Structured stabilization in recurrent neural circuits through inhibitory synaptic plasticity

Dylan Festa* Claudia Cusseddu Julijana Gjorgjieva

School of Life Sciences, Technical University of Munich, Freising, Germany * dylan.festa@tum.de

Inhibitory interneurons play a dual role in recurrently connected biological circuits: they regulate global neural activity to prevent runaway excitation, and contribute to complex neural computations. While the first role can be achieved through unstructured connections tuned for homeostatic rate stabilization, computational tasks often require structured excitatory-inhibitory (E/I) connectivity. Here, we consider a broad class of pairwise inhibitory spike-timing dependent plasticity (iSTDP) rules, demonstrating how inhibitory synapses can self-organize to both stabilize excitation and generate functionally relevant connectivity structures—a process we call "structured stabilization". We show that in both E/I circuit motifs and large spiking recurrent neural networks the choice of iSTDP rule can lead to either mutually connected E/I pairs, or to lateral inhibition, where an inhibitory neuron connects to an excitatory neuron that does not directly connect back to it. In a one-dimensional ring network with two inhibitory subpopulations following these distinct iSTDP rules, the effective connectivity among the excitatory units self-organizes into a Mexican-hat-like profile, with excitatory influence in the center and inhibitory influence away from the center. This leads to emergent dynamical properties such as contextual modulation effects and modular spontaneous activity. Our theoretical work introduces a family of rules that retains the broad applicability and simplicity of spike-timing-based plasticity, while promoting structured, self-organized stabilization. These findings highlight the rich interplay between iSTDP rules, circuit structure, and neuronal dynamics, offering a framework for understanding how inhibitory plasticity shapes network function.

Significance Statement

To function properly and process information, the brain requires a balance between excitatory and inhibitory signals, largely maintained by inhibitory synaptic plasticity. However, how this plasticity simultaneously stabilizes overall activity and creates functional connection patterns remains unclear. We show that a simple spike-timing-dependent inhibitory plasticity rule can achieve both stability and structure in recurrent networks. The shape of this rule dictates the resulting network architecture, favoring either strong connections between reciprocally connected excitatory and inhibitory neurons, or lateral inhibition, where an inhibitory neuron preferentially connects to excitatory neurons that do not connect back. We demonstrate how these different structures lead to emergent computations like context-dependent responses, directly linking synaptic plasticity rules to network-level function.

1 Introduction

In cortical circuits, GABA-ergic interneurons constitute approximately 10% to 20% of the total neural population, and can be further divided into subtypes that exhibit distinct characteristics in connectivity, plasticity and tuning properties (Jiang et al. 2015; Tremblay et al. 2016). Fully understanding how this diverse neural population influences neural circuit function remains a significant challenge.

The dynamic balance between excitation and inhibition is crucial for generating the irregular, Poisson-like firing patterns commonly observed *in vivo* in cortical circuits (**vreeswijkChaoticBalancedState1998**; Brunel 2000). This balance enhances the circuit's sensitivity to input, facilitating rapid information processing (Tsodyks and Sejnowski 1995). Furthermore, it supports essential computations, such as receptive field tuning (Wallenstein and Hasselmo 1997; Froemke et al. 2007; Dorrn et al. 2010; Kleberg et al. 2014; Rubin et al. 2017), neural assembly formation (Sadeh and Clopath 2021), and the storage of precise temporal firing sequences (Maes et al. 2021; Riquelme et al. 2023; Gong and Brunel 2024; Shao et al. 2024).

Excitatory-inhibitory (E/I) balance can be achieved without any specific inhibitory connectivity structures. Often circuit models consider inhibition simply as a "blanket" that regulates excitation, and incorporate it into a global normalization term or a correction to the synaptic weights between excitatory neurons (Jin 2002; Ravid Tannen-

Compiled from LaTeX on:
June 20, 2025

baum and Burak 2016; Montangie et al. 2020). While this simplified approach captures fundamental properties of neural dynamics, it overlooks the possible advantages of more specific E/I connectivity patterns. Incorporating structured inhibitory connections can further refine computations and mediate neural functions such as E/I co-tuning (Clopath et al. 2016; Mackwood et al. 2021; Miehl and Gjorgjieva 2022; Eckmann et al. 2024), memory storage (Mongillo et al. 2018; Bergoin et al. 2023; Miehl et al. 2023; Agnes and Vogels 2024; Meissner-Bernard, Jenkins, et al. 2025; Meissner-Bernard, Zenke, et al. 2025), predictive coding (Schulz et al. 2021), and several other aspects of neural dynamics (Mongillo et al. 2018; Dahmen et al. 2020; Sadeh and Clopath 2020; Hafizi et al. 2022).

Experimental findings demonstrate that inhibitory neurons form non-random connectivity patterns that cannot be explained solely by cell distance or neuron morphology. Additional factors, including brain area, cortical layer, and interneuron subtypes significantly shape these structures (Peng et al. 2021; Bollmann et al. 2023; Reimann et al. 2023; Wu et al. 2023; Kuan et al. 2024; Lakhera et al. 2024; Znamenskiy et al. 2024). In the neonatal mouse cortex, even before the onset of external sensory inputs, inhibitory hub neurons regulate circuit development (Bollmann et al. 2023). Moreover, studies in the neonatal ferret cortex indicate a widespread Mexican-hat-like effective connectivity, with local excitatory effects and longer-range inhibition (Smith et al. 2018). This spatial organization gives rise to modular activity patterns with long-range correlations propagating throughout the sensory cortex. Such structured activity is thought to serve as a scaffold for the emergence of functional neural representation in the early stages of development (Mulholland et al. 2021; Lakhera et al. 2024; Trägenap et al. 2025).

Despite these advances, the mechanisms and principles that enable the formation and maintenance of inhibitory connectivity structures remain elusive. Genetic mechanisms play a fundamental role, particularly in early developmental stages (Duan et al. 2020; Hanganu-Opatz et al. 2021; Wu et al. 2023). Equally important is synaptic plasticity driven by neural activity, which refines inhibitory connectivity through a dynamic, activity-dependent process (Sreenivasan et al. 2022). This mechanism enables a closed-loop regulation of neural dynamics, allowing for precise tuning and adaptability in the face of change.

Although some theoretical studies considered the impact of different forms of iSTDP in feedforward circuit architectures with a single postsynaptic neuron (Kleberg et al. 2014; Luz and Shamir 2014), there is still no general understanding of the effects of pairwise iSTDP on structure formation at the circuit level. Theoretical work proposed a iSTDP rule that guarantees the homeostatic regulation of neural rates (Vogels et al. 2011), becoming the "canonical" choice for plastic circuits. However, this rule tends to over-regulate rates in a homeostatic manner, producing narrow rate distributions in the excitatory population. In other work, inhibitory rules are based on experimental measurements (Lagzi et al. 2021; Bergoin et al. 2024), derived normatively from computation objectives (Mackwood et al. 2021), or are found through numerical optimization on a large parameter space (Confavreux et al. 2024). Here, we propose an integrated approach by analyzing a broad family of pairwise iSTDP rules, dissecting the principles and the mechanisms that contribute to excitatory regulation and structure formation, and testing them on large-scale recurrent neuronal networks (RNNs) of spiking neurons.

We first derive how the choice of the iSTDP rule determines which features of neural activity drive the formation of inhibitory connections. At one extreme, rules dominated by firing rates are weakly influenced by structure, aiming primarily at the homeostatic regulation of firing rates, and thus promoting "blanket" inhibition. At the opposite extreme, we find covariance-dominated rules, strongly influenced by pairwise spike correlations that, in the absence of external inputs, arise from the underlying circuit connectivity. These rules stabilize neural activity, while promoting specific connectivity motifs, such as mutually connected E/I pairs or lateral inhibition, depending on the iSTDP features. We call this effect "structured stabilization". To further explore functional consequences of covariance-dominated rules, we implement an RNN whose excitatory neurons follow a one-dimensional ring topology, and whose inhibitory neurons follow two distinct iSTDP rules. Here, inhibition self-organizes into a Mexican-hat effective connectivity profile, resulting in long-range modular correlations of spontaneous activity. Finally, we demonstrate that the circuit implements contextual modulation, including suppression of responses to matching stimuli and amplification for mismatched stimuli. In summary, our study investigates how distinct inhibitory plasticity rules shape inhibitory connectivity motifs, giving rise to emergent large-scale circuit structures and computations.

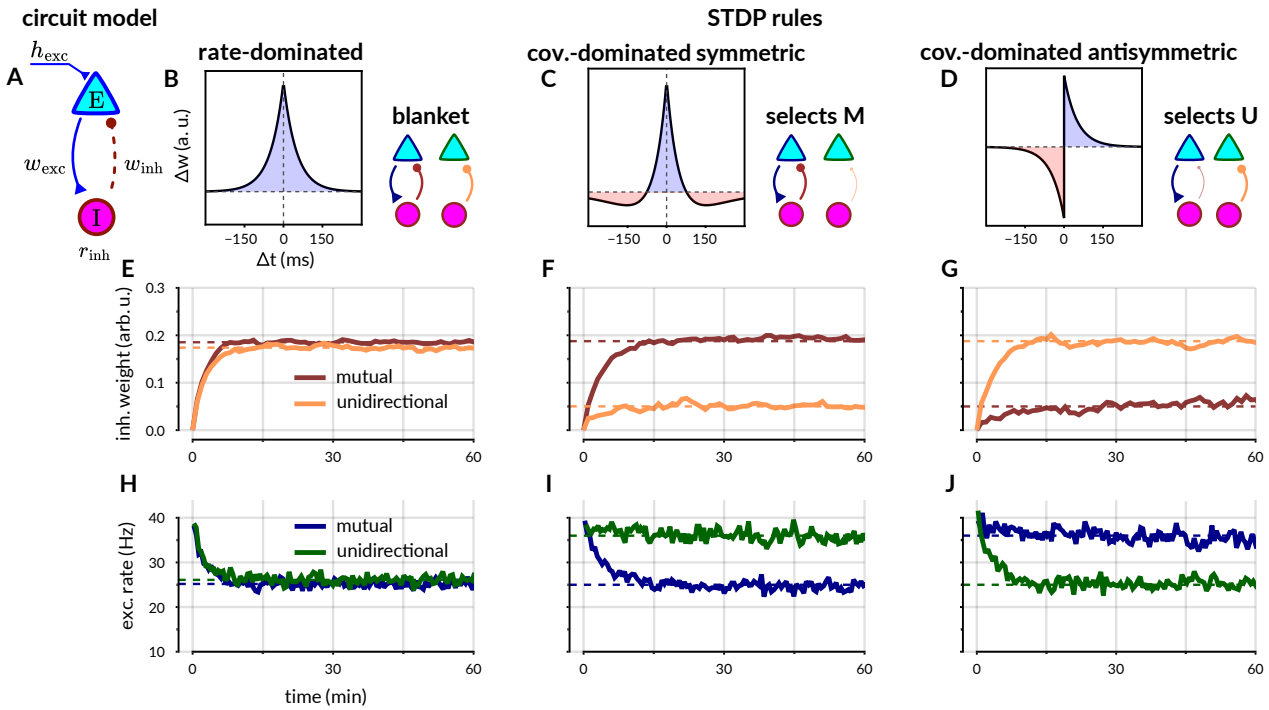

Figure 1: Effects of plasticity for E/I motifs in a reduced model. **A**. Schematics of the reduced circuit model. Two time-varying coupled Poisson units, one excitatory and one inhibitory, are connected by a fixed excitatory weight w_{exc} and a plastic inhibitory weight w_{inh}. The instantaneous rates are computed as the convolution between each incoming spike and an exponential synaptic kernel that either increases (exc) or decreases (inh) the instantaneous firing probability. The input to the excitatory neuron h_{exc} is fixed, whereas the inhibitory input changes dynamically, keeping the inhibitory neuron at a fixed rate r_{inh}. Only the inhibitory synapse (dashed line) is plastic. The circuit is simulated in two configurations: unidirectional (U), with $w_{\text{exc}} = 0$ and mutual (M), with $w_{\text{exc}} = 0.5$. **B,C,D**. Pairwise interaction kernels of the three iSTDP rules tested (Eq. 1). We used the following numerical parameters. **B**. $B = 1$, $\alpha_{\text{pre}} = -25$, $\alpha_{\text{post}} = 0$, $\tau_+ = 50$ ms. **C**. $B = 0$, $\alpha_{\text{pre}} = -0.12$, $\alpha_{\text{post}} = 1.45$, $\tau_+ = 50$ ms, $\tau_- = 500$ ms. **D**. $B = 0$, $\alpha_{\text{pre}} = -0.14$, $\alpha_{\text{post}} = 7.95$, $\tau_+ = 30$ ms, $\tau_- = 30$ ms. **E,F,G**. Change in inhibitory synaptic weight over time for the M (brown) and U (light orange) circuit configurations, under the rules above each column. Dashed lines are analytic predictions for the fixed point (Methods, Eq. 10). **H,I,J**. Change in excitatory firing rate over time for the M (dark blue) and U (dark green) configurations, under the rules above. Dashed horizontal lines are the analytic predictions for the fixed-point (Methods, Eq. 10).

2 Results

2.1 The influence of inhibitory plasticity on circuit structure

We initially considered a general iSTDP rule that depends exclusively on local spike times applied to a synapse from an inhibitory presynaptic to an excitatory postsynaptic neuron (inh-to-exc) (Kempter et al. 1999; Luz and Shamir 2014; Akil et al. 2021). This rule contains two fixed components triggered by each pre- and post-synaptic firing events, denoted as α_{pre} and α_{post}, and a kernel function, $L(\tau)$ triggered by the co-occurrence of pre-post ($\tau > 0$) or post-pre ($\tau < 0$) spiking events (see Methods, Section 4.1). When the spiking activity is modeled as a stochastic process, the mean-field equation for the evolution of the average synaptic weight is:

$$\begin{aligned} \frac{d\langle W\rangle}{d\,t} &= \alpha_{\text{pre}} r_{\text{pre}} + \alpha_{\text{post}} r_{\text{post}} + B\, r_{\text{pre}}\, r_{\text{post}} + Q \\ \text{with}\ \ B &:= \int_{-\infty}^{+\infty} d\tau L(\tau) \quad \text{and} \quad Q := \int_{-\infty}^{+\infty} d\tau L(\tau)\, C(\tau) \end{aligned} \tag{1}$$

where r_{pre} and r_{post} are the pre- and post-synaptic firing rates, B is the net area of the iSTDP kernel $L(\tau)$, and Q captures the effect of the iSTDP kernel on the pre-post pairwise interactions, conveyed by the covariance density curve $C(\tau)$ (see Supporting Information, Section S1.1, or Kempter et al. 1999; Luz and Shamir 2014; Lagzi and Fairhall 2024).

To understand how inhibitory plasticity influences circuit structure, we first developed an analytical framework, defining a family of iSTDP rules. Using these rules, we aimed to derive the closed-form solution for a reduced E/I motif, and link the shape of iSTDP to the structure of learned inhibitory connections.

Several terms in Eq. 1 depend solely on pre- and post-synaptic firing rates. Furthermore, in the general category of rules where B is high, the contribution of the $B\, r_{\text{pre}}\, r_{\text{post}}$ term can easily be an order of magnitude above Q, even in the presence of strong pre-post correlations (see Methods, Section 4.3). Therefore, when $Q \ll B\, r_{\text{pre}}\, r_{\text{post}}$, the plasticity dynamics of Eq. 1 are strongly dominated by the firing rates. We call this plasticity regime "rate-dominated". The previously established and well-studied rate-homeostatic iSTDP rule (Vogels et al. 2011) is a rate-dominated rule with $\alpha_{\text{post}} = 0$. This rule drives the inhibitory weight to adapt continuously so that the postsynaptic rate stays close to the target value $\tilde{r}_{\text{post}} \approx -\alpha_{\text{pre}}/B$, corresponding to a stable fixed point solution in Eq. 1.

While a large value of B simplifies the analysis by making covariance-based terms negligible, it only represents a subset of possible possible plasticity rules. We therefore explored a broader range, including those for which the iSTDP kernel has zero or near-zero net area ($B \approx 0$). These rules amplify the contribution of the second-order correlations between pre- and postsynaptic neuron in the Q term. We refer to them as "covariance-dominated". In choosing specific pairwise interaction kernels, we considered recent experimental findings that reported a symmetric iSTDP rule for parvalbumin-expressing (PV) interneurons, as well as an antisymmetric iSTDP rule for somatostatin-expressing (SST) interneurons, both projecting onto excitatory pyramidal neurons in the orbitofrontal cortex (Lagzi et al. 2021). Consequently, we focus on symmetric kernels with strong short-range potentiation and weak long-range depression (Fig. 1C), or antisymmetric kernels in which pre-post and post-pre interactions balance each other with opposite signs (Fig. 1D).

Analyzing the full dynamics of Eq. 1 is challenging because the covariance $C(\tau)$ generally depends on the synaptic weights themselves. We therefore considered a minimal two-neuron E/I circuit motif (Fig. 1A) with fixed exc-to-inh connection (w_{exc}) but plastic inh-to-exc weight (w_{inh}). We modeled the neurons as time-varying Poisson processes, where each incoming spike changes the instantaneous spiking probability through an excitatory or inhibitory post-synaptic potential (EPSP/IPSP) kernel (internal traces shown in Fig. S1, see also Methods Section 4.4). The inhibitory neuron's firing rate is kept constant by an external homeostatic mechanism. The excitatory neuron receives an excitatory fixed external input in addition to the inhibitory input. Using this simplified model, we derived analytically a closed-form expression for the fixed-point of the iSTDP dynamics (Methods, Eq. 10; Supporting Information).

With this analytical framework in hand, we investigated how different iSTDP rules affect the resulting inhibitory synaptic strength and the excitatory postsynaptic activity. We considered two configurations of our E/I circuit motif: (1) a unidirectional (U) configuration, where the excitatory-to-inhibitory weight w_{exc} is set to zero, and (2) a mutual (M) configuration, featuring a strong excitatory-to-inhibitory connection. We studied three distinct iSTDP

rules in both configurations: a rate-dominated rule with a symmetric interaction kernel (Fig. 1B), a covariance-dominated rule also with a symmetric kernel (Fig. 1C), and a covariance-dominated rule with an antisymmetric kernel (Fig. 1D). As expected, the rate-dominated rule shows minimal sensitivity to the presence of the mutual connection. In both U and M configurations, the inhibitory synapse adjusts the excitatory firing rate toward the target rate dictated by the rate-dependent components, approximately given by $\tilde{r}_{\text{exc}} \approx -\alpha_{\text{pre}}/B$ (Fig. 1E). Minor deviations arise from the non-zero contribution of the Q term.

In contrast, the covariance-dominated rules are sensitive to both the circuit's connectivity structure and the shape of iSTDP kernel. With the symmetric kernel, the final inhibitory weight is significantly stronger in the M configuration compared to the U configuration (Fig. 1F, compare brown and light orange). Conversely, with the antisymmetric kernel, the inhibitory weight is lower in the M configuration (Fig. 1G). Such structural selectivity enable these rules not only to regulate excitatory rates, but also to shape specific E/I connectivity motifs — an effect we termed "structured stabilization". This property emerges from the interaction between the iSTDP kernel and second-order spiking statistics. Namely, the strong excitatory-to-inhibitory connection in the M configuration generates positively-correlated post-pre events. For the symmetric iSTDP, these events occur within the positive region of the kernel, leading to synaptic potentiation. Conversely, with the antisymmetric iSTDP rule, the same positively-correlated events interact with the negative region of the kernel, resulting in synaptic depression. Additionally, the inhibitory-to-excitatory weight generates a negative covariance density, localized instead in the pre-post region of the curve. As both rules are positive here, this introduces a negative feedback in both cases, limiting the value that w_{inh} can reach before plasticity switches into depression. Finally, adjusting the fixed components of the iSTDP rule, α_{pre} and α_{post}, is also crucial for reaching the desired excitatory firing rate (set to 25Hz) in the preferred circuit configuration.

Hence, our general framework provides a link between different iSTDP rules and inhibitory strength and circuit structure.

2.2 The shape of inhibitory STDP affects inhibitory structure

Having established that covariance-dominated rules can achieve structured stabilization, we next investigated how the specific form of the iSTDP kernel shapes the learned inhibitory connectivity. We considered four kernel shapes: symmetric, antisymmetric, and their "inverted" counterparts, simply obtained by reversing the sign of the kernel (Fig. 2A). We analyzed the behavior of these kernels in the reduced two-neuron E/I motif. To explore the parameter space, we varied the rate-dependent parameters B and α_{pre}, while keeping $\alpha_{\text{post}} = 0$. With this parametrization, the covariance-dominated regime corresponds to values of B and α_{pre} close to zero.

Our analysis finds that both symmetric and antisymmetric kernels yield stable fixed-points in the covariance-dominated regime (Fig. 2A,B, top row, and Supporting Information Fig. S2C,D). These kernels, which we term "self-stabilizing" exhibit robustness to parameter fluctuations and reliably converge to a stable inhibitory weight. In contrast, the inverted antisymmetric and the inverted symmetric rule are unstable in the covariance-dominated regime, with inhibitory weights either decaying to zero or saturating depending on the initial conditions. We categorize these as "saturating" kernels (Fig. 2A, bottom panels, and Fig. S2E,F in Supporting Information). Saturating kernels can still reach stability with strong rate-dependent terms, which reduce the overall influence of the kernel shape on the final weight.

This difference in stability arises from how the kernels interact with changes in pre- and postsynaptic correlations. Consider a synapse operating under a covariance-dominated plasticity rule at its fixed point. At equilibrium, potentiation and depression events balance exactly. A perturbation that strengthens the inhibitory weight leads to a reduction in pre-post firing events due to the increased inhibition. For self-stabilizing kernels, which have a positive right branch (corresponding to pre-post events), this reduction in correlation reduces the occurrence of synaptic potentiation events, shifting the balance in favor of synaptic depression. The synapse therefore depresses, eventually compensating the initial perturbation from equilibrium. Conversely, saturating kernels show the opposite effect: since their kernel is negative in the pre-post region, a reduction of pre-post correlation results in fewer synaptic depression events, so the initial balance is further shifted in favor of synaptic potentiation, amplifying the perturbation and leading the synapse towards saturation. Finally, in rate-dominated rules, the specific shape of iSTDP kernel becomes less relevant. Weight fluctuations impact the dynamics through the $B\, r_{\text{pre}}\, r_{\text{post}}$ term rather than the correlation-dependent Q term. Consequently, these rules consistently converge to the classic rate-homeostatic solution $\tilde{r}_{\text{exc}} \approx -\alpha_{\text{pre}}/B$ (Vogels et al. 2011). Note however that in large networks, once these rule fulfill their homeostatic rate target, they can still induce further structural organization at slower

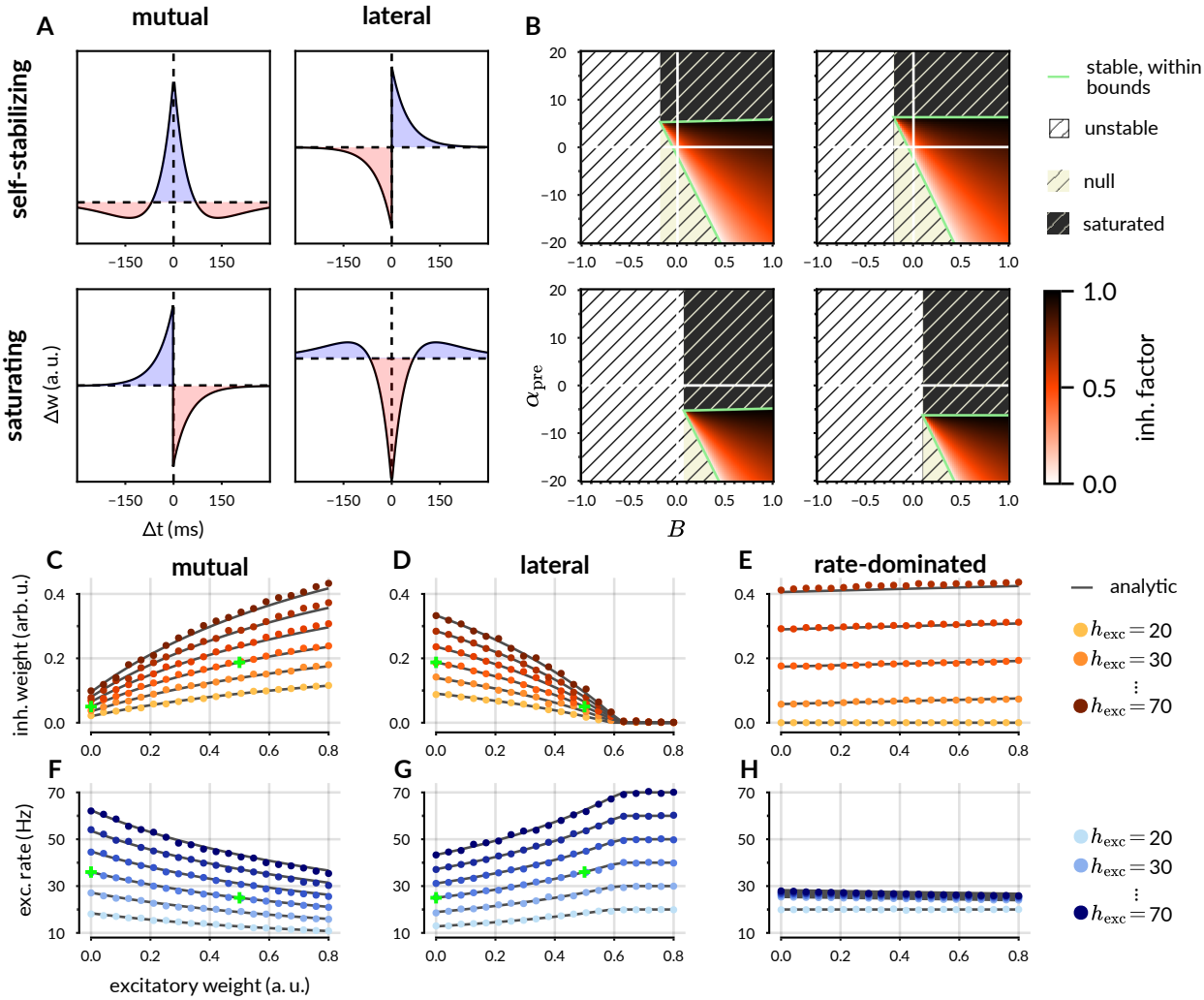

Figure 2: Parametric study of different shapes of iSTDP kernel in the reduced 2D model. **A**. Representation of four possible shapes of iSTDP kernel, labeled by their effects in the covariance-dominated regime (i.e. with near-zero rate-dependent terms). **B**. Parametric study based on the analytic solution for each of the four iSTDP kernels, varying α_{post} and the net area under the iSTDP kernel, B (see Eq. 1). The color map represents the level of inhibition, where 1 indicates that the excitatory neuron is completely silenced, and 0 indicates no inhibition, so that the excitatory firing rate matches the input current **C,D,E**. Final inhibitory weight as a function of the incoming excitatory weight for three different iSTDP rules. Dots are numerical simulations, gray lines are analytic results (Methods, Eq. 10), color shades indicate different input levels. **C**. Results for the symmetric, self-stabilizing rule. The green dots represent the same conditions shown in Fig. 1F. **D**. Results for the antisymmetric, self-stabilizing rule. The green dots represent the same conditions shown in Fig. 1G. **E**. Results for the symmetric rule in a rate-dominated regime. **F,G,H**. Corresponding plots showing the rate of the excitatory neuron after convergence.

timescales (Lagzi and Fairhall 2024).

To further characterize the self-stabilizing rules, we extended our analysis beyond the U and M circuit configurations presented in Fig. 1, exploring a broader range of excitatory-to-inhibitory weights w_{exc} and input currents to the excitatory neuron. For the symmetric kernel, the learned inhibitory weight increases with w_{exc} (Fig. 2C), while for the antisymmetric kernel it decreases (Fig. 2D). Increasing the excitatory external input leads to stronger inhibitory weights in both cases (Fig. 2C,D), however the rate of the excitatory neuron still scales linearly with the input strength (Fig. 2F,G). In contrast, the rate-homeostatic rule exhibits a minimal dependence on w_{exc} (Fig. 2E) and a very tight control on the excitatory rate (Fig. 2H), which restricts the dynamic range of the excitatory neuron, limiting its ability to respond to changes in input.

In summary, our analysis of the reduced E/I motif reveals a spectrum of iSTDP rules with distinct outcomes on circuit structure and dynamics. On the one hand, covariance-dominated rules with self-stabilizing kernels enable structured stabilization. As a consequence, they selectively reinforce connectivity motifs based on pre- and post-synaptic correlations, and allow a broader range of excitatory firing rates. On the other hand, rate-dominated rules prioritize homeostatic control of excitatory activity over the formation of structural motifs. Thus, these rules ensure robust dynamical stability, potentially at the cost of computational flexibility. These results demonstrate a trade-off between the emergence of specific inhibitory structures and the tight regulation of excitatory rates, suggesting that different iSTDP rules may be suited for different computational tasks.

2.3 Structured stabilization in randomly connected RNNs

How do our results on structured stabilization from reduced E/I motifs translate to larger networks? We simulated a randomly connected RNN of conductance-based leaky integrate-and-fire (LIF) neurons, with 900 excitatory and 100 inhibitory units. We modeled the excitatory connections (exc-to-exc and exc-to-inh) as fixed and sparse, while the inh-to-exc connections as dense and initially small, but subject to iSTDP (Table 2).

We first implemented a covariance-dominated plasticity rule with a symmetric kernel (Fig. 3A). Initially, the network produces a pathological high-firing regime, due to the low inhibitory weights. However, inhibitory plasticity quickly stabilizes the dynamics, bringing it to the asynchronous irregular firing regime characteristic of balanced networks (Fig. 3B). To analyze the resulting inhibitory connectivity, we categorized inh-to-exc synaptic weights as either “mutual” or “unidirectional”. We call an inh-to-exc weight mutual when a reciprocal exc-to-inh connection exists between the same two neurons, reflecting a bidirectional link. Conversely, a weight is unidirectional in the absence of such excitatory reciprocal connection. Following the onset of plasticity, we observed that mutual inhibitory weights increased much more than unidirectional ones, up to the saturation threshold (Fig. 3C). This selectivity can be directly visualized through the matrix of synaptic weights after learning: the pattern of learned inh-to-exc weight structure mirrors the fixed exc-to-inh weights (Fig. 3E,F; see Supporting Information for full weight matrix).

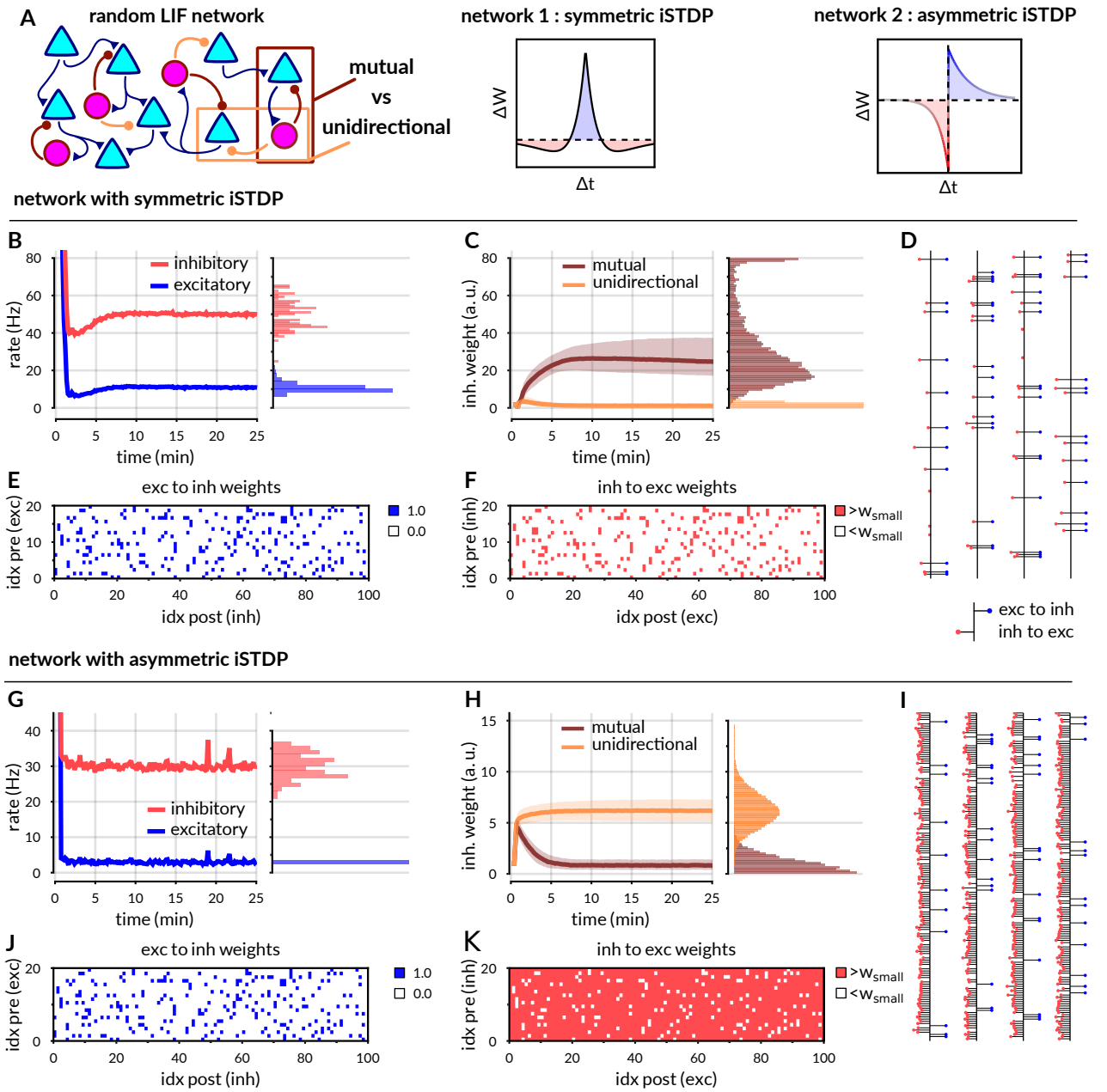

Figure 3: Structure emerges in RNNs of conductance-based LIF neurons under different iSTDP rules. **A** *left*. Representation of the random network with $(N_{\text{exc}} = 900,\ N_{\text{inh}} = 100)$ spiking neurons. The inh-to-exc weights form all-to-all connections and are plastic, all other weights are fixed. Exc-to-exc and exc-to-inh connections are sparse. **A** *middle*. Kernel of the symmetric, covariance-dominated iSTDP rule used in network 1 (panels B-F). **A** *right*. Kernel of the asymmetric, covariance-dominated iSTDP rule used in network 2 (panels G-K). Parameters in Table 2. **B**. Evolution of population mean firing rates over time, with final distribution of rates over the full populations. **C**. Evolution of inh-to-exc synaptic weights over time and final distribution. The weights are split into a "mutual" (brown) and a "unidirectional" (light orange) group. Only few weights reach the saturating value, set to 80. **D**. Synaptic weights of four random inhibitory neurons after convergence. Each neuron is represented by a different column. The y coordinates represent excitatory neuron indices (1 to 150). Excitatory incoming weights are shown as horizontal pulses on the right side (blue), inhibitory outgoing weights are on the left (red). Weights lower than 5% of saturation level are not shown. Matching left and right pulses indicate that the inhibitory neuron both receives and sends a connection to the same excitatory neuron. **E**. Random portion of the fixed and sparse exc-to-inh weights (the full matrix, of size 900×100 is shown in the Supporting Information). **F**. Random portion of the all-to-all and plastic inh-to-exc connections after learning, with indices matching panel **E**. For clarity the matrix has been binarized, with threshold-value w_{small} defined as 0.1 times the 0.9-quantile of the full weight distribution. See Supporting Information for a continuous representation of synaptic weights. **G-K**. Equivalent plots in a second network with identical network dynamics and initial conditions, but using the asymmetric iSTDP.

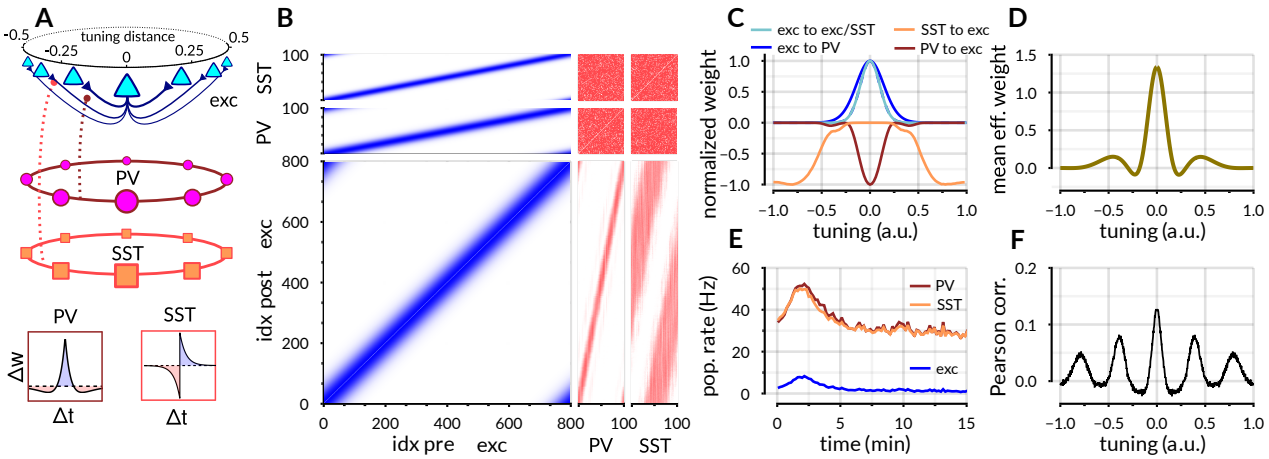

Figure 4: Inhibitory self-organization in a spiking RNN structured as a one-dimensional ring. **A**. Representation of the model. 800 excitatory neurons are arranged in a one-dimensional ring topology, connected to each other and to their inhibitory counterparts with weights that depend to their distance in tuning space (Methods, Eq. 14). 100 "PV" inhibitory neurons connect all-to-all to the excitatory neurons with plasticity following a symmetric iSTDP rule, 100 "SST" neurons connect all-to-all to excitatory neurons with plasticity following an antisymmetric iSTDP. For both rules $B = 0$, $\alpha_{\text{pre}} = -0.1$, $\alpha_{\text{post}} = 0$. **B**. Full weight matrix after weight convergence. The excitatory weights (blue) are fixed, the inh-to-exc weights (red, lower right) are determined by plasticity. See Supporting Information for a version in higher resolution. **C**. Profile of the average outgoing weights as a function of tuning distance. For convenience, the profiles are normalized so that the maximum is either 1 or -1. **D**. Average effective interaction profile between the excitatory neurons, as a function of tuning (see Methods). **E**. Evolution of mean firing rate over time. **F**. Pearson correlation during spontaneous activity between an excitatory neuron and its neighbors, ordered by tuning distance and averaged over the entire population.

Next, we investigated the impact of an asymmetric iSTDP rule within the same random RNN architecture and neural dynamics. As with the symmetric rule, this asymmetric iSTDP stabilizes the network dynamics (Fig. 3G). However, it promotes a markedly different connectivity pattern: purely lateral inhibition. In this configuration, inhibitory neurons strengthen their connections onto excitatory neurons lacking reciprocal connections back to them, thus forming exclusively unidirectional connections (Fig. 3H). This yields an inhibitory-to-excitatory connectivity pattern almost perfectly opposite to the previous configuration, so that existing exc-to-inh synapses carve gaps in the inh-to-exc connections (Fig. 3J,K). Achieving robust lateral inhibition with the asymmetric iSTDP rule requires the introduction of a small rate-dependent component ($B = 0.3$). This deviation from a purely covariance-dominated regime introduces characteristics of rate-homeostatic plasticity, such as rapid convergence to a narrow distribution of excitatory firing rates (Fig. 3G). This requirement is perhaps unsurprising given the challenges of parameter tuning in large RNNs, where closed-form solutions are unavailable (see Discussion for further considerations).

To sum up, our RNN simulations generalize the analytical predictions from the reduced E/I model to larger, more complex networks. Symmetric iSTDP rules reinforce mutual inhibition, mirroring pre-existing excitatory structure, whereas asymmetric kernels favor lateral inhibition.

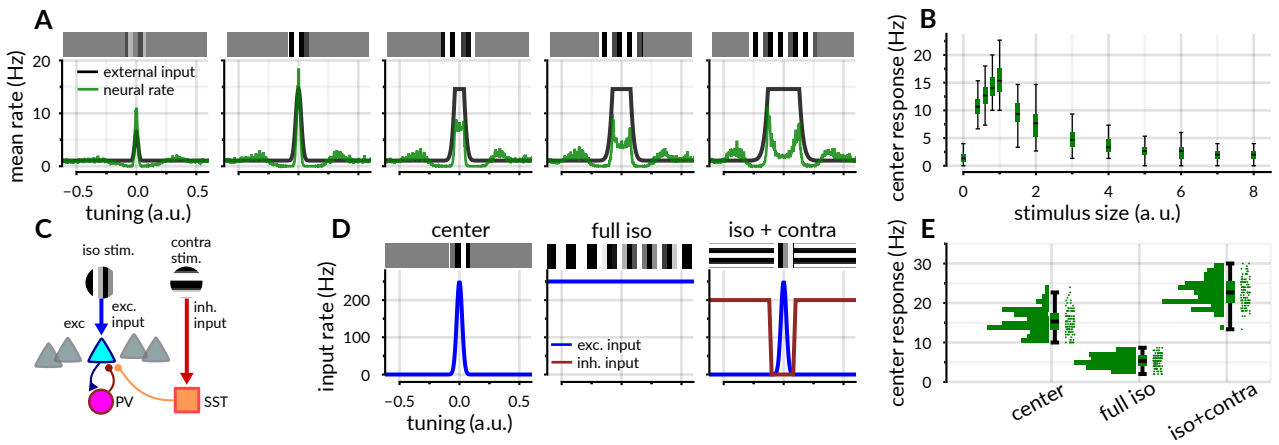

Figure 5: Response properties of the one-dimensional ring network. **A**. Representation of the stimuli (black) and of the network's response (green), as a function of the neuron's location relative to the stimulus center, averaged over 100 trials. **B**. Response to the center neuron only as a function of stimulus size. **C**. Schematics of the iso-contra stimulation protocol. The *iso* direction corresponds to excitatory inputs to excitatory neurons with matching receptive field. The *contra* direction generates instead an inhibitory input on the SST population. **D**. Profiles of the center, full iso and contra stimuli to the network. **E**. Mean response of the center neuron, averaged over 100 trials, for the three stimuli.

2.4 Emergent computations from structured stabilization in a ring network

Having shown that different forms of iSTDP shape E/I connectivity based on local correlations arising from direct connections, we next explored the functional consequences of structured stabilization in a recurrent network with a one-dimensional ring architecture. Ring networks are well-suited to model continuous attractor dynamics, supporting bumps of persistent activity (Amari 1977; Ben-Yishai et al. 1995; Khona and Fiete 2022). This makes them relevant for studying neural computations such as head-direction encoding (Xie et al. 2002; Kim et al. 2019; Petrucco et al. 2023) and orientation selectivity in the primary visual cortex (Goldberg et al. 2004).

We built a ring network of 800 excitatory and 200 inhibitory neurons arranged on a circle, with exc-to-exc and exc-to-inh weights decaying exponentially with the angular distance between neurons, following a rescaled Von Mises (Fig. 4A,B). We further divided the inhibitory population into two subclasses, governed by distinct iSTDP rules, but sharing identical neural dynamics. Inspired by experimental results (Lagzi et al. 2021), one population of "PV" interneurons followed a symmetric kernel, and another population of "SST" interneurons followed an antisymmetric one, both rules operating in the covariance-dominated regime. Initially, we set both PV-to-exc and SST-to-exc synapses as all-to-all, weak, and unbiased (see Fig. 4A for schematics, and Methods Table 3 for numerical parameters).

As the network evolves under the influence of the two plasticity rules, the inhibitory populations develop specialized roles. PV interneurons, governed by symmetric iSTDP, form strong mutual connections with excitatory neurons, mirroring the distance-dependent excitatory structure (Fig. 4B and Fig. 4C, red). Conversely, SST interneurons, driven by antisymmetric iSTDP, establish lateral inhibition across the ring (Fig. 4B and Fig. 4C, orange). This structured stabilization, with both localized and broader components, raises the possibility of nontrivial effective interactions between excitatory neurons, potentially mediated by the different roles of the PV and SST interneurons.

To estimate the effective interaction profile between excitatory neurons, we used linear response theory, equivalent to a Taylor expansion of the network dynamics around its operating point (Sadeh and Clopath 2020, see Methods for details). We found an effective connectivity profile resembling a Mexican-hat, with short-range excitation directly inherited from exc-to-exc connectivity and a broader inhibitory surround resulting from the combined action of PV and SST interneurons (Fig. 4D). This effective connectivity profile motivated us to explore the resulting network dynamics and its response to external stimuli. We examined the emergent spontaneous activity and observed a complex correlation structure, with periodic oscillations in correlations over distance

(Fig. 4E, see Supporting Information Fig. S6 for full correlation matrix). This emergent dynamics is reminiscent of the correlation maps observed in the developing cortex of ferrets, and is considered a signature of modular cortical organization (Smith et al. 2018).

After identifying aspects of emergent cortical dynamics in our network consistent with early developmental patterns, we investigated whether it also displays other fundamental computational properties classically measured in the visual system. A well-established nonlinear phenomenon in the visual cortex is surround suppression, in which gradually increasing the stimulus size initially amplifies the neuron's response up to a peak, but then leads to reduced activation (DeAngelis et al. 1994; Cavanaugh et al. 2002). We therefore introduced patterned stimuli to the network, where plasticity was deactivated after achieving structure. We delivered inputs of different strengths and sizes in the form of excitatory spike trains (Fig. 5A). Each stimulus was presented 100 times in randomized order, with a 0.5 s presentation followed by a 0.5 s blank. We found that larger stimuli elicited lower peak firing rates in their center region, while the response of neurons with receptive field closer to the stimulus' edge was larger (Fig. 5A, raster plot in Supporting Information). Consequently, the center neuron's response as a function of stimulus size exhibited a classic surround suppression profile (Fig. 5B) (DeAngelis et al. 1994).

As a control experiment, we repeated the same analysis in a model where all inhibitory neurons followed a rate-dominated iSTDP. In this case, inh-to-exc connections do not converge to a structured pattern (Fig. S7B). Although the network can still produce surround suppression, the edge effect is reversed, with neurons on the stimulus edge firing at a lower rate compared to neurons in the center (Fig. S7C, note the convex shape of the curve for large stimuli). Moreover, the correlations do not exhibit any spatial structure. In this case, inh-to-exc connections did not converge to a structured pattern (Fig. S7B), and, although the network could still produce surround suppression, the edge effect was reversed, with neurons on the stimulus edge firing at a lower rate compared to neurons in the center (Fig. S7C, note the convex shape of the curve for large stimuli), moreover the correlations did not show any spatial structure (Fig. S7E,F).

Finally, we investigated whether our model could also reproduce contextual facilitating effects, for example when a stimulus in the preferred orientation is surrounded by an orthogonal stimulus (Sillito and Jones 1996). These effects are mediated by a third subclass of interneurons, the vasoactive intestinal peptide (VIP)-expressing cells, that inhibit the SST population and thereby disinhibit excitatory neurons (Keller et al. 2020). We thus modeled the iso-oriented stimulus as excitatory spike trains delivered to the excitatory neurons, and introduced a contra-oriented stimulus as inhibitory inputs from VIP interneurons to the SST population at the corresponding location (Fig. 5C). We then compared the center-neuron response in three conditions: (1) an iso-oriented stimulus restricted to the neuron's receptive field; (2) a broader iso-oriented stimulus; (3) an iso stimulus in the receptive field combined with a contra stimulus in the distal regions (Fig. 5D). In agreement with experimental results (Keller et al. 2020), the distal contra stimulus disinhibits the center neuron and facilitates its response in the iso+contra stimulus conditions (Fig. 5,E).

These results demonstrate how structured stabilization shaped by the interplay of symmetric and antisymmetric iSTDP rules can give rise to complex dynamical and functional properties, such as suppressive and amplifying contextual effects, and to modular organization consistent with large-scale patterns of propagating activity in the developing sensory cortex.

3 Discussion

In this work, we have shown that a family of inhibitory STDP rules can not only stabilize excitation at the circuit level, but also drive the formation of specific connectivity patterns, or motifs, among neurons. In the absence of structured external inputs, or training signals, circuits self-organize purely through their spontaneous internal dynamics. We found that the shape of the iSTDP rule determines which motifs arise, selectively reinforcing structures based on the correlations present in the network.

While other computational studies have addressed the self-organized formation of motifs in large networks through STDP (Ocker et al. 2015; Ravid Tannenbaum and Burak 2016; Triplett et al. 2018; Montangie et al. 2020; Manz and Memmesheimer 2023), they mostly focused on exc-to-exc plasticity. Hence, they do not address how inhibitory structures and E/I motifs emerge, also because experimental evidence of E/I organization and direct measurements of iSTDP rules have become available only recently. Previous work on iSTDP has been largely restricted to feedforward circuits and rules dominated by either firing rates or covariances (Luz and Shamir 2014; Akil et al. 2021; Vignoud and Robert 2022; Lagzi and Fairhall 2024), including exploring multiple forms of iSTDP kernels within a covariance-dominated regime in purely feedforward architectures (Kleberg et al. 2014; Luz

and Shamir 2014). In contrast, our approach focuses on recurrent excitatory-inhibitory circuits and implements a generalized iSTDP framework that combines rate-dependent and covariance-dependent components. The rate-dependent terms guarantee stable neural activity, whereas the covariance-dependent terms support the emergence of E/I structure. Consequently, our framework provides deeper insight into both the regulatory and structural properties of generalized iSTDP rules, expanding our understanding of plasticity in closed-loop recurrent circuits.

A limitation of our work is that we derived analytic solutions only for the two-neurons Poisson model, and relied on numerical simulations for larger networks. By contrast, a recent study demonstrated that it is possible to study how iSTDP shapes and stabilizes recurrent LIF networks through a mean-field approach Akil et al. 2021. To achieve analytical tractability, that work assumed a regime of weak interactions, where internally-generated correlations appear negligible as the size of the neural population grows, and plasticity is driven exclusively by externally-induced correlations. In our approach, finite network size, strong interactions and a dominance of correlation-dependent terms together drive self-organization. It remains an open question whether and to what extent external correlations, when present, would dominate the internal ones, and how these two analytic frameworks could be integrated.

One strength of our framework is that it clarifies how the iSTDP kernel interacts with other plasticity parameters. For example, in rate-dominated regimes the weight evolution mainly depends on the net area of the STDP kernel and the rate-dependent terms, and not on the precise kernel shape. The covariance-dominated regime, however, can pose greater challenges for parameter tuning. A promising way forward is to employ numerical methods, for instance using simulation-based inference and deep learning, which have already been applied to optimize both excitatory and inhibitory STDP rules in recurrent circuits (Confavreux et al. 2024). Our analytical approach can guide and complement these efforts, helping clarify parameter degeneracies and simplifying the parameter search process.

While our analysis has primarily examined how small circuit motifs self-organize under different iSTDP rules, neural circuits can also form larger structures, including feedforward firing sequences and assemblies. Assemblies are groups of strongly connected neurons, often regarded as the fundamental building blocks of neural computations (Buzsáki 2010; Papadimitriou et al. 2020; Miehl et al. 2023). Previous work has examined how various iSTDP rules enable the learning, maintenance and self-repair of neural assemblies (Ocker and Doiron 2019; Lagzi et al. 2021; Bergoin et al. 2023; Bergoin et al. 2024). In particular, recent findings suggest that assemblies are formed and best preserved when inhibitory neurons are divided into subclasses that combine symmetric and rate-based with antisymmetric iSTDP (Lagzi et al. 2021), where the first rule can maintain network stability while the second prevents assemblies from merging through lateral inhibition, as in our work. Related work has also found that a combination of positive symmetric (Hebbian) and negative symmetric (anti-Hebbian) covariance-based rules can form and maintain stable assemblies governed by similar principles (Bergoin et al. 2024). These results on assemblies agree with ours in underscoring the intuition that different rules serve different purposes: some promote E/I balance and internal network stability, while others introduce competition through lateral inhibition.

Here, we address the spontaneous emergence of circuit structure at the sub-assembly level and in the absence of externally patterned inputs. Moreover, we investigate other functionally meaningful circuit configurations, such as ring networks, which self-organize and acquire emergent functional properties, such as contextual input modulation (DeAngelis et al. 1994; Cavanaugh et al. 2002) and the input and long-range correlations reminiscent of modular activity observed in the cortex (Smith et al. 2018).

While we primarily examined how inhibitory plasticity alone can generate E/I motifs, our framework can be naturally extended to incorporate excitatory STDP (eSTDP), extending the implications of self-stabilizing excitatory plasticity (Kempter et al. 1999) to large-scale recurrent networks containing both excitatory neurons and diverse inhibitory subtypes. Many earlier studies imposed additional mechanisms, such as non-local regularization, to stabilize eSTDP (Litwin-Kumar and Doiron 2014; Zenke et al. 2015; Eckmann et al. 2024), but our results suggest that the interplay between eSTDP and iSTDP may suffice to maintain E/I balance while producing a rich variety of connectivity structures, from local E/I motifs to large assemblies. Moreover, excitatory plasticity can enable circuits to learn new representations, refine current ones, and recover after lesions. Overall, these findings highlight how simple, local learning rules can both stabilize global network dynamics and self-organize structured connectivity in recurrent circuits, with broad implications for understanding neural development and adaptive learning.

4 Methods

4.1 General mean-field solution

We consider a class of pairwise iSTDP rules triggered by spiking events and purely based on spike-timing difference between pre- and postsynaptic neuron firing, defined by the following expression.

$$\begin{aligned} \text{on } t_k^{(\text{pre})} \;&\rightarrow\; \Delta W^{(\text{pre})} = \alpha_{\text{pre}} + \sum_{k'<k} L^{\Delta_-}\left(t_k^{(\text{pre})} - t_{k'}^{(\text{post})}\right) \;, \\ \text{on } t_k^{(\text{post})} \;&\rightarrow\; \Delta W^{(\text{post})} = \alpha_{\text{post}} + \sum_{k'<k} L^{\Delta_+}\left(t_k^{(\text{post})} - t_{k'}^{(\text{pre})}\right) \;; \end{aligned} \tag{2}$$

where $t_k^{(\text{pre})}$ and $t_k^{(\text{post})}$ represent pre- and postsynaptic spike events, respectively. In our numerical simulations the instantaneous change of weight is further regulated by a global learning rate A_0, as in $\Delta W^{*\,(\text{pre,post})} = A_0\,\Delta W^{(\text{pre,post})}$. The scaling by A_0 is omitted in our equations for simplicity. Each presynaptic spike triggers a weight change consisting of a fixed component, denoted α_{pre}, and a pairwise interaction term. Similarly, each postsynaptic spike triggers a weight change with a fixed component α_{post} and a pairwise interaction term. The pairwise interaction terms capture the cumulative influence of joint spiking events, incorporating the contribution of all preceding spikes $(k' < k)$ weighted by a kernel $L^{\Delta_\pm}(\cdot)$ that depends on the time differences between the spikes. The two functions $L^{\Delta_-}(\cdot)$ and $L^{\Delta_+}(\cdot)$ refer to the post-pre and the pre-post pairwise interaction, and are merged in a single function, known as the iSTDP pairwise interaction kernel, as follows:

$$L(\tau) := \begin{cases} L^{\Delta_+}(\tau) & \text{if } \tau \geq 0 \\ L^{\Delta_-}(-\tau) & \text{if } \tau < 0\,. \end{cases} \tag{3}$$

By convention, the $\tau > 0$ branch of the curve represents pre-post interactions (presynaptic spike preceding the postsynaptic spike), while the $\tau < 0$ branch represents post-pre interactions (postsynaptic spike preceding the presynaptic spike). In our framework, we consider $L(\tau)$ to be independent of α_{pre} and α_{post}. This is in contrast to previous formulations that incorporate these terms into the kernel representation (e.g., Vogels et al. 2011 and Confavreux et al. 2024). We assume that synaptic plasticity operates on a much slower timescale than neural dynamics, allowing us to treat synaptic weights as constant when calculating first- and second-order firing statistics (Kempter et al. 1999). We further assume variability in the spike-train statistics, so that it can be effectively described as a stochastic process. Thus, we can write the evolution of the mean synaptic weights as:

$$\begin{aligned} \frac{d\langle W\rangle}{d\,t} &= \alpha_{\text{pre}} r_{\text{pre}} + \alpha_{\text{post}} r_{\text{post}} + B\, r_{\text{pre}}\, r_{\text{post}} + Q \quad , \\ \text{with} \quad B &:= \int_{-\infty}^{+\infty} d\tau L(\tau) \quad \text{and} \quad Q := \int_{-\infty}^{+\infty} d\tau L(\tau)\, C(\tau)\;; \end{aligned} \tag{4}$$

where B is the integral of the iSTDP kernel, and Q is the only term that depends on the pairwise covariance density between pre and post-synaptic spike trains, here denoted $C(\tau)$. For a complete derivation see Supporting Information or Kempter et al. (1999) and Luz and Shamir (2014).

4.2 Expressions for the symmetric and antisymmetric exponential kernel

Eq. 4 can be applied to any form of iSTDP kernel, including, e.g., kernels with delayed interactions (Babadi and Abbott 2013). Here, we focus on two specific exponentially decaying shapes: symmetric and antisymmetric, defined as follows:

$$L^{\text{sy}}(\tau) = \frac{1}{2}\left(\frac{1}{\tau_+}\,A_+\,e^{-|\tau|/\tau_+} - \frac{1}{\tau_-}\,A_-\,e^{-|\tau|/\tau_-}\right) \quad \text{and} \quad L^{\text{an}}(\tau) = \begin{cases} -\frac{A_-}{\tau_-}\,e^{\tau/\tau_-} & \text{if } \tau < 0 \\ \frac{A_+}{\tau_+}\,e^{-\tau/\tau_+} & \text{if } \tau \geq 0 \end{cases} \tag{5}$$

With this parametrization, $A_+ = A_-$ implies $B = 0$, corresponding to covariance-dominated rules, whereas for $A_- = 0$ we have $B = A_+$, a rate-dominated regime. This choice of $L(\tau)$ allows us to keep track of the entire spiking history through activity traces updated at every spike time and decaying in time with the same constants as the iSTDP kernel. These traces will determine the amount of potentiation $(+)$ or depression $(-)$ triggered by each spike. For the antisymmetric rule, the presynaptic and postsynaptic traces are respectively defined as

follows:

$$\frac{d}{dt}T_{r+}(t) = -\frac{1}{\tau_+}T_{r+}(t) \quad \text{and} \quad \frac{d}{dt}T_{o-}(t) = -\frac{1}{\tau_-}T_{o-}(t)$$
$$\text{update:} \quad \begin{cases} T_{r+}(t+\varepsilon) = T_{r+}(t) + 1 & \text{on } t = t^{\,(\text{pre})} \\ T_{o-}(t+\varepsilon) = T_{o-}(t) + 1 & \text{on } t = t^{\,(\text{post})} \end{cases} \tag{6}$$

The symmetric rule requires, in addition to $T_{r+}(t)$ and $T_{o-}(t)$, the traces $T_{r-}(t)$ and $T_{o+}(t)$, defined analogously. The implementation of the asymmetric iSTDP rule is then given by:

$$\begin{cases} \Delta W^{\text{an}}(t) = \alpha_{\text{pre}} - \frac{A_-}{\tau_-}T_{o-}(t) & \text{on } t = t^{\,(\text{pre})} \\ \Delta W^{\text{an}}(t) = \alpha_{\text{post}} + \frac{A_+}{\tau_+}T_{r+}(t) & \text{on } t = t^{\,(\text{post})} \end{cases}, \tag{7}$$

and for the symmetric rule:

$$\begin{cases} \Delta W^{\text{sy}}(t) = \alpha_{\text{pre}} + \frac{A_+}{2\tau_+}T_{o+}(t) - \frac{A_-}{2\tau_-}T_{o-}(t) & \text{on } t = t^{\,(\text{pre})} \\ \Delta W^{\text{sy}}(t) = \alpha_{\text{post}} + \frac{A_+}{2\tau_+}T_{r+}(t) - \frac{A_-}{2\tau_-}T_{r-}(t) & \text{on } t = t^{\,(\text{post})} \end{cases}. \tag{8}$$

4.3 Low impact of correlations in rate-dominated rules

To better understand the role of the net area B of the iSTDP kernel in the weight dynamics (Eq. 4), and how it determines whether a rule is rate-dominated or covariance-dominated, we will consider here a simplified scenario, where $L(\tau)$ prescribes a +1 increase in weight for coincident pre-post spikes and no increase otherwise. This choice yields $B = A_+ = 1$. Consider now a pre and a post-synaptic neuron that fire with Poisson statistics at rates $r_{\text{pre}} = r_{\text{post}} = r = 10$ Hz. The term $B\, r_{\text{pre}}\, r_{\text{post}}$ in Eq. 4 takes a value of $100\,\text{s}^{-2}$ regardless of the correlation between the two neurons. But how does it compare to the covariance term Q in the presence of strong correlations between pre and post-synaptic activity? We can consider an regime of (unnaturally) high correlations, such that half of the spikes of the two neurons are synchronous (Pearson correlation 0.5). The term Q, corresponding to the spike-count covariance, is $5\,\text{s}^{-2}$. So even for very strong correlations Q is 20 times smaller than $B\, r_{\text{pre}}\, r_{\text{post}}$. This illustrates that in regimes with strong B, the weight dynamics in Eq. 4 are primarily governed by first-order statistics, with — at best — second-order contributions from correlations in spiking activity.

4.4 The two-dimensional reduced model

Finding a general solution for Eq. 4 can be challenging. For our purposes, we consider a simplified case. First, we model an excitatory and an inhibitory neuron as mutually interacting Poisson processes (Hawkes 1971; Ocker et al. 2017). Each time the presynaptic neuron spikes, the firing probability of the postsynaptic neuron increases (or decreases) proportionally to the excitatory (or inhibitory) post-synaptic potential EPSP (IPSP) kernel:

$$E_e(\tau) = \frac{1}{\tau_e}\, e^{-\tau/\tau_e} \quad \text{and} \quad E_i(\tau) = -\frac{1}{\tau_i}\, e^{-\tau/\tau_i} \quad \text{with } \tau \geq 0 \tag{9}$$

We further assume that the input to the inhibitory neuron is regulated homeostatically to keep its rate constant. In these conditions, Eq. 4 has an approximate fixed-point closed-form solution for the inhibitory connection w_{inh}:

$$\tilde{w}_{\text{inh}} = \frac{\alpha_{\text{pre}}\, r_{\text{inh}} + h_{\text{exc}}\Psi}{r_{\text{inh}}\left(\Psi - M_{ei}^{(10)}\right)} \quad ; \quad \tilde{r}_{\text{exc}} = -\frac{\alpha_{\text{pre}} r_{\text{inh}} + M_{ei}^{(10)} h_{\text{exc}}}{\Psi - M_{ei}^{(10)}} \quad ;$$
$$\text{with } \Psi = \alpha_{\text{post}} + B r_{\text{inh}} + M_{ie}^{(01)} w_{\text{exc}} \quad . \tag{10}$$

The coefficients $M_{ei}^{(10)}$ and $M_{ie}^{(01)}$ are found through the linear response theory (Hawkes 1971; Ocker et al. 2017), and depend on the iSTDP kernel and the EPSP / IPSP parameters (see Supporting Information for full analytic derivation).

For the symmetric iSTDP:

$$M_{ei}^{\text{sy},(10)} = \frac{A_+}{\tau_+ + \tau_i} - \frac{A_-}{\tau_- + \tau_i} \quad ; \quad M_{ie}^{\text{sy},(01)} = \frac{A_+}{\tau_+ + \tau_e} - \frac{A_-}{\tau_- + \tau_e} \tag{11}$$

For the antisymmetric iSTDP instead:

$$M_{ei}^{\text{an},(10)} = \frac{A_+}{\tau_+ + \tau_i} \quad ; \quad M_{ie}^{\text{an},(01)} = -\frac{A_-}{\tau_- + \tau_e} \tag{12}$$

Numerical parameters are in Table 1.

symbol	value	description
N_{exc}	1	number of excitatory neurons
N_{inh}	1	number of inhibitory neurons
τ_e	30 ms	EPSP kernel time constant
τ_i	30 ms	IPSP kernel time constant
h_e	40 Hz	excitatory input current
r_{inh}	80 Hz	inhibitory firing rate
		rate-dominated iSTDP rule
A_0	$5 \cdot 10^{-7}$	global learning rate
B	1	net area under the iSTDP kernel
α_{pre}	-25	constant dependent on presynaptic rate
α_{post}	0	constant dependent on postsynaptic rate
τ_+	50 ms	iSTDP kernel time constant
		covariance-dominated symmetric iSTDP rule
A_0	$3 \cdot 10^{-6}$	global learning rate
B	0	net area under the iSTDP kernel
α_{pre}	-25	constant dependent on presynaptic rate
α_{post}	0	constant dependent on postsynaptic rate
τ_+	50 ms	iSTDP kernel time constant for potentiation
τ_-	500 ms	iSTDP kernel time constant for depression
		covariance-dominated antisymmetric iSTDP rule
A_0	$3 \cdot 10^{-6}$	global learning rate
B	0	net area under the iSTDP kernel
α_{pre}	-25	constant dependent on presynaptic rate
α_{post}	0	constant dependent on postsynaptic rate
τ_+	50 ms	iSTDP kernel time constant for potentiation
τ_-	50 ms	iSTDP kernel time constant for depression

Table 1: Numerical parameters used for Fig. 1. The parameters in Fig. 2 match those of Fig. 1 unless otherwise indicated in the figure caption.

4.5 Randomly connected recurrent neural network

We built a random network of conductance-based leaky integrate-and-fire (LIF) neurons, with 900 excitatory units and 100 inhibitory units. The membrane potential obeys the following dynamics:

$$\begin{gathered} \frac{dV(t)}{dt} = \frac{1}{C_{\text{membrane}}} \left(-g_{\text{leak}}(V - V_{\text{L}}) - g_{\text{AMPA}}\,(V - V_{\text{AMPA}}) - g_{\text{GABA}}\,(V - V_{\text{GABA}}) + \text{input}\right) \\ \frac{dg_{\text{AMPA}}(t)}{dt} = -\frac{g_{\text{AMPA}}(t)}{\tau_{\text{AMPA}}} \quad ; \quad \frac{dg_{\text{GABA}}(t)}{dt} = -\frac{g_{\text{GABA}}(t)}{\tau_{\text{GABA}}} \end{gathered} \tag{13}$$

With $g_{\text{AMPA}}(t)$ and $g_{\text{GABA}}(t)$ updated for each incoming signal proportionally to the excitatory and inhibitory synaptic weights. We assigned a probability of connection for all non-plastic connections (excitatory to excitatory, excitatory to inhibitory, inhibitory to inhibitory). The inhibitory to excitatory connectivity was all-to-all and plastic. The network received an external input in the form of Poisson spike trains at fixed rate. The excitatory neurons received a small amount of shared spikes, along with spikes generated i.i.d. for each neuron. Numerical parameters are in Table 2.

symbol	value	description
N_{exc}	900	number of excitatory neurons
N_{inh}	100	number of inhibitory neurons
C_{membrane}	200 pF	membrane capacitance
τ_{AMPA}	5 ms	glutamatergic synaptic time constant
τ_{GABA}	10 ms	GABAergic synaptic time constant
g_{leak}	10	leak conductivity
V_{L}	-60 mV	membrane leak reverse potential
V_{AMPA}	0 mV	AMPA reverse potential
V_{GABA}	-80 mV	GABA reverse potential
$V_{\text{threshold}}$	-50 mV	firing threshold
$p_{\text{e}\leftarrow\text{e}}$	0.2	exc-to-exc connection probability
$p_{\text{i}\leftarrow\text{e}}$	0.1	exc-to-inh connection probability
$p_{\text{e}\leftarrow\text{i}}$	1.0	inh-to-exc connection probability
$p_{\text{i}\leftarrow\text{i}}$	1.0	inh-to-inh connection probability
$w_{\text{e}\leftarrow\text{e}}$	1.0	exc-to-exc synaptic weight
$w_{\text{i}\leftarrow\text{e}}$	1.0	exc-to-inh synaptic weight
$w_{\text{i}\leftarrow\text{i}}$	0.3	inh-to-inh synaptic weight
		External input
N_{shared}	50	number or shared input neurons for exc and inh
$N_{\text{indep.,exc}}$	250	number of independent input neurons for exc
$N_{\text{indep.,exc}}$	0	number of independent input neurons for inh
r_{input}	50 Hz	rate of input neurons
$w_{\text{input,ext}}$	1	weight of input synapse to exc
$w_{\text{input,inh}}$	0.1	weight of (independent) input synapse to inh
		covariance-dominated symmetric iSTDP rule (panels A-F)
A_0	$2 \cdot 10^{-2}$	global learning rate
B	0	net integral of the STDP kernel
α_{pre}	-0.1	constant dependent on presynaptic rate
α_{post}	0	constant dependent on postsynaptic rate
τ_+	30 ms	iSTDP kernel time constant
τ_-	600 ms	iSTDP kernel time constant
		covariance-dominated asymmetric iSTDP rule (panels G-K)
A_0	$5 \cdot 10^{-3}$	global learning rate
B	0.3	net integral of the iSTDP kernel
α_{pre}	-0.7	constant dependent on presynaptic rate
α_{post}	0.2	constant dependent on postsynaptic rate
τ_+	60 ms	iSTDP kernel time constant for potentiation
τ_-	30 ms	iSTDP kernel time constant for depression

Table 2: Numerical parameters for the random RNN in Fig. 3.

4.6 Ring model

The ring model was implemented with the same neural dynamics as the random network (Eq. 13). Here however both excitatory and inhibitory neurons received a baseline external input, independent for each unit. We used 800 excitatory units and 200 inhibitory units, further divided into a PV and an SST subclass. The synaptic weight between an excitatory neuron and its neighbors decayed with distance, following a bell-shaped curve modeled as a rescaled von Mises distribution:

$$w_{ij}^{(\bullet\leftarrow\text{e})} := C_{\bullet\leftarrow\text{e}} \left(\exp\left(k_{\bullet\leftarrow\text{e}} \cos\left(\theta_i - \theta_j\right)\right)\right) \ ; \tag{14}$$

where θ_i and θ_j represent the locations of the post- and presynaptic neuron on the ring, with each population equally spaced in the $[0,\ 2\pi)$ interval. The factor $k_{\bullet\leftarrow\text{e}}$ regulates the connection range. Finally the scaling factor $C_{\bullet\leftarrow\text{e}}$ is tuned so that the sum of incoming weights to either exc, PV or SST population is at a desired value sum $w_{\bullet\leftarrow\text{e}}$. All inh-to-inh connections were fixed and dense. We considered two distinct inhibitory populations with identical dynamics, but different iSTDP rules. The first population had a symmetric iSTDP kernel, the second an antisymmetric one. The input was different than in the previous model, with a constant external current in the excitatory population, and independent excitatory spike trains for all units.

To measure the network response to external stimuli (Fig. 5), we froze plasticity after learning, and delivered stimuli as uncorrelated spike trains. Iso-oriented stimuli took the form of excitatory spikes delivered to the exc neurons, contra-oriented stimuli were modeled as inhibitory spikes delivered to the SST neurons. Numerical parameters are shown in Table 3.

symbol	**value**	**description**
N_{exc}	800	number of excitatory neurons
N_{PV}	100	number of PV interneurons
N_{SST}	100	number of SST interneurons
sum $w_{\bullet\leftarrow\text{e}}$	145	total incoming exc weights to exc, PV, SST
$k_{\text{e}\leftarrow\text{e}}$	10	width of exc-to-exc connections
$k_{\text{PV}\leftarrow\text{e}}$	5	width of exc-to-pv connections
$k_{\text{SST}\leftarrow\text{e}}$	10	width of exc-to-sst connections
$p_{\text{PV,SST}\leftarrow\text{PV,SST}}$	0.9	connection probability for PV-PV, SST-SST and PV-SST
$w_{\text{PV}\leftarrow\text{PV}}$	2	PV-to-PV synaptic weight
$w_{\text{SST}\leftarrow\text{PV}}$	0.5	PV-to-SST synaptic weight
$w_{\text{SST}\leftarrow\text{SST}}$	2	SST-to-SST synaptic weight
$w_{\text{PV}\leftarrow\text{SST}}$	0.5	SST-to-PV synaptic weight
	External input	
$h_{\text{in,exc}}$	1000 pA	constant input current to exc population
N_{input}	50	number of independent input neurons
w_{input}	1.0	weight of input synapse
$r_{\text{input,exc}}$	80 Hz	rate of input neurons to exc
$r_{\text{input,inh}}$	40 Hz	rate of input neurons to PV,SST
	symmetric iSTDP rule for the PV population	
A_0	$2 \cdot 10^{-3}$	global learning rate
B	0	net integral of the iSTDP kernel
α_{pre}	-0.2	constant dependent on presynaptic rate
α_{post}	1	constant dependent on postsynaptic rate
τ_+	30 ms	STDP kernel time constant for potentiation
τ_-	600 ms	STDP kernel time constant for depression
	antisymmetric iSTDP rule for the SST population	
A_0	$1 \cdot 10^{-3}$	global learning rate
B	0	net integral of the iSTDP kernel
α_{pre}	-0.2	constant dependent on presynaptic rate
α_{post}	1	constant dependent on postsynaptic rate
τ_+	30 ms	iSTDP kernel time constant for potentiation
τ_-	30 ms	iSTDP kernel time constant for depression

Table 3: Numerical parameters for one-dimensional ring model. The parameters that regulate neural dynamics are the same as in Table 2.

5 Software implementation and code availability

All simulations for the reduced models are based on custom code written in Julia (Bezanson et al. 2017, https://julialang.org), also available as separate toolbox. For the large-scale simulations we used Brian2 (Stimberg et al. 2019, https://briansimulator.org). Finally some of the analytic integrals were solved using Maple (https://maplesoft.com). The code is fully open-source and publicly available on GitHub under the Creative Commons Attribution 4.0 International (CC BY 4.0) License.
https://github.com/comp-neural-circuits/structured-stabilization-in-recurrent-neural-circuits

Acknowledgments

This project has received funding from the European Research Council (ERC StG NeuroDevo, Grant agreement No. 804824) and Marie Skłodowska-Curie ITN SmartNets (grant agreement No. 860949) under the European Union's Horizon 2020 research and innovation program, by the International Human Frontier Science Program Organization (RGP0062/2021), and by the Technical University of Munich (TUM). Thanks to the "Computation in Neural Circuits" group for their useful insights and discussions. Special thanks to Alessio Quaresima for his valuable feedback and for independently testing our results.

References

Agnes, Everton J. and Tim P. Vogels (Mar. 20, 2024). "Co-Dependent Excitatory and Inhibitory Plasticity Accounts for Quick, Stable and Long-Lasting Memories in Biological Networks". In: *Nature Neuroscience*, pp. 1–11. DOI: 10.1038/s41593-024-01597-4.

Akil, Alan Eric, Robert Rosenbaum, and Krešimir Josić (May 12, 2021). "Balanced Networks under Spike-Time Dependent Plasticity". In: *PLOS Computational Biology* 17.5, e1008958. DOI: 10.1371/journal.pcbi.1008958.

Amari, Shun'ichi (June 1, 1977). "Dynamics of Pattern Formation in Lateral-Inhibition Type Neural Fields". In: *Biological Cybernetics* 27.2, pp. 77–87. DOI: 10.1007/BF00337259.

Babadi, Baktash and L. F. Abbott (Feb. 21, 2013). "Pairwise Analysis Can Account for Network Structures Arising from Spike-Timing Dependent Plasticity". In: *PLOS Computational Biology* 9.2, e1002906. DOI: 10.1371/journal.pcbi.1002906.

Ben-Yishai, R, R L Bar-Or, and H Sompolinsky (Apr. 25, 1995). "Theory of Orientation Tuning in Visual Cortex." In: *Proceedings of the National Academy of Sciences* 92.9, pp. 3844–3848. DOI: 10.1073/pnas.92.9.3844.

Bergoin, Raphaël, Alessandro Torcini, Gustavo Deco, Mathias Quoy, and Gorka Zamora-López (Apr. 28, 2023). "Inhibitory Neurons Control the Consolidation of Neural Assemblies via Adaptation to Selective Stimuli". In: *Scientific Reports* 13.1 (1), p. 6949. DOI: 10.1038/s41598-023-34165-0.

– (July 17, 2024). *Emergence and Long-Term Maintenance of Modularity in Plastic Networks of Spiking Neurons*. DOI: 10.1101/2024.07.15.603496. Pre-published.

Bezanson, Jeff, Alan Edelman, Stefan Karpinski, and Viral B Shah (2017). "Julia: A fresh approach to numerical computing". In: *SIAM review* 59.1, pp. 65–98.

Bollmann, Yannick et al. (Sept. 2023). "Prominent in Vivo Influence of Single Interneurons in the Developing Barrel Cortex". In: *Nature Neuroscience* 26.9 (9), pp. 1555–1565. DOI: 10.1038/s41593-023-01405-5.

Brunel, N. (2000). "Dynamics of Sparsely Connected Networks of Excitatory and Inhibitory Spiking Neurons". In: *Journal of computational neuroscience* 8.3, pp. 183–208. DOI: 10.1023/A:1008925309027.

Buzsáki, György (Nov. 4, 2010). "Neural Syntax: Cell Assemblies, Synapsembles, and Readers". In: *Neuron* 68.3, pp. 362–385. DOI: 10.1016/j.neuron.2010.09.023.

Cavanaugh, James R., Wyeth Bair, and J. Anthony Movshon (Nov. 2002). "Nature and Interaction of Signals From the Receptive Field Center and Surround in Macaque V1 Neurons". In: *Journal of Neurophysiology* 88.5, pp. 2530–2546. DOI: 10.1152/jn.00692.2001.

Clopath, Claudia, Tim P. Vogels, Robert C. Froemke, and Henning Sprekeler (July 29, 2016). "Receptive Field Formation by Interacting Excitatory and Inhibitory Synaptic Plasticity". In: p. 066589. DOI: 10.1101/066589.

Confavreux, Basile, Everton J. Agnes, Friedemann Zenke, Henning Sprekeler, and Tim P. Vogels (June 17, 2024). *Balancing Complexity, Performance and Plausibility to Meta Learn Plasticity Rules in Recurrent Spiking Networks*. DOI: 10.1101/2024.06.17.599260.

Dahmen, David et al. (Nov. 4, 2020). "Strong Coupling and Local Control of Dimensionality across Brain Areas". In: *bioRxiv*, p. 2020.11.02.365072. DOI: 10.1101/2020.11.02.365072.

DeAngelis, G. C., R. D. Freeman, and I. Ohzawa (Jan. 1994). "Length and Width Tuning of Neurons in the Cat's Primary Visual Cortex". In: *Journal of Neurophysiology* 71.1, pp. 347–374. DOI: 10.1152/jn.1994.71.1.347.

Dorrn, Anja L., Kexin Yuan, Alison J. Barker, Christoph E. Schreiner, and Robert C. Froemke (June 17, 2010). "Developmental Sensory Experience Balances Cortical Excitation and Inhibition". In: *Nature* 465.7300, pp. 932–936. DOI: 10.1038/nature09119.

Duan, Zhe Ran S. et al. (Jan. 8, 2020). "GABAergic Restriction of Network Dynamics Regulates Interneuron Survival in the Developing Cortex". In: *Neuron* 105.1, 75–92.e5. DOI: 10.1016/j.neuron.2019.10.008.

Eckmann, Samuel, Edward James Young, and Julijana Gjorgjieva (June 18, 2024). "Synapse-Type-Specific Competitive Hebbian Learning Forms Functional Recurrent Networks". In: *Proceedings of the National Academy of Sciences* 121.25, e2305326121. DOI: 10.1073/pnas.2305326121.

Froemke, Robert C., Michael M. Merzenich, and Christoph E. Schreiner (Nov. 15, 2007). "A Synaptic Memory Trace for Cortical Receptive Field Plasticity". In: *Nature* 450.7168, pp. 425–429. DOI: 10.1038/nature06289.

Goldberg, Joshua A., Uri Rokni, and Haim Sompolinsky (May 13, 2004). "Patterns of Ongoing Activity and the Functional Architecture of the Primary Visual Cortex". In: *Neuron* 42.3, pp. 489–500. DOI: 10.1016/S0896-6273(04)00197-7.

Gong, Ziyi and Nicolas Brunel (Apr. 10, 2024). *Inhibitory Plasticity Enhances Sequence Storage Capacity and Retrieval Robustness*. DOI: 10.1101/2024.04.08.588573. Pre-published.

Hafizi, Hadi et al. (Aug. 20, 2022). *Inhibition-Dominated Rich-Club Shapes Dynamics in Cortical Microcircuits*. DOI: 10.1101/2021.05.07.443074.

Hanganu-Opatz, Ileana L. et al. (Feb. 3, 2021). "The Logic of Developing Neocortical Circuits in Health and Disease". In: *Journal of Neuroscience* 41.5, pp. 813–822. DOI: 10.1523/JNEUROSCI.1655-20.2020.

Hawkes, Alan G. (1971). "Spectra of Some Self-Exciting and Mutually Exciting Point Processes". In: *Biometrika* 58.1, pp. 83–90. DOI: 10.2307/2334319.

Jiang, Xiaolong et al. (Nov. 27, 2015). "Principles of Connectivity among Morphologically Defined Cell Types in Adult Neocortex". In: *Science* 350.6264, aac9462. DOI: 10.1126/science.aac9462.

Jin, Dezhe (Oct. 2002). "Fast Convergence of Spike Sequences to Periodic Patterns in Recurrent Networks". In: *Physical Review Letters* 89.20. DOI: 10.1103/PhysRevLett.89.208102.

Keller, Andreas J. et al. (Dec. 2020). "A Disinhibitory Circuit for Contextual Modulation in Primary Visual Cortex". In: *Neuron*. DOI: 10.1016/j.neuron.2020.11.013.

Kempter, Richard, Wulfram Gerstner, and J. Leo van Hemmen (Apr. 1, 1999). "Hebbian Learning and Spiking Neurons". In: *Physical Review E* 59.4, pp. 4498–4514. DOI: 10.1103/PhysRevE.59.4498.

Khona, Mikail and Ila R. Fiete (Nov. 3, 2022). "Attractor and Integrator Networks in the Brain". In: *Nature Reviews Neuroscience*. DOI: 10.1038/s41583-022-00642-0.

Kim, Sung Soo, Ann M. Hermundstad, Sandro Romani, L. F. Abbott, and Vivek Jayaraman (Dec. 2019). "Generation of Stable Heading Representations in Diverse Visual Scenes". In: *Nature* 576.7785 (7785), pp. 126–131. DOI: 10.1038/s41586-019-1767-1.

Kleberg, Florence I., Tomoki Fukai, and Matthieu Gilson (May 7, 2014). "Excitatory and Inhibitory STDP Jointly Tune Feedforward Neural Circuits to Selectively Propagate Correlated Spiking Activity". In: *Frontiers in Computational Neuroscience* 8. DOI: 10.3389/fncom.2014.00053.

Kuan, Aaron T. et al. (Feb. 21, 2024). "Synaptic Wiring Motifs in Posterior Parietal Cortex Support Decision-Making". In: *Nature*, pp. 1–7. DOI: 10.1038/s41586-024-07088-7.

Lagzi, Fereshteh, Martha Canto Bustos, Anne-Marie Oswald, and Brent Doiron (Sept. 7, 2021). "Assembly Formation Is Stabilized by Parvalbumin Neurons and Accelerated by Somatostatin Neurons". In: *bioRxiv*. DOI: 10.1101/2021.09.06.459211.

Lagzi, Fereshteh and Adrienne L. Fairhall (Mar. 20, 2024). "Emergence of Co-Tuning in Inhibitory Neurons as a Network Phenomenon Mediated by Randomness, Correlations, and Homeostatic Plasticity". In: *Science Advances* 10.12, eadi4350. DOI: 10.1126/sciadv.adi4350.

Lakhera, Shreya, Elizabeth Herbert, and Julijana Gjorgjieva (June 10, 2024). "Modeling the Emergence of Circuit Organization and Function during Development". In: *Cold Spring Harbor Perspectives in Biology*, a041511. DOI: 10.1101/cshperspect.a041511.

Litwin-Kumar, Ashok and Brent Doiron (Nov. 14, 2014). "Formation and maintenance of neuronal assemblies through synaptic plasticity". In: *Nature Communications* 5, p. 5319. DOI: 10.1038/ncomms6319.

Luz, Yotam and Maoz Shamir (July 7, 2014). "The Effect of STDP Temporal Kernel Structure on the Learning Dynamics of Single Excitatory and Inhibitory Synapses ". In: *PLOS ONE* 9.7, e101109. DOI: 10.1371/journal.pone.0101109.

Mackwood, Owen, Laura B Naumann, and Henning Sprekeler (Apr. 26, 2021). "Learning Excitatory-Inhibitory Neuronal Assemblies in Recurrent Networks". In: *eLife* 10. Ed. by Stephanie E Palmer and Richard B Ivry, e59715. DOI: 10.7554/eLife.59715.

Maes, Amadeus, Mauricio Barahona, and Claudia Clopath (Mar. 25, 2021). "Learning Compositional Sequences with Multiple Time Scales through a Hierarchical Network of Spiking Neurons". In: *PLOS Computational Biology* 17.3, e1008866. DOI: 10.1371/journal.pcbi.1008866.

Manz, Paul and Raoul-Martin Memmesheimer (Apr. 12, 2023). "Purely STDP-based Assembly Dynamics: Stability, Learning, Overlaps, Drift and Aging". In: *PLOS Computational Biology* 19.4, e1011006. DOI: 10.1371/journal.pcbi.1011006.

Meissner-Bernard, Claire, Bethan Jenkins, et al. (Mar. 25, 2025). "Computational Functions of Precisely Balanced Neuronal Microcircuits in an Olfactory Memory Network". In: *Cell Reports* 44.3. DOI: 10.1016/j.celrep.2025.115330.

Meissner-Bernard, Claire, Friedemann Zenke, and Rainer W Friedrich (Jan. 13, 2025). "Geometry and Dynamics of Representations in a Precisely Balanced Memory Network Related to Olfactory Cortex". In: *eLife* 13. Ed. by Julijana Gjorgjieva and Panayiota Poirazi, RP96303. DOI: 10.7554/eLife.96303.

Miehl, Christoph and Julijana Gjorgjieva (Dec. 2, 2022). "Stability and Learning in Excitatory Synapses by Nonlinear Inhibitory Plasticity". In: *PLOS Computational Biology* 18.12, e1010682. DOI: 10.1371/journal.pcbi.1010682.

Miehl, Christoph, Sebastian Onasch, Dylan Festa, and Julijana Gjorgjieva (2023). "Formation and Computational Implications of Assemblies in Neural Circuits". In: *The Journal of Physiology* 601.15, pp. 3071–3090. DOI: 10.1113/JP282750.

Mongillo, Gianluigi, Simon Rumpel, and Yonatan Loewenstein (Oct. 2018). "Inhibitory Connectivity Defines the Realm of Excitatory Plasticity". In: *Nature Neuroscience* 21.10, pp. 1463–1470. DOI: 10.1038/s41593-018-0226-x.

Montangie, Lisandro, Christoph Miehl, and Julijana Gjorgjieva (May 8, 2020). "Autonomous Emergence of Connectivity Assemblies via Spike Triplet Interactions". In: *PLOS Computational Biology* 16.5, e1007835. DOI: 10.1371/journal.pcbi.1007835.

Mulholland, Haleigh N, Bettina Hein, Matthias Kaschube, and Gordon B Smith (Dec. 8, 2021). "Tightly Coupled Inhibitory and Excitatory Functional Networks in the Developing Primary Visual Cortex". In: *eLife* 10, e72456. DOI: 10.7554/eLife.72456.

Ocker, Gabriel K. et al. (Oct. 1, 2017). "From the Statistics of Connectivity to the Statistics of Spike Times in Neuronal Networks". In: *Current Opinion in Neurobiology*. Computational Neuroscience 46, pp. 109–119. DOI: 10.1016/j.conb.2017.07.011.

Ocker, Gabriel Koch and Brent Doiron (Mar. 1, 2019). "Training and Spontaneous Reinforcement of Neuronal Assemblies by Spike Timing Plasticity". In: *Cerebral Cortex* 29.3, pp. 937–951. DOI: 10.1093/cercor/bhy001.

Ocker, Gabriel Koch, Ashok Litwin-Kumar, and Brent Doiron (Aug. 20, 2015). "Self-Organization of Microcircuits in Networks of Spiking Neurons with Plastic Synapses". In: *PLOS Comput Biol* 11.8, e1004458. DOI: 10.1371/journal.pcbi.1004458.

Papadimitriou, Christos H., Santosh S. Vempala, Daniel Mitropolsky, Michael Collins, and Wolfgang Maass (June 23, 2020). "Brain Computation by Assemblies of Neurons". In: *Proceedings of the National Academy of Sciences* 117.25, pp. 14464–14472. DOI: 10.1073/pnas.2001893117.

Peng, Yangfan et al. (June 16, 2021). "Spatially Structured Inhibition Defined by Polarized Parvalbumin Interneuron Axons Promotes Head Direction Tuning". In: *Science Advances* 7.25, eabg4693. DOI: 10.1126/sciadv.abg4693.

Petrucco, Luigi et al. (Apr. 24, 2023). "Neural Dynamics and Architecture of the Heading Direction Circuit in Zebrafish". In: *Nature Neuroscience*, pp. 1–9. DOI: 10.1038/s41593-023-01308-5.

Ravid Tannenbaum, Neta and Yoram Burak (Aug. 12, 2016). "Shaping Neural Circuits by High Order Synaptic Interactions". In: *PLOS Computational Biology* 12.8, e1005056. DOI: 10.1371/journal.pcbi.1005056.

Reimann, Michael W., Daniela Egas Santander, András Ecker, and Eilif B. Muller (Dec. 22, 2023). *Specific Inhibition and Disinhibition in the Higher-Order Structure of a Cortical Connectome*. DOI: 10.1101/2023.12.22.573036.

Riquelme, Juan Luis, Mike Hemberger, Gilles Laurent, and Julijana Gjorgjieva (Feb. 13, 2023). "Single Spikes Drive Sequential Propagation and Routing of Activity in a Cortical Network". In: *eLife* 12. Ed. by Peter Latham, Laura L Colgin, and Abigail Morrison, e79928. DOI: 10.7554/eLife.79928.

Rubin, Ran, L. F. Abbott, and Haim Sompolinsky (Oct. 17, 2017). "Balanced Excitation and Inhibition Are Required for High-Capacity, Noise-Robust Neuronal Selectivity". In: *Proceedings of the National Academy of Sciences*, p. 201705841. DOI: 10.1073/pnas.1705841114.

Sadeh, Sadra and Claudia Clopath (Oct. 27, 2020). "Theory of Neuronal Perturbome in Cortical Networks". In: *Proceedings of the National Academy of Sciences* 117.43, pp. 26966–26976. DOI: 10.1073/pnas.2004568117.

– (Nov. 5, 2021). "Excitatory-Inhibitory Balance Modulates the Formation and Dynamics of Neuronal Assemblies in Cortical Networks". In: *Science Advances* 7.45, eabg8411. DOI: 10.1126/sciadv.abg8411.

Schulz, Auguste, Christoph Miehl, Michael J Berry II, and Julijana Gjorgjieva (Oct. 14, 2021). "The Generation of Cortical Novelty Responses through Inhibitory Plasticity". In: *eLife* 10. Ed. by Maria N Geffen, Joshua I Gold, and Maria N Geffen, e65309. DOI: 10.7554/eLife.65309.

Shao, Shuai, Juan Luis Riquelme, and Julijana Gjorgjieva (Sept. 30, 2024). *Spontaneous Emergence and Drifting of Sequential Neural Activity in Recurrent Networks*. DOI: 10.1101/2024.09.27.615499. URL: https://www.biorxiv.org/content/10.1101/2024.09.27.615499v1 (visited on 10/15/2024).

Sillito, AM and HE Jones (Jan. 1, 1996). "Context-Dependent Interactions and Visual Processing in V1". In: *Journal of Physiology-Paris* 90.3, pp. 205–209. DOI: 10.1016/S0928-4257(97)81424-6.

Smith, Gordon B., Bettina Hein, David E. Whitney, David Fitzpatrick, and Matthias Kaschube (Nov. 2018). "Distributed Network Interactions and Their Emergence in Developing Neocortex". In: *Nature Neuroscience* 21.11 (11), pp. 1600–1608. DOI: 10.1038/s41593-018-0247-5.

Sreenivasan, Varun et al. (May 17, 2022). "Input-Specific Control of Interneuron Numbers in Nascent Striatal Networks". In: *Proceedings of the National Academy of Sciences* 119.20, e2118430119. DOI: 10.1073/pnas.2118430119.

Stimberg, Marcel, Romain Brette, and Dan FM Goodman (Aug. 20, 2019). "Brian 2, an Intuitive and Efficient Neural Simulator". In: *eLife* 8. Ed. by Frances K Skinner, Ronald L Calabrese, Frances K Skinner, Fleur Zeldenrust, and Richard C Gerkin, e47314. DOI: 10.7554/eLife.47314.

Trägenap, Sigrid, David E. Whitney, David Fitzpatrick, and Matthias Kaschube (Feb. 2025). "The Developmental Emergence of Reliable Cortical Representations". In: *Nature Neuroscience* 28.2, pp. 394–405. DOI: 10.1038/s41593-024-01857-3.

Tremblay, Robin, Soohyun Lee, and Bernardo Rudy (July 20, 2016). "GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits". In: *Neuron* 91.2, pp. 260–292. DOI: 10.1016/j.neuron.2016.06.033.

Triplett, Marcus A., Lilach Avitan, and Geoffrey J. Goodhill (Sept. 28, 2018). "Emergence of Spontaneous Assembly Activity in Developing Neural Networks without Afferent Input". In: *PLOS Computational Biology* 14.9, e1006421. DOI: 10.1371/journal.pcbi.1006421.

Tsodyks, M. V. and T. Sejnowski (May 1995). "Rapid State Switching in Balanced Cortical Network Models". In: *Network: Computation in Neural Systems* 6.2, p. 111. DOI: 10.1088/0954-898X/6/2/001.

Vignoud, Gaëtan and Philippe Robert (May 10, 2022). "Spontaneous Dynamics of Synaptic Weights in Stochastic Models with Pair-Based Spike-Timing-Dependent Plasticity". In: *Physical Review E* 105.5, p. 054405. DOI: 10.1103/PhysRevE.105.054405.

Vogels, Tim P., H. Sprekeler, F. Zenke, C. Clopath, and W. Gerstner (Nov. 10, 2011). "Inhibitory Plasticity Balances Excitation and Inhibition in Sensory Pathways and Memory Networks". In: *Science* 334.6062, pp. 1569–1573. DOI: 10.1126/science.1211095.

Wallenstein, Gene V. and Michael E. Hasselmo (July 1997). "GABAergic Modulation of Hippocampal Population Activity: Sequence Learning, Place Field Development, and the Phase Precession Effect". In: *Journal of Neurophysiology* 78.1, pp. 393–408. DOI: 10.1152/jn.1997.78.1.393.

Wu, Sherry Jingjing et al. (Sept. 6, 2023). "Cortical Somatostatin Interneuron Subtypes Form Cell-Type-Specific Circuits". In: *Neuron* 111.17, 2675–2692.e9. DOI: 10.1016/j.neuron.2023.05.032.

Xie, Xiaohui, Richard H. R. Hahnloser, and H. Sebastian Seung (Oct. 9, 2002). "Double-Ring Network Model of the Head-Direction System". In: *Physical Review E* 66.4, p. 041902. DOI: 10.1103/PhysRevE.66.041902.

Zenke, Friedemann, Everton J. Agnes, and Wulfram Gerstner (Apr. 21, 2015). "Diverse Synaptic Plasticity Mechanisms Orchestrated to Form and Retrieve Memories in Spiking Neural Networks". In: *Nature Communications* 6. DOI: 10.1038/ncomms7922.

Znamenskiy, Petr et al. (Mar. 20, 2024). "Functional Specificity of Recurrent Inhibition in Visual Cortex". In: *Neuron* 112.6, 991–1000.e8. DOI: 10.1016/j.neuron.2023.12.013.

Supporting Information for:

Structured stabilization in recurrent neural circuits through inhibitory synaptic plasticity

S1 Analytic results

S1.1 Mean-field equation for weight dynamics

In this section, we adapt the results introduced by Kempter et al. (1999) to our notation and our specific needs. We first describe a sufficiently general expression for the pairwise STDP rules. Given two neurons, one presynaptic and one postsynaptic, an excitatory or inhibitory STDP rule is a weight update triggered by either a pre- or a postsynaptic spike. The weight update solely depends on the time differences between the previous spike times, relative to the current spike time:

$$\begin{aligned} \text{on } \; t = t_k^{(\text{pre})} \; &\rightarrow \; \Delta W^{(\text{pre})} = \alpha_{\text{pre}} + \sum_{k'<k} L^{\Delta_-}\left(t - t_{k'}^{(\text{post})}\right) \\ \text{on } \; t = t_k^{(\text{post})} \; &\rightarrow \; \Delta W^{(\text{post})} = \alpha_{\text{post}} + \sum_{k'<k} L^{\Delta_+}\left(t - t_{k'}^{(\text{pre})}\right) \end{aligned} , \tag{S1}$$

where $t_k{}^{(\text{pre})}$ and $t_k{}^{(\text{post})}$ represent pre- and postsynaptic spike events, respectively. We can merge the functions L^{Δ_-} and L^{Δ_+} into a single STDP kernel defined as:

$$L(\tau) := \begin{cases} L^{\Delta_+}(\tau) & \text{if } \; \tau \geq 0 \\ L^{\Delta_-}(-\tau) & \text{if } \; \tau < 0 \, . \end{cases} \tag{S2}$$

Therefore $\tau > 0$ (the Δ_+ branch) represents pre-post interactions (presynaptic spike preceding the postsynaptic spike), while $\tau < 0$ (the Δ_- branch) represents post-pre interactions (postsynaptic spike preceding the presynaptic spike).

It is useful to express each spike train as a sum of Dirac delta functions centered at each spike time:

$$S^i(t) := \sum_k \delta(t - t_k^i) \tag{S3}$$

where i denotes the neuron index. With this notation, the number of spikes in the $[t_a, t_b]$ interval is:

$$N^i_{[t_a,t_b]} = \int_{t_a}^{t_b} dt \; S^i(t) \tag{S4}$$

Assuming stationary conditions, we can therefore define the firing rate of neuron i as:

$$r_i := \frac{1}{T} \int_0^T dt \; S^i(t) = \left\langle S^i(t) \right\rangle_t . \tag{S5}$$

Using the definition in Eq. S3, we can then rewrite the pairwise interaction terms in Eq. S1 as:

$$\sum_{k'} L^{\Delta_\pm}(t - t^i_{k'}) = \int_{-\infty}^{t} d\tau \, S^i(\tau) \, L^{\Delta_\pm}(t - \tau) = \int_0^{\infty} d\tau^* \; S^i(t - \tau^*) \, L^{\Delta_\pm}(\tau^*) \tag{S6}$$

Where in the last step we performed the change of variable $\tau \leftarrow t - \tau^*$. This leads to the following expression for the instantaneous change in synaptic weights, numerically equivalent to Eq. S1:

$$\begin{aligned} \text{on } \; t = t_k^{(\text{pre})} \; &\rightarrow \; \Delta W^{(\text{pre})}(t) = \alpha_{\text{pre}} + \int_0^{\infty} d\tau \, S^{\text{post}}(t - \tau) \, L^{\Delta_-}(\tau) \\ \text{on } \; t = t_k^{(\text{post})} \; &\rightarrow \; \Delta W^{(\text{post})}(t) = \alpha_{\text{post}} + \int_0^{\infty} d\tau \, S^{\text{pre}}(t - \tau) \, L^{\Delta_+}(\tau) \end{aligned} , \tag{S7}$$

Now, using again the same definition of spike train in Eq. S3, an event-based update can be expressed as an integral over time over the pre- and post-synaptic spiking events. The total weight change occurring in an interval $[t_a, t_b]$ is:

$$\Delta W_{[t_a,\,t_b]} = \int_{t_a}^{t_b} dt\; S^{\text{pre}}(t)\; \Delta W^{(\text{pre})}(t) + \int_{t_a}^{t_b} S^{\text{post}}(t)\; \Delta W^{(\text{post})}(t)\;. \tag{S8}$$

We can then express the average weight change per second, considering a time interval $[0, T]$ large enough to ensure precise sampling in second-order firing statistics, but still small compared to the rate of changes in ΔW. Under this assumption of separation of timescales between neuronal dynamics and plasticity, we can substitute Eq. S7 into Eq. S8, and write:

$$\begin{aligned}\langle \Delta W(t)\rangle_t = \frac{1}{T}\int_0^T dt\, S^{\text{pre}}(t)\;\alpha_{\text{pre}} + \frac{1}{T}\int_0^T dt\, S^{\text{post}}(t)\;\alpha_{\text{post}} + \\ + \frac{1}{T}\int_0^T dt \int_0^\infty d\tau\;\; S^{\text{post}}(t)\; S^{\text{pre}}(t-\tau)\, L^{\Delta_+}(\tau)\; + \frac{1}{T}\int_0^T dt \int_0^\infty d\tau\;\; S^{\text{pre}}(t)\; S^{\text{post}}(t-\tau)\, L^{\Delta_-}(\tau).\end{aligned} \tag{S9}$$

Since α_{pre} and α_{post} are constants, we can factor them out. The first two terms of the r.h.s. of Eq. S9 become:

$$\frac{1}{T}\int_0^T dt\, S^{\text{pre}}(t)\alpha_{\text{pre}} = r_{\text{pre}}\,\alpha_{\text{pre}} \quad \text{and} \quad \frac{1}{T}\int_0^T dt\, S^{\text{post}}(t)\alpha_{\text{post}} = r_{\text{post}}\,\alpha_{\text{post}} \tag{S10}$$

The pairwise interaction terms instead can be simplified as follows:

$$\frac{1}{T}\int_0^T dt \int_0^\infty d\tau\;\; S^{\text{post}}(t)\; S^{\text{pre}}(t-\tau)\, L^{\Delta_+}(\tau) = r_{\text{pre}}\, r_{\text{post}} \int_0^\infty d\tau\, L^{\Delta_+}(\tau) + \int_0^\infty d\tau\, C^{\Delta_+}(\tau) L^{\Delta_+}(\tau), \tag{S11}$$

and similarly for the last term in Eq. S9. In the above equation, $C^{\Delta_+}(\tau)$ denotes the covariance density between the pre- and the postsynaptic neuron, following the convention that Δ_+ represents the pre-post order and Δ_- the post-pre order:

$$C^{\Delta_+}(\tau) := \frac{1}{T}\int_0^T \left(S^{\text{post}}(t) - r_{\text{post}}\right)\left(S^{\text{pre}}(t-\tau) - r_{\text{pre}}\right)\;. \tag{S12}$$

An analogous definition holds for $C^{\Delta_-}(\tau)$, with the pre- and post- temporal order inverted. Therefore

$$\begin{aligned}\langle \Delta W(t)\rangle_t = r_{\text{pre}}\,\alpha_{\text{pre}} + r_{\text{post}}\,\alpha_{\text{post}} + r_{\text{pre}}\, r_{\text{post}} \int_0^\infty d\tau\, \left(L^{\Delta_-}(\tau) + L^{\Delta_+}(\tau)\right) + \\ + \int_0^\infty d\tau\, C^{\Delta_-}(\tau) L^{\Delta_-}(\tau) + \int_0^\infty d\tau\, C^{\Delta_+}(\tau) L^{\Delta_+}(\tau)\;.\end{aligned} \tag{S13}$$

Now using Eq. S2, we can write:

$$r_{\text{pre}}\, r_{\text{post}} \int_0^\infty d\tau\, \left(L^{\Delta_-}(\tau) + L^{\Delta_+}(\tau)\right) = r_{\text{pre}}\, r_{\text{post}} \int_{-\infty}^\infty d\tau\, L(\tau)\;; \tag{S14}$$

and given that $C^{\Delta_+}(\tau) = C^{\Delta_-}(-\tau)$ we can perform the following variable substitution:

$$\int_0^\infty d\tau\, C^{\Delta_-}(\tau) L^{\Delta_-}(\tau) = \int_{-\infty}^0 d\tau\, C^{\Delta_+}(\tau) L^{\Delta_+}(\tau) \tag{S15}$$

Finally, substituting Eq. S14 and Eq. S15 in Eq. S13, and defining $C(\tau) := C^{\Delta_+}(\tau)$, we have:

$$\begin{aligned}\langle \Delta W(t)\rangle_t = \alpha_{\text{pre}} r_{\text{pre}} + \alpha_{\text{post}} r_{\text{post}} + B\, r_{\text{pre}}\, r_{\text{post}} + Q \quad , \\ \text{with} \quad B := \int_{-\infty}^{+\infty} d\tau\, L(\tau) \quad \text{and} \quad Q := \int_{-\infty}^{+\infty} d\tau\, L(\tau)\; C(\tau)\;;\end{aligned} \tag{S16}$$

Corresponding to Eq. 1 in the main text.

S1.2 Derivation of the closed-form solution for the E/I motif

In this section, we derive a closed-form solution for a reduced model of one inhibitory and one excitatory neuron (Eq. S16). We model the two neurons as mutually interacting Hawkes processes (Main Text, Fig. 1A, Hawkes 1971; Pernice et al. 2011). The excitatory unit receives a constant input current h_{exc}, a plastic inhibitory connection $w_{\text{inh}}(t)$ and it projects a fixed excitatory connection w_{exc} to the inhibitory neuron. For simplicity, we assume that the inhibitory unit is kept at a constant rate r_{inh} by an adaptive external current. In this framework, the average firing rates behave linearly, therefore, the excitatory rate can be written as:

$$r_{\text{exc}}(t) = h_{\text{exc}} - w_{\text{inh}}(t)\, r_{\text{inh}} \ . \tag{S17}$$

As we focus on the connection $w_{\text{inh}}(t)$ from the inhibitory to the excitatory neuron, we rename $r_{\text{inh}} = r_{\text{pre}}$ and $r_{\text{exc}} = r_{\text{post}}$ in Eq. S16 and we get:

$$\begin{aligned}\left\langle \frac{dw_{\text{inh}}(t)}{dt} \right\rangle &= \alpha_{\text{pre}} r_{\text{inh}} + \alpha_{\text{post}} \left(h_{\text{exc}} - w_{\text{inh}} r_{\text{inh}}\right) + B\, r_{\text{inh}} \ \left(h_{\text{exc}} - w_{\text{inh}} r_{\text{inh}}\right) + Q \\ &= \alpha_{\text{pre}} r_{\text{inh}} - w_{\text{inh}} r_{\text{inh}} \left(\alpha_{\text{post}} + B\, r_{\text{inh}}\right) + h_{\text{exc}} \left(\alpha_{\text{post}} + B\, r_{\text{inh}}\right) + Q.\end{aligned} \tag{S18}$$

The only nontrivial quantity to compute is the term Q, as it requires the inhibitory-to-excitatory covariance density curve. To avoid ambiguities, we define the Fourier transform operator acting on a function $f(t)$ and the convolution operator acting on two functions $f(t)$ and $g(t)$ as follows:

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} dt \ \ f(t)\, e^{-j\,\omega\, t} \quad \text{and} \quad (f * g)(\tau) = \int_{-\infty}^{+\infty} dt \ \ f(t)\ g(\tau - t) \ ; \tag{S19}$$

where j denotes the imaginary unit.

Note that the total integral Q of the product $C(\tau)\, L(\tau)$ (as defined in Eq. S16) is equivalent to its Fourier transform evaluated at $\omega = 0$. But for the inverse convolution theorem, the Fourier transform of the product $C(\tau)\, L(\tau)$ corresponds to the convolution of the two transforms $\hat{C}(\omega)$ and $\hat{L}(\omega)$. In equations:

$$\begin{aligned}Q = \int_{-\infty}^{\infty} d\tau\, C(\tau)\, L(\tau) = \hat{f}\left(C(\tau)\, L(\tau)\right)(\omega)\Big|_{\omega=0} = \frac{1}{2\pi} \ \left(\hat{C}(\tilde{\omega}) * \hat{L}(\tilde{\omega})\right)(\omega)\Big|_{\omega=0} = \\ = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\tilde{\omega}\ \hat{C}(\tilde{\omega})\, \hat{L}(-\tilde{\omega})\, .\end{aligned} \tag{S20}$$

The advantage of re-defining Q in the Fourier space is that, for Hawkes processes, the covariance $\hat{C}(\omega)$ can be expressed in a closed form as a function of network connectivity and neural activity (Hawkes 1971; Pernice et al. 2011; Trousdale et al. 2012; Jovanović et al. 2015). More specifically, given the matrix of post-synaptic potential kernels (Eq. 9 in Methods)

$$\mathbf{E}(\tau) = \begin{pmatrix} \frac{1}{\tau_e}\, e^{-\tau/\tau_e} & 0 \\ 0 & -\frac{1}{\tau_i}\, e^{-\tau/\tau_i} \end{pmatrix} \tag{S21}$$

and the connectivity matrix

$$\mathbf{W} = \begin{pmatrix} 0 & w_{\text{inh}} \\ w_{\text{exc}} & 0 \end{pmatrix}, \tag{S22}$$

we can first write the vector of firing rates $\mathbf{r} = [r_{\text{exc}}\, ,\, r_{\text{inh}}]^{\top}$ as

$$\mathbf{r} = \mathbf{h_0} + \int_{-\infty}^{+\infty} d\tau \ \ \mathbf{G}(\tau)\, S(\tau - t), \tag{S23}$$

where $\mathbf{h_0}$ is the baseline input and $\mathbf{G}(\tau) = \mathbf{E}(\tau)\, \mathbf{W}$ is referred to as the matrix of "interaction kernels". At the equilibrium state,

$$\mathbf{r} = \mathbf{h_0} + \mathbf{G}\, \mathbf{r}\, , \quad \text{with } \mathbf{G} := \int_{-\infty}^{+\infty} d\tau \ \ \mathbf{G}(\tau), \tag{S24}$$

where the integral in the r.h.s. corresponds to the Fourier transform evaluated at zero, i.e., $\hat{\mathbf{G}}(\omega)|_{\omega=0}$. Equivalently, the rate vector can be written in an explicit form:

$$\mathbf{r} = (\mathbf{I} - \mathbf{G})^{-1}\,\mathbf{h_0}, \tag{S25}$$

where $\mathbf{I}$ represents the identity matrix. Moreover, if we write the covariance matrix (cf. Eq. S12) as

$$\hat{\mathbf{C}}(\tau) = \langle \mathbf{S}(t)\,\mathbf{S}(t-\tau)^T\rangle - \mathbf{r}\,\mathbf{r}^T \tag{S26}$$

then we have the following expression in the Fourier domain, at the equilibrium state:

$$\hat{\mathbf{C}}(\omega) = \left(\mathbf{I} - \hat{\mathbf{G}}(\omega)\right)^{-1}\,\mathbf{D}\,\left(\mathbf{I} - \hat{\mathbf{G}}^T(\omega)\right)^{-1} \tag{S27}$$

where $\mathbf{D}$ is a diagonal matrix with the firing rates on its diagonal. Finally, recall that, for a matrix $\mathbf{G}$ with spectral radius < 1,

$$(\mathbf{I} - \mathbf{G})^{-1} = \sum_{\alpha}^{\infty} \mathbf{G}^{\alpha}. \tag{S28}$$

Therefore, the covariance matrix can be written as the following expansion

$$\hat{\mathbf{C}}(\omega) = \sum_{\alpha=0}^{N}\sum_{\beta=0}^{N} \hat{\mathbf{E}}^{\alpha}(\omega)(\hat{\mathbf{E}}^T)^{\beta}(-\omega)\,\left[(\mathbf{W}^{\alpha})\mathbf{D}(\mathbf{W}^T)^{\beta})\right]\,, \tag{S29}$$

where the indices α, β indicate the number of connections between any spiking neuron in the network to the post- and to the presynaptic neuron, respectively. This captures the influence of each spike propagated in the network through existing synaptic connections on the correlated firing of the pre- and postsynaptic neurons. Within this formalism, in our reduced model we consider only pairwise interactions, that is, $\alpha + \beta = 1$. The inhibitory-to-excitatory covariance is the $(1, 2)$ element of $\hat{\mathbf{C}}(\omega)$. We can compute it as:

$$\hat{C}(\omega) = \hat{E}_e(\omega)\,w_{\text{inh}}\,r_{\text{inh}} + \hat{E}_i(-\omega)\,r_{\text{exc}}\,w_{\text{exc}} \tag{S30}$$

By combining Eq. S20 and Eq. S30, we get the following expression for Q:

$$\begin{aligned} Q &= M_{ei}^{(10)} w_{\text{inh}}\,r_{\text{inh}} + M_{ie}^{(01)} w_{\text{exc}}\,r_{\text{exc}} \\ \text{with} \quad M_{jk}^{(\alpha\beta)} &:= \int_{-\infty}^{+\infty} d\omega\,\hat{E}_j^{\alpha}(\omega)\hat{E}_k^{\beta}(-\omega)\hat{L}(-\omega). \end{aligned} \tag{S31}$$

The $M_{ei}^{(10)}$, $M_{ie}^{(01)}$ coefficients can be computed analytically depending on the iSTDP rule (see next section). Considering them as constants, we can then write Eq. S18 as:

$$\begin{aligned} \left\langle \frac{dw_{\text{inh}}(t)}{dt} \right\rangle = \alpha_{\text{pre}} r_{\text{inh}} - w_{\text{inh}}\,r_{\text{inh}}\left(\alpha_{\text{post}} + B\,r_{\text{inh}}\right) + h_{\text{exc}}\,\left(\alpha_{\text{post}} + B\,r_{\text{inh}}\right) + \\ M_{ei}^{(10)} w_{\text{inh}}\,r_{\text{inh}} + M_{ie}^{(01)} w_{\text{exc}}\,r_{\text{exc}} \end{aligned} \tag{S32}$$

or equivalently,

$$\left\langle \frac{dw_{\text{inh}}(t)}{dt} \right\rangle = \alpha_{\text{pre}} r_{\text{inh}} - w_{\text{inh}}\,r_{\text{inh}}\left(\Psi - M_{ei}^{(10)}\right) + h_{\text{exc}}\,\Psi \tag{S33}$$

with $\Psi := \alpha_{\text{post}} + B r_{\text{inh}} + M_{ie}^{(01)} w_{\text{exc}}$.

We then find the stationary point (Eq. 10 in Methods)

$$\tilde{w}_{\text{inh}} = \frac{\alpha_{\text{pre}}\,r_{\text{inh}} + h_{\text{exc}}\Psi}{r_{\text{inh}}\left(\Psi - M_{ei}^{(10)}\right)} \tag{S34}$$

The equilibrium point $\tilde{w}_{\text{inh}}$ is a stable fixed point attractor, and positive, that is $\tilde{w}_{\text{inh}} \in [0, w_{\text{max}}]$, if the following conditions hold:

$$\begin{aligned} \Psi - M_{ei}^{(10)} &> 0 \\ \alpha_{\text{pre}}\,r_{\text{inh}} + \Psi\,h_{\text{exc}} &> 0, \end{aligned} \tag{S35}$$

If a solution for $\tilde{w}_{\text{inh}}$ exists, we can also express the steady-state postsynaptic rate as follows:

$$\tilde{r}_{\text{exc}} = h_{\text{exc}} - \tilde{w}_{\text{inh}}\, r_{\text{inh}} = -\frac{\alpha_{\text{pre}} r_{\text{inh}} + M_{ei}^{(10)} h_{\text{post}}}{\Psi - M_{ei}^{(10)}} \quad , \tag{S36}$$

The solution is valid only for positive rates. This adds one more condition for a stable and positive solution. The constraints are therefore:

$$\begin{aligned} \Psi - M_{ei}^{(10)} &> 0 \\ \alpha_{\text{pre}}\, r_{\text{inh}} + \Psi\, h_{\text{exc}} &> 0 \\ r_{\text{inh}} &> \frac{M_{ei}^{(10)}\, h_{\text{exc}}}{-\alpha_{\text{pre}}} \quad \text{with } \alpha_{\text{pre}} < 0. \end{aligned} \tag{S37}$$

S1.3 Derivation of first-order motif coefficients

To calculate the motif coefficients

$$M_{ei}^{(10)} = \int_{-\infty}^{+\infty} d\omega\, \hat{L}(-\omega)\, \hat{E}_i(\omega) \quad \text{and} \quad M_{ie}^{(01)} = \int_{-\infty}^{+\infty} d\omega\, \hat{L}(-\omega)\, \hat{E}_e(-\omega) \tag{S38}$$

for the connection W, we need to compute the Fourier transform of the postsynaptic potential E_e, E_i and the Fourier transform of the iSTDP kernel.

The Fourier transforms of the excitatory and inhibitory postsynaptic potential (Eq. 9 in Methods) are respectively given by:

$$\hat{E}_e(\omega) = \frac{1}{1 + j\tau_e\omega}, \quad \text{and} \quad \hat{E}_i(\omega) = -\frac{1}{1 + j\tau_i\omega}, \tag{S39}$$

where j is the imaginary unit.

Consider first the Hebbian symmetric STDP kernel, defined via the function

$$L^{\text{sy}}(\tau) = \frac{1}{2}\left(\frac{1}{\tau_+}\, A_+\, e^{-|\tau|/\tau_+} - \frac{1}{\tau_-}\, A_-\, e^{|\tau|/\tau_-}\right) \tag{S40}$$

with A_+, $A_- > 0$. The Fourier transform is given by

$$\hat{L}^{\text{sy}}(\omega) = \hat{L}^{\text{sy}}(-\omega) = \frac{A_+}{1 + \tau_+^2\omega^2} - \frac{A_-}{1 + \tau_-^2\omega^2}. \tag{S41}$$

By substituting Eq. S39 and Eq. S41 in Eq. S38, we can calculate the first-order motif coefficients $M_{ei}^{\text{sy},(10)}$, $M_{ie}^{\text{sy},(01)}$:

$$M_{ei}^{\text{sy},(10)} = \frac{A_+}{\tau_+ + \tau_i} - \frac{A_-}{\tau_- + \tau_i} \quad ; \quad M_{ie}^{\text{sy},(01)} = \frac{A_+}{\tau_+ + \tau_e} - \frac{A_-}{\tau_- + \tau_e}, \tag{S42}$$

and therefore find an analytical solution for the integral Q in Eq. S31.

Consider now the antisymmetric iSTDP rule, with kernel:

$$L^{\text{an}}(\tau) = \begin{cases} -\frac{A_-}{\tau_-}\, e^{\tau/\tau_-} & \text{if } \tau < 0 \\ \frac{A_+}{\tau_+}\, e^{-\tau/\tau_+} & \text{if } \tau \geq 0 \end{cases} \tag{S43}$$

whose Fourier transform is given by

$$\hat{L}^{\text{an}}(\omega) = \frac{A_+}{1 + j\tau_+\omega} - \frac{A_-}{1 - j\tau_-\omega}. \tag{S44}$$

Again, we can calculate the first order motif coefficients $M_{ei}^{\text{an},(10)}$, $M_{ie}^{\text{an},(01)}$ for the antisymmetric rule by substituting Eq. S39 and Eq. S44 in Eq. S38:

$$M_{ei}^{\text{an},(10)} = \frac{A_+}{\tau_+ + \tau_i} \quad ; \quad M_{ie}^{\text{an},(01)} = -\frac{A_-}{\tau_- + \tau_e}. \tag{S45}$$

and find an analytical solution for Eq. S31.

S2 Supporting Figures

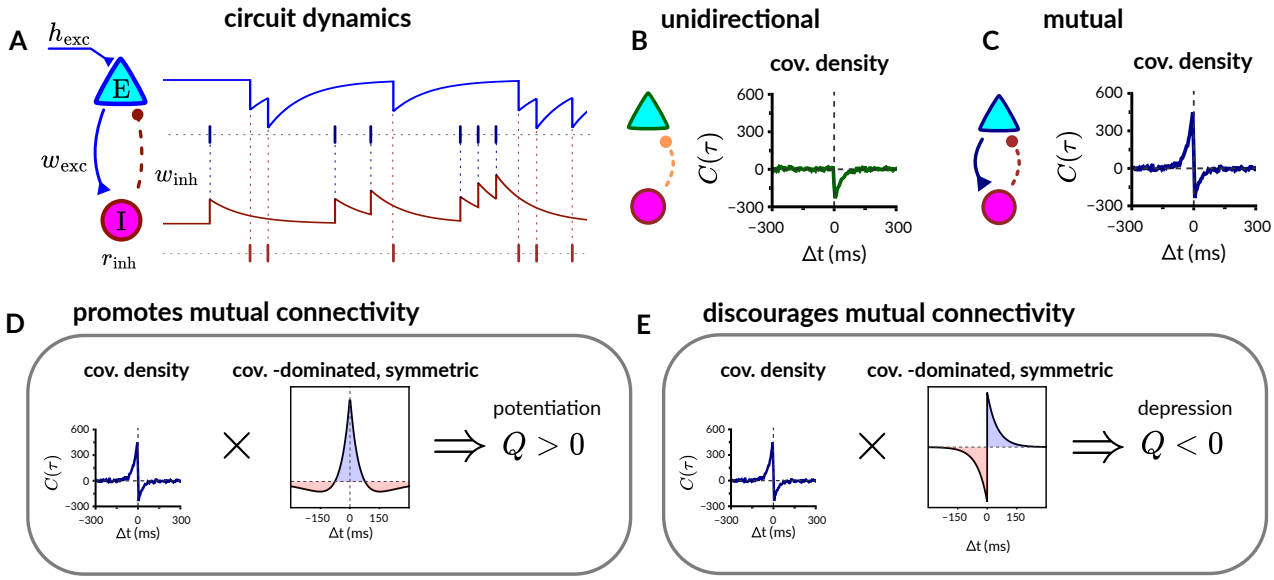

Figure S1: Schematics of the reduced circuit model and of interactions between covariance density and iSTDP kernel. **A**. Representation of the model used for Fig. 1 in main text. Two time-varying coupled Poisson units, one excitatory and one inhibitory, are connected by a non-plastic excitatory weight w_{exc} and a plastic inhibitory weight w_{inh}. The instantaneous rates, here shown as traces, are computed as the convolution between each incoming spike and an exponential synaptic kernel that either increases (exc) or decreases (inh) the instantaneous firing probability. The input to the excitatory neuron h_{exc} is fixed, whereas the inhibitory input changes dynamically, keeping the inhibitory neuron at a fixed rate r_{inh}. Only the inhibitory synapse (dashed line) is plastic. **B**. Unidirectional configuration. Here $w_{\text{exc}} = 0$ and the covariance density between the pre (inh) and post (exc) neuron can only be negative in the pre-post ($\Delta t > 0$) section of the curve, due to the inhibitory interaction given by w_{inh}. **C**. Mutual configuration, $w_{\text{exc}} = 0.5$. Here the covariance density is positive in the post-pre ($\Delta t < 0$) section of the curve due to w_{exc}, and negative for ($\Delta t > 0$) for the effect of w_{inh}. **D**. Effect of a symmetric, covariance-dominated iSTDP kernel on a circuit in the M configuration: the positive left branch of the covariance density interacts with the kernel, resulting in a positive Q factor in the weight update. (Eq. 1). **E**. Effect of an antisymmetric iSTDP kernel. Here the positive covariance density interacts with the negative branch of the kernel, resulting into synaptic depression. Hence, the Q term is negative, and mutual connections are discouraged.

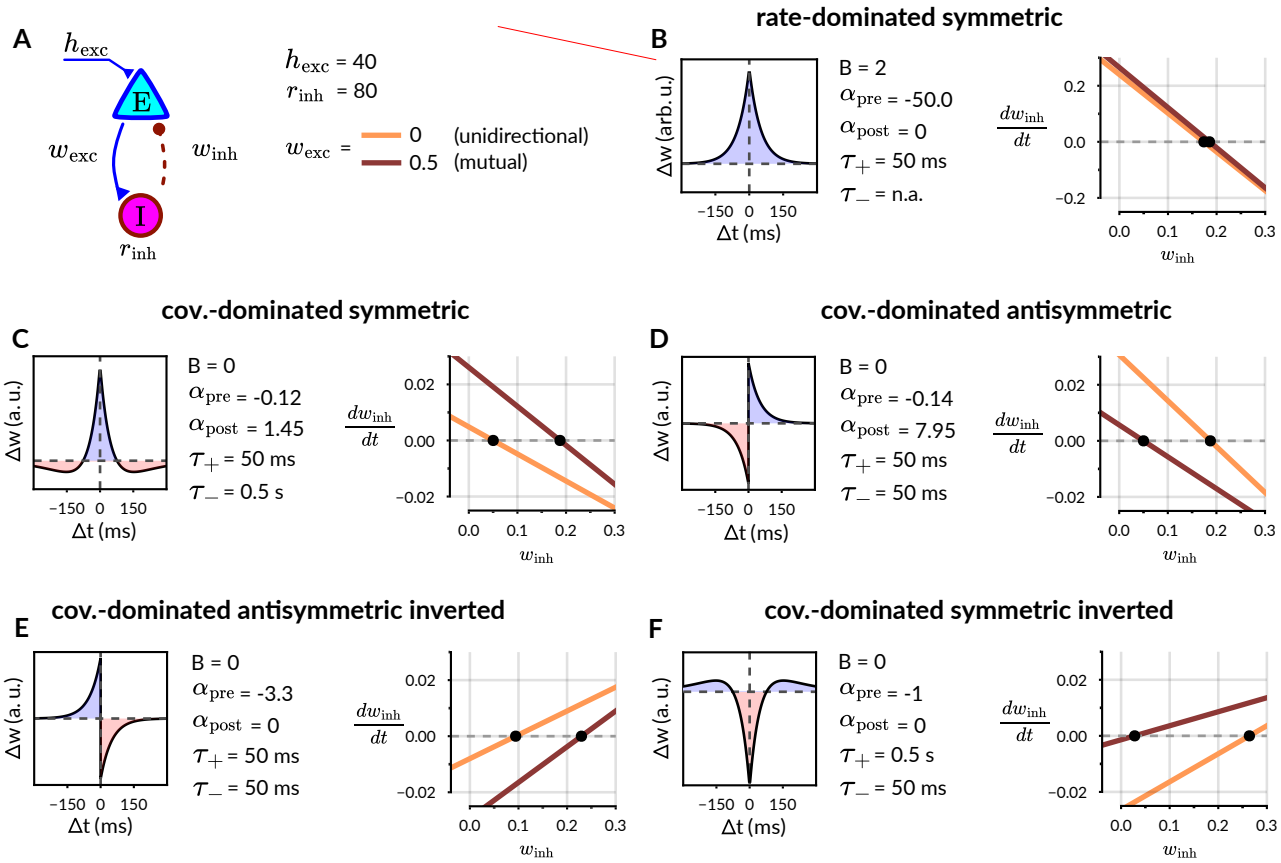

Figure S2: Analytic table solutions for different iSTDP interaction kernels (complements Fig. 2 in main text). **A**. Representation of the model (see also Fig. 1A). **B-E**. iSTDP rules, represented by their kernel and their additional parameters (left panels) and study of the associated weight dynamics as a one-dimensional differential equation (right panels), black dots represent fixed points. **B** Rate-dominated iSTDP rule. **C-E**. covariance dominated rules presented in the same order as in Fig. 2A, main text. Note that the configurations in **E** and **F** produce unstable fixed points.

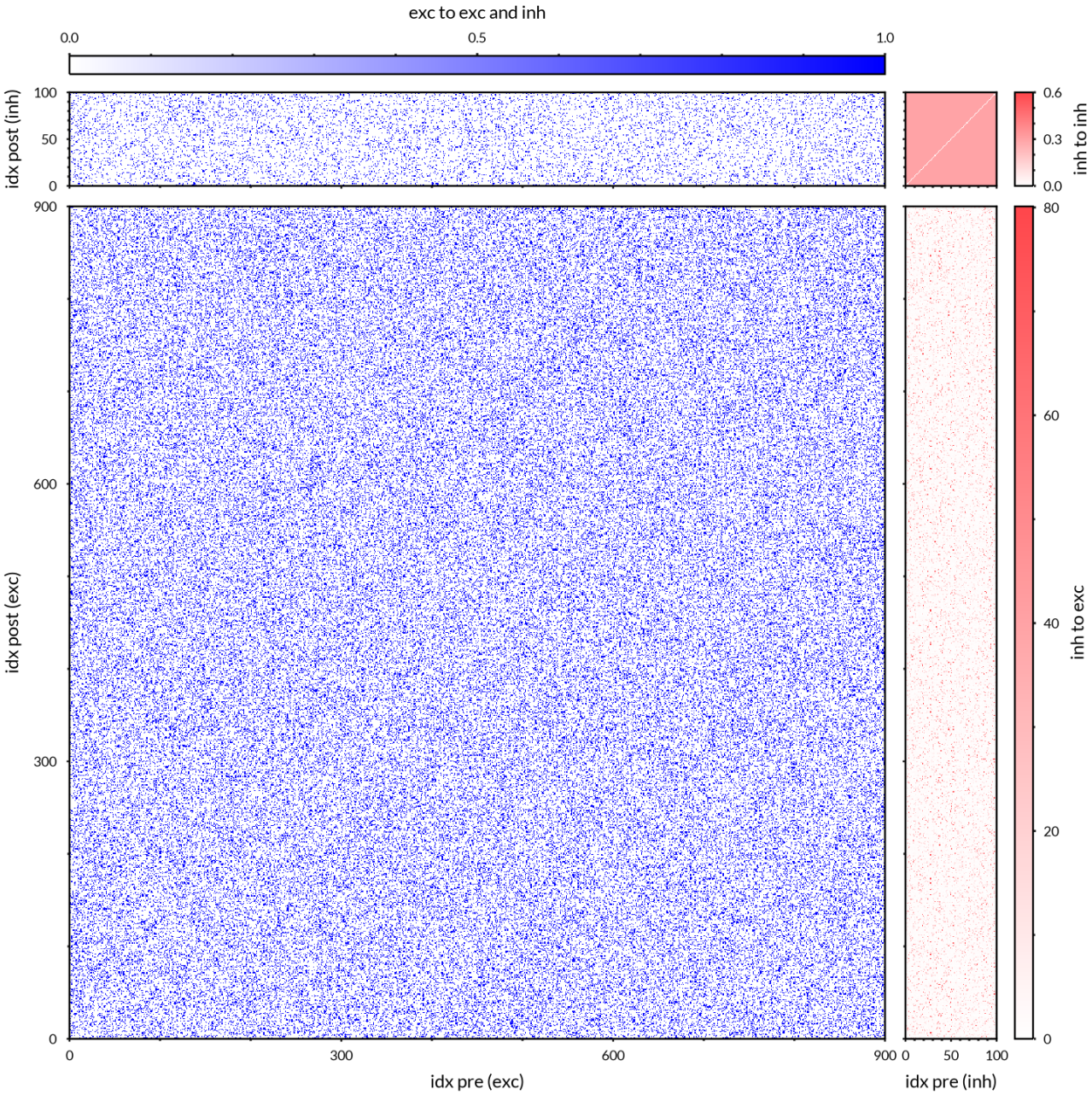

Figure S3: Full weight matrix for the random sparse RNN simulation with symmetric iSTDP (see Main Text, Fig. 3B-F).

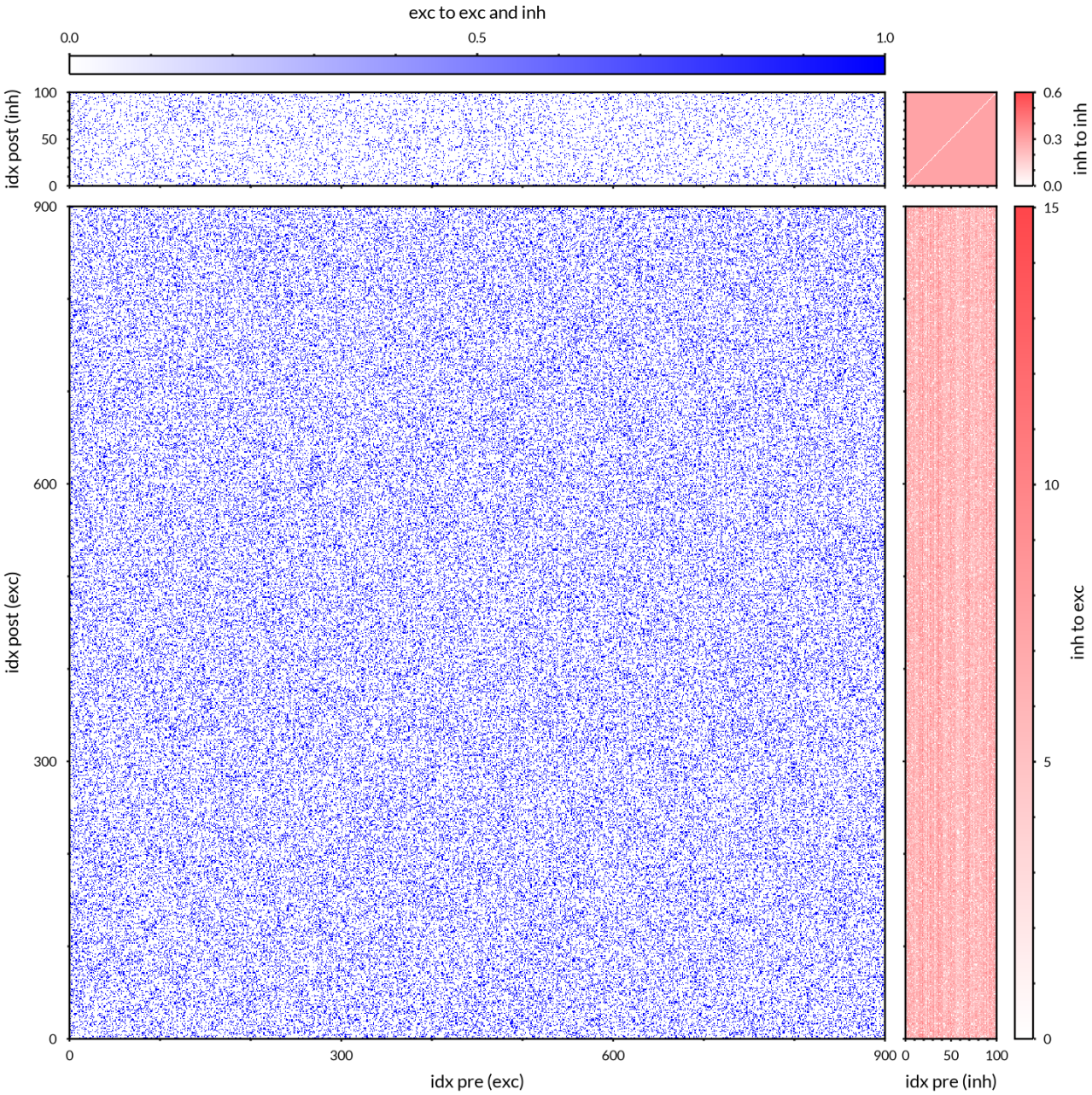

Figure S4: Full weight matrix for the random sparse RNN simulation with antisymmetric iSTDP (see Main Text, Fig. 3G-K).

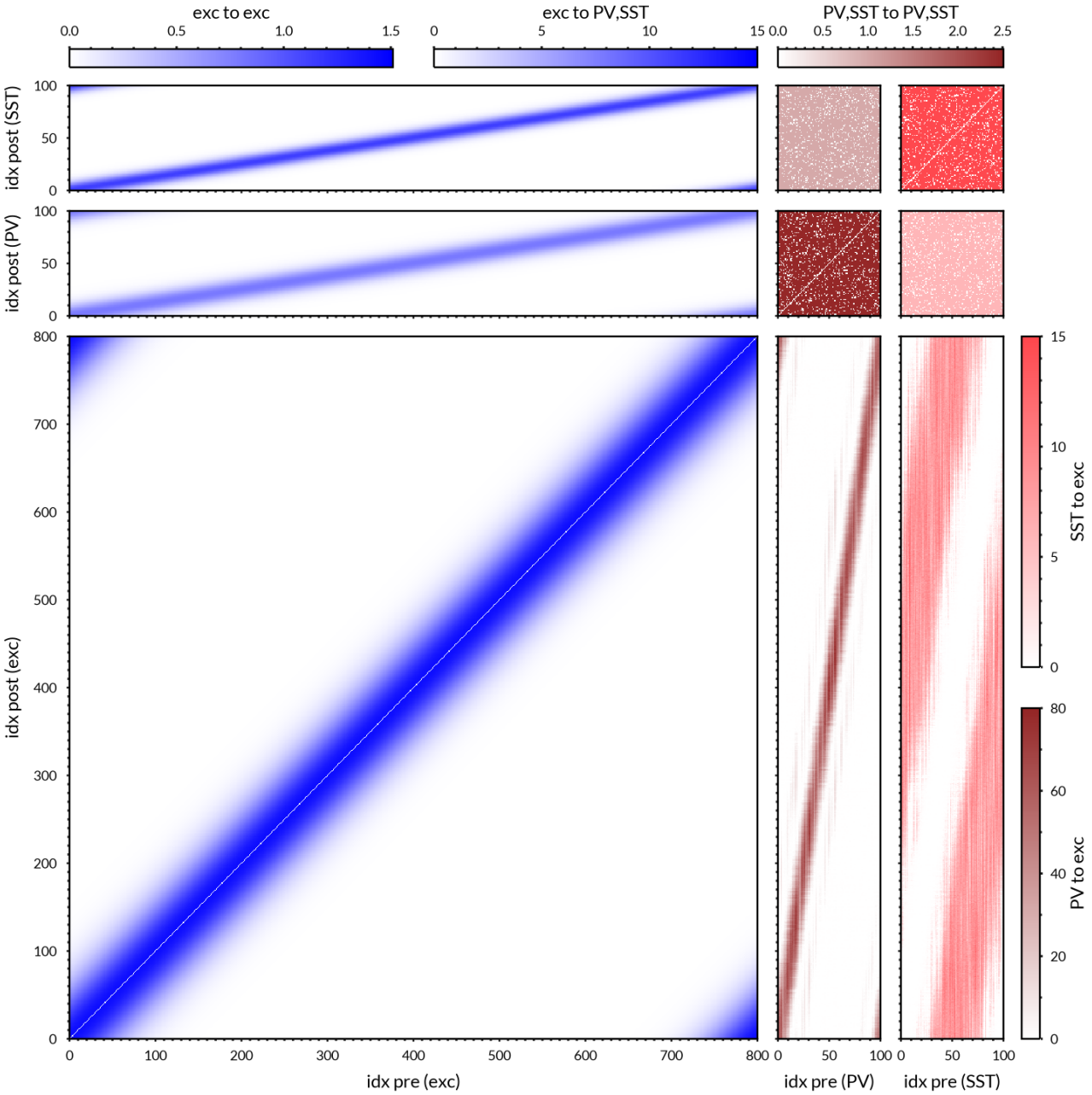

Figure S5: Full weight matrix for the recurrent network with 1D ring connectivity structure after weight convergence with two inhibitory populations.

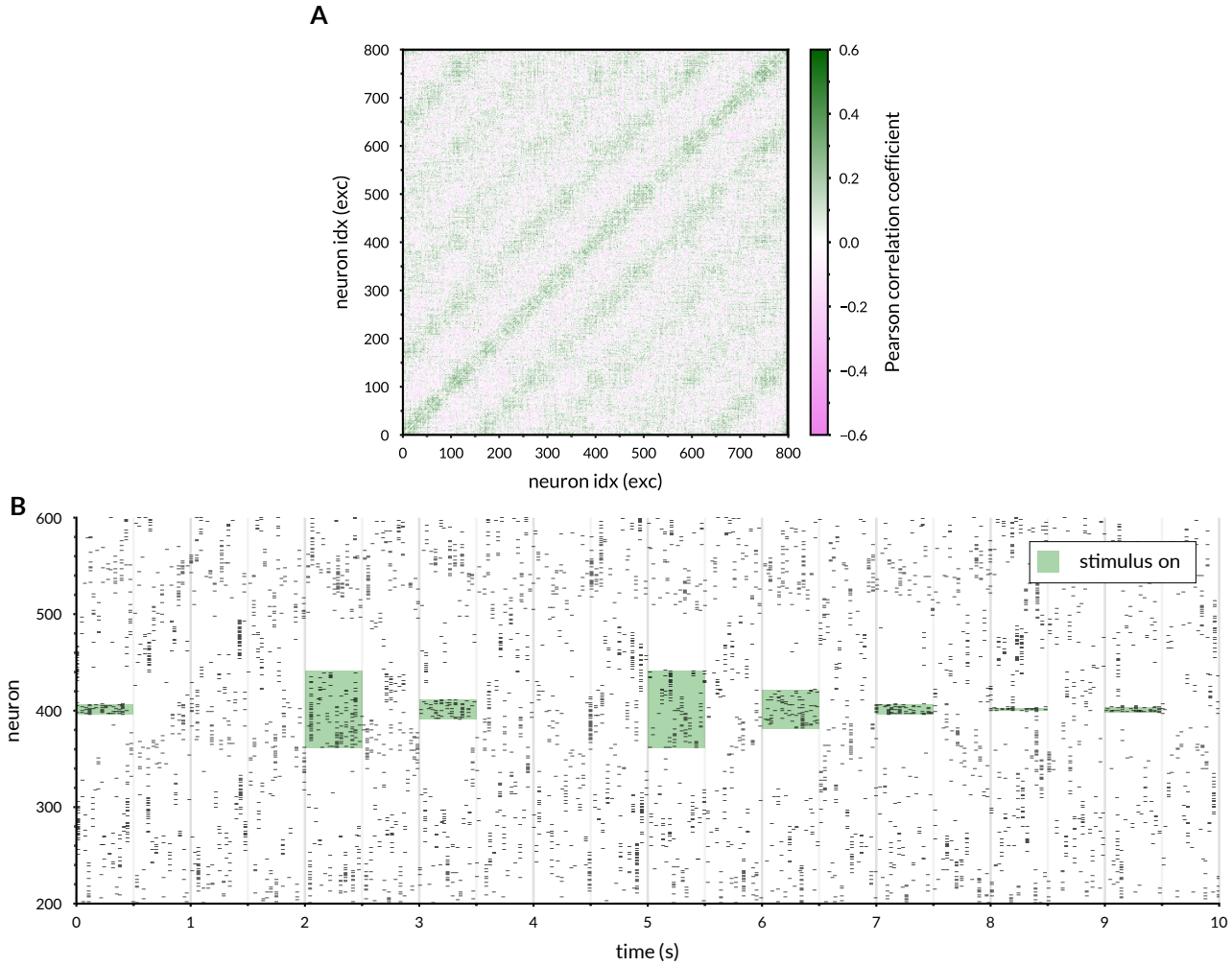

Figure S6: **A**. Full matrix of Pearson correlations between excitatory neurons in the one-dimensional ring model during spontaneous activity, after convergence of plastic weights. **B** Spike raster showing the response of the 1D ring model to localized stimuli (see Fig. 4 in main text).

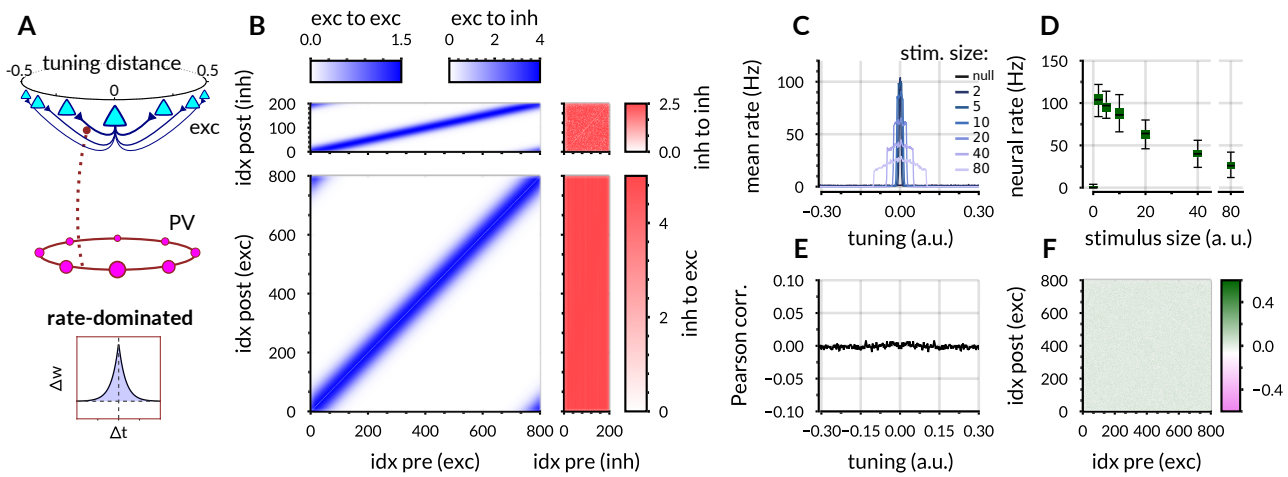

Figure S7: Spontaneous and evoked activity in a 1D ring structure with rate-dominated iSTDP. **A**. Model schematics, compare with Fig. 4A. The excitatory component is identical to the previous case, but the inhibitory population follows a single, rate-dominated iSTDP. **B**. Full weight matrix after weight convergence, compare with Fig. 4B. **C**. Mean responses of the excitatory population to stimuli of different sizes, averaged over 100 trials (compare withFig. 5A). **D**. Response of the center neuron to stimuli of different sizes (compare withFig. 5B). **E**. Pearson correlation during spontaneous activity between an excitatory neuron and its neighbors, ordered by tuning distance and averaged over the entire population (compare withFig. 4F). **F**. Pearson correlation of spontaneous activity for every pair of excitatory neurons (compare withFig. S6A).

References for Supporting Information

Hawkes, Alan G. (1971). "Spectra of Some Self-Exciting and Mutually Exciting Point Processes". In: *Biometrika* 58.1, pp. 83–90. DOI: 10.2307/2334319.

Jovanović, Stojan, John Hertz, and Stefan Rotter (Apr. 7, 2015). "Cumulants of Hawkes Point Processes". In: *Physical Review E* 91.4, p. 042802. DOI: 10.1103/PhysRevE.91.042802.

Kempter, Richard, Wulfram Gerstner, and J. Leo van Hemmen (Apr. 1, 1999). "Hebbian Learning and Spiking Neurons". In: *Physical Review E* 59.4, pp. 4498–4514. DOI: 10.1103/PhysRevE.59.4498.

Pernice, Volker, Benjamin Staude, Stefano Cardanobile, and Stefan Rotter (May 19, 2011). "How Structure Determines Correlations in Neuronal Networks". In: *PLOS Computational Biology* 7.5, e1002059. DOI: 10.1371/journal.pcbi.1002059.

Trousdale, James, Yu Hu, Eric Shea-Brown, and Krešimir Josić (Mar. 22, 2012). "Impact of Network Structure and Cellular Response on Spike Time Correlations". In: *PLoS Comput Biol* 8.3, e1002408. DOI: 10.1371/journal.pcbi.1002408.