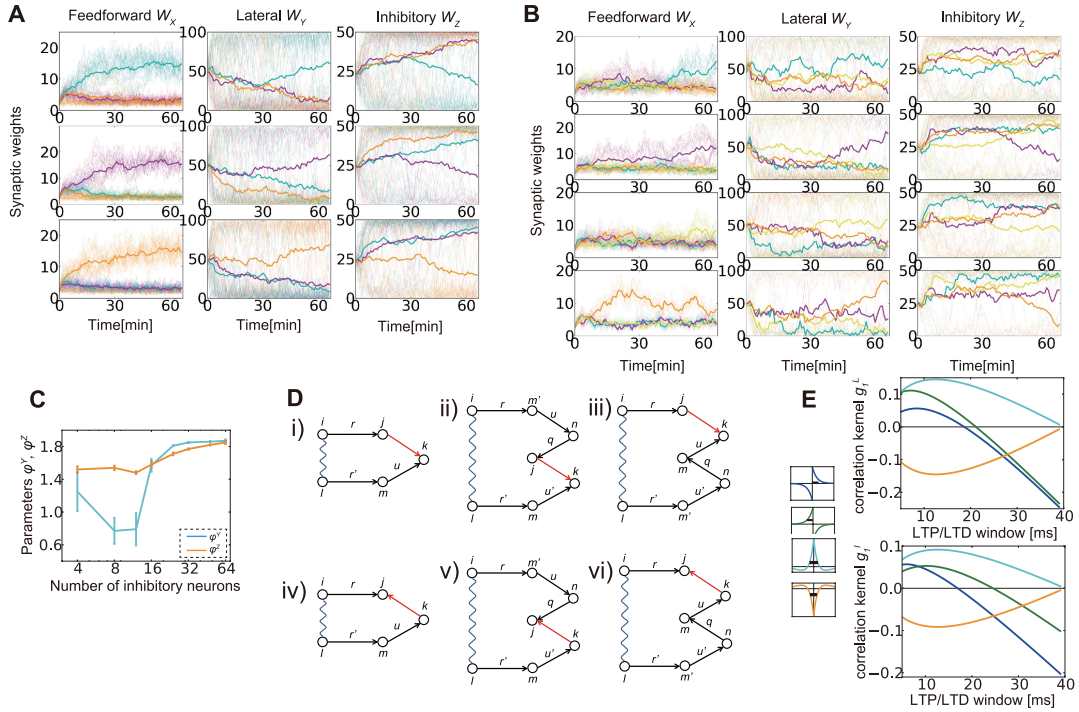
Table 2. Parameter settings

T	Simulation time	3000 s (for Figures 4.5C-E, 4.6, 4.7: $T = 4000$ s)
L, M, N	Neural population	400, 20, 20 (for Figures 4.7, 4.8: M = N = 40)
L_a, M_a, N_a	Neural subpopulation	100, 10, 10
τ_A^X, τ_B^X, τ_A^Y, τ_B^Y, τ_A^Z, τ_B^Z	EPSP/IPSP time constants	5.0, 1.0, 4.0, 0.8, 2.5, 0.5 ms
w_o^X, w_o^Y, w_o^Z	Synaptic weights	2.5, 100.0, 50.0 (for Figures 4.7, 4.8: $w_o^Z = 80.0$)
d_{min}^{Xa}, d_{max}^{Xa}	Axonal delays	2.0, 4.0 ms
d_{min}^{Xd}, d_{max}^{Xd}	Dendritic delays	0.5, 1.5 ms
d_{min}^Y, d_{max}^Y, d_{min}^Z, d_{max}^Z	Synaptic (axonal) delays	0.2, 1.2, 0.2, 1.2 ms
θ_t	Correlation timescale	2.0 ms
ν_o^S, ν_o^X	Firing rates	10, 10 Hz
η_X	Learning rate	$0.05 w_o^X$
σ_{sig}	Noise amplitude of plasticity	0.3
τ_p, τ_d	STDP time windows	17, 34 ms
α, β	Parameters for log-STDP	20.0, 50.0
σ_W^{init}	Initial variance of synaptic weights	0.1
γ^Y, γ^Z	LTD/LTP balance	1.4, 0.7

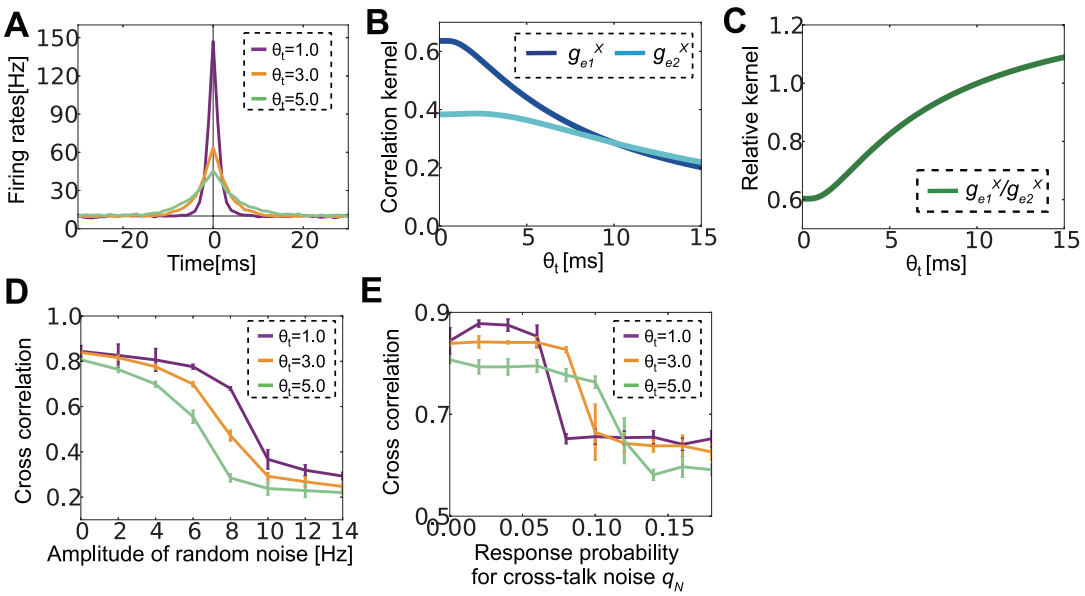
Supplementary Figures



Supplementary Figure 1. Simulations with the leaky integrate-and-fire model. (**A**) Synaptic weight developments at the feedforward connection. (**B**) Cross-correlation and mutual information calculated for various delays. Both values were calculated by averaging five independent simulation results. (**C**) Development of two values for the simulation shown in (**A**). (**D**) PSTH of the membrane potential calculated for gray areas in (**A**). (**E**) Peristimulus time histogram (PSTH) of the firing probability for the same simulation.



Supplementary Figure 2. Spike-timing-dependent plasticity (STDP) at lateral connections shapes network structure. (**A, B**) Synaptic weight development when the number of external inputs is three (**A**) and four (**B**). Thick lines represent averages over all synapses, and thin lines represent individual synaptic weights. Colors represent detected sources for output neurons (left) and inhibitory neurons (middle right). (**C**) Relationship between the number of inhibitory neurons and the lateral structure. (**D**) Propagation of structure. i to iii correspond to lateral excitatory connections, and iv to vi correspond to feedback inhibitory connections. (**E**) Analytic results for various types of STDP.



Supplementary Figure 3. The effects of noise in the model with exponential correlation kernel. (**A**) Cross-correlations among input neurons responding to the same source calculated from simulated data for three different correlation timescale parameters θ_t. Note that in **Figure 4.3**, I used θ_t = 0.5, 2.0, 4.0 ms, while here I used θ_t =1.0, 3.0, 5.0ms. (**B, C**) The correlation kernels g_{e1}^X, g_{e2}^X (**B**) and their ratio g_{e1}^X/g_{e2}^X (**C**) are shown for the kernels g_{e1}^X and g_{e2}^X that were calculated from equation (30) with $\phi_e(t) = e^{-t/\theta_t}/\theta_t$. (**D,E**) The effects of random noise (**D**) and crosstalk noise (**E**) at various correlation timescales.

Chapter 5

A Spiking Neuron Model of Cell Assembly Modulation

Introduction

Learning and memory are fundamental brain functions supported by hippocampal neural circuits, and long-term potentiation (LTP) and depression (LTD) of synapses are considered to underlie activity-dependent modifications of hippocampal circuits during memory processes. According to the cell-assembly hypothesis [96] [30], memory traces may be represented by functionally grouped assemblies of neurons. Although the mechanism to generate memory traces remains elusive, experimental evidence suggests that the groups of neurons activated during behavior are reactivated and reorganized in the awake-quiet and sleep states of animals [179] [55]. These results indicate that memory traces are not static entities driven solely by external stimuli as often assumed in previous theoretical studies, but are actively retained and modulated by spontaneous network dynamics. Moreover, latent modulations, especially selective retention and integration, of memory traces are important in various cognitive tasks [140]. Especially, recent experiments found spontaneous flickering of cell assemblies in the quiet states [111] [112] [59], but their functional roles and circuit mechanism are not yet known.

In order to explore the spontaneous modulation of memory traces, we need to model spontaneous activity states with activity-dependent synaptic plasticity, such as spike-timing-dependent plasticity (STDP), in which synaptic weights are modified depending on pre- and post-synaptic spike events occurring in a millisecond-range timescale [152] [20]. Along with long-term plasticity, cortical synapses also undergo short-term plasticity [1] [223]. Short-term plasticity, especially short-term depression (STD), can induce dramatic changes in the characteristic dynamics of recurrent network models such as spontaneous transitions among point attractors [183] [155] or rotational motions in ring attractors [249]. Because STDP depends on spiking activity within a timescale comparable with that of the complex network dynamics, short-term plasticity may significantly influence the processes of cell-assembly for-

mation and retention in recurrent neural networks. In fact, recent experimental results suggest strong influences of short-term synaptic plasticity on memory function [219] [2]. Nevertheless, little is known about interplay between short-term and long-term synaptic plasticity in activity-dependent structuring of recurrent neural networks.

Motivated by the cell-assembly hypothesis [96] [30], here I investigate how STD and STDP may cooperatively generate and modulate cell assemblies in response to external stimuli to a recurrent network model also equipped with homeostatic plasticity [224]. I ask whether and how the network retains the memory traces of stimuli for a significantly long period of seconds and minutes in the absence of the stimuli. I explore interactions between multiple cell assemblies during their formation and retention. The model reveals several conditions on the properties of STD and STDP for the robust maintenance of memory traces in noisy background network activity. In particular, I show that STD plays a crucial role in the retention process. Moreover, my results indicate that spontaneous flickering support cell assembly retention, by controlling synaptic efficiency change due to STD. Furthermore, I show that modifications of STDP time window, such as observed in hippocampal synapses under dopaminergic modulations [253] or in some neocortical synapses during the development [109], enable the model to dynamically combine multiple cell assemblies into stable clusters with a finite memory capacity.

Results

I construct a recurrent circuit model consisting of 2500 excitatory neurons and 500 inhibitory neurons that are randomly connected with each other. I introduce short-term plasticity and long-term plasticity into synaptic connections between excitatory neurons, where long-term plasticity is implemented as a combination of log-STDP (Fig. 5.1A) and homeostatic plasticity (Methods). I focus on the effect of short-term depression on the generation and retention of cell assemblies by long-term plasticity.

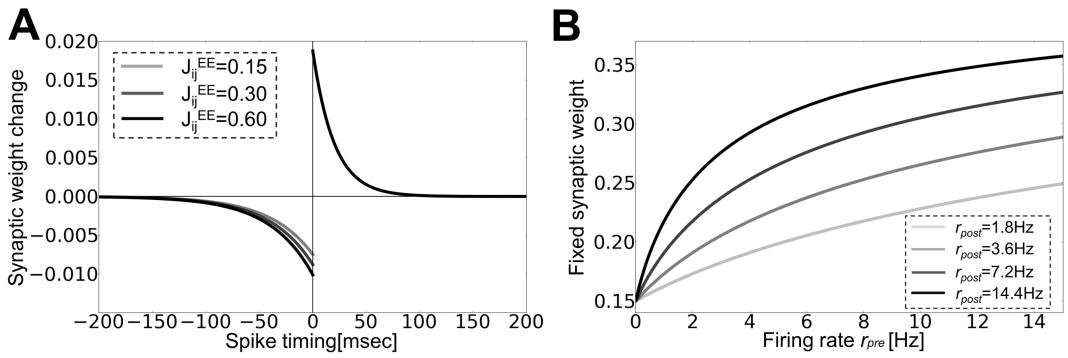


Figure 5.1. Rate-dependent plasticity through STDP and homeostatic plasticity. (**A**) Spike timing dependence of log-STDP was calculated from equation (7) for given synaptic weights (inset). See Methods for details. (**B**) Firing rate dependence of synaptic weights at the fixed-point of equation (1) representing synaptic dynamics of STDP and homeostatic plasticity. The fixed weights are analytically calculated for various firing rates of pre-neuron r_{pre} at given firing rates of post-neuron r_{post}.

Cell assembly formation

If we neglect the effect of synaptic noise, the weight change of synapse J_{ij}^{EE} is approximately written as

$$\frac{dJ_{ij}^{EE}}{dt} \cong r_{pre} r_{post} \left(C_p \tau_p - f_d \left(J_{ij}^{EE} \right) C_d \tau_d \right) + \frac{J_{EE} - J_{ij}^{EE}}{\tau_h} \tag{5.1}$$

where r_{pre} and r_{post} are the firing rates of pre- and post-synaptic neurons, respectively. The first term expresses the effect of STDP, whereas the latter term describes the effect of homeostatic plasticity. When LTP slightly outbalances LTD on average, at its steady state, weights have positive correlations with the firing rates of both pre and post neurons (Fig. 5.1B, for a given r_{post}) due to relatively strong homeostatic plasticity. If a synaptic weight is large, on average it decreases not only for low input/output rates but also for high firing rates due to the weight dependence of LTD term, so the network tends to be stabilized at a finite firing rate with robustly configured synaptic weights.

First, I consider the effect of STD on cell assembly formation by selectively stimulating an excitatory neuron group (Fig. 5.2A). The weights of synaptic connections are initially random (Fig. 5.2C left), and the network shows an irregular spontaneous activity state with low firing rates ($r_E = 1.5 - 2.0$Hz, $r_I = 10 - 15$Hz) (Fig. 5.2D left). Then, I apply a constant external current $I_p = 1.0$ to randomly chosen 20% of excitatory neurons for 30 seconds. During this external stimulation, those 20% of excitatory neurons constantly fire at a high firing rate of 10-15Hz, and as a result synaptic connections among these neurons become strong (Fig. 5.2B, blue shadow indicates the neurons receiving the external stimulus) due to long-term potentiation caused by the high firing rates of presynaptic and postsynaptic neurons (as shown in Fig. 5.1B). After the stimulus is turned off, the average connection strength between stimulated neurons is significantly larger than other excitatory connections (Fig. 5.2C right), and the firing rates of these neurons are also higher than others (Fig. 5.2D). Thus, a cell assembly can be formed in a stimulus-dependent manner. The average weight of synapses belonging to the assembly becomes larger for stronger input current (Fig. 5.2E). The observed phenomena are qualitatively the same for simulations at different values of the release probability parameters (Fig. 5.2F), implying that the details of STD are not essential for the generation of cell assemblies.

Cell assembly retention

Because neurons belonging to a cell assembly interact with neurons outside it, the stability of cell assembles in the absence of external stimuli is not trivially ensured. In fact, this stability crucially relies on the properties of STD, as shown below. After the termination of stimuli, the average synaptic weights in general return slowly toward the initial values, although they eventually converge to certain values that may not coincide with the initial ones. When the release probability is small ($u_{sd} = 0.1$), the weights inside the cell assembly is distinctly larger than other weights (Fig. 5.3A left), and the trace of the cell assembly remains visible even after 30 minutes in both synaptic weights (Fig. 5.3B left) and neural activity (Fig. 5.3C left). Synaptic weights between neurons inside the cell assembly

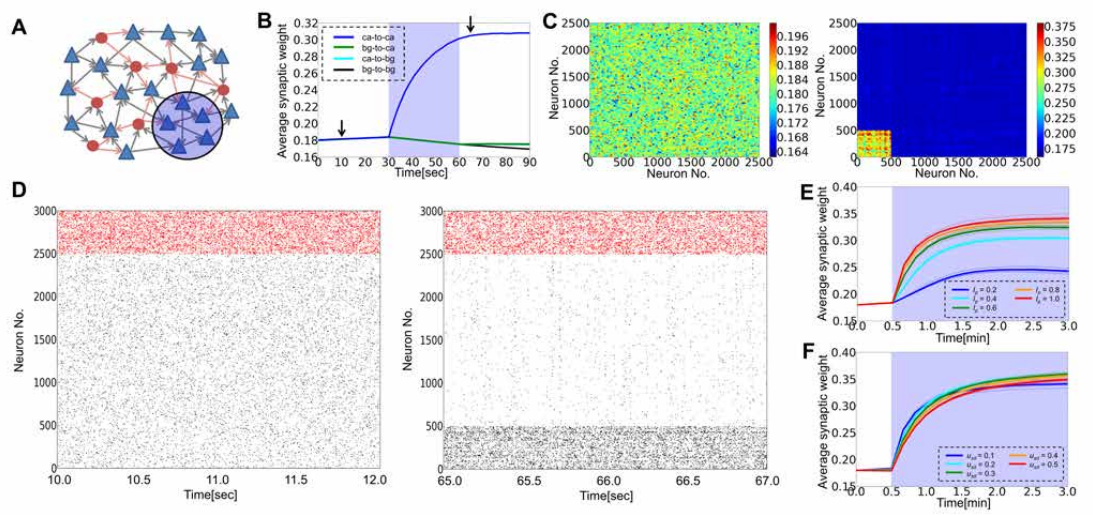


Figure 5.2. Cell assembly formation by external input for arbitrary strength of STD. In all panels, "ca" stands for a cell assembly and "bg" for background neurons that do not belong to the assembly. The strength of STD was set as $u_{sd} = 0.1$ in simulations from panel **B** to **E**. (**A**) Schematic illustration of the model. I stimulate some of excitatory neurons (blue shaded area) in a randomly connected recurrent neural circuit. Triangles indicate excitatory neurons, whereas circles represent inhibitory neurons. (**B**) Time evolution of the average synaptic weights within the selected cell assembly (blue), from background excitatory neurons to the assembly (green), from the assembly to background excitatory neurons (cyan), and outside the cell assembly (black). (**C**) Synaptic weight matrices of excitatory connections are shown before (left) and after (right) the application of external input (arrows in **B**). Excitatory neurons are separated into 100 bins to calculate the average weights. (**D**) Raster plots of spiking activity before (left) and after (right) the application of external input, where red dots represent inhibitory spikes and black dots show excitatory spikes. The temporal position of dots are represents the update timing of the spiking state. Neurons 1 to 500 belong to the cell assembly. (**E**) Dynamics of the average synaptic weight within the cell assembly calculated for various magnitudes of external input I_p. Thin lines are the results from individual simulation trials, and the thick lines are the averages of five simulation trials at each parameter value. (**F**) Dynamics of the average synaptic weight within the cell assembly calculated at $I_p = 1.0$ for various values of the release probability usd.

and background cells (i.e., cells not belonging to the assembly) are somewhat larger than those among background cells, as the high rate of presynaptic (postsynaptic) firing enhances synaptic weights due to the firing-rate dependency of STDP. Background neurons also change their firing pattern because the balance condition of the network changes after learning. On each excitatory neuron belonging to the cell assembly, synaptic weights from other cells in the assembly remain large showing large fluctuations, whereas the weights from background cells stay small (Fig. 5.3D). Eventually, the synaptic weights on assembly cells obey a long tailed distribution in which the long-tail mainly consists of synapses from other neurons in the assembly, while that of background neurons constitutes a more Gaussian-like distribution (Fig. 5.3E). In contrast, for strong STD ($u_{sd} = 0.5$), spontaneous activity gradually erases the cell assembly (Fig. 5.3A right), and both neural activity and the synaptic weight matrix become nearly random after several minutes (Fig. 5.3B right, Fig. 5.3C right). These results indicate that STD is highly influential on the cell assembly retention: especially strong STD disturbs the retention.

Fig. 5.4A shows the average synaptic weight inside the cell assembly observed after 30 minutes. The value decreases monotonically as the release probability increases. When the release probability is larger than 0.2, the assembly becomes indistinguishable from other synaptic weights. I studied whether the above results are a direct consequence of STD or merely reflect the effect of other parameters modulated by STD. I first checked the effect of inhibitory inputs. When STD is strong, each excitatory neuron generate fewer spikes for the same inputs, thus the excitatory-inhibitory balance of the recurrent network shifts to an inhibition-dominant regime. I calculated the average firing rate of excitatory neurons for various inhibitory connection weights J_{EI} and release probabilities u_{sd} at a fixed value of J_{EE} (Fig. 5.4B). Then, I adjusted the values of J_{EI} such that excitatory neurons fire at a similar average firing rate (of 1.8Hz) for simulations at different release probabilities, and calculated the average synaptic weight in the cell assembly after 30 minutes of exposure to long-term synaptic plasticity. If the weight dependence on u_{sd} arises from differences in the excitation-inhibition balance in Fig. 5.4A, the weights would not change their values in these simulations. However, the average weight almost monotonically decreases as the release probability increases (Fig. 5.4C), indicating that inhibitory inputs are unlikely to cause the decrease of synaptic weights.

Next, I considered the effect of input duration. For $u_{sd} = 0.1$, longer input duration resulted in slightly larger synaptic weights in the cell assembly. In contrast, the weights were not retained for $u_{sd} = 0.5$ even when the input duration was as long as three minutes (Fig. 5.4D). Therefore, a robust retention of cell assemblies is possible only if STD is sufficiently weak. If LTP is sufficiently strong compared to LTD ($C_p\tau_p/C_d\tau_d > 1.6$) cell assemblies also remain stable for large u_{sd} (Fig. 5.4E). However, such a strong LTP is highly unlikely for cortical synapses. Here, I defined the relative weight w_1 as $w_1 = \langle J_{ij}^{EE} \rangle_{cellassembly} - \langle J_{ij}^{EE} \rangle_{all}$ to evaluate the robustness of cell assemblies.

Finally, I numerically solved equation (10) to study the effect of STD on the stability of cell assemblies. I calculated the fixed points of equation (10) for given value of J_{ca}, and then calculated the weight velocity shown in equation (1) at various values of J_{ca}. I found that for given release probability

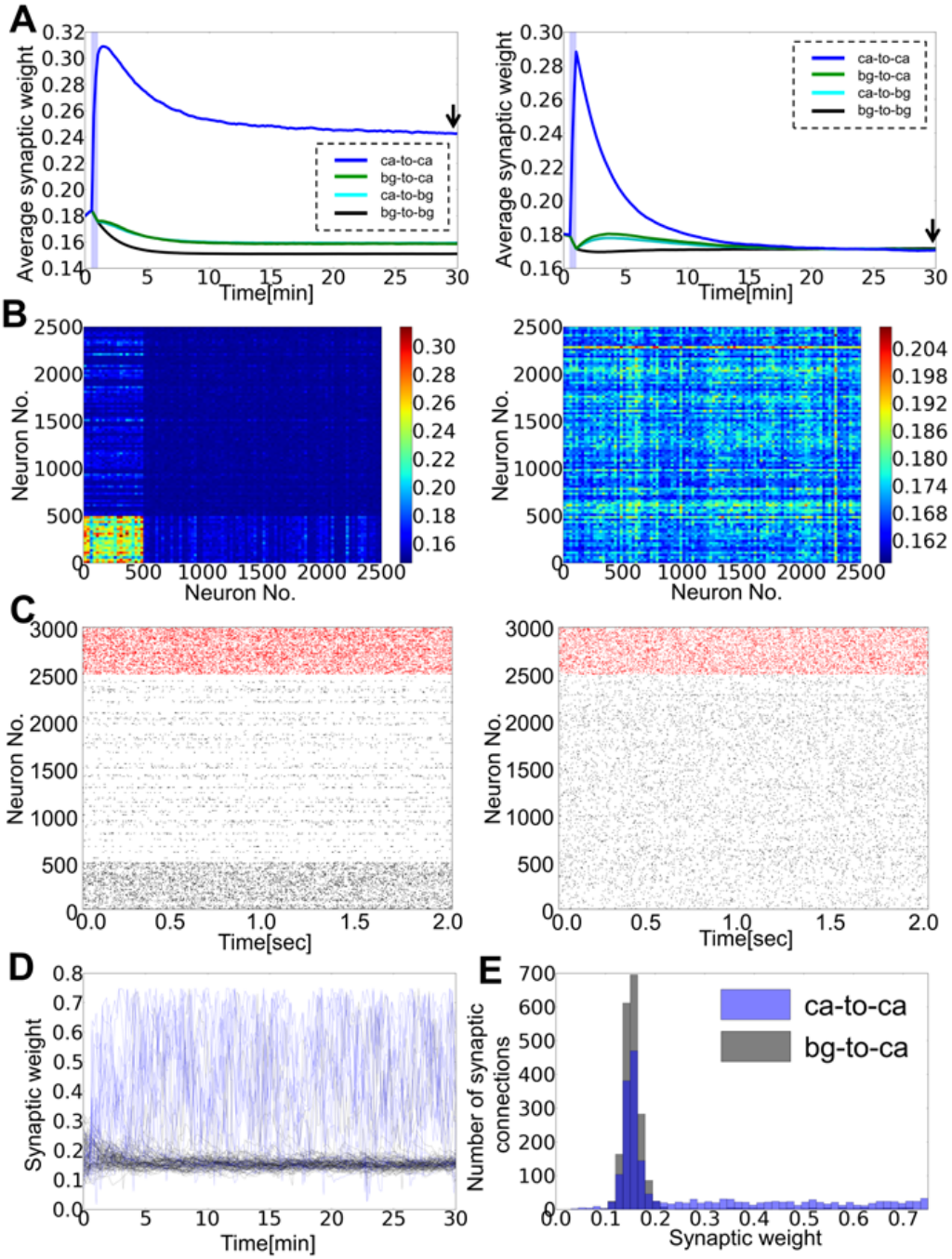


Figure 5.3. Strong STD disturbs cell assembly retention. (**A**) Time evolution of average synaptic weights within the selected cell assembly (blue), from background excitatory neurons to the assembly (green), from the assembly to background neurons (cyan), and between background excitatory neurons (black). The left and right panel show results for $u_{sd} = 0.1$ and $u_{sd} = 0.5$, respectively. (**B**) Weight matrices of excitatory synaptic connections calculated at t = 30 min are shown for $u_{sd} = 0.1$(left) and $u_{sd} = 0.5$(right). (**C**) Raster plots are displayed for the weight matrices shown in **B**. (**D**) Dynamics of individual synaptic weights is shown on one excitatory neuron in the assembly. Blue lines correspond to weights from neurons belonging to the assembly, whereas gray lines to those from background excitatory neurons. (**E**) Distributions of input synaptic weights were calculated from simulation data at t = 26.7-30 min for the neuron shown in **D**.

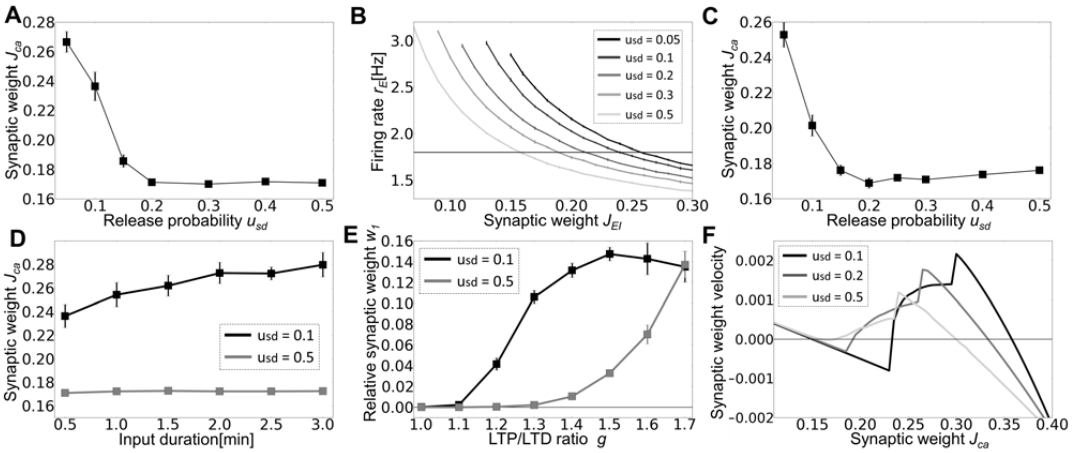


Figure 5.4. Crucial effects of STD on cell assembly retention. Unless otherwise mentioned, error bars represent the standard deviation obtained by five simulation trials. The results shown in panel **A** and **C** to **E** were calculated at t = 30 min. (**A**) Relationship between the release probability usd and the average synaptic weight within the cell assembly. The results were averaged over five simulation trials. The weights of synapses other than J_{EE} were constant. (**B**) Relationship between inhibitory-to-excitatory synaptic weights J_{EI} and the average firing rates of excitatory neurons is shown in a network model without long-term synaptic plasticity. Horizontal line indicates $r_E = 1.8$ Hz. (**C**) Release probability dependence of the average synaptic weight within the assembly is shown. Each plot was calculated using the value of J_{EI} which sets the average firing rate of excitatory neurons to 1.8Hz. (**D**) Relationship between the average synaptic weight within the assembly and input duration is shown. (**E**) The dependence of the relative synaptic weight w_1 to LTP/LTD ratio $g = C_p\tau_p/C_d\tau_d$, which I varied by changing the value of C_p between 0.015 and 0.0255. (**F**) Mean-field approximation gives the velocity of weight change as a function of the synaptic weight. Each line is calculated from equation (10) using the steepest descent method from various initial conditions.

u_{sd}, the numerical solution typically has two stable points corresponding to a state (with small J_{ca}) in which background neurons are most active and a state (with large J_{ca}) in which neurons belonging to a cell assembly are almost exclusively active (Fig. 5.4F). As the release probability is increased, the stable fixed point with large J_{ca} moves to the left side, while the stable point with small J_{ca} eventually disappears in the analytic treatment. In numerical simulations of the network model, however, the two states become closer and less distinguishable (data not shown), implying that they should merge together at a critical value of u_{sd} in Fig. 5.4F. This discrepancy around a critical point is considered to arise from the approximations I employed for making the neural dynamics and weight dynamics analytically tractable. For example, I used mean synaptic weights in analyzing neural and synaptic dynamics although the weight distribution is far from a Gaussian (Fig. 5.3E). These approximations presumably oversimplify the dynamics of my network model with highly heterogeneous synaptic weights.

Interferences between cell assemblies

The results shown in the previous section have revealed that STD has strong influences on the retention of a cell assembly, but not much on its formation. To further demonstrate the effects of STD on the formation and retention of multiple cell assemblies, I stimulated a randomly chosen 20% of excitatory neurons in a recurrent network that initially had random synaptic weights. Directly after the first stimulation, I stimulated another 20% of excitatory neurons that do not overlap with the first group (Fig. 5.5A). I applied the first stimulus for 90 seconds and the second stimulus for 30 seconds because the application of the second one rapidly weakened recurrent synapses in the first neuron group. During the second stimulus, inhibitory neurons suppress the activity of the first neuron group, and then homeostatic plasticity weakens synaptic connections between these inactive neurons. Under these conditions, the external stimuli generated two cell assemblies in the recurrent network. Here, I ask whether these cell assemblies survive separately, disappear or merge with one another when they undergo spontaneous network activity.

To quantify the different wiring patterns emergent in the network, I define the relative synaptic weight w_2 as

$$\tilde{w}_2 = \left(J_{11} - \frac{1}{2}(J_{12}+J_{21})\right)\left(J_{22} - \frac{1}{2}(J_{12}+J_{21})\right),\ w_2 = \tilde{w}_w/\sqrt{|\tilde{w}_2|}$$

where $J_{\mu\nu}$ is the average weight of synaptic connections from cell assembly ν to cell assembly μ. The relative weight is normalized such that it has the dimension of synaptic weights. If the two assemblies survive independently, J_{11} and J_{22} should be much larger than J_{12} and J_{21}, making w_2 strongly positive. On the contrary, if the first assembly survives and the second one disappears, w_2 may take a negative value. If the two assemblies merge into one or both of them disappear, w_2 will be close to zero.

Depending on the value of the release probability, the relative weight acquires positive, negative or almost vanishing values when the network undergoes spontaneous activity (Fig. 5.5B). For small

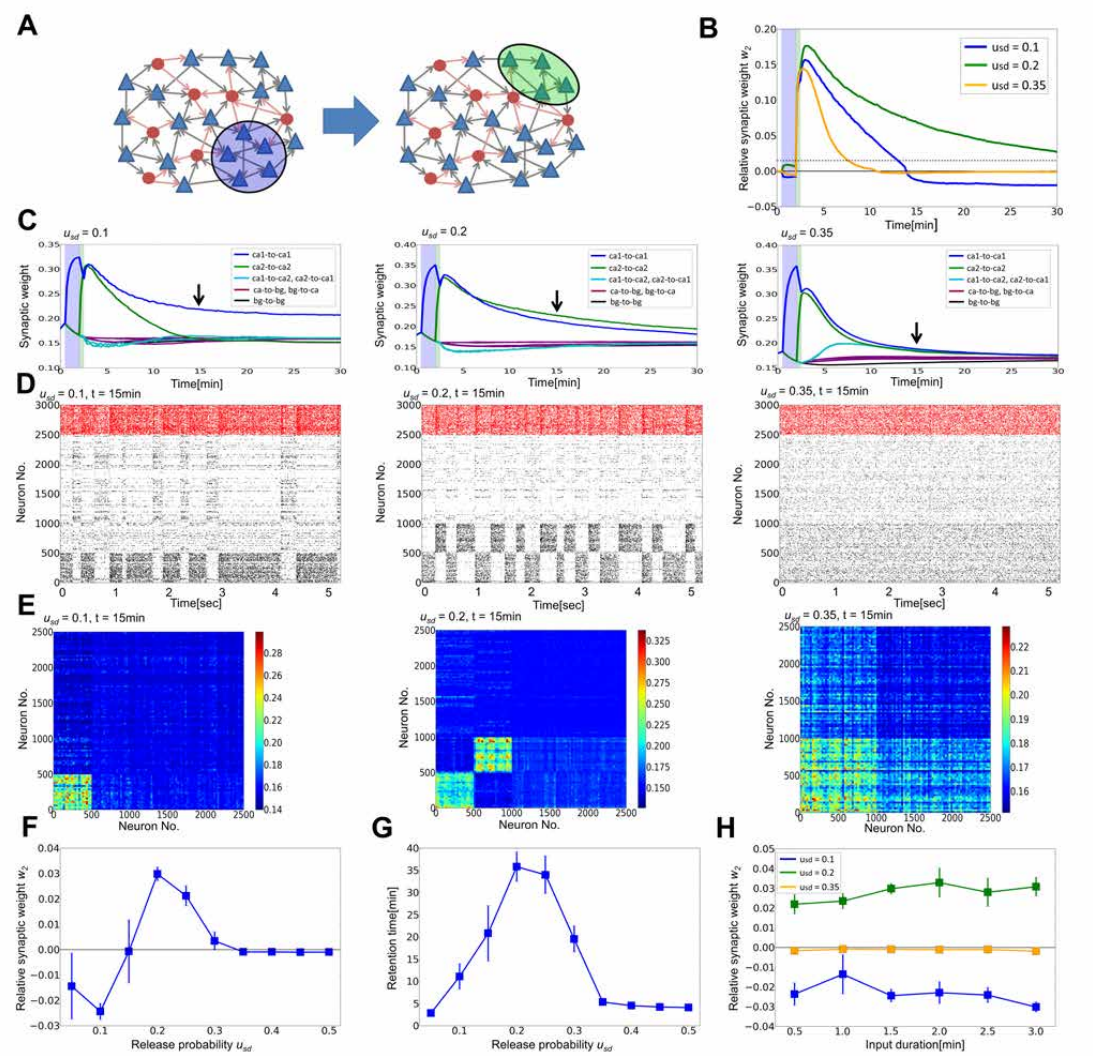


Figure 5.5. Retention of cell assemblies by weak STD. (**A**) A first external input activates 20% of excitatory neurons (ca1, blue shaded area), and then a second input successively activates other 20% of excitatory neurons (ca2, green area). Neurons not stimulated by the external inputs are regarded as background (bg). (**B**) Time evolution of relative synaptic weight w_2. Blue shade indicates the interval of the first stimulus, and the green shade denotes the second one. I defined the retention time of a cell assembly as the time at which w_2 crosses threshold from above ($w_2 = 0.015$: dotted line). (**C**) Time evolution of the average synaptic weight for three values of u_{sd}. The weights were separately averaged over synapses within and between different cell assemblies and background neurons. In the left and middle panels, black lines for bg-to-bg connections are hidden behind purple lines. (**D**) Raster plots of spiking activity corresponding to the three cases shown in **C**. Color codes are the same as in **Fig. 5.2C**. First 500 neurons belong to the first assembly and the second 500 neurons to the second assembly. (**E**) Synaptic weight matrices of excitatory connections are shown for the above three cases. (**F**), (**G**) The relative synaptic weight w_2 and the retention time of ca2 are shown as functions of the release probability u_{sd}. (**H**) Relationship between the input duration to ca1 and the relative synaptic weight w_2 at t = 30 min.

release probability ($u_{sd} = 0.1$) both assemblies exhibit high firing rates after the two stimuli, but only one of them remains active after several minutes (Fig. 5.5D, left). Accordingly, the synaptic weight matrix retains memory traces only for the surviving assembly, but not for the other (Fig. 5.5C and 5E, left). Interestingly, the transient state of cell assemblies can show slow oscillations at 0.5-2 Hz (Fig. 5.5D, left), unlike in the previous case with a single cell assembly. If STD is slightly stronger ($u_{sd} = 0.2$), the two assemblies are kept activated alternately even 15 minutes (biological time) after the termination of external stimuli (Fig. 5.5D, middle), and the synaptic weight matrix indicates clearly distinct memory traces of these assemblies (Fig. 5.5E, middle). However, I note that these assemblies are not permanently stable and eventually disappear, typically after 30 to 60 minutes (Fig. 5.5E, middle). If STD is further strengthened ($u_{sd} = 0.35$), the average synaptic weights rapidly decrease in both assemblies (Fig. 5.5C, E, right) and connections become stronger between the assemblies. As a result, they merge into a large assembly (Fig. 5.5D, right) though this assembly is also unstable and eventually disappears (Fig. 5.5C right).

The relative weight w_2 at 30 minutes takes negative values for weak STD ($u_{sd} < 0.15$), positive values for intermediate strength of STD ($0.15 < u_{sd} < 0.35$), and vanishes for stronger STD (Fig. 5.5F). If we define the lifetime of assemblies as the time at which w_2 becomes smaller than $0.1 J_{EE}$, the lifetime is maximized when STD is modestly strong (Fig. 5.5G). Therefore, adequately strong STD is necessary for a prolonged retention of stimulus-induced cell assemblies. Varying the duration of the first stimulus does not essentially change these results (Fig. 5.5H), suggesting that the internal dynamics of synapses and neurons determines the lifetime of cell assembles. At $u_{sd} = 0.1$, the winning assembly changes from the second to the first if the duration of the first stimulus is about 1-1.5 minutes (data not shown). I also performed simulation with Poisson neuron model to ensure the universality of the results (Supplementary Text S1 and Supplementary Fig. 5.S1).

Stability analysis for cell assemblies

I next investigate the stability conditions for dual cell assemblies. Because the synaptic weight matrix changes much more slowly than the membrane potentials, I first study the dynamics of average firing rates for a given weight configuration by the mean-field approximation. I derived the null-clines $\dot{r}_{ca1}$, $\dot{r}_{ca2}$ of firing rates by numerically solving equation (9) for a network containing two cell assemblies, that is for a synaptic weight matrix given as: $J_{ca1} = J_{ca2} = 0.3$, and all other excitatory weights as 0.17. The intersections of the two null-clines correspond to the fixed points of the network dynamics. In general, the network has an unstable fixed point and two stable fixed points in which one of the two assemblies displays a non-vanishing firing rate (Fig. 5.6A). Making an approximation that a smaller variable between r_{ca1} and r_{ca2} is slaved to a bigger one, for the case when $r_{ca1} > r_{ca2}$, we obtain the

potential function

$$\begin{aligned} \frac{dr_{ca1}}{dt} &= \frac{\partial U}{\partial r_{ca1}} \\ U(r_{ca1} - r_{ca2}) &\cong \int_0^{r_{ca1}} dr'_{ca1} \frac{1}{\tau_E^{ud}} H\left[u_{ca1}\left(r'_{ca1}, r^*_{ca2}(r'_{ca1})\right)/\sigma_{ca1}\left(r'_{ca1}, r^*_{ca2}(r'_{ca1})\right)\right] \\ &\quad + \frac{1}{2} r_{ca1}^2 + U_0 \end{aligned} \tag{5.2}$$

The indices "ca1" and "ca2" are reversed when $r_{ca1} < r_{ca2}$. Note that in general we cannot derive a one-dimensional potential function for a dynamical system of more than two variables without such an approximation. I adjust the constant term U_0 such that $U(0) = 0$ for different values of the release probability.

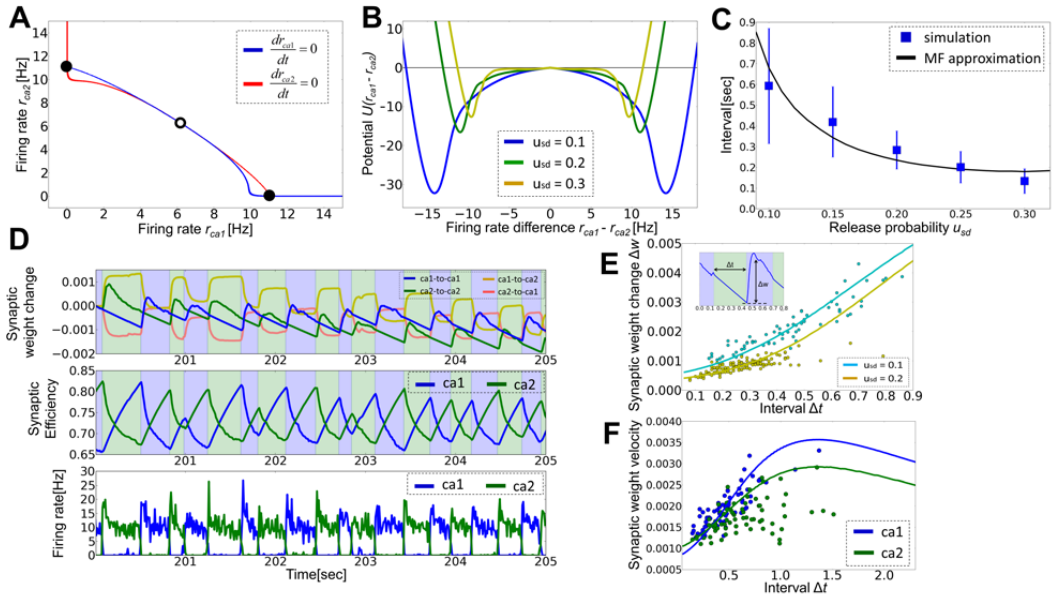


Figure 5.6. STD induces alternate excitations of assemblies, which enlarges synaptic weights within the assemblies. (**A**) Null-clines of firing rates for a synaptic weight matrix calculated from equation (9). (**B**) Potential function U is calculated for the difference in firing rate between two assemblies. The normalization factor U_0 is determined to ensure $U(0) = 0$. (**C**) A monotonic relationship between the release probability and the average interval of the alternation of cell assemblies. The interval was defined as a duration in which one assembly continuously shows higher firing rates than the other. Firing rates were calculated in 10 milliseconds-long time bins. Error bars are the standard derivation of intervals observed during 80 seconds after the stimulus termination in a simulation trial. (**D**), Typical behavior of the average synaptic weights (above), synaptic efficiency for STD (middle), and neuronal firing rates (below). The first (blue) and second (green) cell assemblies show high firing rates alternately. (**E**) Relationship between the interval and synaptic weight change for $u_{sd} = 0.1$ (cyan) and $u_{sd} = 0.2$ (yellow). Inset illustrates the two quantities shown. The ordinate shows synaptic weight change Δw in an interval (Δt_w milliseconds) starting from the activation of the corresponding cell assembly. Dots are data points obtained from simulation, while solid curves indicate analytic results. (**F**) Interval dependence of the synaptic weight velocity is shown, which was defined as an expected synaptic weight change in a second. Solid curves show the analytic results calculated at $J_{ca1} = 0.311$, $J_{ca2} = 0.287$, $J_{bg} = 0.156$, $r_{ca1} = 13.38$ Hz and $r_{ca2} = 12.82$ Hz.

For a given synaptic weight matrix, the potential barrier separating the two stable states becomes lower as the release probability gets larger (Fig. 5.6B). Driven by random noise, therefore the network

state tends to oscillate between the two stable points, each corresponding to one active cell assembly, more frequently for larger release probability. We have already observed this alternation between active cell assemblies in the previous simulations. I confirmed this result by numerical calculations of the average periods of these oscillations following the stimulus termination and a regression analysis with function $Ae^{\beta U(u_{sd})}$ ($A = 0.0679$, $\beta = 0.0691$), where $U(u_{sd})$ is the potential calculated at $u = u_{sd}$ (Fig. 5.6C). Note that the average interval is shorter when the amplitude of noise is larger, which typically occurs when the average firing rate of excitatory neurons is high.

I next consider how the evolution of firing rate controls the dynamics of synaptic weights. Synaptic weights within a cell assembly rapidly increase when the assembly is active, and gradually decrease otherwise (Fig. 5.6D above). Correspondingly, the synaptic efficiencies for STD drop sharply at the beginning of the active epoch, and they recover slowly in the silent epoch (Fig. 5.6D middle). In contrast, synaptic weights between the two assemblies undergo significant changes only when a post-synaptic assembly is transiently active (Fig. 5.6D above). To analyze how STD influences this active maintenance of synaptic weights, I investigate the relationship between the interval of cell-assembly activation (i.e. the duration of the silent epoch), Δt, and the change in intra-assembly synaptic weights at the beginning of an active epoch, ΔJ. The two quantities are positively correlated (dots in Fig. 5.6E), and ΔJ tends to be larger for weaker STD (i.e., smaller u_{sd}), as explained analytically below. When a cell assembly is active, the efficiency of synapses decreases in the assembly until it reaches the equilibrium value $\tilde{y}_{ca} = 1/(1 + u_{sd}\tau_{sd}r_{ca})$. In contrast, during the silent period of an assembly, the efficiencies gradually recover toward an initial level,

$$\tilde{y}'_{ca}(\Delta t) = \tilde{y}_{ca} + (1 - \tilde{y}_{ca})(1 - e^{\Delta t/\tau_{sd}}),$$

which depends nonlinearly on the value of u_{sd}. After the silent epoch of length Δt, the average firing rate $r'_{ca}(\Delta t)$ of the assembly becomes higher than the average firing rate r_{ca} in the equilibrium state, because the synaptic efficiency $\tilde{y}'_{ca}(\Delta t)$ is larger than the equilibrium efficiency $\tilde{y}_{ca}$. We can calculate the firing rate $r'_{ca}(\Delta t)$ by substituting $\tilde{y}'_{ca}(\Delta t)$ into y_{ca} in equation (9) (Method). From equation (1), we can then calculate the average weight increase $\Delta J(\Delta t)$ between the neurons in the initial Δt_w milliseconds of the active epoch as

$$\Delta J(\Delta t) = \left[\left(r'_{ca}(\Delta t)\right)^2 \left(C_p\tau_p - f_d(J_{ca})C_d\tau_d\right) + (J_{EE} - J_{ca})/\tau_h \right] \Delta t_w.$$

This function calculated from the numerical data observed in simulations ($J_{ca1} = 0.311$, $J_{ca2} = 0.287$, $J_{bg} = 0.156$, $r_{ca1} = 13.38$ Hz for $u_{sd} = 0.1$; $J_{ca1} = 0.317$, $J_{ca2} = 0.309$, $J_{bg} = 0.155$, $r_{ca1} = 10.14$ Hz for $u_{sd} = 0.2$) fits the actual values well (Fig. 5.6E, solid lines).

I found that the firing rate $r'_{ca}(\Delta t)$ generally increases with Δt. However, this does not imply that longer Δt, which typically occurs for weaker STD, is advantageous for the retention of cell assemblies

because the velocity of weight change per unit time, $\Delta J(\Delta t)/(\Delta t+T_{active})$, where T_{active} is the average interval of an active epoch, does not increase monotonically with Δt. In Fig. 5.6F, I show the weight velocity calculated by using the average intervals obtained numerically ($T^{ca1}_{active} = 0.65$, $T^{ca2}_{active} = 0.53$ for $u_{sd} = 0.1$). Thus, although longer intervals generate larger weight changes, they also generate more robust stable states of the potential function (Fig. 5.6B), and the alternate activation of two cell assemblies becomes more difficult (see Fig. 5.5D). In contrast, if the strength of STD is in an appropriate range, the two assemblies are alternately activated by noise, enabling the synaptic weights in a resting assembly to increase during its following active period. Although a rigorous analysis of the stability of cell assemblies at relatively strong STD is difficult, we can provide intuition for the observed effects. If STD is weak, an active assembly has a relatively long lifetime. In this case, active assemblies switch only infrequently and the alternate activation can be stable. In contrast, if STD is strong and an active assembly has a short lifetime, active cell assemblies switch frequently and synaptic connections are reciprocally strengthened between the two assemblies, implying that they eventually merge together.

Crucial effects of STDP time window on the stability of cell assemblies

The results shown in the preceding section reveal that cell assemblies are metastable and can survive synaptic bombardment in spontaneous activity only for a few tens of minutes. Although the storage of episodic memory can be as long as hours and days, biological processes responsible for this are considered to involve cellular and molecular mechanisms [192]. Results explained above demonstrate how cell assemblies may be maintained against noise through a network mechanism for minutes to hours. The lifetime of assemblies observed in the previous section is much longer than the characteristic time scales of synaptic and neuronal dynamics. However, the lifetime may not be long enough to induce molecular and cellular processes to stabilize patented synapses. Especially, as we will see later, cell assemblies are less stable when more metastable states exist in the network. In this section, I explore a possible solution to this problem.

As in the previous section, I define the relative weight w_p as

$$\tilde{w}_p = \min_{\mu\neq\nu}\left(J_{\mu\mu} - \frac{1}{2}\left[J_{\mu\nu}+J_{\nu\mu}\right]\right)\left(J_{\nu\nu} - \frac{1}{2}\left[J_{\mu\nu}+J_{\nu\mu}\right]\right),\ w_p = \tilde{w}_p/\sqrt{|w_p|},$$

for general cases with more than two cell assemblies, where $J_{\mu\nu}$ is the average synaptic weight from cell assembly μ to ν. Because it is time-consuming to train the network with many cell assemblies, hereafter I construct a synaptic weight matrix by hand such that it contains p assemblies each consisting of $N_E a$ excitatory neurons (Methods). I examine what STDP rule may retain stable cell assemblies.

I first investigate models with a relatively small number of assemblies ($p = 3$ or 5). When STDP is asymmetric-Hebbian and u_{sd} has an adequate value (Fig. 5.7A, B), the cell assemblies are activated independently and randomly for a while. However, the transient network state switches between different activation patterns of cell assemblies until it displays a sequential activation pattern of assemblies, which

in turn evolves into synfire-like activity (Fig. 5.7C, at t=60-70 sec). However, this activity is unstable and does not persist. Thus, the network eventually returns to random firing states. The lifetime of cell assemblies is longest at a moderate release probability (Fig. 5.7B). I found that such a transient state evolution is typical for the asymmetric STDP window.

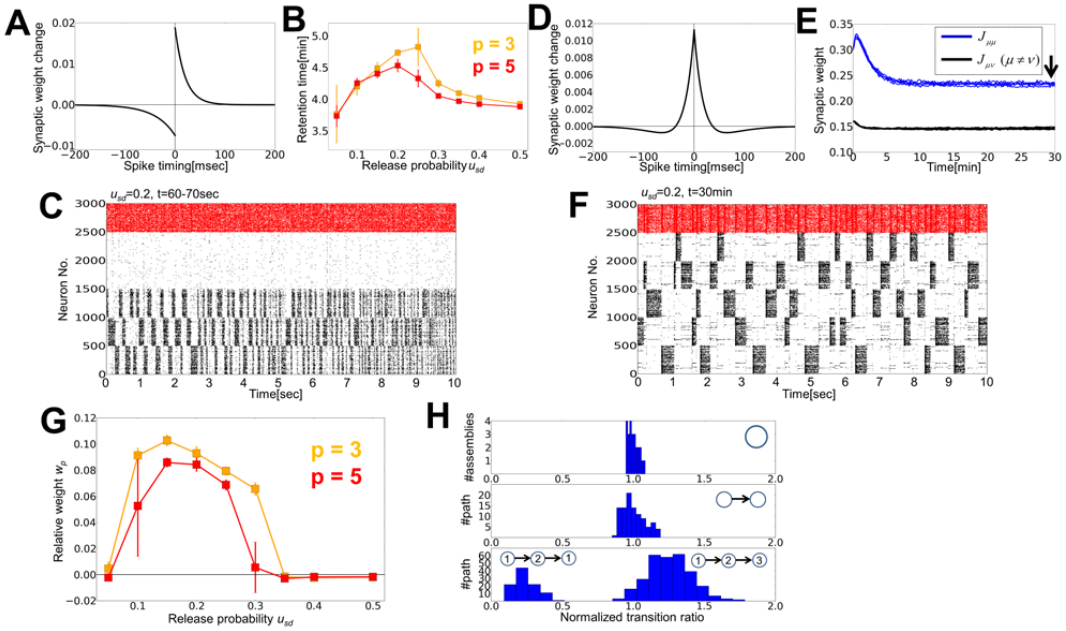


Figure 5.7. The retention of cell assemblies with Hebbian and symmetric STDP windows. (**A**) An asymmetric STDP window was calculated for $J_{ij}^{EE} = 0.15$. (**B**) The retention time significantly varies with the release probability of STD. I defined the retention time as a period with a sufficiently large relative weights: $w_p > 0.1 J_{EE}$. (**C**) Raster plot of spiking activity is shown for the Hebbian STDP rule shown in **A**. (**D**) A symmetric STDP window was calculated for $J_{ij}^{EE} = 0.15$. (**E**) Dynamics of the average synaptic weights at $u_{sd} = 0.2$ within (blue) and between (black) assemblies. (**F**) Raster plot of spiking activity for the symmetric STDP rule shown in **D**. (**G**) Relationship between the release probability u_{sd} and relative weight w_p at t = 30 min. (**H**) (top) I constructed a histogram of the number of activation over all cell assemblies shown in **F**. The abscissa shows the number of activation of each assembly normalized by the average number of activation of all assemblies. (middle) I calculated a histogram for the occurrence of all possible 20 (54) sequential transitions between two assemblies. The occurrence number of each transition was normalized by the average occurrence number over all transitions. (bottom) Histograms of triplet transitions, such as assembly $1 \to 2 \to 1$ (left) and $1 \to 2 \to 3$ (right), are shown after a normalization by all possible 80 (54+543) triplet transition patterns. All three histograms are obtained from the results of five simulation trials.

Cortical synapses are known to change their STDP rules [207] [34]. In particular, under the presence of dopamine, the STDP window of glutamate synapses turns nearly symmetric in rat hippocampus [253]. Moreover, during the developmental stage, excitatory connections from layer 4 to layer 2/3 display symmetric STDP [109]. So, I investigated whether a symmetric window function may change the stability of cell assemblies with the following STDP window (Fig. 5.7D):

$$\Delta J_{ij} = C_p \exp\left(|t_{pre} - t_{post}|/\tau_p\right) - f_d(J_{ij}) C_d \exp\left(|t_{post} - t_{pre}|/\tau_d\right). \tag{5.3}$$

I performed numerical simulations of this network for $p = 3$ or 5 and $u_{sd} = 0.2$. The average weights within cell assemblies converge to stable values after several minutes (Fig. 5.7E). The network persis-

tently and irregularly activates all cell assemblies one by one, and this state remains stable even after 30 minutes (Fig. 5.7F). Consistent with out previous results, such irregular stable states appear only when the strength of STD is in an adequate range (Fig. 5.7G). I next examined whether the activation pattern is random or biased by analyzing spike data taken from 10 to 30 minutes after the initiation of spontaneous activity. I found that all assemblies are activated for nearly the same amount of time (Fig. 5.7H, top). The frequencies of sequential transitions between two assemblies show no statistically significant bias (Fig. 5.7H, middle). In contrast, sequences involving the reactivation of an assembly, such as $1 \to 2 \to 1$, are less likely to occur because STD of mutual excitation in an active assembly suppresses the immediate reactivation of the same assembly. Therefore, the frequencies show some bias among triplets of assemblies (Fig. 5.7H, bottom). The occurrence of monotonous short sequences of cell assemblies is a typical problem in recurrent networks with STDP [65]. It is noteworthy that excitatory weight matrices do not develop short sequences in the present model because synaptic efficiency does not recover in a short time.

Does the retention of cell assemblies sustained by random activation shown above in neural networks with small numbers of assemblies hold for large-scale network models? To answer this, I performed simulations of a network containing a large number of cell assemblies. I set model parameters as $u_{sd} = 0.2$, $p = 32$, $a = 0.03$, $J_{ca} = 0.7$, and $J_{bg} = 0.15$. Note that the size of this network is the same as the previous ones, but each cell assembly now consists of 75 neurons while 500 in previous models. The network initially retains all assemblies by randomly visiting them (Fig. 5.8A, left). After 30 minutes passed, however, some cell assemblies survived stably, but others simply disappeared or merged into bigger stable assemblies (Fig. 5.8A, right). Activity-dependent reorganization of synaptic weight matrix $J_{\mu\nu}$ underlies these changes in the spontaneous activity pattern (Fig. 5.8B). We may define "the storage capacity" of the recurrent network as the number of independent assemblies surviving the reorganization process. This definition can be considered as a natural extension of the storage capacity defined for associative memory model [106]. To this end, I define a binary matrix $\tilde{A}_{\mu\nu}$ as

$$\tilde{A}_{\mu\nu} = \begin{cases} 1, & \text{if } J_{\mu\nu} > 1.5\langle J_{\mu\nu} \rangle \\ 0, & \text{otherwise.} \end{cases}$$

I remove the columns and rows that give vanishing diagonal elements $\tilde{A}_{\mu\mu} = 0$ because cell assembly μ no longer exists in such a case. I then counted the number of disconnected subgraphs in the graph generated from the resultant adjacency matrix (Fig. 2.8C: in this case the storage capacity is 12), which should be equivalent to the storage capacity. I found that the storage capacity depends on the strength of STD, and vanishes for too strong STD (Fig. 5.8D). Furthermore, whether a particular cell assembly survives or merges into a larger assembly strongly depends on the initial weight matrix (Methods). If some initial cell assemblies have weak intra-assembly connections, they are unlikely to survive (Fig. 5.8E). Two assemblies are likely to merge into a single assembly if one or both directions of

the inter-assembly connections are strong (Fig. 5.8F). Thus, when excitatory connections obey STDP and STD, the network has a limited capacity that is maintained by eliminating "weak" assemblies and integrating strongly linked assemblies into single assemblies.

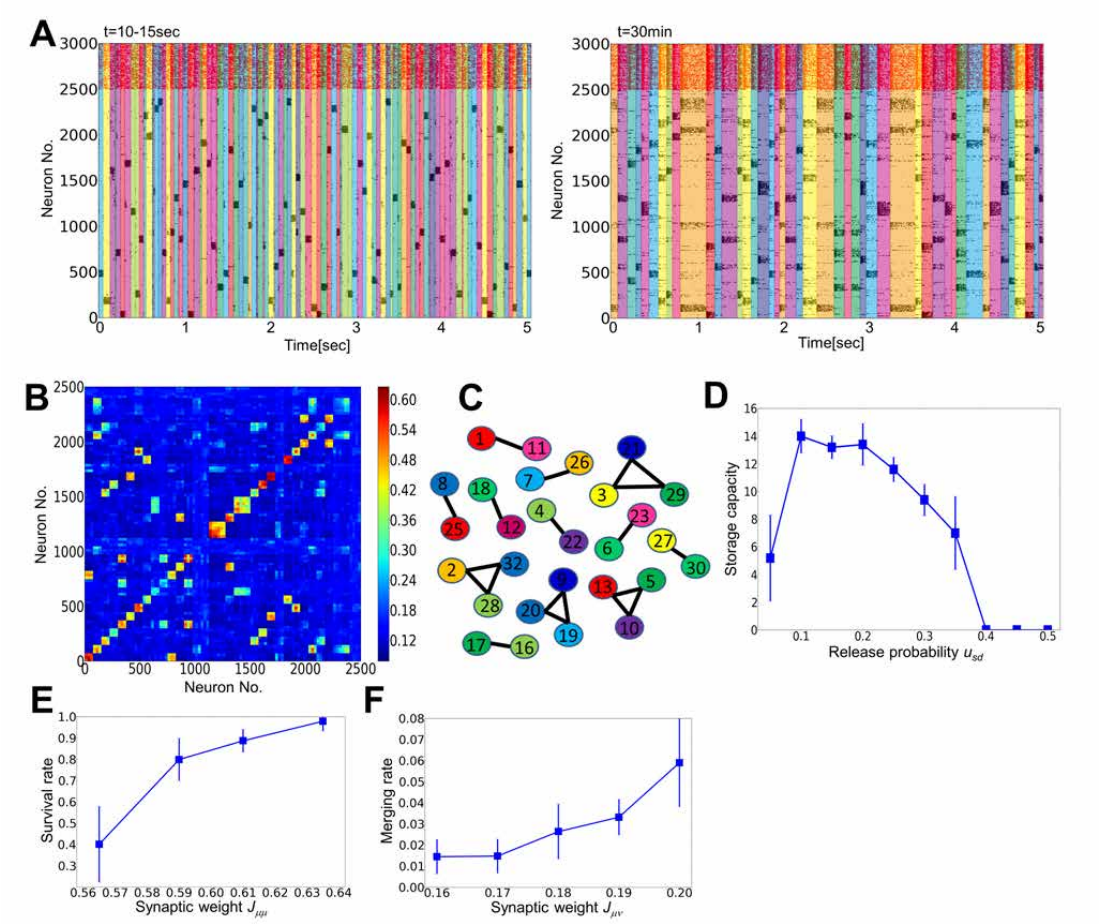


Figure 5.8. Merging and oblivion of cell assemblies through spontaneous activity. (**A**) Raster plot of spiking activity in a network embedding 32 cell assemblies. Active epochs of initial assemblies are shown by different colors in the left panel, while those of merged assemblies are shown in the right panel. (**B**) Synaptic weight matrix after 30 minutes of spontaneous activity. (**C**) A graphical representation of the merged connection matrix, where each numbered circle corresponds to an initial assembly. (**D**) Relationship between the storage capacity and the release probability. (**E**) The survival rate of each assembly depends on the initial magnitudes of intra-assembly synaptic weights. I separated cell assemblies into four groups according to the initial weight values ($0.55 < J_{\mu\mu} \leq 0.58$, $0.58 < J_{\mu\mu} \leq 0.60$, $0.60 < J_{\mu\mu} \leq 0.62$, $0.62 < J_{\mu\mu} \leq 0.65$: the boundaries were decided such that each group contains 5 to 15 assemblies) and calculated the fraction of the assemblies that survived in the reorganization. See Methods for other details of the simulations. (**F**) The rate of merging of a cell assembly as a function of the initial synaptic weight. As in **E**, I separated 992 inter-assembly connections into five groups ($0.155 < J_{\mu\nu} \leq 0.165$,$0.165 < J_{\mu\nu} \leq 0.175$,$0.175 < J_{\mu\nu} \leq 0.185$,$0.185 < J_{\mu\nu} \leq 0.195$,$0.195 < J_{\mu\nu} \leq 0.205$) so that each group contains more than 100 assemblies.

Discussion

I have shown that interplays between STDP and STD enrich synaptic weight dynamics in recurrent neural networks, and cause critical effects on the cell assembly retention and modulation in the timescales

of seconds and minutes. Some cell assemblies merge into a larger assembly or others are eliminated, and the resultant neuronal circuit is able to retain a finite number of memory traces. In these processes, STD crucially influences the stability of modifiable synapses against noisy background activity.

Implications in cortical memory processing

The model proposes a possible circuit mechanism for the long-term retention of selective memory traces encoded by external stimuli into subnetworks of highly connected neurons. In a long time scale, molecular and cellular mechanisms are necessary to maintain synaptic memory traces [192], and it is unlikely that constant reactivation of synapses is permanently necessary for retaining memory. Nevertheless, many experimental results indicate the importance of reactivation of memory traces in learning [179] [55]. My results suggest that these memory traces undergo flexible modifications through the internal network dynamics, and consequently only strong memory traces are preserved in the circuits (Fig. 5.8E). Moreover, if some assemblies are initially linked with stronger excitatory connections, where the initial connection strength is determined by the strength of external stimuli (Fig. 5.2E), the internal dynamics likely integrate these assemblies into one large assembly to co-activate them in the equilibrium network state. These results seem to be consistent with some properties of episodic memory processing by the brain. It is known in humans that sleep enhances the formation of relational memory [60] and false memory [56]. Though my model is too oversimplified to replicate characteristic neural activity during sleep, it explains that initially correlated memory traces can merge together through a repeated reactivation of the corresponding cell assemblies (Fig. 5.8F). Direct experimental evidence supporting this result is awaited.

Possible implications in memory deficits and cortical development

A recent study shows that mice lacking cbl-b, a cell signaling related gene widely expressed in the hippocampus of rodents, display an improved performance in long-term memory retention tasks. In these mice, paired-pulse facilitation at Schaffer collateral-CA1 synapses is enhanced, but long-term synaptic potentiation shows no difference [219]. Because paired-pulse facilitation is enhanced at low release probabilities [51], my model with weaker STD may account for the enhanced memory retention of cbl-b null mice observed in experiments. The model may also explain the relationship between the accumulation of amyloid-β and pathological memory dysfunction. Accumulated amyloid-β is known to disturb long-term potentiation in the hippocampus [234] and this disturbance is often considered as the potential mechanism of dysfunction. My model implies that an enhanced short-term depression, which actually occurs in the presence of an excess amount of amyloid-β [2], may disturb memory retention. It is also known that corticosterone, a hormone controlling stress-induced memory improvement and impairment [200], modifies the probability of presynaptic glutamate releases in the hippocampus of mice [115]. Thus, my model suggests that modifications in short-term plasticity may provide a universal

mechanism to control the stability of memory traces in pathological neural circuits.

The results are possibly relevant to developmental plasticity as well. It is known that in the primary sensory cortex of rodents, glutamatergic synapses show a weakened short-term depression as the animal grows up. The timing of this change typically coincides with the critical period [181] [36] in which the maturation of GABAergic synapses also occurs [98]. A possible explanation of this coincident timing is that the reduction of STD occurs in order to provide more excitatory current, so that the network can keep a balanced state, despite the growth of inhibitory current. As shown in Fig. 5.4B, my model supports this view. Moreover, my model may explain why the strength of STD has to change with successive developmental stages. If STD were strong in immature animals, STDP would not organize any input-dependent structure in cortical circuits: STD may effectively decouple cortical networks from the influence of afferent inputs from thalamocortical pathways until they are well organized.

Limitations of the model

Although I pursued biologically plausibility in the present modeling, some assumptions of the model remain to be confirmed by experiment. I assumed that LTD of excitatory synapses has a logarithmic weight dependence, implying that synaptic weights only sublinearly influence the LTD of strong synapses. However, the weight dependence for strong synapses is still unknown. I also implicitly assumed that synaptic weights are solely modified by STDP and homeostatic plasticity within 30 minutes to 1 hour from the application of external stimuli and molecular processes for the consolidation of memory trace occur later. However, the actual synaptic mechanism of memory consolidation is more complicated and remains elusive [192]. In addition, synaptic weights displayed large fluctuations in Fig. 5.3D, which has not been observed in previous experiments. The large-amplitude fluctuations were partly due to my choice of a relatively large learning rate and partly due to the inherent nature of the present log-STDP model. Nevertheless, these fluctuations are unlikely to be harmful to the practical function of synapses because the oscillation amplitude of the mean weight change was less than 1% of the mean synaptic weight (Fig. 5.6D).

Related previous studies

There are a few recurrent network models that consider both STDP and short-term plasticity. Del Giudice and Mattia showed that a recurrent network with short-term depression is able to robustly organize working memory activity by STDP without destabilizing spontaneous activity [53]. My results are consistent with this result because STD generates a shallow potential well for memory traces (Fig. 5.6B). I have further investigated recurrent circuits embedding multiple cell assemblies, and found that moderate STD is beneficial to the memory retention through interactions. The model proposes that interplay between STD and STDP is a possible mechanism of selective retention and integration of memory traces in recurrent neural networks. The role of STD was also demonstrated in recurrent

neural networks with STDP for the improvement of pattern separation and pattern completion [69].

As for the role of STDP in cell assembly formation without retention or modulation, already many studies exist [80]. While weight-dependent STDP degrades memory retention compared to additive STDP [21], the log-STDP rule (a variant of multiplicative STDP) used in this study improves the stability of learned network structure, reproducing experimentally observed long-tailed unimodal synaptic weight distributions [81]. Log-normal weight distribution can also be reproduced by network effect [254]. A recent theoretical study showed that stable learning is also possible by considering meta-plasticity in addition to the conventional additive STDP [24]. Multiple cell assemblies were created by inducing symmetry breaking through synchronous spikes [139], correlated inputs [**?**] [79], or synaptic delays caused by topological network structure [110]. Other models made use of additional mechanisms such as oscillatory dynamics [137], voltage-dependence [43], triplet STDP [29], or specific network configurations [121]. In some works short-term plasticity was also introduced [110] [121], though its functional role was not intensively discussed in these studies. The effects of neuromodulation were also considered, in which neuromodulators scaled up the learning speed and scaled down the synaptic weight [29]. Recently, some models even consider cell assembly retention [142] [252], yet in these studies, assemblies are simply retained without ant active modulation.

Methods

Model configuration

I construct a recurrent circuit model based on the chaotic balance network model [228] [229] and extend it to include both short-term and long-term plasticity. The network consists of N_E excitatory neurons and N_I inhibitory neurons ($N_E = 2500$, $N_I = 500$), connected randomly with connection probability c_{XY} ($X, Y = E \text{or} I$). I defined connection matrix $\{c_{ij}^{XY}\}_{i=1,\ldots,N_X, j=1,\ldots,N_Y}^{X,Y=E \text{or} I}$ in which $c_{ij}^{XY} = 1s$ if there is a synaptic connection from j to i, otherwise $c_{ij}^{XY} = 0$. For simplicity, I consider the case where only synaptic connections between excitatory neurons show both types of plasticity, while the weights of excitatory to inhibitory, inhibitory to excitatory, and inhibitory to inhibitory connections are kept at constant values J_{IE}, J_{EI}, and J_{II}, respectively. In the main result, I used binary neurons taking only two states, 0 or 1. In the binary model, the states of the i-th excitatory and inhibitory neurons are defined as $x_i^E(t), x_i^I \in \{0, 1\}$. The state of each neuron is updated at time $\{t_{i,k}^E\}_{k=1,2,\ldots}$ or $\{t_{i,k}^I\}_{k=1,2,\ldots}$ according to a random process with the average intervals t_{ud}^E and t_{ud}^I, respectively. In the simulation, I implemented this update procedure by updating $N_E h/\tau_E^{ud}$ excitatory and $N_I h/\tau_I^{ud}$ inhibitory neurons at every h milliseconds ($h = 0.01$ milliseconds; τ_{ud}^E, $\tau_{ud}^I = 5.0$ and 2.5 milliseconds, respectively). The use of binary neurons and discrete update rule reduces the computational load of the simulation of a large recurrent network model with long-term plasticity, and similar results are also

observable in Poisson Model (Fig. S1). The update rules are written as

$$x_i^E(t_{i,k}^E) = \theta\left[\sum_{j\neq i}^{N_E} c_{ij}^{EE} J_{ij}^{EE} y_j(t_{i,k}^E) x_j^E(t_{i,k}^E) - \sum_j^{N_I} c_{ij}^{EI} J_{EI} x_j^I(t_{i,k}^E) + I_E^{ex}(m_{ex} + \sigma_{ex}\zeta_{i,k}^E) + I_p^i(t_{i,k}^E) - h_E\right] \quad (5.4)$$

$$x_i^I(t_{i,k}^I) = \theta\left[\sum_j^{N_E} c_{ij}^{IE} J_{IE} x_j^E(t_{i,k}^I) - \sum_{j\neq i}^{N_I} c_{ij}^{II} J_{II} x_j^I(t_{i,k}^I) + I_I^{ex}(m_{ex} + \sigma_{ex}\zeta_{i,k}^I) - h_I\right], \quad (5.5)$$

where $\theta[]$ is a step function, and $y_j(t)$ is the synaptic efficiency, representing the effect of short-term depression. The terms $I_E^{ex} m_{ex}$ and $I_I^{ex} m_{ex}$ are the fixed components of the amplitudes of random external inputs to excitatory and inhibitory neurons, respectively, while $I_E^{ex}\sigma_{ex}\zeta_{i,k}^E$ and $\zeta_I^{ex}\sigma_{ex}\zeta_{i,k}^I$ are the random components of those external inputs. The noise terms $\{\zeta_{i,k}^E\}, \{\zeta_{i,k}^I\}$ are Gaussian random variables with mean 0 and variance 1. The additional external current $I_p^i(t_{i,k}^E)$ is I_p only for excitatory neurons in the stimulated assembly during the external stimulation, and otherwise remains zero. In the present simulation, I typically applied $I_p = 1.0$ to 500 selected excitatory neurons for tens of seconds. The variables h_E, h_I are the thresholds of the neurons. Once updated, each neuron keeps its state until the next update. For instance, if $t_{j,l}^E \leq t_{i,k}^E < t_{j,l+1}^E$, then $x_j^E(t_{i,k}^E) = x_j^E(t_{j,l}^E)$. I did not introduce a reset procedure mimicking a repolarization process after spiking, because inputs to a neuron are refreshed by every update of the neuron. Excitatory neurons stay in the spiking state for 5 msec on average, while inhibitory ones continue to fire typically for 2.5 msec. Thus, neurons rarely stay in the spiking state for a long time due to the randomness of update. Note J_{ij}^{EE} is normalized such that the size of the first EPSP is the same ($=J_{ij}^{EE}$) for different release probabilities. This means that the total synaptic weight $J_{ij,max}^{EE}$ is given as $J_{ij,max}^{EE} = J_{ij}^{EE}/u_{sd}$. Under this normalization, we can investigate the effect of STD without interference from absolute synaptic weights.

Short-term plasticity is approximately described by the spiking activity of presynaptic neuron [223]. Namely, synaptic efficiency y_j is described with the differential equation

$$\frac{dy_j}{dt} = \frac{1-y_j}{\tau_{sd}} - u_{sd} y_j \sum_k x(t_{j,k}^E)\delta(t - t_{j,k+i}^E), \quad (5.6)$$

where u_{sd} is the release probability and τ_{sd} is the recovery time constant ($\tau_{sd} = 0.6$ seconds). In numerical simulations, I discretize the time variable such that the synaptic efficiency decreases at the next update when a presynaptic neuron fires.

For long-term plasticity, I consider log-STDP [81] and homeostatic plasticity. Log-STDP is a spike-pair-based STDP-model with a logarithmic weight dependence of LTD (Fig. 5.1A). It was modeled to account for the long-tailed, typically lognormal, distributions of the strength of excitatory synapses in the hippocampus and neocortex [209] [31]. The synaptic weight change for two spikes at tpre and

tpost is written as

$$\Delta J_{ij} = \begin{cases} C_p \exp\left((t_{pre} - t_{post})/\tau_p\right) & (\text{if } t_{pre} \leq t_{post}) \\ f_d(J_{ij}) C_d \exp\left((t_{post} - t_{pre})/\tau_d\right) & (\text{if } t_{post} < t_{pre}), \end{cases} \tag{5.7}$$

where $f_d(J_{ij}) = \log(1 + \alpha J_{ij}/J_{EE})/\log(1+\alpha)$, and τ_p, τ_d are the decay time constants of LTP and LTD respectively ($\tau_p = 20$, $\tau_d = 40$ milliseconds). In calculating the time differences between pre- and post-synaptic firing for STDP, I define the time of firing of a neuron as the time of update at which its state becomes 1. Conduction delays between neurons were not taken into account. If a neuron remains in the spiking state for two consecutive bins, those events are regarded as the generation of two spikes. In addition, I consider the effect of homeostatic synaptic plasticity as

$$\frac{dJ_{ij}^{EE}}{dt} = \frac{J_{EE} - J_{ij}^{EE}(t)}{\tau_h} + \sigma_h \zeta_{ij}(t), \tag{5.8}$$

with Gaussian random noise $\zeta_{ij}(t)$. Time constant τ_h of homeostatic plasticity need to be sufficiently short in order to stabilize the network with STDP, while that should be long enough not to erase learned structure rapidly [251]. I set τ_h in order of minutes in the simulation.

Finally, to ensure the stability of the recurrent network, I set boundary conditions for excitatory synapses as $0 < J_{ij}^{EE} < J_{max}$ and for the mean excitatory synaptic weight on individual excitatory cells as $0 < \frac{1}{K_i^E} \sum_{j \neq i}^{N_E} J_{ij}^{EE} < J_{max}^{tot}$, where K_i^E is the total number of excitatory inputs to neuron i. When the mean excitatory synaptic weight exceeds the upper limit, I subtract the excess amount from all synapses equally.

I used discrete update rule for spiking to reduce the computational cost, and employed differential equations only for slow variables (i.e., synaptic efficacies and homeostatic plasticity). This heterotic update procedure makes simulations faster and more robust in a broad range of parameter values without changing the essential features of network dynamics. However, because the exact spike timing depends on the random update of binary neurons, the update of synapses by STDP undergoes additional noise. This large noise seems reasonable because the in vitro synaptic modification by STDP is often highly noisy [20], and is expected to be more noisy in vivo. To justify the heterotic update procedure, I performed simulations in a similar network of Poisson neuron model. The details of this model are explained below and Supplementary Figure S1.

Spiking neuron model

In the main article, I used a binary model for modeling neuron. In order to support the generality of the model, I reproduce the main results of the model with a Poisson neuron model [74] [76]. Excitatory and inhibitory neurons follow spiking dynamics defined as below. Synaptic depression is added only for

E-to-E connections.

$$
\begin{aligned}
u_i^E(t) &= \sum_{j\neq i}^{N_E} J_{ij}^{EE} \int_0^\infty \varepsilon_E(\tau) y_j(t-\tau) x_j^E(t-\tau-d_{ij}^{EE}) d\tau - \sum_j^{N_I} J_{EI} \int_0^\infty \varepsilon_I(\tau) x_j^I(t-\tau-d_{ij}^{EI}) d\tau \\
u_i^I(t) &= \sum_j^{N_E} J_{IE} \int_0^\infty \varepsilon_E(\tau) x_j^E(t-\tau-d_{ij}^{IE}) d\tau - \sum_{j\neq i}^{N_I} J_{II} \int_0^\infty \varepsilon_I(\tau) x_j^I(t-\tau-d_{ij}^{II}) d\tau \\
\varepsilon_E(t) &= \frac{\exp(-t/\tau_E^A) - \exp(-t/\tau_E^B)}{\tau_E^A - \tau_E^B}, \varepsilon_I(t) = \frac{\exp(-t/\tau_I^A) - \exp(-t/\tau_I^B)}{\tau_I^A - \tau_I^B} \\
\frac{dy_i(t)}{dt} &= \frac{1-y_i(t)}{\tau_{sd}} - u_{sd} y_i(t) x_i^E(t)
\end{aligned}
$$

u_i^E, u_i^I are membrane potentials of excitatory/inhibitory neurons calculated by a sum of excitatory and inhibitory currents of a neuron. Synaptic currents are given by convolution of input spikes with EPSP/IPSP curves given as $\epsilon_E(t)$, $\epsilon_I(t)$. I assumed that synaptic delays d_{ij}^{EE}, d_{ij}^{IE}, d_{ij}^{EI}, d_{ij}^{II} are uniformly distributed in 0.5-1.5milliseconds for all connections. Synaptic depression is controlled by synaptic efficiency y_i. By membrane dynamics described in equations above, spiking process of neurons is given as below.

$$
\begin{aligned}
\rho_i^E(t) &= \rho_i^{E,ext}(t) + g_E\left(u_i^E(t)\right), g_E(u) = \frac{A_E}{1+\exp(-\lambda u + h_E)} \\
\rho_i^I(t) &= g_I\left(u_i^I(t)\right), g_I(u) = \frac{A_I}{1+\exp(-\lambda u + h_I)} \\
x_i^E(t) &\leftarrow Poisson\left(\rho_i^E(t)\right), x_i^I(t) \leftarrow Poisson\left(\rho_i^I(t)\right)
\end{aligned}
$$

Spikes x_i^E,x_i^I are probabilistically generated with sigmoidal response functions $g_E(u)$, $g_I(u)$. I added external inputs $\rho_i^{E,ext}(t) = 10Hz$ to ignite the spiking process at first 100 milliseconds of simulation. After that, external input terms $\rho_i^{E,ext}(t)$ are kept as zero. Synaptic weights of E-to-E connections are modified by STDP and homeostatic plasticity as below.

$$
\begin{aligned}
\frac{dJ_{ij}^{EE}}{dt} &= x_j^E(t-d_{ij}^{EE}) \int_0^\infty F_d(s, J_{ij}^{EE}) x_i^E(t-s) ds + x_i^E(t) \int_0^\infty F_p(s) x_i^E(t-s-d_{ij}^{EE}) ds \\
&\quad + \frac{J_{EE} - J_{ij}^{EE}}{\tau_h} + \sigma_h \zeta \\
F_d(s, J_{ij}^{EE}) &= C_d\left(1+\sigma_{stdp}\zeta\right) \frac{\log(1+\alpha J_{ij}^{EE}/J_{EE})}{\log(1+\alpha)} \exp(-s/\tau_d), F_p(s) = C_p\left(1+\sigma_{stdp}\zeta\right)\exp(-s/\tau_p)
\end{aligned}
$$

To guarantee stability of the model, I set lower/upper boundaries ($0 < J_{ij}^{EE} < 10 J_o^{EE}$) to E-to-E connections. I chose the same parameter with the model in the main text for time constant of STD, STDP, and homeostatic plasticity. Parameters used in the simulation are summarized in Table 2. All differential equations are solved with Runge-Kutta method with interval h = 0.1milliseconds.

As the simulation tends to take a long time, I created relatively small network with 300 excitatory neurons and 60 inhibitory neurons. Also, because the robustness in parameter space is relatively limited [166], I simulated only one configuration corresponding to Fig. 5.5, at a given parameter set. I

introduced two cell assemblies each consists of 100 non-overlapping excitatory neurons by hands with following equations for a Gaussian random variable ζ_{ij}.

$$J_{ij}^{EE}(t=0) = \begin{cases} 4J_{init}^{EE}\,(1+\sigma_J\zeta_{ij})\,(\mathrm{insidecellassemblies}) \\ J_{init}^{EE}\,(1+\sigma_J\zeta_{ij})\,(\mathrm{otherwise}) \end{cases}$$

Then, observed dynamics change of synaptic weights and neural activity after a dozen minutes of spontaneous activity. As a result, the network showed similar phenomena with those we observed in Fig. 5.5. When STD is weak (i.e., u_{sd} is small), two assemblies show competition, then eventually one of them become dominant (Figure S1-left, $u_{sd} = 0.15$). On the other hand, at strong STD, two assemblies tend to merge each other (Figure S1-right, $u_{sd} = 0.25$). At the adequate level of STD, both of them survive by alternative excitation (Figure S1-center, $u_{sd} = 0.2$).

In order to obtain the results shown for the Poisson neuron model, $g_E(u)$ needs to be a sigmoid-type function. When $g_E(u)$ is linear, bi-stable state is not robustly attained, while $g_E(u)$ is exponential, the network tends to display epileptic states. In addition, synaptic weight changes by STDP need to be noisy. On the other hand, in the original model σ_{step} was zero because the model has intrinsic noise due to probabilistic updating.

Mean-field (MF) approximation of cell-assembly dynamics

When the firing rate of presynaptic neuron j is constant, we find from the fixed point of equation (6) that synaptic efficiency y_j converges to $y_j = \frac{1}{1+u_{sd}\tau_{sd}r_j}$. With this relation, we may use a mean-field approximation for a given synaptic weight configuration [183] [196]. When excitatory neurons are separated into p number of non-overlapping cell assemblies with the sparseness a_1, a_2,..., a_p ($\sum_{\mu=1}^{p} a_\mu = 1$), the mean-field equations are calculated as follows:

$$\begin{aligned}
r_\mu &= H(u_\mu/\sigma_\mu)/\tau_E^{ud},\ r_I = H(u_I/\sigma_I)/\tau_I^{ud},\ H(x) = \frac{1}{2}\mathrm{erfc}(-x/\sqrt{2}),\ y_\mu = 1/(1+\gamma r_\mu \tau_E^{ud}), \\
u_\mu &= c_{EE}N_E\sum_{\nu=1}^{p} a_{nu}J_{\mu\nu}y_\nu r_\nu \tau_E^{sd} - c_{EI}N_I J_{EI} r_I \tau_I^{ud} + I_E^{ex}m_{es} - h_E, \\
u_I &= c_{IE}N_E J_{IE}\sum_{\mu=1}^{p} a_\mu r_\mu \tau_E^{ud} - c_{II}N_I J_{II} r_I \tau_I^{ud} + I_I^{ex}m_{ex} - h_I, \\
\sigma_\mu^2 &\cong c_{EE}N_E(1+\sigma_J^2)\sum_{\nu=1}^{p} J_{\mu\nu}^2 y_{nu}^2 r_\nu \tau_E^{ud} + c_{EI}N_I J_{EI}^2 r_I \tau_i^{ud} + (I_E^{ex}\sigma_{ex})^2, \\
\sigma_I^2 &\cong c_{IE}N_E J_{IE}^2(1+\sigma_J^2)\sum_{\mu=1}^{p} a_\mu r_\mu \tau_E^{ud} + c_{II}N_I J_{II}^2 r_I \tau_I^{ud} + (I_I^{ex}\sigma_{ex})^2
\end{aligned} \tag{5.9}$$

where parameter σ_J is the relative variance of synaptic weight, and $J_{\mu\nu}$ is the average synaptic weight from cell assembly ν to μ. When the synaptic weight distribution is not Gaussian, as in the case for log-STDP, the mean-field approximation is not accurate unless the correction terms representing the effect of strong synapses are added [220] [100]. However, here I use the above equations for simplicity.

In Fig. 5.6A-C, I calculate the fixed points of equation (9) for two cell assemblies, $ca1$ and $ca2$, by substituting $p = 3$, $a_1 = 0.2$, $a_2 = 0.2$, $a_3 = 0.6$ (a_3 corresponds to the background neurons) to equation (9) and by setting synaptic weights as

$$J_{\mu\nu} = \begin{cases} J_{ca1} & (\text{if } \mu = \nu = 1) \\ J_{ca2} & (\text{if } \mu = \nu = 2) \\ J_{bg} & (\text{otherwise}). \end{cases}$$

In the calculation, I assume that variables r_I and $r_3 (= r_{bg})$ are slaved to $r_1 (= r_{ca1})$ and $r_2 (= r_{ca2})$. As shown in Fig. 5.6E-F, I calculate the average firing rate $r_{ca}(\Delta t)$ after Δt milliseconds of a silent epoch, by substituting the post-silent-epoch efficiency $\tilde{y}'_{ca}(\Delta t)$ into the corresponding y_μ in equation (9). For instance, in the derivation of r'_{ca1}, I used $\tilde{y}'_{ca1}(\Delta t) = \frac{1}{1+\gamma r_{ca1}\tau_E^{ud}} + \left(1 + \frac{1}{1+\gamma r_{ca1}\tau_E^{ud}}\right)(1 - e^{\Delta t/\tau_{sd}})$ instead of $y_1 = \frac{1}{1+\gamma r_1 \tau_E^{ud}}$, then calculate the fixed point. Note that I set r_{ca1} equal to a fixed value estimated from simulations (in Fig. 5.6E, $r_{ca1} = 13.38$ [Hz] for $u_{sd} = 0.1$ and $r_{ca1} = 10.14$ [Hz] for $u_{sd} = 0.2$. In Fig. 5.6F, $r_{ca1} = 13.38$ [Hz] and $r_{ca2} = 12.82$ [Hz]), while r_1 was kept as a free variable.

MF approximation of weight dynamics

I extend the MF approximation to consider the weight dynamics under long-term synaptic plasticity. For simplicity, I assume that the average synaptic weight from a cell assembly to a background neuron pool is the same as the average weight from the background to the cell assembly. In this case, from the MF approximation, the stable point of the network is described by the three parameters r_I, r_{ca}, and r_{bg} corresponding to the average firing rates of inhibitory neurons, excitatory neurons belonging to a cell assembly, and other excitatory neurons (background neurons), and the three parameters J_{ca}, J_m, and J_{bg} representing the average weights of connections inside the cell assembly, between the assembly and the background, and among the background neurons, respectively. Thus, the equilibrium firing rates are expressed as

$$\begin{aligned} & r_I = H(u_I/\sigma_I)/\tau_I^{ud},\ r_{ca} = H(u_{ca}/\sigma_{ca})/\tau_E^{ud}/\tau_E^{ud},\ r_{bg} = H(u_{bg}/\sigma_{bg})/\tau_E^{ud}, \\ & r_{ca}^2 \left(C_p\tau_p - f_d(J_{ca})C_d\tau_d\right) + (J_{EE} - J_{ca})/\tau_h = 0, \\ & r_{ca}r_{bg} \left(C_p\tau_p - f_d(J_m)C_d\tau_d\right) + (J_{EE} - J_m)/\tau_h = 0, \\ & r_{bg}^2 \left(C_p\tau_p - f_d(J_{bg})C_d\tau_d\right) + (J_{EE} - J_{bg})/\tau_h = 0. \end{aligned} \tag{5.10}$$

Note that the above approximation is only applicable under the assumption that the firing rates are uniquely determined for the given synaptic weights. When the firing rates show bi-stability for given synaptic weights, an analytic approach to the synaptic weight dynamics is very hard.

Initial conditions

I set the initial synaptic weight matrix for simulations as $J_{ij}^{EE}(t=0) = J_{EE}^{init}(1+\sigma_J\zeta_{ij})$ in simulations shown in Figures 5.2 to 5.6. Those in Fig. 5.7 and Fig. 5.8A-D, the initial synaptic weight matrix is given as

$$J_{ij}^{EE}(t=0) = \begin{cases} J_{ca}(1+\sigma_J\zeta_{ij}) & \text{(inside cel assemblies)} \\ J_{bg}(1+\sigma_J\zeta_{ij}) & \text{(otherwise),} \end{cases}$$

where each cell assembly contains $N_E a$ neurons and ζ_{ij} is a Gaussian random variable. Parameter values are chosen as $J_{ca} = 0.70$, $J_{bg} = 0.16$, $a = 0.03$ and $p = 32$ for the model with a large number of cell assemblies, while $J_{ca} = 0.30$, $J_{bg} = 0.16$, $a = 0.2$ and $p = 3\text{or}5$ for the models with a small number of assemblies. In Fig. 5.8E, I introduce an initial bias in the weights within cell assemblies as

$$J_{ij}^{EE}(t=0) = \begin{cases} J_{ca}(1-0.2\eta_\mu)(1+\sigma_J\zeta_{ij}) & \text{(inside cel assemblies)} \\ J_{bg}(1+\sigma_J\zeta_{ij}) & \text{(otherwise),} \end{cases}$$

where η_μ is an uniform random variable drawn from $\eta_\mu \in [0,1)$ for each cell assembly μ. Similarly in Fig. 5.8F, I bias the weights within assemblies as

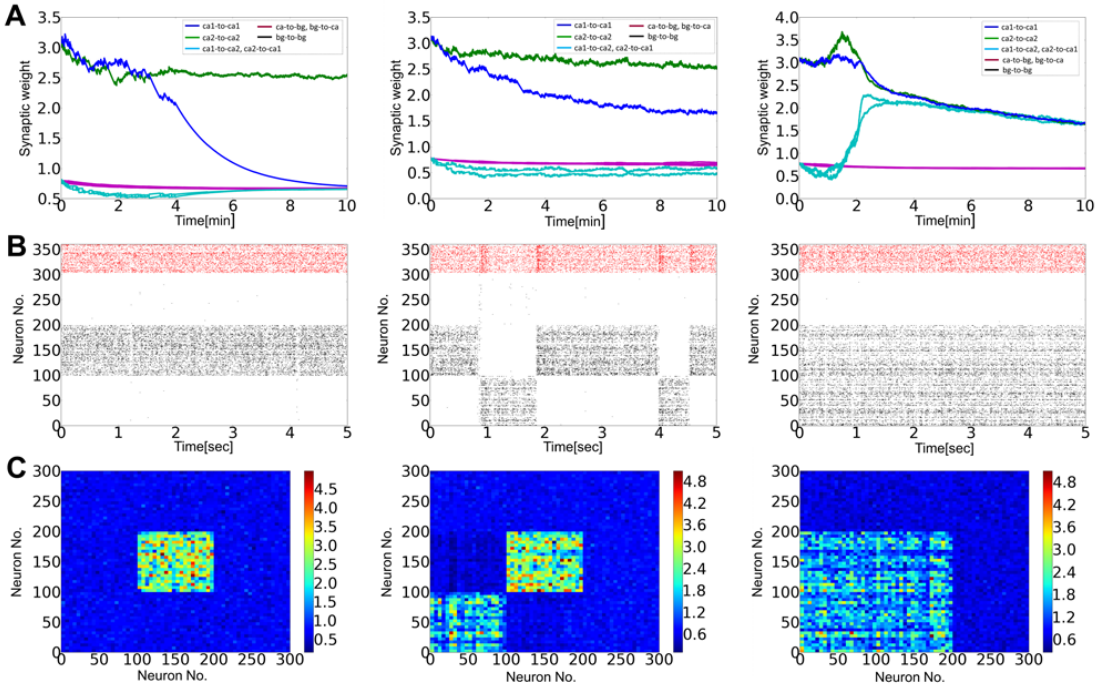
$$J_{ij}^{EE}(t=0) = \begin{cases} J_{ca}(1+\sigma_J\zeta_{ij}) & \text{(inside cel assemblies)} \\ J_{bg}(1+0.25\eta_{\mu\nu})(1+\sigma_J\zeta_{ij}) & \text{(otherwise).} \end{cases}$$

In all simulations, I set other initial conditions as $y_j(t=0) = 1/(1+6u_{sd})$, Prob$[x_i^E(t=0)=1] = 0.02$, and Prob$[x_i^I(t=0)=1] = 0.01$.

Details of simulation

In the presented simulations, every 0.01 milliseconds, 5 excitatory and 2 inhibitory randomly selected neurons are updated. STDP is calculated for neighboring spikes within 500 milliseconds. The differential equations of synaptic efficiency for STD is solved by Runge-Kutta method with 0.1 ms time steps, while homeostatic plasticity is calculated by Runge-Kutta method with 10.0 milliseconds time step in which values are updated at every t = 10.0 milliseconds for t = 0, 10, 20 ms,... This approximation is reasonable as homeostatic plasticity generates negligibly small changes in synaptic weights at each time step. The parameters used in the present simulations are summarized in Table 1. Code for simulations is written with C++ and Python, and is performed on a cluster machine.

Supplementary Figures



Supplementary Figure 1. Figure S1. The model with Poisson neuron model (**A**) Time evolution of the average synaptic weight for three values of u_{sd} (u_{sd} =0.15,0.20,0.25 from the left side). (**B**) Raster plots of spiking activity corresponding to the three cases shown in **A**. (**C**) Synaptic weight matrices of excitatory connections are shown for the above three cases. Configuration of graphs are the same with Fig. 5.5(**C**),(**D**),(**E**). Details of the model are summarized in the Method.

Table

Table 1: Parameters used in the simulations.

N_E, N_I	Number of excitatory/inhibitory neurons	2500, 500
$c_{EE}, c_{EI}, c_{IE}, c_{II}$	Connection probabilities	0.2,0.5,0.2,0.5
J_{IE}, J_{EI}, J_{II}	Synaptic weights	0.15,0.15,0.06 (In Fig. 5.2A and 3, $J_{EI} = 0.20$)
J_{EE}	Standard synaptic weight	0.15
J_{EE}^{init}, σ_J	Initial conditions of synaptic weight	0.18, 0.3
I_E^{ex}, I_I^{ex}	Amplitude of steady external input	2.0, 0.5
m_{ex}, σ_{ex}	Mean and variance of external input	0.3, 0.1
h_E, h_I	Thresholds of update	1.0, 1.0
t_E^{ud}, t_I^{ud}	Average intervals of update	5.0, 2.5 milliseconds
h	Interval of state update	0.01 milliseconds
τ_{sd}	Decay time constant of STD	600 milliseconds
u_{sd}	Release probability of synapse	0.05-0.5
C_p, C_d	Coefficients of STDP	0.01875, 0.0075
τ_p, τ_d	Decay time constants of STDP	20, 40 milliseconds
α	Degree of log-STDP	50.0
τ_h	Decay time of homeostatic plasticity	100 seconds
σ_h	Noise amplitude of homeostatic plasticity	0.00015 per 10 milliseconds
J_{max}, J_{max}^{tot}	Boundary conditions	0.75, 0.25

Table 2: Parameters used in the Poisson model.

N_E, N_I	Number of excitatory/inhibitory neurons	300, 60
c_{EE}, c_{EI}, c_{IE}, c_{II}	Connection probabilities	0.5,1.0,1.0,1.0
J_{IE}, J_{EI}, J_{II}	Synaptic weights	1.333,0.600,0.333
J_{EE}	Standard synaptic weight	0.667
d_{ij}^{EE}, d_{ij}^{EI}, d_{ij}^{IE}, d_{ij}^{II}	Synaptic delays	0.5-1.5 milliseconds
J_{EE}^{init}, σ_J	Initial conditions of synaptic weight	$1.15J_{EE}$, 0.1
A_E ,A_I	Maximal firing rates	100, 200Hz
h_E, h_I	Thresholds of f-I curve	0.5,2.0
τ_E^A, τ_E^B	EPSP-curve	5.0, 1.0 milliseconds
τ_I^A, τ_I^B	IPSP-curve	2.5, 1.0 milliseconds
τ_{sd}	Decay time constant of STD	600 milliseconds
u_{sd}	Release probability of synapse	0.15-0.25
C_p, C_d	Coefficients of STDP	$0.125J_{EE}$, $0.05J_{EE}$
τ_p, τ_d	Decay time constants of STDP	20, 40 milliseconds
α	Degree of log-STDP	50.0
σ_{step}	Noise amplitude of STDP	1.0
τ_h	Decay time of homeostatic plasticity	100 seconds
σ_h	Noise amplitude of homeostatic plasticity	$0.0001J_{EE}$ per 0.1 milliseconds

Chapter 6

Conclusion

How biological mechanisms of plasticity provide efficient learning schemes for neural computation?

In this thesis, I investigated synaptic dynamics and learning in various spatial and temporal scales, through both dynamic systems perspective and information-theoretic or machine learning perspectives. Due to this integrative approach, my studies provide several insights on how biological mechanisms of plasticity provide efficient learning schemes for neural computation.

First, on h-STDP, I found that h-STDP is effective for detecting a change in the environment, but not for maximization of neuronal excitability, because h-STDP robustly causes the detailed balance in dendritic branches. In particular, due to branch specificity of h-STDP, each branch is specialized for certain change, as a result, single neuron can detect change in a large domain (Chapter 2).

Secondly, I demonstrated that functional advantages of spine turnover depend on the sparseness of connectivity in the considered circuits. When connections are sparsely organized, creation and elimination of spines can yield a connection structure which is able to perform robust inference from given inputs, because functional connection structures tend to reduce signal variability. On the other hand, if there are dense connections between two layers, connection structure should capture the time-invariant components of the stimuli (Chapter 3).

Furthermore, I found that, in feedback-type neural circuits, correlation-based STDP learning mimics Bayesian ICA algorithms. To achieve the learning, spike correlation should not be too precise, because spike correlation does not propagate effectively in the circuit in that case. Moreover, my study also revealed potential functions of excitatory-to-inhibitory STDP and inhibitory-to-excitatory STDP in feedback-type circuits. These plasticity can cooperatively shape the lateral circuit for signal detection. In particular, through STDP, the lateral circuits is self-organized into a suitable structure depending on the number of independent signals projected to the circuit (Chapter 4).

Finally, my study on cell assembly modulation proposes a functional role for dopaminergic modulation of STDP. Cell assemblies are potentially better retained under dopaminergic modulation, and

bi-directional merging is enhanced because of the change in STDP time window. In addition, small but non-zero synaptic release probability supports these retention and merging process by enriching the neural dynamics (Chapter 5).

Relationship between studies

It should be noted that, chronologically speaking, four works are conducted in the opposite way. In my Master's thesis on a spiking neuron model of associative memory, I revealed the condition in which attractor states and spontaneous activity [100], but both analytical and simulation study suggested that such multistable states are only attainable in some finely-tuned parameter regions. I hypothesized that if attractors states are automatically recalled in the spontaneous activity, by activity-dependent synaptic plasticity, the neural circuit could be able to stay in the finely tuned state, and consequently retain memory traces. As a result, I developed a spiking neuron model of cell assembly modulation discussed in Chapter 5.

In that study, I developed an analytical techniques to analyze interaction between neural dynamics and synaptic weight dynamics, but the correspondence with simulation results and analytical predictions was limited partly due to complexity of the fully recurrent neural model. Thus, I was motivated to do analytical works on some simpler network motifs, such as feedback-type circuits. In addition, in the model used in Chapter 5 and many other previous studies on STDP in recurrent circuits, the learning was mainly driven by firing rates, not by spike-correlation, although STDP learning should be performed though spike correlation, because otherwise STDP is not necessary. From these two motivations, I next studied STDP learning based on spike-correlation propagation in a feedback-type circuit (Chapter 4).

The work in chapter 3 was conducted from a little different motivation. For one thing, I hypothesized that synaptic rewiring can be well described from the perspective of optimality, partly motivated from the result about Bayesian ICA in Chapter 4. For the other thing, considering the learning beyond local neural circuits, connection structure is expected to play a crucial role, but very few theoretical results were known on that topic, especially, the relationship between connection structure and synaptic weight plasticity remains elusive. Motivated from these two perspectives, the study in Chapter 3 was developed.

The work in chapter 2 was motivated from works in Chapter 3 and 4. In the study in Chapter 4, I focused on the influence of somatic inhibition for synaptic plasticity, although many inhibitory inputs are projected to the dendritic tree. Therefore, I was motivated to perform complementary study on the dendritic inhibition, especially, on their functional roles in excitatory synaptic plasticity. In addition, the simple model of spine turnover in Chapter 3 had a limited prediction power over experimental study. Thus, I developed a model of dendritic plasticity in Chapter 2 which proposed several experimentally testable predictions.

Future direction

Functional roles of redundant synaptic connections

Due to technical advance, growing numbers of new kinds of data are available in neuroscience nowadays. One remarkable attempt is connectomics, which is a study on detailed structure of neural circuits. Although, so far they have reconstructed a tiny portion of the brain ($\sim 10\mu m^3$), some of their results are already insightful. For instance, in their recent paper [118], they revealed that there are many multiple connections between identified axon-dendrite pairs, though they only constructed a small portion of a dendritic tree.

Another interesting yet highly criticized attempt is the blue brain project. In the project, Markram and his colleagues are conducting reconstruction of rodent or hopefully human brain in a supercomputer. Their reconstruction is still limited to a single column of the barrel cortex, yet some of their data accumulated for reconstruction are again quite insightful. In particular, From morphological reconstruction and algorithmic estimation, they claimed that in the barrel cortex, most interneuronal connections are actually realized by multiple synapses, and mean number of synapses per connection is estimated to be around 10 [154].

Thus, both of these two new studies indicate that synaptic connections are much more redundant than we previously though they were. However, little is known about their functional roles. In particular, synaptic connections are often created sporadically on the dendritic tree, thus each synapse in a single pre-post pair may play different roles in dendritic computation. By extending Bayesian method employed in Chapter 3, I am planning to give a insight on this issue.

Beyond local circuits

In this thesis, I mainly considered local circuits, or small fractions of local circuits for simplicity. Indeed, most of previous studies in theoretical neuroscience are focused on local circuit such as feedforward networks, or randomly connected recurrent networks [212]. However, to fully understand the brain, especially its higher-order functions, it is inevitable to study global circuits, such as cortical microcircuits, or hippocampal-entorhinal circuits. For example, we do not know how information is routed from a circuit to other circuits, how neural circuits learn to select relevant information from bombardment of incoming spikes, and how innate or learned connection structures guide neural computation. These questions should be fully investigated in the next decade, for further understanding of the brain.

Bibliography

[1] Abbott LF, Varela JA, Sen K, Nelson SB (1997) Synaptic Depression and Cortical Gain Control. Science 275:221-224.

[2] Abramov E, Dolev I, Fogel H, Ciccotosto GD, Ruff E, Slutsky I (2009) Amyloid-β as a positive endogenous regulator of release probability at hippocampal synapses. Nat Neurosci 12:1567-1576.

[3] Adesnik H, Scanziani M. Lateral competition for cortical space by layer-specific horizontal circuits. Nature. 2010;464: 1155-1160. doi:10.1038/nature08935

[4] Alonso J-M, Usrey WM, Reid RC. Precisely correlated firing in cells of the lateral geniculate nucleus. Nature. 1996;383: 815-819. doi:10.1038/383815a0

[5] Amari S. Dynamics of pattern formation in lateral-inhibition type neural fields. Biol Cybern. 1977;27: 77-87.

[6] Amari S. Natural Gradient Works Efficiently in Learning. Neural Comput. 1998;10: 251-276. doi:10.1162/089976698300017746

[7] Araya R, Vogels TP, Yuste R. Activity-dependent dendritic spine neck changes are correlated with synaptic strength. Proc Natl Acad Sci. 2014;111: E2895-E2904. doi:10.1073/pnas.1321869111

[8] Asari H, Pearlmutter BA, Zador AM. Sparse Representations for the Cocktail Party Problem. J Neurosci. 2006;26: 7477-7490. doi:10.1523/JNEUROSCI.1563-06.2006

[9] Arevian AC, Kapoor V, Urban NN. Activity-dependent gating of lateral inhibition in the mouse olfactory bulb. Nat Neurosci. 2008;11: 80-87. doi:10.1038/nn2030

[10] Atencio CA, Schreiner CE. Spectrotemporal Processing Differences between Auditory Cortical Fast-Spiking and Regular-Spiking Neurons. J Neurosci. 2008;28: 3897-3910. doi:10.1523/JNEUROSCI.5366-07.2008

[11] Babadi B, Sompolinsky H. Sparseness and Expansion in Sensory Representations. Neuron 2014;83, 1213-1226. doi:10.1016/j.neuron.2014.07.035

[12] Bair W, Zohary E, Newsome WT. Correlated Firing in Macaque Visual Area MT: Time Scales and Relationship to Behavior. J Neurosci. 2001;21: 1676-1697.

[13] Barbieri F and Brunel N (2007) Irregular persistent activity induced by synaptic excitatory feedback. Front Comput Neurosci 1:5.

[14] Bar-Ilan, L., Gidon, A., Segev, I., 2013. The role of dendritic inhibition in shaping the plasticity of excitatory synapses. Front. Neural Circuits 6. doi:10.3389/fncir.2012.00118

[15] Bartol TM, Bromer C, Kinney JP, Chirillo MA, Bourne JN, Harris KM, et al. Nanoconnectomic upper bound on the variability of synaptic plasticity. eLife. 2015; e10778. doi:10.7554/eLife.10778

[16] Bartsch AP, van Hemmen JL. Combined Hebbian development of geniculocortical and lateral connectivity in a model of primary visual cortex. Biol Cybern. 2001;84: 41-55.

[17] Beck JM, Ma WJ, Kiani R, Hanks T, Churchland AK, Roitman J, et al. Probabilistic Population Codes for Bayesian Decision Making. Neuron. 2008;60: 1142-1152. doi:10.1016/j.neuron.2008.09.021

[18] Beck JM, Ma WJ, Pitkow X, Latham PE, Pouget A. Not Noisy, Just Wrong: The Role of Suboptimal Inference in Behavioral Variability. Neuron. 2012;74: 30-39. doi:10.1016/j.neuron.2012.03.016

[19] Bell AJ, Sejnowski TJ. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. Neural Comput. 1995;7: 1129-1159. doi:10.1162/neco.1995.7.6.1129

[20] Bi GQ, Poo MM. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci Off J Soc Neurosci. 1998;18: 10464-10472.

[21] Billings G, van Rossum MCW (2009) Memory Retention and Spike-Timing-Dependent-Plasticity. J Neurophysiol 101:2775-2788.

[22] Blaise, J.H., Bronzino, J.D., 2003. Effects of stimulus frequency and age on bidirectional synaptic plasticity in the dentate gyrus of freely moving rats. Exp. Neurol. 182, 497-506. doi:10.1016/S0014-4886(03)00136-5

[23] Bliss TV, Collingridge GL. A synaptic model of memory: long-term potentiation in the hippocampus. Nature. 1993;361: 31-39. doi:10.1038/361031a0

[24] Boustani SE, Yger P, Frgnac Y, Destexhe A (2012) Stable learning in stochastic network states. J Neurosci 32:194-214.

[25] Branco, T., Clark, B.A., Husser, M., 2010. Dendritic Discrimination of Temporal Input Sequences in Cortical Neurons. Science 329, 1671-1675. doi:10.1126/science.1189664

[26] Brunel N, Hakim V, Isope P, Nadal J-P, Barbour B. Optimal Information Storage and the Distribution of Synaptic Weights: Perceptron versus Purkinje Cell. Neuron. 2004;43: 745-757. doi:10.1016/j.neuron.2004.08.023

[27] Buesing L, Bill J, Nessler B, Maass W. Neural Dynamics as Sampling: A Model for Stochastic Computation in Recurrent Networks of Spiking Neurons. PLoS Comput Biol. 2011;7: e1002211. doi:10.1371/journal.pcbi.1002211

[28] Buracas, G.T., Zador, A.M., DeWeese, M.R., Albright, T.D., 1998. Efficient Discrimination of Temporal Patterns by Motion-Sensitive Neurons in Primate Visual Cortex. Neuron 20, 959-969. doi:10.1016/S0896-6273(00)80477-8

[29] Bush D, Philippides A, Husbands P, O'Shea M (2010) Dual coding with STDP in a spiking recurrent neural network model of the hippocampus. PLoS Comput Biol 6: e1000839.

[30] Buzski G (2010) Neural syntax: cell assemblies, synapsembles, and readers. Neuron 68:362-385.

[31] Buzsáki G, Mizuseki K. The log-dynamic brain: how skewed distributions affect network operations. Nat Rev Neurosci. 2014;15: 264-278. doi:10.1038/nrn3687

[32] Caporale, N., Dan, Y., 2008. Spike Timing-Dependent Plasticity: A Hebbian Learning Rule. Annu. Rev. Neurosci. 31, 25-46. doi:10.1146/annurev.neuro.31.060407.125639

[33] Caroni P, Donato F, Muller D. Structural plasticity upon learning: regulation and functions. Nat Rev Neurosci. 2012;13: 478-490. doi:10.1038/nrn3258

[34] Cassenaer S, Laurent G (2012) Conditional modulation of spike-timing-dependent plasticity for olfactory learning. Nature 482:47-52.

[35] Chechik G, Meilijson I, Ruppin E. Synaptic Pruning in Development: A Computational Account. Neural Comput. 1998;10: 1759?1777. doi:10.1162/089976698300017124

[36] Cheetham CEJ, Fox K (2010) Presynaptic Development at L4 to L2/3 Excitatory Synapses Follows Different Time Courses in Visual and Somatosensory Cortex. J Neurosci 30:12566-12571.

[37] Chen BL, Hall DH, Chklovskii DB. Wiring optimization can relate neuronal structure and function. Proc Natl Acad Sci U S A. 2006;103: 4723-4728. doi:10.1073/pnas.0506806103

[38] Chen, X., Leischner, U., Rochefort, N.L., Nelken, I., Konnerth, A., 2011. Functional mapping of single spines in cortical neurons in vivo. Nature 475, 501-505. doi:10.1038/nature10193

[39] Chen, J.L., Villa, K.L., Cha, J.W., So, P.T.C., Kubota, Y., Nedivi, E., 2012. Clustered Dynamics of Inhibitory Synapses and Dendritic Spines in the Adult Neocortex. Neuron 74, 361-373. doi:10.1016/j.neuron.2012.02.030

[40] Chen, J.-Y., Lonjers, P., Lee, C., Chistiakova, M., Volgushev, M., Bazhenov, M., 2013. Heterosynaptic Plasticity Prevents Runaway Synaptic Dynamics. J. Neurosci. 33, 15915-15929. doi:10.1523/JNEUROSCI.5088-12.2013

[41] Cherry EC. Some Experiments on the Recognition of Speech, with One and with Two Ears. J Acoust Soc Am. 1953;25: 975-979. doi:10.1121/1.1907229

[42] Chklovskii DB, Mel BW, Svoboda K. Cortical rewiring and information storage. Nature. 2004;431: 782-788. doi:10.1038/nature03012

[43] Clopath C, Bsing L, Vasilaki E, Gerstner W. Connectivity reflects coding: a model of voltage-based STDP with homeostasis. Nat Neurosci. 2010;13: 344-352. doi:10.1038/nn.2479

[44] Comon P. Independent component analysis, A new concept? Signal Process. 1994;36: 287-314. doi:10.1016/0165-1684(94)90029-9

[45] Couey JJ, Witoelar A, Zhang S-J, Zheng K, Ye J, Dunn B, et al. Recurrent inhibitory circuitry as a mechanism for grid formation. Nat Neurosci. 2013;16: 318-324. doi:10.1038/nn.3310

[46] Cutsuridis, V., 2011. GABA inhibition modulates NMDA-R mediated spike timing dependent plasticity (STDP) in a biophysical model. Neural Netw. 24, 29-42. doi:10.1016/j.neunet.2010.08.005

[47] Dan Y, Alonso J-M, Usrey WM, Reid RC. Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. Nat Neurosci. 1998;1: 501-507. doi:10.1038/2217

[48] Dayan P, Hinton GE, Neal RM, Zemel RS. The Helmholtz machine. Neural Comput. 1995;7: 889-904.

[49] Dayan P, Abbott LF. Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems. 1 edition. Cambridge, Mass.: The MIT Press; 2005.

[50] Deger M, Helias M, Rotter S, Diesmann M. Spike-timing dependence of structural plasticity explains cooperative synapse formation in the neocortex. PLoS Comput Biol. 2012;8: e1002689.

[51] Debanne D, Gurineau NC, Ghwiler BH, Thompson SM (1996) Pair-pulse facilitation and depression at unitary synapses in rat hippocampus: quantal fluctuation affects subsequent release. J Physiol 491:163-176.

[52] deCharms RC, Merzenich MM. Primary cortical representation of sounds by the coordination of action-potential timing. Nature. 1996;381: 610-613. doi:10.1038/381610a0

[53] Del Guidice P, Mattia M (2001) Long and short-term synaptic plasticity and the formation of working memory: A case study. Neurocomputing 38-40:1175-1180.

[54] DeWeese, M.R., Wehr, M., Zador, A.M., 2003. Binary Spiking in Auditory Cortex. J. Neurosci. 23, 7940-7949.

[55] Diekelmann S, Born J (2010) The memory function of sleep. Nat Rev Neurosci 11: 114-126.

[56] Diekelmann S, Born J, Wagner U (2010) Sleep enhance false memories depending on general memory performance. Behav Brain Res 208:425-429.

[57] Donoho DL. Compressed sensing. IEEE Trans Inf Theory. 2006;52: 1289?1306. doi:10.1109/TIT.2006.871582

[58] Dorrn, A.L., Yuan, K., Barker, A.J., Schreiner, C.E., Froemke, R.C., 2010. Developmental sensory experience balances cortical excitation and inhibition. Nature 465, 932-936. doi:10.1038/nature09119

[59] Dupret D, O'Neill J, Csicsvari J. Dynamic reconfiguration of hippocampal interneuron circuits during spatial learning. Neuron. 2013;78: 166-180. doi:10.1016/j.neuron.2013.01.033

[60] Ellenbogen JM, Hu PT, Payne JD, Titone D, Walker WP (2007) Human relational memory requires time and sleep. Proc Natl Acad Sci U S A 104:7723-7728.

[61] Espinosa, J.S., Stryker, M.P., 2012. Development and Plasticity of the Primary Visual Cortex. Neuron 75, 230-249. doi:10.1016/j.neuron.2012.06.009

[62] Faisal AA, Selen LPJ, Wolpert DM. Noise in the nervous system. Nat Rev Neurosci. 2008;9: 292-303. doi:10.1038/nrn2258

[63] Fauth M, Wörgötter F, Tetzlaff C. The Formation of Multi-synaptic Connections by the Interaction of Synaptic and Structural Plasticity and Their Functional Consequences. PLoS Comput Biol. 2015;11. doi:10.1371/journal.pcbi.1004031

[64] Feldman DE. Synaptic Mechanisms for Plasticity in Neocortex. Annu Rev Neurosci. 2009;32: 33-55. doi:10.1146/annurev.neuro.051508.135516

[65] Fiete IR, Senn W, Wang CZH, Hahnloser RHR (2010) Spike-timing-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. Neuron 65:563-576.

[66] Fino E, Paille V, Deniau J-M, Venance L. Asymmetric spike-timing dependent plasticity of striatal nitric oxide-synthase interneurons. Neuroscience. 2009;160: 744-754. doi:10.1016/j.neuroscience.2009.03.015

[67] Froemke, R.C., 2015. Plasticity of Cortical Excitatory-Inhibitory Balance. Annu. Rev. Neurosci. 38, 195-219. doi:10.1146/annurev-neuro-071714-034002

[68] Fukai T, Tanaka S. A simple neural network exhibiting selective activation of neuronal ensembles: from winner-take-all to winners-share-all. Neural Comput. 1997;9: 77-97.

[69] Fukai T, Kanemura S (2001) Noise-tolerant stimulus discrimination by synchronization with depressing synapses. Biol Cybern 85:107-116.

[70] Fusi S, Asaad WF, Miller EK, Wang X-J. A neural circuit model of flexible sensorimotor mapping: learning and forgetting on multiple timescales. Neuron. 2007;54: 319-333. doi:10.1016/j.neuron.2007.03.017

[71] Gambino, F., Pags, S., Kehayas, V., Baptista, D., Tatti, R., Carleton, A., Holtmaat, A., 2014. Sensory-evoked LTP driven by dendritic plateau potentials in vivo. Nature 515, 116-119. doi:10.1038/nature13664

[72] Ganguli S, Sompolinsky H. Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis. Annu Rev Neurosci. 2012;35: 485-508. doi:10.1146/annurev-neuro-062111-150410

[73] Gerstner W, Kempter R, van Hemmen JL, Wagner H. A neuronal learning rule for sub-millisecond temporal coding. Nature. 1996;383: 76-81. doi:10.1038/383076a0

[74] Gerstner W, Kistler WK (2002) Spiking Neuron Models. Cambridge, UK: Cambridge University Press.

[75] Gidon, A., Segev, I., 2012. Principles Governing the Operation of Synaptic Inhibition in Dendrites. Neuron 75, 330-341. doi:10.1016/j.neuron.2012.05.015

[76] Gilson M, Burkitt AN, Grayden DB, Thomas DA, van Hemmen JL. Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. I. Input selectivity–strengthening correlated input pathways. Biol Cybern. 2009;101: 81-102. doi:10.1007/s00422-009-0319-4

[77] Gilson M, Burkitt AN, Grayden DB, Thomas DA, van Hemmen JL. Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. II. Input selectivity–symmetry breaking. Biol Cybern. 2009;101: 103-114. doi:10.1007/s00422-009-0320-y

[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 IV: Structuring synaptic pathways among recurrent connections. Biol Cybern 101: 427-444.

[79] Gilson M, Burkitt AN, Grayden DB, Thomas DA, van Hemmen JL (2010a) Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks V: self-organization schemes and weight dependence. Biol Cybern 103:365-386.

[80] Gilson M, Burkitt A, van Hemmen JL (2010b) STDP in Recurrent Neuronal Networks. Front Comput Neurosci 4:23.

[81] Gilson M, Fukai T. Stability versus neuronal specialization for STDP: long-tail weight distributions solve the dilemma. PloS One. 2011;6: e25339. doi:10.1371/journal.pone.0025339

[82] Gilson M, Fukai T, Burkitt AN. Spectral analysis of input spike trains by spike-timing-dependent plasticity. PLoS Comput Biol. 2012;8: e1002584. doi:10.1371/journal.pcbi.1002584

[83] Girolami M, Gyfe C. Extraction of independent signal sources using a deflationary exploratory projection pursuit network with lateral inhibition. Vis Image Signal Process IEE Proc -. 1997;144: 299-306. doi:10.1049/ip-vis:19971418

[84] Gjorgjieva J, Clopath C, Audet J, Pfister J-P. A triplet spike-timing-dependent plasticity model generalizes the Bienenstock-Cooper-Munro rule to higher-order spatiotemporal correlations. Proc Natl Acad Sci. 2011;108: 19383-19388. doi:10.1073/pnas.1105933108

[85] Govindarajan A, Kelleher RJ, Tonegawa S. A clustered plasticity model of long-term memory engrams. Nat Rev Neurosci. 2006;7: 575?583. doi:10.1038/nrn1937

[86] Graupner, M., Brunel, N., 2012. Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. Proc. Natl. Acad. Sci. 109, 3991-3996. doi:10.1073/pnas.1109359109

[87] Graupner, M., Brunel, N., 2007. STDP in a Bistable Synapse Model Based on CaMKII and Associated Signaling Pathways. PLoS Comput Biol 3, e221. doi:10.1371/journal.pcbi.0030221

[88] Gtig R, Aharonov R, Rotter S, Sompolinsky H. Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. J Neurosci Off J Soc Neurosci. 2003;23: 3697-3714.

[89] Haas JS, Nowotny T, Abarbanel HDI. Spike-timing-dependent plasticity of inhibitory synapses in the entorhinal cortex. J Neurophysiol. 2006;96: 3305-3313. doi:10.1152/jn.00551.2006

[90] Habenschuss S, Puhr H, Maass W. Emergence of Optimal Decoding of Population Codes Through STDP. Neural Comput. 2013;25, 1371-1407. doi:10.1162/NECO_a_00446

[91] Haefner RM, Gerwinn S, Macke JH, Bethge M. Inferring decoding strategies from choice probabilities in the presence of correlated variability. Nat Neurosci. 2013;16: 235-242. doi:10.1038/nn.3309

[92] Hansel D and Mato H (2013) Short-Term Plasticity Explains Irregular Persistent Activity in Working Memory Tasks. J Neurosci 33: 133-149.

[93] Harvey, C.D., Svoboda, K., 2007. Locally dynamic synaptic learning rules in pyramidal neuron dendrites. Nature 450, 1195-1200. doi:10.1038/nature06416 09-1416. doi:10.1038/nn.3496

[94] Hayama T, Noguchi J, Watanabe S, Takahashi N, Hayashi-Takagi A, Ellis-Davies GCR, et al. GABA promotes the competitive selection of dendritic spines by controlling local Ca2+ signaling. Nat Neurosci. 2013;16: 1409-1416. doi:10.1038/nn.3496

[95] Haykin S, Chen Z. The Cocktail Party Problem. Neural Comput. 2005;17: 1875-1902. doi:10.1162/0899766054322964

[96] Hebb DO (1949) The organization of behavior: a neuropsychological theory. New York: Wiley.

[97] Hennequin G, Gerstner W, Pfister J-P (2010) STDP in Adaptive Neurons Gives Close-To-Optimal Information Transmission. Front Comput Neurosci 4:143.

[98] Hensch TK (2005) Critical period plasticity in local cortical circuits. Nat Rev Neurosci 6:877-888.

[99] Higley, M.J., Sabatini, B.L., 2012. Calcium Signaling in Dendritic Spines. Cold Spring Harb. Perspect. Biol. 4, a005686. doi:10.1101/cshperspect.a005686

[100] Hiratani N, Teramae J-N, Fukai T. Associative memory model with long-tail-distributed Hebbian synaptic connections. Front Comput Neurosci. 2013;6. doi:10.3389/fncom.2012.00102

[101] Hiratani N, Fukai T. Interplay between short- and long-term plasticity in cell-assembly formation. PloS One. 2014;9: e101535. doi:10.1371/journal.pone.0101535

[102] Hiratani, N., Fukai, T., 2015. Mixed Signal Learning by Spike Correlation Propagation in Feedback Inhibitory Circuits. PLoS Comput Biol 11, e1004227. doi:10.1371/journal.pcbi.1004227

[103] Hofer SB, Ko H, Pichler B, Vogelstein J, Ros H, Zeng H, et al. Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. Nat Neurosci. 2011;14: 1045-1052. doi:10.1038/nn.2876

[104] Holtmaat AJGD, Trachtenberg JT, Wilbrecht L, Shepherd GM, Zhang X, Knott GW, et al. Transient and persistent dendritic spines in the neocortex in vivo. Neuron. 2005;45: 279-291. doi:10.1016/j.neuron.2005.01.003

[105] Holtmaat A, Svoboda K. Experience-dependent structural synaptic plasticity in the mammalian brain. Nat Rev Neurosci. 2009;10: 647-658. doi:10.1038/nrn2699

[106] Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci U S A 79:2554-2558.

[107] Huang S, Huganir RL, Kirkwood A. Adrenergic gating of Hebbian spike-timing-dependent plasticity in cortical interneurons. J Neurosci Off J Soc Neurosci. 2013;33: 13171-13178. doi:10.1523/JNEUROSCI.5741-12.2013

[108] Ikegaya Y, Sasaki T, Ishikawa D, Honma N, Tao K, Takahashi N, et al. Interpyramid Spike Transmission Stabilizes the Sparseness of Recurrent Network Activity. Cereb Cortex. 2013;23: 293-304. doi:10.1093/cercor/bhs006

[109] Itami C, and Kimura F (2012) Developmental Switch in Spike Timing-Dependent Plasticity at Layers 4-2/3 in the Rodent Barrel Cortex. J Neurosci 32: 15000-15011.

[110] Izhikevich EM, Gally JA, Edelman GM (2004) Spike-timing dynamics of neuronal groups. Cereb Cortex 14:933-944.

[111] Jackson J, Redish AD. Network dynamics of hippocampal cell-assemblies resemble multiple spatial maps within single tasks. Hippocampus. 2007;17: 1209-1229. doi:10.1002/hipo.20359

[112] Jezek K, Henriksen EJ, Treves A, Moser EI, Moser M-B. Theta-paced flickering between place-cell maps in the hippocampus. Nature. 2011;478: 246-249. doi:10.1038/nature10439

[113] Jia, H., Rochefort, N.L., Chen, X., Konnerth, A., 2010. Dendritic organization of sensory input to cortical neurons in vivo. Nature 464, 1307-1312. doi:10.1038/nature08947

[114] Jutten C, Herault J. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. Signal Process. 1991;24: 1-10. doi:10.1016/0165-1684(91)90079-X

[115] Karst H, Berger S, Turiault M, Tronche F, Schtz G, and Jols M (2005) Mineralocorticoid receptors are indispensable for nongenomic modulation of hippocampal glutamate transmission by corticosterone. Proc Natl Acad Sci U S A 102: 19204-19207.

[116] Kasai H, Hayama T, Ishikawa M, Watanabe S, Yagishita S, Noguchi J. Learning rules and persistence of dendritic spines. Eur J Neurosci. 2010;32: 241-249. doi:10.1111/j.1460-9568.2010.07344.x

[117] Kastellakis G, Cai DJ, Mednick SC, Silva AJ, Poirazi P. Synaptic clustering within dendrites: An emerging theory of memory formation. Prog Neurobiol. 2015;126: 19?35. doi:10.1016/j.pneurobio.2014.12.002

[118] Kasthuri N, Hayworth KJ, Berger DR, Schalek RL, Conchello JA, Knowles-Barley S, et al. Saturated Reconstruction of a Volume of Neocortex. Cell. 2015;162: 648?661. doi:10.1016/j.cell.2015.06.054

[119] Kempter R, Gerstner W, van Hemmen JL. Hebbian learning and spiking neurons. Phys Rev E. 1999;59: 4498-4514. doi:10.1103/PhysRevE.59.4498

[120] King PD, Zylberberg J, DeWeese MR. Inhibitory interneurons decorrelate excitatory cells to drive sparse code formation in a spiking model of V1. J Neurosci Off J Soc Neurosci. 2013;33: 5475-5485. doi:10.1523/JNEUROSCI.4188-12.2013

[121] Klampfl S, Maass W (2013) Emergence of dynamic memory traces in cortical microcircuit models through STDP. J Neurosci 33:11515-11529.

[122] Kleindienst, T., Winnubst, J., Roth-Alpermann, C., Bonhoeffer, T., Lohmann, C., 2011. Activity-Dependent Clustering of Functional Synaptic Inputs on Developing Hippocampal Dendrites. Neuron 72, 1012-1024. doi:10.1016/j.neuron.2011.10.015

[123] Koch, C., 1998. Biophysics of Computation: Information Processing in Single Neurons. Oxford University Press.

[124] Kohn A, Smith MA. Stimulus Dependence of Neuronal Correlation in Primary Visual Cortex of the Macaque. J Neurosci. 2005;25: 3661-3673. doi:10.1523/JNEUROSCI.5106-04.2005

[125] Kremkow, J., Aertsen, A., Kumar, A., 2010. Gating of Signal Propagation in Spiking Neural Networks by Balanced and Correlated Excitation and Inhibition. J. Neurosci. 30, 15760-15768. doi:10.1523/JNEUROSCI.3874-10.2010

[126] Knoblauch A, Palm G, Sommer FT. Memory capacities for synaptic and structural plasticity. Neural Comput. 2010;22: 289-341. doi:10.1162/neco.2009.08-07-588

[127] Knott GW, Holtmaat A, Wilbrecht L, Welker E, Svoboda K. Spine growth precedes synapse formation in the adult neocortex in vivo. Nat Neurosci. 2006;9: 1117-1124. doi:10.1038/nn1747

[128] Knuth KH (2002) A Bayesian approach to source separation. arXiv:physics/0205032.

[129] Ko H, Cossell L, Baragli C, Antolik J, Clopath C, Hofer SB, et al. The emergence of functional microcircuits in visual cortex. Nature. 2013;496: 96?100. doi:10.1038/nature12015

[130] Kuhlman, S.J., Tring, E., Trachtenberg, J.T., 2011. Fast-spiking interneurons have an initial orientation bias that is lost with vision. Nat. Neurosci. 14, 1121-1123. doi:10.1038/nn.2890

[131] Kumar A, Rotter S, Aertsen A. Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. Nat Rev Neurosci. 2010;11: 615-627. doi:10.1038/nrn2886

[132] Lampl I, Reichova I, Ferster D. Synchronous Membrane Potential Fluctuations in Neurons of the Cat Visual Cortex. Neuron. 1999;22: 361-374. doi:10.1016/S0896-6273(00)81096-X

[133] Lamsa KP, Kullmann DM, Woodin MA. Spike-timing dependent plasticity in inhibitory circuits. Front Synaptic Neurosci. 2010;2: 8. doi:10.3389/fnsyn.2010.00008

[134] Lee S-H, Kwan AC, Zhang S, Phoumthipphavong V, Flannery JG, Masmanidis SC, et al. Activation of specific interneurons improves V1 feature selectivity and visual perception. Nature. 2012;488: 379-383. doi:10.1038/nature11312

[135] Lefort S, Tomm C, Floyd Sarria J-C, Petersen CCH. The Excitatory Neuronal Network of the C2 Barrel Column in Mouse Primary Somatosensory Cortex. Neuron. 2009;61: 301-316. doi:10.1016/j.neuron.2008.12.020

[136] Legenstein R, Naeger C, Maass W. What can a neuron learn with spike-timing-dependent plasticity? Neural Comput. 2005;17: 2337-2382. doi:10.1162/0899766054796888

[137] Lengyel M, Kwag J, Paulsen O, Dayan P (2005) Matching storage and recall: hippocampal spike timing-dependent plasticity and phase response curves. Nat Neurosci 8:1677-1683.

[138] Letzkus JJ, Kampa BM, Stuart GJ. Learning Rules for Spike Timing-Dependent Plasticity Depend on Dendritic Synapse Location. J Neurosci. 2006;26: 10420-10429. doi:10.1523/JNEUROSCI.2650-06.2006

[139] Levy N, Horn D, Meilijson I, Ruppin E (2001) Distributed synchrony in a cell assembly of spiking neurons. Neural Netw 14:815-824.

[140] Lewis PA, Durrant SJ (2011) Overlapping memory replay during sleep builds cognitive schemata. Trends Neurosci 15: 343-351.

[141] Litwin-Kumar A, Doiron B (2012) Slow dynamics and high variability in balanced cortical networks with clustered connections. Nat Neurosci 15:1498-1505.

[142] Litwin-Kumar A, Doiron B. Formation and maintenance of neuronal assemblies through synaptic plasticity. Nat Commun. 2014;5. doi:10.1038/ncomms6319

[143] Liu, G., 2004. Local structural balance and functional interaction of excitatory and inhibitory synapses in hippocampal dendrites. Nat. Neurosci. 7, 373-379. doi:10.1038/nn1206

[144] Lochmann T, Deneve S. Neural processing as causal inference. Curr Opin Neurobiol. 2011;21: 774-781. doi:10.1016/j.conb.2011.05.018

[145] London, M., Husser, M., 2005. Dendritic Computation. Annu. Rev. Neurosci. 28, 503-532. doi:10.1146/annurev.neuro.28.061604.135703

[146] Lu J, Li C, Zhao J-P, Poo M, Zhang X. Spike-timing-dependent plasticity of neocortical excitatory synapses on inhibitory interneurons depends on target cell type. J Neurosci Off J Soc Neurosci. 2007;27: 9711-9720. doi:10.1523/JNEUROSCI.2513-07.2007

[147] Lüscher, C., Malenka, R.C., 2012. NMDA Receptor-Dependent Long-Term Potentiation and Long-Term Depression (LTP/LTD). Cold Spring Harb. Perspect. Biol. 4, a005710. doi:10.1101/cshperspect.a005710

[148] Ma, W., Liu, B., Li, Y., Huang, Z.J., Zhang, L.I., Tao, H.W., 2010. Visual Representations by Cortical Somatostatin Inhibitory Neurons-Selective But with Weak and Delayed Responses. J. Neurosci. 30, 14371-14379. doi:10.1523/JNEUROSCI.3248-10.2010

[149] Maass W, Natschläger T, Markram H. Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations. Neural Comput. 2002;14: 2531-2560. doi:10.1162/089976602760407955

[150] Mainen ZF, Sejnowski TJ. Reliability of spike timing in neocortical neurons. Science. 1995;268: 1503-1506. doi:10.1126/science.7770778

[151] Malenka, R.C., Bear, M.F., 2004. LTP and LTD: An Embarrassment of Riches. Neuron 44, 5-21. doi:10.1016/j.neuron.2004.09.012

[152] Markram H, Lbke J, Frotscher M, Sakmann B. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science. 1997;275: 213-215.

[153] Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., Wu, C., 2004. Interneurons of the neocortical inhibitory system. Nat. Rev. Neurosci. 5, 793-807. doi:10.1038/nrn1519

[154] Markram H, Muller E, Ramaswamy S, Reimann MW, Abdellah M, Sanchez CA, et al. Reconstruction and Simulation of Neocortical Microcircuitry. Cell. 2015;163: 456?492. doi:10.1016/j.cell.2015.09.029

[155] Marro J, Torres JJ (2007) Chaotic hopping between attractors in neural networks. Neural Netw. 20:230-235.

[156] Masamizu Y, Tanaka YR, Tanaka YH, Hira R, Ohkubo F, Kitamura K, et al. Two distinct layer-specific dynamics of cortical ensembles during learning of a motor task. Nat Neurosci. 2014;17: 987-994. doi:10.1038/nn.3739

[157] Masquelier T, Guyonneau R, Thorpe SJ. Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains. PloS One. 2008;3: e1377. doi:10.1371/journal.pone.0001377

[158] Masquelier T, Guyonneau R, Thorpe SJ. Competitive STDP-based spike pattern learning. Neural Comput. 2009;21: 1259-1276. doi:10.1162/neco.2008.06-08-804

[159] Matsui A, Tran M, Yoshida AC, Kikuchi SS, U M, Ogawa M, et al. BTBD3 controls dendrite orientation toward active axons in mammalian neocortex. Science. 2013;342: 1114-1118. doi:10.1126/science.1244505

[160] Matsuzaki M, Honkura N, Ellis-Davies GCR, Kasai H. Structural basis of long-term potentiation in single dendritic spines. Nature. 2004;429: 761-766. doi:10.1038/nature02617

[161] McClelland JL, McNaughton BL, O'Reilly RC (1995) Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. Psychol Rev 102:419-457.

[162] McDermott JH. The cocktail party problem. Curr Biol CB. 2009;19: R1024-1027. doi:10.1016/j.cub.2009.09.005

[163] Mel, B.W., Schiller, J., 2004. On the Fight Between Excitation and Inhibition: Location Is Everything. Sci. Signal. 2004, pe44-pe44. doi:10.1126/stke.2502004pe44

[164] Molgedey L, Schuster HG. Separation of a mixture of independent signals using time delayed correlations. Phys Rev Lett. 1994;72: 3634-3637.

[165] Moore, A.K., Wehr, M., 2013. Parvalbumin-Expressing Inhibitory Interneurons in Auditory Cortex Are Well-Tuned for Frequency. J. Neurosci. 33, 13713-13723. doi:10.1523/JNEUROSCI.0663-13.2013

[166] Morrison A, Aertsen A, Diesmann M (2007) Spike-Timing-Dependent Plasticity in Balanced Random Networks. Neural Comput 19:1437-1467.

[167] Moussaoui S, Brie D, Mohammad-Djafari A, Carteret C. Separation of Non-Negative Mixture of Non-Negative Sources Using a Bayesian Approach and MCMC Sampling. IEEE Trans Signal Process. 2006;54: 4133-4145. doi:10.1109/TSP.2006.880310

[168] Müllner, F.E., Wierenga, C.J., Bonhoeffer, T., 2015. Precision of Inhibition: Dendritic Inhibition by Individual GABAergic Synapses on Hippocampal Pyramidal Cells Is Confined in Space and Time. Neuron 87, 576-589. doi:10.1016/j.neuron.2015.07.003

[169] Munz M, Gobert D, Schohl A, Poquérusse J, Podgorski K, Spratt P, et al. Rapid Hebbian axonal remodeling mediated by visual stimulation. Science. 2014;344: 904-909. doi:10.1126/science.1251593

[170] Navlakha S, Barth AL, Bar-Joseph Z. Decreasing-Rate Pruning Optimizes the Construction of Efficient and Robust Distributed Networks. PLoS Comput Biol. 2015;11. doi:10.1371/journal.pcbi.1004347

[171] Nessler B, Pfeiffer M, Buesing L, Maass W. Bayesian Computation Emerges in Generic Cortical Microcircuits through Spike-Timing-Dependent Plasticity. PLoS Comput Biol. 2013;9: e1003037. doi:10.1371/journal.pcbi.1003037

[172] Nissen W, Szabo A, Somogyi J, Somogyi P, Lamsa KP. Cell Type-Specific Long-Term Plasticity at Glutamatergic Synapses onto Hippocampal Interneurons Expressing either Parvalbumin or CB1 Cannabinoid Receptor. J Neurosci. 2010;30: 1337-1347. doi:10.1523/JNEUROSCI.3481-09.2010

[173] Niessing J, Friedrich RW. Olfactory pattern classification by discrete neuronal network states. Nature. 2010;465: 47-52. doi:10.1038/nature08961

[174] Nishiyama, J., Yasuda, R., 2015. Biochemical Computation for Spine Structural Plasticity. Neuron 87, 63-75. doi:10.1016/j.neuron.2015.05.043

[175] O'Donnell C, Nolan MF, van Rossum MCW. Dendritic Spine Dynamics Regulate the Long-Term Stability of Synaptic Plasticity. J Neurosci. 2011;31: 16142-16156. doi:10.1523/JNEUROSCI.2520-11.2011

[176] Oh WC, Parajuli LK, Zito K. Heterosynaptic Structural Plasticity on Local Dendritic Segments of Hippocampal CA1 Neurons. Cell Rep. 2015;10: 162-169. doi:10.1016/j.celrep.2014.12.016

[177] Oja E. Simplified neuron model as a principal component analyzer. J Math Biol. 1982;15: 267-273. doi:10.1007/BF00275687

[178] Oja E. Neural networks, principal components, and subspaces. Int J Neural Syst. 1989;01: 61-68. doi:10.1142/S0129065789000475

[179] O'Neill J, Pleydell-Bouverie B, Dupret D, Csicsvari J (2010) Play it again: reactivation of waking experience and memory. Trends Neurosci 33:220-229.

[180] O'Reilly, R.C., 1996. Biologically Plausible Error-Driven Learning Using Local Activation Differences: The Generalized Recirculation Algorithm. Neural Comput. 8, 895-938. doi:10.1162/neco.1996.8.5.895

[181] Oswald AMM, Reyes AD (2008) Maturation of Intrinsic and Synaptic Properties of Layer 2/3 Pyramidal Neurons in Mouse Auditory Cortex. J Neurophysiol 99:2998-3008.

[182] Paille, V., Fino, E., Du, K., Morera-Herreras, T., Perez, S., Kotaleski, J.H., Venance, L., 2013. GABAergic Circuits Control Spike-Timing-Dependent Plasticity. J. Neurosci. 33, 9353-9363. doi:10.1523/JNEUROSCI.5796-12.2013

[183] Pantic L, Torres JJ, Kappen HJ, Gielen SCAM (2002) Associative memory with dynamic synapses. Neural Comput 14:2903-2923.

[184] Perin R, Berger TK, Markram H. A synaptic organizing principle for cortical neuronal groups. Proc Natl Acad Sci. 2011;108: 5419-5424. doi:10.1073/pnas.1016051108

[185] Petersen, C.C.H., Malenka, R.C., Nicoll, R.A., Hopfield, J.J., 1998. All-or-none potentiation at CA3-CA1 synapses. Proc. Natl. Acad. Sci. 95, 4732-4737.

[186] Petrovici MA, Bill J, Bytschok I, Schemmel J, Meier K (2013) Stochastic inference with deterministic spiking neurons. ArXiv13113211 Cond-Mat Physicsphysics Q-Bio.

[187] Pfister J-P, Gerstner W. Triplets of Spikes in a Model of Spike Timing-Dependent Plasticity. J Neurosci. 2006;26: 9673-9682. doi:10.1523/JNEUROSCI.1425-06.2006

[188] Poirazi P, Mel BW. Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. Neuron. 2001;29: 779-796.

[189] Poirazi, P., Brannon, T., Mel, B.W., 2003. Pyramidal Neuron as Two-Layer Neural Network. Neuron 37, 989-999. doi:10.1016/S0896-6273(03)00149-1

[190] Potjans TC, Diesmann M. The Cell-Type Specific Cortical Microcircuit: Relating Structure and Activity in a Full-Scale Spiking Network Model. Cereb Cortex. 2014;24: 785?806. doi:10.1093/cercor/bhs358

[191] Ryan TJ, Roy DS, Pignatelli M, Arons A, Tonegawa S. Engram cells retain memory under retrograde amnesia. Science. 2015;348: 1007?1013. doi:10.1126/science.aaa5542

[192] Redondo RL and Morris RGM (2011) Making memories last: the synaptic tagging and capture hypothesis. Nat Rev Neurosci 12:17-30.

[193] Roberts SJ. Independent component analysis: source assessment and separation, a Bayesian approach. Vis Image Signal Process IEE Proc -. 1998;145: 149-154. doi:10.1049/ip-vis:19981928

[194] Rodgers CC, DeWeese MR. Neural correlates of task switching in prefrontal cortex and primary auditory cortex in a novel stimulus selection task for rodents. Neuron. 2014;82: 1157-1170. doi:10.1016/j.neuron.2014.04.031

[195] Rokni D, Hemmelder V, Kapoor V, Murthy VN. An olfactory cocktail party: figure-ground segregation of odorants in rodents. Nat Neurosci. 2014;17: 1225-1232. doi:10.1038/nn.3775

[196] Romani S, Amit DJ, and Mongillo G (2006) Mean-field analysis of selective persistent activity in presence of short-term synaptic depression. J Comput Neurosci 20:201-217.

[197] Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533-536. doi:10.1038/323533a0

[198] Sacramento J, Wichert A, van Rossum MCW. Energy Efficient Sparse Connectivity from Imbalanced Synaptic Plasticity Rules. PLoS Comput Biol. 2015;11: e1004265. doi:10.1371/journal.pcbi.1004265

[199] Salinas E, Romo R. Conversion of Sensory Signals into Motor Commands in Primary Motor Cortex. J Neurosci. 1998;18: 499-511.

[200] Sandi C (2011) Glucocorticoids act on glutamatergic pathways to affect memory processes. Trends Neurosci 34: 165-176.

[201] Sanger TD. Optimal unsupervised learning in a single-layer linear feedforward neural network. Neural Netw. 1989;2: 459-473. doi:10.1016/0893-6080(89)90044-0

[202] Santamaria, F., Wils, S., De Schutter, E., Augustine, G.J., 2006. Anomalous Diffusion in Purkinje Cell Dendrites Caused by Spines. Neuron 52, 635-648. doi:10.1016/j.neuron.2006.10.025

[203] Savin C, Joshi P, Triesch J. Independent component analysis in spiking neurons. PLoS Comput Biol. 2010;6: e1000757. doi:10.1371/journal.pcbi.1000757

[204] Sayer RJ, Friedlander MJ, Redman SJ. The time course and amplitude of EPSPs evoked at synapses between pairs of CA3/CA1 neurons in the hippocampal slice. J Neurosci. 1990;10: 826?836.

[205] Schreiner CE, Read HL, Sutter ML. Modular Organization of Frequency Integration in Primary Auditory Cortex. Annu Rev Neurosci. 2000;23: 501-529. doi:10.1146/annurev.neuro.23.1.501

[206] Sengupta B, Stemmler MB, Friston KJ. Information and Efficiency in the Nervous System-A Synthesis. PLoS Comput Biol. 2013;9: e1003157. doi:10.1371/journal.pcbi.1003157

[207] Seol GH, Ziburkus J, Huang SY, Song L, Kim IT, Takamiya K, Huganir RL, Lee H-K, Kirkwood A (2007) Neuromodulators Control the Polarity of Spike-Timing-Dependent Synaptic Plasticity. Neuron 55:919-929.

[208] Shouval, H.Z., Bear, M.F., Cooper, L.N., 2002. A unified model of NMDA receptor-dependent bidirectional synaptic plasticity. Proc. Natl. Acad. Sci. 99, 10831-10836. doi:10.1073/pnas.152343099

[209] Sjöström PJ, Turrigiano GG, Nelson SB. Rate, Timing, and Cooperativity Jointly Determine Cortical Synaptic Plasticity. Neuron. 2001;32: 1149-1164. doi:10.1016/S0896-6273(01)00542-6

[210] Sjöström PJ, Husser M. A Cooperative Switch Determines the Sign of Synaptic Plasticity in Distal Dendrites of Neocortical Pyramidal Neurons. Neuron. 2006;51: 227-238. doi:10.1016/j.neuron.2006.06.017

[211] Smaragdis P. Blind separation of convolved mixtures in the frequency domain. Neurocomputing. 1998;22: 21-34. doi:10.1016/S0925-2312(98)00047-2

[212] Sompolinsky H. Computational neuroscience: beyond the local circuit. Curr Opin Neurobiol. 2014;25: xiii?xviii. doi:10.1016/j.conb.2014.02.002

[213] Song, S., Miller, K.D., Abbott, L.F., 2000. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. Nat. Neurosci. 3, 919-926. doi:10.1038/78829

[214] Song S, Sjöström PJ, Reigl M, Nelson S, Chklovskii DB. Highly Non-random Features of Synaptic Connectivity in Local Cortical Circuits. PLoS Biol. 2005;3: e68. doi:10.1371/journal.pbio.0030068

[215] Stepanyants A, Hof PR, Chklovskii DB. Geometry and Structural Plasticity of Synaptic Connectivity. Neuron. 2002;34: 275-288. doi:10.1016/S0896-6273(02)00652-9

[216] Sul JH, Jo S, Lee D, Jung MW. Role of rodent secondary motor cortex in value-based action selection. Nat Neurosci. 2011;14: 1202-1208. doi:10.1038/nn.2881

[217] Takahashi, N., Kitamura, K., Matsuo, N., Mayford, M., Kano, M., Matsuki, N., Ikegaya, Y., 2012. Locally Synchronized Synaptic Inputs. Science 335, 353-356. doi:10.1126/science.1210362

[218] Takesian, A.E., Hensch, T.K., 2013. Balancing plasticity/stability across brain development. Prog. Brain Res. 207, 3-34. doi:10.1016/B978-0-444-63327-9.00001-1

[219] Tan DP, Liu QY, Koshiya N, Gu H, Alkon D (2006) Enhancement of long-term memory retention and short-term synaptic plasticity in cbl-b null mice. Proc Natl Acad Sci U S A 103:5125-5130.

[220] Teramae J-N, Tsubo Y, and Fukai T (2012) Optimal spike-based communication in excitable networks with strong-sparse and weak-dense links. Sci Rep 2:485.

[221] Toyama K, Kimura M, Tanaka K. Cross-Correlation Analysis of Interneuronal Connectivity in cat visual cortex. J Neurophysiol. 1981;46: 191-201.

[222] Toyoizumi T, Pfister J-P, Aihara K, Gerstner W. Optimality Model of Unsupervised Spike-Timing-Dependent Plasticity: Synaptic Memory and Weight Distribution. Neural Comput. 2007;19: 639-671. doi:10.1162/neco.2007.19.3.639

[223] Tsodyks MV and Markram H (1997) The neural code between neocortical pyramidal neurons depends on neurotransmitter release-probability. Proc Natl Acad Sci U S A 94:719-723.

[224] Turrigiano GG, Nelson SB. Homeostatic plasticity in the developing nervous system. Nat Rev Neurosci. 2004;5: 97-107. doi:10.1038/nrn1327

[225] van Rossum MCW, Bi GQ, Turrigiano GG. Stable Hebbian Learning from Spike Timing-Dependent Plasticity. J Neurosci. 2000;20: 8812-8821.

[226] van Rossum MCW, Turrigiano GG. Correlation based learning from spike timing dependent plasticity. Neurocomputing. 2001;38-40: 409-415. doi:10.1016/S0925-2312(01)00360-5

[227] Varshney LR, Sjstrm PJ, Chklovskii DB. Optimal Information Storage in Noisy Synapses under Resource Constraints. Neuron. 2006;52: 409?423. doi:10.1016/j.neuron.2006.10.017

[228] Vreeswijk CV and Sompolinsky H (1996) Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity. Science 274(5293):1724-1726.

[229] Vreeswijk CV and Sompolinsky H (1998) Chaotic Balanced State in a Model of Cortical Circuits. Neural Comput 10:1321-71.

[230] Vogels, T.P., Sprekeler, H., Zenke, F., Clopath, C., Gerstner, W., 2011. Inhibitory Plasticity Balances Excitation and Inhibition in Sensory Pathways and Memory Networks. Science 334, 1569-1573. doi:10.1126/science.1211095

[231] Vogels TP, Froemke RC, Doyon N, Gilson M, Haas JS, Liu R, et al. Inhibitory synaptic plasticity: spike timing-dependence and putative network function. Front Neural Circuits. 2013;7: 119. doi:10.3389/fncir.2013.00119

[232] von der Malsburg C, Schneider W. A neural cocktail-party processor. Biol Cybern. 1986;54: 29-40. doi:10.1007/BF00337113

[233] von der Malsburg C (1994) The Correlation Theory of Brain Function. In: Domany PE, Hemmen PDJL van, Schulten PK, editors. Models of Neural Networks. Physics of Neural Networks. Springer New York. pp. 95-119.

[234] Walsh DM, Klyubin I, Fadeeva JV, Cullen WK, Anwyl R, Wolfe MS, Rowan MJ, Selkoe DJ (2002) Naturally secreted oligomers of amyloid bold beta protein potently inhibit hippocampal long-term potentiation in vivo. Nature 416:535-539.

[235] Wang, B.-S., Sarnaik, R., Cang, J., 2010. Critical Period Plasticity Matches Binocular Orientation Preference in the Visual Cortex. Neuron 65, 246-256. doi:10.1016/j.neuron.2010.01.002

[236] Wang, B.-S., Feng, L., Liu, M., Liu, X., Cang, J., 2013. Environmental Enrichment Rescues Binocular Matching of Orientation Preference in Mice that Have a Precocious Critical Period. Neuron 80, 198-209. doi:10.1016/j.neuron.2013.07.023

[237] Wang L, Maffei A. Inhibitory plasticity dictates the sign of plasticity at excitatory synapses. J Neurosci Off J Soc Neurosci. 2014;34: 1083-1093. doi:10.1523/JNEUROSCI.4711-13.2014

[238] Wenisch OG, Noll J, van Hemmen JL. Spontaneously emerging direction selectivity maps in visual cortex through STDP. Biol Cybern. 2005;93: 239-247. doi:10.1007/s00422-005-0006-z

[239] Wiechert MT, Judkewitz B, Riecke H, Friedrich RW. Mechanisms of pattern decorrelation by recurrent neuronal circuits. Nat Neurosci. 2010;13: 1003-1010. doi:10.1038/nn.2591

[240] Wiegert JS, Oertner TG. Long-term depression triggers the selective elimination of weakly integrated synapses. Proc Natl Acad Sci U S A. 2013;110: E4510-4519. doi:10.1073/pnas.1315926110

[241] Wilson RI, Mainen ZF. Early events in olfactory processing. Annu Rev Neurosci. 2006;29: 163-201. doi:10.1146/annurev.neuro.29.051605.112950

[242] Wilson, N.R., Ty, M.T., Ingber, D.E., Sur, M., Liu, G., 2007. Synaptic Reorganization in Scaled Networks of Controlled Size. J. Neurosci. 27, 13581-13589. doi:10.1523/JNEUROSCI.3863-07.2007

[243] Woodin MA, Ganguly K, Poo M. Coincident pre- and postsynaptic activity modifies GABAergic synapses by postsynaptic changes in Cl- transporter activity. Neuron. 2003;39: 807-820.

[244] Xu T, Yu X, Perlik AJ, Tobin WF, Zweig JA, Tennant K, et al. Rapid formation and selective stabilization of synapses for enduring motor memories. Nature. 2009;462: 915-919. doi:10.1038/nature08389

[245] Yang G, Pan F, Gan W-B. Stably maintained dendritic spines are associated with lifelong memories. Nature. 2009;462: 920-924. doi:10.1038/nature08577

[246] Yang G, Lai CSW, Cichon J, Ma L, Li W, Gan W-B. Sleep promotes branch-specific formation of dendritic spines after learning. Science. 2014;344: 1173-1178. doi:10.1126/science.1249098

[247] Yasumatsu N, Matsuzaki M, Miyazaki T, Noguchi J, Kasai H. Principles of long-term dynamics of dendritic spines. J Neurosci Off J Soc Neurosci. 2008;28: 13592-13608. doi:10.1523/JNEUROSCI.0603-08.2008

[248] Yazaki-Sugiyama, Y., Kang, S., Cteau, H., Fukai, T., Hensch, T.K., 2009. Bidirectional plasticity in fast-spiking GABA circuits by visual experience. Nature 462, 218-221. doi:10.1038/nature08485

[249] York LC, van Rossum MCW (2009) Recurrent networks with short-term synaptic depression. J Comput Neurosci 27:607-620.

[250] Yoshimura Y, Dantzker JLM, Callaway EM. Excitatory cortical neurons form fine-scale functional networks. Nature. 2005;433: 868-873. doi:10.1038/nature03252

[251] Zenke F, Hennequin G, Gerstner W (2013) Synaptic plasticity in neural networks needs homeostasis with a fast rate detector. PLoS Comput Biol 9:e1003330.

[252] Zenke, F., Agnes, E.J., Gerstner, W., 2015. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. Nat. Commun. 6. doi:10.1038/ncomms7922

[253] Zhang J-C, Lau P-M, Bi G-Q. Gain in sensitivity and loss in temporal contrast of STDP by dopaminergic modulation at hippocampal synapses. Proc Natl Acad Sci. 2009;106: 13028-13033. doi:10.1073/pnas.0900546106

[254] Zheng P, Dimitrakakis C, Triesch J. Network self-organization explains the statistics and dynamics of synaptic connection strengths in cortex. PLoS Comput Biol. 2013;9: e1002848. doi:10.1371/journal.pcbi.1002848

[255] Zuo Y, Lin A, Chang P, Gan W-B. Development of Long-Term Dendritic Spine Stability in Diverse Regions of Cerebral Cortex. Neuron. 2005;46: 181-189. doi:10.1016/j.neuron.2005.04.001

Publications

2 **Hiratani N**, Fukai T (2015) Mixed Signal Learning by Spike Correlation Propagation in Feedback Inhibitory Circuits, *PLOS Computational Biology* 11(4): e1004227. doi:10.1371/journal.pcbi.1004227

1 **Hiratani N**, Fukai T (2014) Interplay between short- and long-term plasticity in cell-assembly formation, *PLOS ONE* 9(7):e101535. doi:10.1371/journal.pone.0101535.

Refereed conference presentations

2 **Hiratani N**, Fukai T, "Structural plasticity generates efficient network structure for synaptic plasticity" Mar. 7 (2015), Computational and Systems Neuroscience (*Cosyne*) 2015, Salt Lake City, USA, Mar. 5- Mar. 8 (2015)

1 **Hiratani N**, Fukai T, "Interplay between short- and long-term plasticity in cell-assembly formation" Feb. 27 (2014), Computational and Systems Neuroscience (*Cosyne*) 2014, Salt Lake City, USA, Feb. 27- Mar. 2 (2014)

Non-Refereed conference presentations

6 **Hiratani N**, Fukai T, "GABA driven circuit formation through heterosynaptic spike-timing-dependent plasticity" Oct. 18 (2015) 45th Annual Meeting of Society for Neuroscience (*Neuroscience 2015*), McCormick Place, USA, Oct. 17-Oct. 21 (2015)

5 **Hiratani N**, Fukai T, "GABA driven circuit formation through heterosynaptic spike-timing-dependent plasticity" Neural Coding, Computation and Dynamics (*NCCD*), Bilbao, Spain, Aug. 30 - Sep. 2 (2015)

4 **Hiratani N**, Fukai T, "Network structure generates priors for internal probabilistic model" Nov. 15 (2014), 44th Annual Meeting of Society for Neuroscience (*Neuroscience 2014*), Walter E. Washington Convention Center, USA, Nov. 15-Nov. 19 (2014)

3 **Hiratani N**, Fukai T, "Learning higher order structure of correlated input by excitatory and inhibitory spike timing dependent plasticity" Nov. 15 (2014), 44th Annual Meeting of Society for Neuroscience (*Neuroscience 2014*), Walter E. Washington Convention Center, USA, Nov. 15-Nov. 19 (2014)

2 **Hiratani N**, Fukai T, "Learning higher order structure of correlated input by excitatory and inhibitory spike timing dependent plasticity" Sep. 11 (2014), The 37th Annual Meeting of the Japan Neuroscience Society. Pacifico Yokohama, Japan, Sep. 11-Sep. 13 (2014).

1 **Hiratani N**, Fukai T, "The effect of short-term depression on cell assembly formation" Nov. 12 (2013), 43rd Annual Meeting of Society for Neuroscience (*Neuroscience 2013*), San Diego Convention Center, USA, Nov. 9 - Nov. 13 (2013)