Assembly formation is stabilized by Parvalbumin neurons and accelerated by Somatostatin neurons

Fereshteh Lagzi^{1,3,4,6}, Martha Canto Bustos^{2,3}, Anne-Marie Oswald^{2,3,4} & Brent Doiron^{1,3,4,5,6}

¹Department of Mathematics, University of Pittsburgh, Pittsburgh, PA, USA

²Department of Neuroscience, University of Pittsburgh, Pittsburgh, PA, USA

³Center for the Neural Basis of Cognition, Pittsburgh, PA, USA

⁴Department of Neurobiology, University of Chicago, Chicago, IL, USA

⁵Department of Statistics, University of Chicago, Chicago, IL, USA

⁶Grossman Center for Quantitative Biology and Human Behavior, University of Chicago, Chicago, IL, USA

Abstract

Learning entails preserving the features of the external world in the neuronal representations of the brain, and manifests itself in the form of strengthened interactions between neurons within assemblies. Hebbian synaptic plasticity is thought to be one mechanism by which correlations in spiking promote assembly formation during learning. While spike timing dependent plasticity (STDP) rules for excitatory synapses have been well characterized, inhibitory STDP rules remain incomplete, particularly with respect to sub-classes of inhibitory interneurons. Here, we report that in layer 2/3 of the orbitofrontal cortex of mice, inhibition from parvalbumin (PV) interneurons onto excitatory (E) neurons follows a symmetric STDP function and mediates homeostasis in E-neuron firing rates. However, inhibition from somatostatin (SOM) interneurons follows an asymmetric, Hebbian STDP rule. We incorporate these findings in both large scale simulations and mean-field models to investigate how these differences in plasticity impact network dynamics and assembly formation. We find that plasticity of SOM inhibition builds lateral inhibitory connections and increases competition between assemblies. This is reflected in amplified correlations between neurons within assembly and anti-correlations between assemblies. An additional finding is that the emergence of tuned PV inhibition depends on the interaction between SOM and PV STDP rules. Altogether, we show that incorporation of differential inhibitory STDP rules promotes assembly formation through competition, while enhanced inhibition both within and between assemblies protects new representations from degradation after the training input is removed.

Introduction

A fundamental link between systems and circuits neuroscience is that the recurrent wiring of neuronal networks is functionally organized [1], and can reflect sensory experience [2]. For example, excitatory neurons in the visual cortex show strong reciprocal wiring when neuron pairs have similar stimulus tuning [3, 4, 5, 6], and neurons in the hippocampus form groups of co-activated assemblies that may be used in memory formation [7, 8]. A group of neurons with strong recurrent wiring and shared stimulus or task selectivity is often termed a *neuronal assembly* [9]. A signature of strong assembly structure is that the dynamic and trial-to-trial variability in the joint spiking activity of neuron pairs also reflects their functional structure – neuron pairs within a common assembly show correlated fluctuations in their activity when compared to neuron pairs that are members of distinct assemblies [10, 5, 11, 12]. Thus, tracking how correlated variability across a neuronal population changes with repeated experiences offers a window into the learning mechanisms which carve assembly structure within a network.

Assembly wiring can be learned through an unsupervised Hebbian plasticity rule between excitatory neurons, whereby co-activated neurons form strong reciprocal synaptic connections [13, 11, 12, 14]. However, such strong recurrent excitatory connections can in principle destabilize network activity through unchecked positive feedback. In cortex such network instabilities are alleviated by having strong recurrent excitation dynamically tracked and

balanced by an opposing strong recurrent inhibition [15, 16, 17, 18, 19, 20]. In networks that have unstructured wiring such balanced excitation and inhibition promote an asynchronous network state [21, 22, 23, 24, 25]. However, when wiring structure is embedded within an excitatory network, the resulting network dynamics are heavily shaped by how any structure in the inhibitory wiring relates to the excitatory wiring [25, 26, 27, 28, 29, 30]. The specifics of structured inhibition present distinct problems for assembly learning within the excitatory network, as we introduce with a simple motivating example.

Figure 1: Stability versus competition. **A**: Schematic of a network with bulk inhibition. The scalars w and g change the relative strength of the weights within the excitatory assemblies, and the projection weights from the inhibitory population onto the excitatory assemblies. **B**: Schematic of a network with lateral inhibition. The inhibitory weights which form lateral inhibition are 3 times as strong as the weights from the inhibitory populations to the excitatory assemblies. **C₁**, **D₁**: Raster plots of the networks in **A** and **B** for $g = 6$ and $w = 1$. **C₂**, **D₂**: Raster plots of the networks in **A** and **B** for $g = 6$ and $w = 3.5$. There is a stronger competition in the network with lateral inhibition. **C₃**, **D₃**: Raster plots of the networks in **A** and **B** for $g = 6$ and $w = 4.5$. The network with lateral inhibition demonstrates episodes of synchronous regular dynamics due to lack of stability in this regime.

Consider a network with a baseline excitatory strength J and an embedded pair of assemblies (E1 and E2) where the within assembly excitatory strength is wJ ($w > 1$ reflects a Hebbian assembly; Figure 1). We first consider structured recurrent inhibition whereby each excitatory assembly receives strong inhibition from a paired inhibitory assembly (I1 and I2; Figure 1A). In this case as w is increased the recurrent inhibition maintains approximate asynchronous spiking dynamics (Figure 1C1 and C2). Thus, while network stability is maintained it now comes at a price – the assembly structure ($w > 1$) is not reflected in the spiking dynamics of the network i.e competition between assemblies is weak. The opposite structure in inhibitory wiring would be lateral connections that mediate competition between the excitatory assemblies (Figure 1B). In this network as w is increased a clear competition between the assemblies manifests (Figure 1D1 and D2). This dynamic certainly reflects the excitatory assembly structure, yet now with the price of pathological behavior when w is large, where each assembly either shows high activity with very coordinated spiking or near complete silence (Figure 1D2). In total, this exercise exposes a concrete dilemma. Inhibition can either significantly mask or greatly exaggerate the dynamical signature of embedded excitatory assemblies. This contrasts with the reality that structured excitatory wiring is clearly, yet moderately, reflected in the network dynamics [5, 31, 32]. The central goal of this study is to uncover how real cortical networks use inhibition to ensure network stability yet permit reasonable within assembly correlated

activity and between assembly competition.

It is now well established that inhibition is diverse within most neuronal networks, both in terms of its cellular structure and position within the cortical circuit [33]. Over the past decade a popular hypothesis has emerged whereby this diversity allows distinct inhibitory classes to participate in distinct cortical functions such as providing stability and controlling population-wide correlated activity [34, 35, 36, 37]. In a parallel series of studies inhibition has also been shown to be plastic [38, 39, 40, 41, 42, 43, 44, 45], and whose activity dependent learning mirrors the Hebbian learning in excitatory networks. In our work we explore the hypothesis that cell specific inhibitory plasticity onto excitatory neurons allows a diverse inhibitory network to simultaneously confer both stability and assembly based network competition. In an *in vitro* preparation of mouse orbital frontal cortex (OFC) we use a spike timing dependent plasticity (STDP) protocol and uncover that parvalbumin-expressing (PV) interneurons have a homeostatic and symmetric STDP learning rule, while somatostatin-expressing (SOM) interneurons have a causal Hebbian learning STDP rule. Armed with these inhibitory STDP rules we use a combination of simulations of a model cortical network and an associated mean-field theory to show how a cortical network with PV and SOM neurons support the rapid training of stable excitatory assemblies whose structure is reflected in the ongoing spiking dynamics. In particular, the different inhibitory plasticity rules result in a division of labor: the PV network confers a homeostatic stability of the network dynamics (Figure 1A), thereby permitting the SOM inhibitory network to support cross assembly competition (Figure 1B), however, SOM populations provide competition between excitatory assemblies as a result of the formation of lateral inhibition. In addition to this, we also show that SOM neurons could also potentially result in a subtle emergence of tuned PV to excitatory connections, by correlating PV and excitatory neurons given the lateral connections to the excitatory subnetwork and heterogeneous connections to PV neurons. Our work thus provides an important mechanistic insight into how diverse inhibitory learning supports the stable creation of learned excitatory networks.

Results

Experimental findings

Previous studies have established the importance of the temporal window between synaptic input and evoked action potentials in the modification of excitatory synaptic strength[46]. However, fewer studies have investigated the STDP of inhibitory synapses[43, 39, 47] particularly between specific subclasses of interneurons and excitatory pyramidal neurons (PNs). We expressed channelrhodopsin (ChR2) in PV or SOM interneurons then investigated the STDP of light-evoked inhibition onto PNs in L2/3 of orbitofrontal cortex (OFC). Briefly, STDP was induced by pairing light-evoked inhibitory post synaptic potentials (IPSPs) with evoked action potentials in the PN (See Methods, Figure 2A). For each recorded neuron, inhibition was evoked at a single temporal interval defined as the difference between the time of the AP (t_{spike}) and the time of the IPSP (t_{IPSP}) (Figure 2A1). Negative interval values correspond to AP's arriving prior to the IPSP, while positive values correspond to APs following the IPSP (Figure 2A2). To assess the time course of synaptic plasticity, baseline IPSCs (Figure 2B1, C1) were collected in voltage clamp for 5 min prior to the STDP pairing protocol and test IPSCs were collected for up to 1 hour following (Figure 2B2,C2). Synaptic plasticity was quantified as the average test IPSC amplitude 25-30 min post induction (gray shading, Figure 2B2,C2) normalized to baseline IPSC amplitude.

We find that inhibitory STDP rules differ between PV and SOM interneurons. Inhibition mediated by PV interneurons is significantly potentiated for short intervals within ± 30 ms of the post-synaptic action potential (red shaded regions, Figure 2B3) and declines as intervals increase from ± 30 -60 ms (blue shaded regions). We binned normalized amplitude in 30 ms bins and found that synaptic potentiation was significantly >1 for short-intervals (-30-0 ms: 1.79 ± 0.69 , $p=0.004$, $n=11$; 0-30 ms: 1.60 ± 0.62 , $p=0.010$, $n=10$, single distribution t-test, Figure 2B4). This relationship was symmetric with respect to the timing of IPSPs pre versus post the AP as synaptic enhancement did not significantly differ for positive versus negative intervals ($p=0.532$, unpaired t-test; Figure 2B4). Conversely, the temporal dependence for STDP of SOM inhibition was asymmetric. Inhibition was significantly potentiated (Norm. Amp. >1) for positive intervals between 0-50 ms when inhibition preceded spiking (1.66 ± 0.44 , $p=0.001$, $n=10$, Figure 2C3,C4). Whereas inhibition was significantly depressed (Norm. Amp. <1) for negative intervals (-50-0 ms: 0.78 ± 0.22 , $p=0.004$, $n=13$). Synaptic plasticity significantly differed between positive and negative intervals <50 ms ($p=1.5E-5$, unpaired t-test, Figure 2C4).

Figure 2: Inhibitory STDP differs between PV and SOM interneurons **A:** Schematic of STDP pairing protocol. Action potentials (AP) are evoked in the post-synaptic pyramidal neurons (E_{Cell}) every 4 seconds (0.25 Hz) (a1,a2 red) while light evoked IPSPs are evoked at a constant interval (ms) relative to the AP (a2,black). **B:** Inhibitory STDP from PV interneurons is symmetric. B1: Light evokes a single AP in a Chr2+ presynaptic PV interneuron (blue trace). IPSCs (red trace) are recorded in the post-synaptic E_{Cell} . B2: Baseline IPSCs are collected 5 min prior to the STDP protocol (time =0) and up to 60 min post pairing. IPSC amplitude is normalized to the average baseline amplitude. B3: Normalized IPSC amplitude recorded 25-30 min post STDP induction (B2, gray shaded) plotted against temporal window for STDP pairing (ms). Negative intervals correspond to post-synaptic spikes before the IPSP (open circles), positive intervals correspond to IPSPs before the spike (closed circles). B4: Normalized amplitude was binned in 30 ms intervals for comparison between positive and negative intervals. Intervals between 0 and +/- 30m ms show significant potentiation, * $p < 0.05$, but do not significantly differ between positive and negative intervals. **C:** Inhibitory STDP for SOM interneurons is asymmetric. C1-C3: As for b1-b3. C4) Normalized amplitude was binned in 50 ms intervals for comparison between positive and negative intervals. Short positive intervals 0-50 ms show significant potentiation (** $p < 0.01$) whereas short negative intervals -50-0 ms show significant depression (** $p < 0.01$). Short intervals significantly differ between positive and negative intervals.

Correlated population spiking activity induced by SOM and PV plasticity

To explore the emerging network properties as a result of PV and SOM plasticity, we include the learning rules of these inhibitory subtypes in networks of model neurons with spiking dynamics (leaky integrate-and-fire; see Methods). We model assembly formation by separating the excitatory neuron (E) into two subnetworks (E1 and E2; Figure 3A-C) based on their exposure to a fluctuating training signal that will interact with a Hebbian $E \rightarrow E$ STDP plasticity to form E assembly structure [11, 13, 48]. In what follows there is no initial $E \rightarrow E$ structure (i.e. E1 and E2 have identical within and between connectivity strengths), and any final structure must emerge through assembly training.

We consider three different network models, distinguished by their inhibitory subnetworks. The first is a '1PV network' (Figure 3A) with two excitatory populations and one global inhibitory population, without any tuned connections. The second is a '2PV network' (Figure 3B), composed of E and PV cells where neurons in the E1 population connect more strongly and with a higher probability to half of the PV cells in the PV population (the E2

population connects more strongly to the other half). This network is composed of the same number of PV neurons as the 1PV network. Finally, the ‘SOM-PV’ network (Figure 3C) has two different inhibitory cell types, PV and SOM class interneurons. The connections from E to PV are not structured, but we assume a tuned connection from E to SOM cells, such that half of the excitatory neurons are more strongly and with a higher probability are connected to a subset of SOM population (similar to the 2PV network). In all cases the $E \rightarrow E$, $PV \rightarrow E$, and $SOM \rightarrow E$ synaptic connections have the appropriate STDP learning rules (see Methods; colored connections in Figure 3A-C denote plastic connections).

Before training, all networks express asynchronous, temporally irregular spiking activity (Figure 3E-F). SOM neurons in the SOM-PV network received synaptic inputs only from the E population, along with an external static current to bring up the resting membrane potentials in SOM neurons; nevertheless, before training, SOM neurons have negligible firing rates (Figure 3F). During training, each E neuron receives external excitatory synaptic trains with Poisson interval statistics. There are two sources of inputs to each neuron. The major inputs are unshared sources that feed each neuron independently with 70 % of the total external input firing rate. In addition to this unshared input, there are also two independent sources of Poisson processes, each of which provide a shared input (the remaining 30 % of the total external input) to two individual subgroups of the excitatory population. The goal of the shared input is to correlate neurons in each excitatory sub-population independently of the other population. The shared source could for example represent a different odor which acts as a different sensory stimulus. During training, neurons in each excitatory population that receive shared input are highly correlated. This correlation manifests itself in the correlation function between average neuron spontaneous firing rates which reside in each population (the average PSTH function for each excitatory population is shown on the raster plot for reference). The amplitude of the auto-correlation function for the average firing rate in each assembly is maximum for the SOM-PV network, and minimum for the 2PV network (Figure 3G). Moreover, the cross-correlation function between the average firing rates of the two excitatory populations is more negative for the SOM-PV network than the cross-correlations for other networks under study (Figure 3H). This implies that competition between cell assemblies during training is stronger in the SOM-PV network. Moreover, the bursty nature of activity of SOM neurons in the SOM-PV network is due to lack of connections between SOM neurons and the fact that they only receive excitation from the activity of the excitatory populations. An experimental observation reporting a similar phenomenon can be found in [49].

After 2000 seconds of training, we removed the shared source and replaced it with a global independent unshared stimulus source. It is noticeable that the activity of excitatory neurons in each assembly remains highly correlated in the SOM-PV network, and this correlation structure is much less impressive in the 1PV and 2PV network (Figure 3I). Similar to the competition structure between cell assemblies during training, the cross-correlation between assembly firing rates after training is more impressive in the SOM-PV network (Figure 3J). This hints to the fact that after training, the assembly structure in the SOM-PV network has maintained itself to a larger degree. Overall, we conclude that the existence of SOM neurons with Hebbian STDP, during and after training, has resulted in a bigger amplitude of auto and cross correlation function between average neuronal firing rates in each assembly.

Figure 3: Different network architectures and the resulting correlation structure during and after training. **A:** 1PV network has 2 separate groups of excitatory neurons, and one shared group of PV cells. There are no tuned connections in this case. **B:** 2pv network is composed of 2 distinct groups of excitatory neurons which are tuned to distinct groups of PV neurons. **C:** SOM-PV network is similar to the 1pv network, but in addition, it has 2 groups of distinct SOM neurons, each of which receive tuned connections from excitatory neurons. **D, E, F:** raster plots of the networks in panels A, B, C, respectively, when there is no training, during training, and after training. Only 3 seconds of the network activity has been shown for each case. **G, I:** Autocorrelation function for the average population firing rate of a single excitatory assembly during training, and after training, respectively, for each network architecture. **H, J:** Cross-correlation function between the average firing rates of each excitatory assembly during and after training, respectively.

Excitatory weight evolution in different networks

To explain how the connection weight between two neurons changes, we consider a pre-synaptic neuron j impinging on a postsynaptic neuron i (Figure 4A). Each neuron fires according to a pattern. In order to calculate the average weight evolution between the pre-synaptic and a post-synaptic neuron, it is conceivable to first calculate the cross-correlation function between the spike trains of the pre-synaptic and the post-synaptic neuron (Figure 4B). The

integral of the product of this function with the STDP curve will yield the average rate of change in the connection weight between the two neurons [50]. This integral, after being scaled by a learning rate η is the sum of all average potentiation (indicated by $U1$ in Figure 4C) and depression (indicated by $U2$ in Figure 4C) that the synapse undergoes. We use this idea throughout the paper to justify the average weight change between populations of neurons. For this, we assume that the average spontaneous firing rate of neurons in each population stands for the firing rate of individual neurons in a mean-field sense.

During training, the average auto-correlation function for the excitatory populations in all networks under study has a peak for zero lag time shifts (Figures 4D,E, 5G). The integral of the product of these functions with the STDP curve for the excitatory connections results in positive scalars, which indicates potentiation of the synapse. This product has been illustrated in Figures 4D1, E1, 5G1 for the 1pv, SOM-PV and the 2PV network, respectively. The area under the curves are positive in darker shades, and negative in lighter shades for each corresponding synapse. The overall integral, the sum of the positive and negative areas, are 1.18×10^{-5} , 1.41×10^{-5} , and 7.45×10^{-6} mV/s, for the 1pv, SOM-PV and 2PV network, respectively (These results were obtained by considering neuronal average PSTH signals from $t = 200$ to $t = 700$ seconds during training). The auto-correlation function for the SOM-PV network grows in amplitude for progressive points in time, and this growth in amplitude also slightly changes the growth rate of the weight through a positive feedback loop. At $t = 2000$ seconds during training, the average weights within an excitatory assembly is slightly above 0.12 mV for the 1pv network (Figure 4D3). This quantity reaches 0.146 mV for the SOM-PV network, and 0.095 mV for the 2PV network (Figures 4E3, 5H, respectively).

However, the integral of the product of the STDP curve and the cross-correlation functions between the average firing rates of neurons in different assemblies is negative. This quantity is equal to -3.33×10^{-6} mV/s for the 1pv network (Figure 4D2), and is equal to -3.04×10^{-6} and -2.21×10^{-6} mV/s for the SOM-PV network (Figure 4E2) and the 2PV network (Figures 5G2), respectively. This results in the depression of the coupling weights between neurons in different assemblies (Figures 4D3, E3, H, pink and red curves). This depression of the average weights between assemblies improves assembly segregation, and facilitates distinct assembly formations corresponding to different input signals.

Role of PV homeostasis in the evolution of excitatory weights

The STDP curves for the excitatory connections had a net negative integral. This indicates that if there was no correlation structure between the pre-synaptic and post-synaptic activity of neurons, the average synapse between those neurons would depress. The fact that the assemblies are being formed due to the potentiation of the synapses within assemblies is because the product between the “cross-covariance” function between neurons and the STDP curve is a positive scalar and bigger than the product of the mean firing rates and the STDP curve (In this argument we rely on the fact that a cross-correlation function is the sum of the cross-covariance function and the product of the mean firing rates of individual neurons). This picture is different for the synapses between neuron in two distinct assemblies. Due to competition between neurons in different assemblies, the cross-covariance between those neurons is mostly negative. Hence, the total integral of the product of the STDP curve and the correlation function is negative, and as a result, the synapses between assemblies decay with time. In all networks under study, the homeostasis provided by the connections from PV to E cells guaranteed that the firing rate of the excitatory neurons stay fixed and close to 3 Hz. Therefore, the difference in growth rate of the excitatory synapses in the networks under study is mainly due to the different cross-covariance structure between the activities of the excitatory neurons. We will show in the rest of the paper that this covariance is shaped by the SOM plasticity in favor of speeding up the growth of the synapses between neurons in excitatory assemblies.

SOM plasticity results in lateral inhibition

The tuned connections from each excitatory assembly to their corresponding SOM population results in a causal activity of SOM after its corresponding excitatory activity visits the high state of its spontaneous fluctuations. This causality, due to the Hebbian nature of the plasticity, results in a left-skewed correlation function between SOM_1 and E_1 , as shown in Figure 4F. As a result, the integral of the superposition of this function on the STDP curve will be a negative quantity (Figure 4F1, -2.39×10^{-7} mV/s for correlations between signals that represent average neuronal activities from $t = 200$ to $t = 300$ seconds), which implies the depression of the synaptic weight between SOM_1 and E_1 (and similarly the weight between SOM_2 and E_2). On the other hand, the activity of neurons in

SOM₁ suppresses the neurons in E₂. This results in a trough for positive and small time lags in the cross-correlation function (Figure 4F). However, since the integral of the STDP curve is positive (Figure 4F2, 6.2×10^{-6} mV/s for correlations between signals that represent average neuronal activities from $t = 200$ to $t = 300$ seconds), the overall weight change would result in a positive quantity and consequently the potentiation of the synapse. Over time, this potentiation forms lateral inhibition from the SOM population that is tuned to the opponent excitatory assembly onto the other excitatory assembly which is suppressed as a result of the activity of those SOM cells. Subsequently, the lateral inhibition amplifies the negative cross-correlation between the excitatory populations, and hence stronger competition between assemblies. Moreover, as a result of this competition, single population auto-correlation functions also increase in magnitude over time. As inferred from [50], this increase in auto-correlation function causes faster formation of cell assemblies. Also, the competition between assemblies which reflects in the cross-correlation function between the excitatory assemblies results in a weak coupling between the two assemblies.

In summary, the emergence of lateral inhibition shapes the correlation functions in such a way that both the auto- and the cross-correlation functions increase in amplitude over time, and this facilitates and speeds up assembly formation. We compared the weights within and between the excitatory assemblies during training after 2000 seconds. The weights between assemblies were the minimum, and the weights within an assembly were the maximum for the SOM-PV network (Figure 4J1). This means that the 1pv and the 2PV network, in comparison with the SOM-PV network, show a slower formation of assemblies. Also, the least segregation between assemblies was observed in the 2PV network. Moreover, 3800 seconds after training, we checked those weights and observed that all weights had potentiated. This potentiation was stronger for the weights within assemblies (Figure 4J2). This is because in all these networks 2000 seconds of training was enough time to form assemblies with correlation functions that were big enough to maintain assemblies (see [11] for more details). The structure of the correlations were such that the assembly structure was the best preserved and grown in the SOM-PV network, and the growth was the slowest for the 2PV network. In the rest of the paper, SOM-PV and 1PV networks will be compared with each other. We will refer to the 2PV network later in the Results section.

Figure 4

Figure 4: Correlation functions and weigh evolution. **A**: Spiking activities of a presynaptic neuron j which projects onto a postsynaptic neuron i . On average, the presynaptic neuron fires before the postsynaptic neuron. **B**: STDP curve is plotted in gray, and the correlation function between the spike trains of the presynaptic neuron with the postsynaptic neuron is plotted in red. **C**: The integral of the cross-correlation function and the STDP curve results in a scalar with a positive part (U_1), and a negative part (U_2). The sign of the sum of these two quantities determines whether the synaps potentiates or depresses. **D, E, G**: Within assembly autocorrelation function (solid lines) and between assebmly cross-correlation function (dashed lines) for the 1pv network, SOM-PV network, and 2PV network, respectively. The Hebbian STDP curve which rules the weight update between excitatory neuron is plotted in gray. **D1, D2**: The integral of the product of the STDP curve with the single assembly auto-correlation function and the cross-correlation function between assemblies in panel D, respectively. **D3**: Within assembly (cyan) and between assembly (pink) weight changes as a function of training time. At $t = 2000$ s, the average weights within an excitatory assembly is slightly above 0.12 mV. **E1, E2**: The integral of the product of the STDP curve with the auto-correlation function and the cross-correlation function in panel E, respectively. **E3**: The excitatory weight evolution for the connections within (blue) and between (red) assemblies. At the end of the training time ($t = 2000$ s), the average weight within an excitatory assembly is slightly above 0.14 mv. **F**: Cross-Correlation function between SOM1 and E1 in solid lines, and cross-correlation between SOM1 and E2 in dashed lines. **F1, F2**: The integral of the product of the STDP curve with the cross-correlation function between SOM1 and E1 (F1), and the cross-correlation function between SOM1 and E2 (F2) in panel E, respectively. **F3**: Development of the lateral and mutual inhibitory weights from SOM1 to E2 and to E1, during training, plotted in green and yellow respectively. The lateral inhibition is stronger than the mutual inhibition. **J1**: During training, at $t=2000$ sec, the weights within assembly are the largest, and the weights between assemblies are the weakest for the SOM-PV network. **J2**: After 3800 seconds post training, the weights within the excitatory assemblies are the largest for the SOM-PV network, and the smallest for the 2PV network. The weights between assemblies is bigger in the 2PV network.

Figure 5: Correlation functions and weigh evolution for the 2PV network. **G1, G2**: The integral of the product of the STDP curve and the correlations in panel G. **H**: Excitatory weight evolutions for the connections within (cyan) and between (pink) assemblies, for the 2PV network.

SOM facilitates assembly formation and slows down assembly deformation

For more realistic parameters of the STDP curve (integral of the total STDP curve comparable to the integral of the curve for positive lag) and slightly lower input firing rates, we simulated the 1pv and SOM-PV network for different duration of the training time. Since the STDP curve is depression dominant, we did not provide a training signal to the network for 1000 seconds until all transient dynamics in the weight evolution brought the excitatory weights to their minimum, and similar initial conditions for both 1pv and SOM-PV network. For similar initial conditions in the excitatory weights for both networks, a training signal was provided to both networks. After 1000 seconds of training, the SOM-PV network developed a slightly stronger assembly structure compared to the 1pv network (Figure 6A). At $t = 0$, the training signal was removed. Right after training, the weights within assemblies started to decay, and there was a slight potentiation in the weights between assemblies. The depression for the 1pv network was stronger and happened faster for the 1pv network. The SOM-PV network, however, regained the assembly structure about 1500 seconds after training. In a separate simulation, we increased the input training time to 4000 seconds, and observed that assemblies were formed much faster in the SOM-PV network (Figure 6B). After the training signal was removed at $t = 0$, assemblies continued growing after a slight deformation. This growth can be attributed to the remaining correlation structure which can maintain itself and get amplified even when the training signal does not exist. A similar phenomenon was reported in [51, 11], where internally generated correlations as a result of training were sufficient to drive the weight evolution even after training. At $t = 5000$ s post training, the weights within individual assemblies are much bigger for the SOM-PV network.

In order to quantify the training time necessary to maintain the assembly structure similar to the existing one at $t = 0$ post training, we referred to a measure which was the difference between the weights within and between assemblies, called dw_0 . We also calculated this measure for $t = 4000$ for different network simulations with different

training times and called this dw_{4000} . Then, we plotted $dw_{4000} - dw_0$ for different training times for the SOM-PV and the 1pv network (Figure 6C). The training time for which this measure crosses zero is the minimum training time necessary to recover the same assembly structure at $t = 4000$ s that existed right after training ($t = 0$). As our results suggest, this training time is much smaller for the SOM-PV network compared to the 1pv network. This concludes that it is easier to train and maintain an assembly structure when SOM inhibitory plasticity contributes to assembly formation.

Figure 6: Faster training and slower forgetting in the SOM-PV network. **A:** For a training period of 1000 seconds, the excitatory weights within each assembly grow faster in the SOM-PV network compared to the 1pv network. After training, the weights within and between assemblies are almost identical in the 1pv network. **B:** For a prolonged training period of 4000 seconds, assemblies continue to grow their structure after training. The growth rate is faster in the SOM-PV network. **C:** 4000 seconds after the training signal was removed ($t = 4000$ s), the difference between the weights within each assembly and between assemblies (dw_t) were compared to that quantity right after training was stopped (dw_0 at $t=0$ s). The index t in dw_t represents the training time. When the difference crosses 0, it means the same assembly structure right after training is recovered at $t = 4000$ s. Zero-level crossing happens in less training time for the SOM-PV network.

To understand the dynamics of the excitatory weights during assembly formation, we considered a dynamical system analysis in the weight state space. To characterize assembly formation, we were interested in two states: weights within assemblies (w_{11}) and weights between assemblies (w_{12}). We conducted a mean-field analysis for the average neuron's correlation functions within individual populations. These correlations are functions of the plastic and static weights. Given that the SOM-PV network has 5 interacting populations (E_1 , E_2 , I , SOM_1 , and SOM_2), and 5 plastic weights (w_{11} , w_{12} , w_i , w_{s1} , and w_{s2}), the entire set of coupled systems of ODEs is 10 dimensional (see the Methods section for details). A similar analysis for the 1pv network results in a 6 dimensional system. In order to get a phase plane analysis in the parameter space of interest (w_{11} and w_{12}), we used the Laplace transform to relate the dynamics of the weights to the firing rates, and performed a dimensionality reduction in the Laplace domain such that the entire system could be described as a function of w_{11} and w_{12} only.

The phase plane analysis for the dynamics of w_{11} and w_{12} for the 1pv and the SOM-PV network suggests that in a given parameter range close to the origin, the nullcline associated with w_{11} is closer to the origin for the SOM-PV network (Figure 7A,B). This means that the area in the parameter space in which learning takes place (arrows that point to the right and downwards) is bigger for the SOM-PV network. Also, the velocity of the vectors associated to the w_{11} component is bigger for the SOM-PV network compared to the 1pv network. This automatically implies that training assemblies is faster in the SOM-PV network.

The phase portraits for both networks after training suggest that if w_{11} and w_{12} are both small, the flow in both directions decreases until the weights converge to zero. However, if the state of the weights within and between assemblies are such that the nullcline for w_{11} is passed, then the flow can continue in the direction of increasing w_{11} (Figures 7C,D). This means that the assembly can sustain itself, only if the training time was sufficient enough and as a result the initial conditions post-training were left in the region of interest where assemblies could potentially grow. This region of interest is highlighted in gray in Figure 7E,F for both networks. For an exemplary initial condition (w_{11}, w_{12}) = (0.5, 0.2), right after training, the flow will move towards the origin for the 1pv case, however, for the SOM-PV network, the flow continues in the direction of increasing w_{11} and decreasing w_{12} . This is all due

to the leftward shift of the nullclines for the SOM-PV case which suggests maintaining the assembly structure after training is relatively easy if SOM plasticity contributes to learning. Also, the region of interest is relatively bigger in this case.

Figure 7: Mean-field phase portraits for the excitatory weight evolution for the two networks. **A**: When the training signal is present ($c=0.3$), the nullcline for the 1pv network is farther from the origin than the nullcline for the SOM-V network. The flow is bigger in the SOM-PV network case. **B**: After training, the nullcline for $c = 0$ is closer to the origin in the case of SOM-PV network. **C**: The region in the parameter space in which assembly growth continues after training is highlighted in gray. For an initial condition of $(w_{11}, w_{12}) = (0.5, 0.2)$, which could result from a training phase, the trajectories after the training signal was removed have been plotted in red.

Emergent properties of interaction between inhibitory subtype plasticity

Thus far, we have discussed the connection weights within and between excitatory assemblies in the 3 networks we considered. We also discussed the emergence of lateral inhibition from SOM to E cells in the SOM-PV network. In this section, we investigate the evolution of the inhibitory weights from PV onto E cells.

For all networks, we summed up all the weights that any individual PV neuron impinged on the E cells in E_1 , and called this sum $PV \rightarrow E_1$. Similarly, we summed up all the projecting weights from any PV cell onto all neurons in E_2 , and called this quantity $PV \rightarrow E_2$. We then sorted the index of all PV cells such that the neurons with lower indices had the largest $PV \rightarrow E_1$. The connectivity matrix from the PV cells to the entire E cells for the 1pv network is shown in Figure 8A. The first left block of this matrix is sorted as explained before. The second half on the right does not follow any specific pattern, and therefore, we conclude that for this network, there is no systematic preferential bias for PV cells to connect more strongly to E cells in either of the assemblies.

This picture, however, is different for the SOM-PV network. As it turns out, PV cells which are more strongly connected to E cells in E_1 (have a high value of $PV \rightarrow E_1$), connect weakly to neurons in E_2 , as the right side of the matrix is almost the inverted version of the left side of this matrix (Figure 8B). This fact indicates that the PV population has split into two sub-populations, each of which have more preference to connect strongly to one of the excitatory cell assemblies. We call the first half of the PV cells in the connectivity matrix- with the higher value of $PV \rightarrow E_1$ - PV_1 population. The second half with higher indices of PV cells- which in this case connect more strongly to E_2 cells- are called members of the PV_2 population. For the 1pv network, there was no distinction

between PV_1 and PV_2 , but nevertheless, we used the same labeling for this network as well.

To understand the origin of the emergent tuned PV to E weights, we checked the correlation functions between PV_1 average firing rate and E_1 and E_2 average firing rates respectively, and we super-imposed it on the symmetric STDP curve that we used in our network simulations. The cross-correlation function between PV_1 and E_2 has a big negative trough which makes the entire integral of the product of the correlation function and the STDP curve negative (Figure 8E). This results in the depression of the connection weights from PV_1 to E_2 . However, the cross-correlation function between PV_1 and E_1 is mainly positive, and the interaction of this function with the STDP curve yields a positive quantity for the rate of change of the connections from PV_1 to E_1 . This justifies the emergence of tuned weights in the SOM-PV network. These correlation functions for the 1pv network are almost identical (8D). Hence, there is no preference for the connections from PV onto E_1 or E_2 .

The 2PV network already has two distinct PV populations. Therefore, we did not sort the indices of the PV cells in this case. The PV to E connectivity matrix has a clear block structure, in which the weights from PV_1 onto E_1 and the weights from PV_2 onto E_2 are bigger than the other counterpart weights (Figure 8C). This tuned connection from PV_x to E_x can similarly be explained by examining the correlation functions between PV and E activity. The correlation function between PV_1 and E_1 is mostly positive, and this results in the potentiation of the inhibitory synapse connecting these two populations. However, the correlation function between PV_1 and E_2 is mainly negative, and this explains weaker connections from PV_1 to E_2 .

So far, we have examined the correlation structure between the PV sub-populations and the excitatory cell assemblies. To shed light on the reason behind the difference in correlation functions between PV populations and the E populations in the SOM-PV network, we dissected the projection weights from the SOM populations onto the PV sub-populations. It turned out that SOM_1 projected weakly onto PV_1 , but more strongly onto PV_2 (Figure 8G). The same pattern of connectivity was observed for the total projection weights from SOM_2 onto PV_2 and PV_1 , respectively. This observations explains that when SOM_1 neurons are active, the strong projection weights from this population onto E_2 and PV_2 suppresses both populations simultaneously. When SOM_1 is relatively silent, both E_2 and PV_2 will get the chance to be activated together. As a result, PV_2 and E_2 will be correlated. The same holds true for the correlation structure between PV_1 and E_1 . Since E_1 and E_2 are anti-correlated, PV_1 and E_2 (also PV_2 and E_1) will also represent negative correlation. This explains the correlation structure in Figure 8F.

Since the initial bias from E_1 to PV_1 , and from E_2 to PV_2 was the origin of the emergence of reciprocal inhibitory connections in the 2PV network, we checked those random connections in the SOM-PV network as well. There was a slight bias in the overall excitatory connections from E_1 to PV_1 (Figure 8H), however, the distributions of the overall excitatory weights onto each subgroup of PV were not as distinguishable as the distributions of the inhibitory projection weights from SOM onto the PV subpopulations (Figure 8G). Hence, for the SOM-PV network, the main cause of the emergence of tuned PV to E connections were the heterogeneity in random connections from SOM to PV populations.

The distribution of the projection weights from SOM_1 onto PV_1 is not identical to the distribution of projection weights from SOM_1 onto PV_2 because of the random Bernoulli distribution of connections (Figure 8G). This fact can by chance break the symmetry in the total projection weights between these two distinct inhibitory subtype. This slight break in symmetry associates the PV neurons to either of the E assemblies. For the SOM-PV network considered here, slight symmetry breaking in the projections from E to PV does not cause a big difference in the distribution of the weights, and therefore, should not play as important role as the SOM to PV projections. In our mean field analysis, we assumed a slight bias in the probability of connections from SOM_1 to PV_2 , and a bias from SOM_2 to PV_1 , represented as ϵ in equation (30). As derived in the methods section (equation (38)), the difference between the weights from PV_1 to E_1 and the weights from PV_1 to E_2 scales linearly and positively with $w_{11} - w_{12}$ (difference between the weights within an excitatory assembly and between the two excitatory assemblies). Our network simulations result confirms this prediction, as illustrated in Figure 8I, where the difference between the tuned and untuned PV to E weights are plotted against the difference between the weights within and between the excitatory assemblies. This plot suggests an almost linear function approximation for the relation between these two variables. This implies that as long as assemblies are being learned ($w_{11} - w_{12} > 0$), there should emerge a tuned connection from the PV sub-population- that receives slightly stronger projection from the antagonistic SOM population- onto its corresponding excitatory assembly which also receives a strong projection weight from the same SOM population.

Figure 8: Emergence of tuned PV to E connections in the existence of tuned excitatory connections. **A:** Connectivity matrix from presynaptic PV neurons to postsynaptic E neurons for the 1pv network and the SOM-PV network, respectively. The index of the PV neurons were sorted such that lower indexed neurons project strongly onto neurons in E₁. After this sorting, the first half of the PV neurons were called PV₁, and the second half PV₂. **B:** Connectivity matrix representing the PV to E connections. After sorting the indices of the PV neurons, it appears that PV cells mostly tend to connect strongly either to E₁ or E₂, but not both. **C:** PV to E connectivity matrix for the 2PV network without sorting any index. **D, E, F:** Cross-covariance function between the average spontaneous firing rates of PV₁ and E₁, and also between PV₁ and E₂ for the networks represented on the left side of each figure. The symmetric STDP curve for the connections from PV to E neurons is shown in gray. The right axis represents the relative scales of the symmetric STDP curve for the connections between PV and E cells. **G:** Histograms of the cumulative weight projections from SOM₁ onto PV₁ and PV₂ neurons in the SOM-PV network. **H:** Histograms of the cumulative weight projections from E₁ onto PV₁ and PV₂ neurons in the SOM-PV network. **I:** The difference between tuned and untuned PV to E weights scales almost linearly and positively with the difference between the weights “within” and “between” assemblies.

Discussion

Cell assemblies are thought to be the building blocks of neuronal dynamics which map neuronal activities to behaviour [10, 7]. These coactivated groups of neurons are responsive both during spontaneous and stimulus conditions [6], and are believed to appear as a result of neuronal correlations under Hebbian plasticity [10] (also see [52]). While there has been some theoretical work to understand the role of inhibition in the formation of assemblies [13, 12, 11], there is an emerging experimental interest in understanding the contribution of inhibition to assembly formation [26, 53, 54]. It is particularly interesting to study the role of inhibition in correlative learning because it is known that inhibition decorrelates the activity of excitatory neurons in the balanced condition [23, 24]. However, the structure of correlation when different inhibitory subtypes are involved is not clear. Since inhibitory plasticity plays a role in the formation of assemblies, and since inhibition plays a crucial role in shaping network correlations, it is important to understand how different inhibitory subtypes shape their connections to facilitate learning.

In [55], the authors reported a symmetric STDP function for the connections from SOM to pyramidal neurons, and a depressive STDP curve for the connections from PV to CA1 pyramidal neurons in the Hippocampus. Our experimental findings demonstrate that in the OFC, the STDP curve for the connections from PV to E cells follows a symmetric rule with potentiation of the synapse for small time lags between pre and post synaptic spikes. It is easy to show that this kind of STDP curve can result in a homeostasis rule to regulate the firing rate of the post-synaptic neuron, conditioned that upon the firing of the inhibitory neuron, the synapse depresses. Regulating the postsynaptic firing rate brings stability to the system, and prevents runaway dynamics, particularly during training.

We showed, for the first time, that the STDP curve for the connections from SOM to E cells in the OFC follow an asymmetric Hebbian learning rule. In our large scale network simulations, we demonstrated that if the excitatory neurons have a tuned connection to SOM cells, then the causality in the STDP rule can weaken the synapse from SOM to its driving excitatory population, and potentiate the synapse from SOM to the competitor excitatory population. This results in a di-synaptic pathway with lateral inhibition between the excitatory assemblies. Lateral inhibition results in more pronounced competition between assemblies [56, 57]. We demonstrated both in network simulations, and in our theoretical framework, that the existence of lateral inhibition can amplify the correlation functions between excitatory neurons.

Whether excitatory neurons have a tuned connectivity profile to inhibitory neurons in the OFC is not clear. In other cortices such as the visual cortex, there are some evidence for the existence of tuned excitatory to PV connections [58], and tuned excitatory to SOM connections [59, 60]. Also, whether inhibitory subtypes in the OFC are tuned to a specific input feature is not known. In the visual cortex of mice, it has been claimed that despite the fact that excitatory neurons are sharply selective for certain input features, Gabaergic interneurons are less selective and broadly tuned [61, 62], and they do not form subnetworks of coactivated groups [63]. However, other studies report that PV cells in the visual cortex are integrated into subnetworks of excitatory cells with similar visual selectivity [58, 64]. Considering this ambiguity, we considered 3 different networks: the 1pv network with a shared pool of PV neurons, without any tuned connections, the 2PV network with two distinct groups of PV cells that received stronger projection weights from their feature-specific excitatory population, and the SOM-PV network with tuned connections from the excitatory populations to their specific SOM population. The SOM-PV network also had a PV population which received no tuned input from the excitatory subnetworks. In our network simulations, we observed that during and after training, the correlation structure for neurons in the same assembly was the strongest in the SOM-PV network, and hence assemblies were formed faster. Correlations between excitatory neurons within each group were weak in the 2PV network. The correlation structure was reversed between the excitatory neurons that belonged to different assemblies. Increased correlations can come about not only through lateral inhibition, but also through reciprocal inhibitory to excitatory populations. The heterogeneity in random connections from SOM to PV cells provided a substrate for correlations between a group of PV and individual excitatory assemblies such that neurons which received strong inhibition from SOM were more correlated. This resulted in tuned PV to cell assembly connections due to the symmetric shape of the STDP curve for PV to E connections, and the fact that these connections should maintain a constant firing rate. These reciprocal connections also play an important role in shaping correlations between excitatory assemblies. The 2PV network with biased connections from E to PV cells also developed reciprocal connections from PV

back to the driving excitatory assemblies. However, due to lack of strong competition and weaker correlations, assemblies did not form as quickly as those in the SOM-PV network. Existence of tuned PV to E connections and subnetworks of excitatory and PV cells with relatively strong connections were reported earlier in the visual cortex [65, 64, 58], and our network simulations and theoretical analysis confirms this finding. Interestingly, in a network with homeostasis but no tuned excitatory to inhibitory connections, like the 1pv network we studied here, no tuned PV to E connections emerge. The reason is that the correlation structure between PV and excitatory populations is not different across PV neurons.

In our study, the amplified cross-correlation function between excitatory cell assemblies implies promoted competition, which results from SOM plasticity onto excitatory neurons. While the homeostasis provided by PV plasticity can also impose competition between assemblies [13, 12, 11], we showed that the strength of competition is enhanced when SOM plasticity contributes to learning. However, the amplitude of fluctuation can increase significantly, and beyond control, if there is no homeostatic mechanism to maintain a fixed firing rate for the excitatory neurons. We conclude that PV plasticity is responsible for providing circuit stability, while SOM plasticity promotes competition. This division of labor is necessary because one source of inhibition alone cannot be responsible for both functions in an adequate way.

We showed throughout the paper that increased correlation between excitatory neurons due to SOM plasticity contribution can result in faster formation of assemblies. Increased speed of learning is critical for animal survival, and this improved cognitive function results from SOM inhibitory subtype plasticity. Also, considering different network structures, we showed that after the training signal was removed, in the presence of SOM inhibitory subtype, assemblies could maintain their structure for a longer time. More precisely, for a given short training period, the network without SOM inhibitory subtype lost its assembly structure completely, however, the network including this subtype recovered the assembly structure shortly after training. This indicates that SOM plasticity helps assemblies to be more robust to perturbations after training, and protects them against decorrelative factors, such as global input projections, or spontaneous brain dynamics.

In summary, based on our experimental results for the spike timing dependent plasticity curves for the connections from different inhibitory subtypes- mainly SOM and PV neurons- onto excitatory cells, we conclude that while PV cells provide circuit stability, SOM cells are the main contributors to provide competition between excitatory cells, which in turn accelerate assembly formation and protect assemblies from degeneration. One way to test our predictions would be to block the activity of SOM cells and compare the correlation structure between excitatory neurons before and after this blockage. According to our study, if SOM activity is removed from the network dynamics, and if the firing rate of PV does not change much as a result of this dysfunction, then the correlation between the excitatory neurons should decrease. Another prediction is that animals' behavior in odor discrimination tasks should be impaired if the activity of SOM cells are blocked (either during or after training the animal with new odors) because the structure of assemblies fade away very quickly when SOM activity is blocked.

Methods

Experimental details

Mice: SST-Cre (B6:Sst^{tm2.1}(cre)Zjh>/J) and PV-Cre (B6.129P2-Pvalbtm1(cre)Arbr/J) mice [66] that express cre-recombinase were crossed with Ai32 mice (B6;129S-Gt ROSA)26Sortm32 (CAG-COP4*H134R/EYFP)Hze/J) [67] to express channelrhodopsin (ChR2) in somatostatin or parvalbumin interneurons, respectively. All mice were from Jackson Laboratory. Slice preparation: Coronal brain slices of orbitofrontal cortex (OFC) located 2.845-2.245 mm anterior to Bregma (Allen Brain Atlas) were prepared from mice of both sexes that were P20-28 days old. The mice were anesthetized with isoflurane and decapitated. The brain was removed from the skull and immersed in ice-cold oxygenated (95% O₂-5% CO₂) ACSF solution (in mM: 125 NaCl, 2.5 KCl, 25 NaHCO₃, 1.25 NaH₂PO₄, 1.0 MgCl₂, 25 Dextrose, 2.5 CaCl₂) (Sigma-Aldrich, USA). Coronal slices were 300 μ m thick and cut using a vibratome (Leica Biosystems) in ice cold ACSF. The slices were transferred to warm ACSF (36 C) for 30 min and then rested at 20-22 C for 45-60 min prior to recording in the same conditions. All procedures used in this study were approved by The University of Pittsburgh IACUC.

Electrophysiology: Recordings were obtained from L2/3 excitatory neurons with input resistance <250 M Ω visualized using infrared differential interference contrast microscopy (IR-DIC, Olympus). For all recorded neurons,

intrinsic subthreshold properties such as input resistance, and time constant were obtained using a series of hyperpolarizing and depolarizing current steps (10 pA, 1 s duration). Whole cell, voltage and current clamp recordings were performed using a MultiClamp 700B amplifier (Molecular Devices, Union City, CA). Data were low pass filtered (4 kHz) and digitized at 10 kHz using an ITC-18 (Instrutech) controlled by custom software (Recording Artist, <https://bitbucket.org/rgerkin/recording-artist>) written in IgorPro (Wavemetrics). Recording pipettes (5-10 M Ω) were pulled from borosilicate glass (1.5 mm, outer diameter) on a Flaming/Brown micropipette puller (Sutter Instruments). The series resistance (<20 M Ω) was not corrected. The intracellular solution consisted of (in mM) 130 K-gluconate, 5 KCl, 2 MgCl₂, 4 ATP-Mg, 0.3 GTP, 10 HEPES, and 10 phosphocreatine, 0.05% biocytin. Since post-synaptic spike responses are required for STDP induction, we could not use Cesium or QX-314 in the recording pipette. The reversal potential for inhibition was found for each neuron to be between -75 to -70 mV. All voltage clamp recordings were performed with a holding potential between -50 to -60 mV to maximize inhibitory currents but maintain a subthreshold membrane potential in the PN. To isolate GABA_A receptor mediated inhibition, the GABA_B receptor antagonist, Saclofen (20-40 M), was included in the bath.

STDP protocol: The STDP induction protocol was applied with 10 min of patching the PN. Baseline IPSCs were light-evoked from either PV or SOM interneurons every 30 s and recorded in voltage clamp for 5 min prior to STDP induction. Recordings were then switched to current clamp to evoke action potentials in the post synaptic PN during the STDP protocol. To induce STDP, a single action potential (AP) was evoked by current injection at the soma (50 μ s pulse width). The AP was paired with light-evoked IPSP at a consistent time interval with respect to the AP. We applied this pairing protocol every 4 sec (0.25 Hz) for 4 min. The interval is defined as the difference in time (ms) between the AP and the light evoked IPSP. Following the STDP protocol, IPSCs were again evoked every 30 s for up to 1 hour. In all cases, series resistance and input resistance were monitored throughout baseline, induction and post-induction testing. Neurons with deviations greater than 25% from baseline were excluded from analysis.

Optogenetic Stimulation: Shutter controlled full field light stimulation of blue light (473 nm) provided by a mercury lamp was delivered through the epifluorescence pathway of the microscope (Olympus) using a water-immersion objective (40x). The duration of the light pulses was 5 ms, light intensity (3-4 mW) was adjusted to induce reliable, single spike responses in neurons expressing ChR2 (Figure 2B1,C1).

Data analysis: Inhibitory strength was taken as the peak amplitude of the IPSC within 10 ms of IPSC onset. Since these are population IPSCs, the rising slope of the IPSC and the integral of the IPSC were also analyzed. The results did not differ from the main findings reported in this study (data not shown). To assess the time course of plasticity, IPSC amplitudes pre and post induction were normalized to the average baseline amplitude. To assess the change in synaptic strength following the STDP protocol, we took the average normalized IPSC amplitude between 25-30 min post induction (10 IPSCs). We chose this time point to minimize any influence of recording drift across neurons. However, in the majority of neurons the level of synaptic plasticity measured at 30 min was maintained until the end of recording (1 hr).

Statistics: All data is presented as mean SE. Statistical tests were performed using two tailed, one or two-sample, unpaired Students t-test as appropriate. In cases of small sample sizes the non-parametric tests Mann-Whitney U-test (MWU) test was used.

Network simulations

In all networks we studied (SOM-PV, 1pv, and 2PV networks), 80% of the neurons were excitatory and the remaining 20 % were inhibitory. We used Leaky Integrate-and-Fire (LIF) neuron models with delta shape PSC. The membrane potential dynamics for the LIF neuron was a low pass filter of first order with the dynamics

$$\tau \dot{v} = -v + \mu + \sigma \sqrt{\tau} \xi(t) \quad (1)$$

where μ and σ represent the mean and the standard deviation of the fluctuations imposed on the neuron as a result of external input. The variable $\xi(t)$ is a normal Gaussian white noise.

Each excitatory population was composed of N_e neurons, and each Somatostatin population had N_s neurons. The inhibitory PV population was composed of N_i neurons for the SOM-PV network, and it had 2.5 times more inhibitory neurons for both 1pv and 2PV networks. The probability of connection between excitatory neurons was

p , and every other probability of connection was $4p$. As such, every two neurons were connected according to a Bernoulli distribution with parameters p or $4p$.

The non-plastic excitatory weight to PV neurons is represented by J , which in our study took the value of 0.06 mV, and the fixed inhibitory weights are $-gJ$.

For the SOM-PV network, we chose tuned excitatory to SOM connections by strengthening the weights by a factor of 2, meaning that projections from the excitatory populations to their corresponding SOM population was $2J$. Also, the probability of connection from each excitatory assembly to their corresponding SOM population was $p_t = 2(4p)/3 = 8p/3$, in which p_t represents tuned probability of connections. The tuned excitatory to PV connections in the 2PV network also followed the same method.

Each neuron in the excitatory and PV population received spikes from an independent Poisson process with a firing rate of σ^2 when there was no training. During training however, neurons received spikes from independent external sources with a firing rate of $0.7\sigma^2$, and a shared source of spikes with a rate equal to $0.3\sigma^2$.

The network simulation time resolution and also the synaptic transmission delay between neurons were chosen to be 0.1ms.

Plasticity equations

The STDP curve for the excitatory connections was

$$e_{stdp}(\hat{t}) = \eta_e \begin{cases} \exp(-\hat{t} / \tau_{+}) & \text{if } \hat{t} > 0 \\ -\lambda \exp(\hat{t} / \tau_{-}) & \text{if } \hat{t} < 0 \end{cases}, \quad (2)$$

The STDP curve for the connections from SOM to excitatory neurons was

$$som_{stdp}(\hat{t}) = \eta_s \begin{cases} \exp(-\hat{t} / \tau_{+}) & \text{if } \hat{t} > 0 \\ -\lambda \exp(\hat{t} / \tau_{-}) & \text{if } \hat{t} < 0 \end{cases}, \quad (3)$$

where we chose $\tau_{+} = 30$ ms and $\tau_{-} = 30$ ms for SOM stdp, and $\tau_{+} = 15$ ms and $\tau_{-} = 30$ ms for the excitatory stdp. The factor $\lambda = 0.53$.

The symmetric STDP curve for the connections from PV to E was

$$PV_{stdp}(\hat{t}) = \eta_p \begin{cases} \exp(-\hat{t} / \tau_{+}) & \text{if } \hat{t} > 0 \\ \exp(\hat{t} / \tau_{+}) & \text{if } \hat{t} < 0 \end{cases}, \quad (4)$$

The time constant of the STDP curve for the PV to E connections was $\tau_{+} = 20$ ms. The time constant of learning for the excitatory to excitatory connections were η_e , and the equivalent quantity for the connections from SOM to E were η_s .

Mean-field equations

The dynamic mean-field equations for individual neurons for the SOM-PV network are

$$\begin{aligned} \tau \dot{E}_1 &= -E_1 + \tau p k_e(\mu, \sigma) (N_e w_{11} E_1 + N_e w_{12} E_2 - 4 N_i w_i I - 4 N_s w_{s1} S_1 - 4 N_s w_{s2} S_2) \\ &\quad + k_e(\mu, \sigma) \tau \sigma (\sqrt{c} \xi_{c_1}(t) + \sqrt{1-c} \xi(t)) \\ \tau \dot{E}_2 &= -E_2 + \tau p k_e(\mu, \sigma) (N_e w_{12} E_1 + N_e w_{11} E_2 - 4 N_i w_i I - 4 N_s w_{s1} S_1 - 4 N_s w_{s2} S_2) \\ &\quad + k_e(\mu, \sigma) \tau \sigma (\sqrt{c} \xi_{c_2}(t) + \sqrt{1-c} \xi(t)) \\ \tau \dot{I} &= -I + \tau p k_i(\mu, \sigma) (4 N_e J E_1 + 4 N_e J E_2 - 4 N_i g J I - 4 N_s g J S_1 - 4 N_s g J S_2) + k_i(\mu, \sigma) \tau \sigma \xi(t) \\ \tau \dot{S}_1 &= -S_1 + \tau p k_s(\mu, \sigma) (8/3 N_e 2J E_1 + 4/3 N_e J E_2) \\ \tau \dot{S}_2 &= -S_2 + \tau p k_s(\mu, \sigma) (4/3 N_e J E_1 + 8/3 N_e 2J E_2) \end{aligned} \quad (5)$$

for the details of the derivation see the supporting information.

Supporting Information

Coupled rate and weight dynamics

Here, we study the interactions between dynamics of leaky integrate-and-fire (LIF) neurons and the evolution of weights between neurons in different assemblies, as well as the inhibitory weights to the excitatory populations. The membrane potential dynamics for an LIF neuron is a low pass filter of first order with the dynamics

$$\tau \dot{v} = -v + \mu + \sigma \sqrt{\tau} \xi(t) \quad (6)$$

where $\xi(t)$ is a normal Gaussian white noise.

In order to get neuronal correlation functions, it is required to linearize the dynamics of individual neurons around their operating stationary firing rates. A transfer function that relates the stationary input to the firing rate of the neuron is obtained from the stationary solution of the Fokker-Planck equation to solve a first passage time problem. The stationary firing rate r is obtained from $r = f(\mu, \sigma)$ where the transfer function $f(\cdot)$ follows

$$f^{-1}(\mu, \sigma) = \tau_{ref} + \tau \sqrt{\pi} \int_{\frac{v_r - \mu}{\sigma}}^{\frac{\theta - \mu}{\sigma}} e^{u^2} (1 + \operatorname{erf}(u)) du = \tau_{ref} + \tau \sqrt{\pi} \int_{y_r}^{y_\theta} g(u) du \quad (7)$$

To linearize the transfer function around the operating point r^* , we take the derivative of both sides with respect to μ . This results in

$$\frac{\partial f}{\partial \mu} = f^2(\mu, \sigma) \frac{\tau \sqrt{\pi}}{\sigma} [g(y_r) - g(y_\theta)] = k(\mu, \sigma), \quad (8)$$

where $k(\mu, \sigma)$ represents the slope of the $f - I$ curve at the linearization point, and $y_r = \frac{v_r - \mu}{\sigma}$ and $y_\theta = \frac{\theta - \mu}{\sigma}$.

We choose E_1 and E_2 to represent the dynamic firing rates of individual neurons in the excitatory assemblies, I to represent the fluctuating activity of one neuron in the PV inhibitory population, and S_1 and S_2 to stand for the dynamic firing rates of the somatostatin neurons. The operating point for excitatory, PV and SOM neurons are r_e^* , r_i^* , and r_s^* , respectively (the stationary solution comes from solving coupled Fokker-Planck equations). The connection weights from population j to i are w_{ij} . The variable w_i indicates the weight from PV to excitatory neurons. The variable ws_i represents the connection weight from a SOM neuron to an excitatory neuron in assembly i .

The dynamic mean-field equations for individual neurons are

$$\begin{aligned} \tau \dot{E}_1 &= -E_1 + \tau p k_e(\mu, \sigma) (N_e w_{11} E_1 + N_e w_{12} E_2 - 4 N_i w_i I - 4 N_s ws_1 S_1 - 4 N_s ws_2 S_2) \\ &\quad + k_e(\mu, \sigma) \tau \sigma (\sqrt{c} \xi_{c_1}(t) + \sqrt{1 - c} \xi(t)) \\ \tau \dot{E}_2 &= -E_2 + \tau p k_e(\mu, \sigma) (N_e w_{12} E_1 + N_e w_{11} E_2 - 4 N_i w_i I - 4 N_s ws_1 S_1 - 4 N_s ws_2 S_2) \\ &\quad + k_e(\mu, \sigma) \tau \sigma (\sqrt{c} \xi_{c_2}(t) + \sqrt{1 - c} \xi(t)) \\ \tau \dot{I} &= -I + \tau p k_i(\mu, \sigma) (4 N_e J E_1 + 4 N_e J E_2 - 4 N_i g J I - 4 N_s g J S_1 - 4 N_s g J S_2) + k_i(\mu, \sigma) \tau \sigma \xi(t) \\ \tau \dot{S}_1 &= -S_1 + \tau p k_s(\mu, \sigma) (8/3 N_e 2J E_1 + 4/3 N_e J E_2) \\ \tau \dot{S}_2 &= -S_2 + \tau p k_s(\mu, \sigma) (4/3 N_e J E_1 + 8/3 N_e 2J E_2) \end{aligned} \quad (9)$$

which in matrix form can be represented as

$$\begin{pmatrix} \dot{E}_1 \\ \dot{E}_2 \\ \dot{I} \\ \dot{S}_1 \\ \dot{S}_2 \end{pmatrix} = -\frac{1}{\tau} \begin{pmatrix} E_1 \\ E_2 \\ I \\ S_1 \\ S_2 \end{pmatrix} + p \begin{pmatrix} k_e N_e w_{11} & k_e N_e w_{12} & -4 k_e N_i w_i & -4 k_e N_s w_{s1} & -4 k_e N_s w_{s2} \\ k_e N_e w_{12} & k_e N_e w_{11} & -4 k_e N_i w_i & -4 k_e N_s w_{s2} & -4 k_e N_s w_{s1} \\ 4 k_i N_e J & 4 k_i N_e J & -4 k_i N_i gJ & -4 k_i N_s gJ & -4 k_i N_s gJ \\ 8/3 k_s N_e 2J & 4/3 k_s N_e J & 0 & 0 & 0 \\ 4/3 k_s N_e J & 8/3 k_s N_e 2J & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} E_1 \\ E_2 \\ I \\ S_1 \\ S_2 \end{pmatrix} + \begin{pmatrix} k_e(\mu, \sigma) \sqrt{c} \sigma & 0 & 0 & 0 & 0 \\ 0 & k_e(\mu, \sigma) \sqrt{c} \sigma & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \xi_{c1} \\ \xi_{c2} \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (10)$$

The fluctuation of the rate dynamics around the fixed point can be written in the following general form:

$$\dot{X} = -A X + B \Xi$$

$$X = \begin{pmatrix} E_1 \\ E_2 \\ I \\ S_1 \\ S_2 \end{pmatrix}; \quad A = p \begin{pmatrix} k_e N_e w_{11} - \frac{1}{\tau p} & k_e N_e w_{12} & -4 k_e N_i w_i & -4 k_e N_s w_{s1} & -4 k_e N_s w_{s2} \\ k_e N_e w_{12} & k_e N_e w_{11} - \frac{1}{\tau p} & -4 k_e N_i w_i & -4 k_e N_s w_{s2} & -4 k_e N_s w_{s1} \\ 4 k_i N_e J & 4 k_i N_e J & -4 k_i N_i gJ - \frac{1}{\tau p} & -4 k_i N_s gJ & -4 k_i N_s gJ \\ 8/3 k_s N_e 2J & 4/3 k_s N_e J & 0 & -\frac{1}{\tau p} & 0 \\ 4/3 k_s N_e J & 8/3 k_s N_e 2J & 0 & 0 & -\frac{1}{\tau p} \end{pmatrix} \quad (11)$$

The auto-covariance matrix for X is

$$C(\hat{t}) = \langle X(t), X(t + \hat{t}) \rangle = \begin{cases} \exp(A\hat{t}) Q & \text{if } \hat{t} > 0 \\ Q \exp(-A^T \hat{t}) & \text{if } \hat{t} < 0 \end{cases} \quad (12)$$

and Q is the solution to the following Lyapunov equation

$$A Q + Q A^T = B B^T$$

The dynamics of the weight evolution is slow. In order to obtain a governing equation for individual weight dynamics, we need to calculate the integral of the product between the stdp curve of the synapse, and the cross correlation between the pre and post synaptic firing rates.

From Kempter et.al, the dynamics of the weights follow

$$\begin{aligned} \dot{w}_{11} &= \int_{-\infty}^{+\infty} e_{stdp}(\hat{t}) \{C(E_1(t), E_1(t + \hat{t})) + r_e^{*2}\} d\hat{t} \\ \dot{w}_{12} &= \int_{-\infty}^{+\infty} e_{stdp}(\hat{t}) \{C(E_1(t), E_2(t + \hat{t})) + r_e^{*2}\} d\hat{t} \\ \dot{w}_i &= \int_{-\infty}^{+\infty} p v_{stdp}(\hat{t}) \{C(I(t), E_1(t + \hat{t})) + (r_e^* - r_e^{target}) r_i^*\} d\hat{t} \\ \dot{w}_{s1} &= \int_{-\infty}^{+\infty} som_{stdp}(\hat{t}) \{C(S_1(t), E_1(t + \hat{t})) + r_e^* r_s^*\} d\hat{t} \\ \dot{w}_{s2} &= \int_{-\infty}^{+\infty} som_{stdp}(\hat{t}) \{C(S_1(t), E_2(t + \hat{t})) + r_e^* r_s^*\} d\hat{t} \end{aligned} \quad (13)$$

We assume that the operating point does not change when the weights are changing, i.e $r^* = \text{constant}$. The challenge is to write the weight dynamics (13) in a self-consistent way, such that it only depends on the weights of the network. The correlation function (12) can be calculated semi-analytically (not analytically due to the existence of Q), using the Laplace transform. Since the impulse response of the general system $\dot{y} = A y$ is $\exp(A t)$ in time domain, and $(s\mathbf{I} - A)^{-1}$ in the Laplace domain (s is the complex variable defined in the Laplace transform, and $\mathbf{I}$ is the identity matrix), the matrix exponentials in equation (12) can easily be replaced by $(s\mathbf{I} - A)^{-1}$ and $(s\mathbf{I} - A^T)^{-1}$, respectively, when needed. This can also help to evaluate the integrals in equation (13) due to the fact that the product of a linear system and an exponential function causes a shift in the Laplace domain. Recruiting this transform, equations (12) and (13) can be rewritten as

$$C(s) = \mathcal{L}\{< X(t), X(t + \hat{t}) >\} = \begin{cases} (s\mathbf{I} - A)^{-1} Q & \text{if } \hat{t} > 0 \\ Q (s\mathbf{I} - A^T)^{-1} & \text{if } \hat{t} < 0 \end{cases} \quad (14)$$

and the element on row i and column j of matrix $C(s)$ will be represented by $C_{i,j}(s)$. Now, equation series (13) can be written as the following form, where the variable s will be replaced by 0 in order to get the values of the integrals in (13):

$$\begin{aligned} \dot{w}_{11} &= [((s + \frac{1}{\tau_+^e})\mathbf{I} - A)^{-1} Q]_{1,1} - \lambda[Q ((s + \frac{1}{\tau_-^e})\mathbf{I} - A^T)^{-1}]_{1,1} + r_e^{*2} (\tau_+^e - \lambda\tau_-^e) \\ \dot{w}_{12} &= [((s + \frac{1}{\tau_+^e})\mathbf{I} - A)^{-1} Q]_{2,1} - \lambda[Q ((s + \frac{1}{\tau_-^e})\mathbf{I} - A^T)^{-1}]_{2,1} + r_e^{*2} (\tau_+^e - \lambda\tau_-^e) \\ \dot{w}_i &= [((s + \frac{1}{\tau_+^i})\mathbf{I} - A)^{-1} Q]_{3,1} + [Q ((s + \frac{1}{\tau_-^i})\mathbf{I} - A^T)^{-1}]_{3,1} + 2 \tau_+^p (r_e^* - r_e^{\text{target}}) r_i^* \\ \dot{w}_{s1} &= [((s + \frac{1}{\tau_+^s})\mathbf{I} - A)^{-1} Q]_{4,1} - \lambda[Q ((s + \frac{1}{\tau_-^s})\mathbf{I} - A^T)^{-1}]_{4,1} + r_e^* r_s^* (\tau_+^s - \lambda\tau_-^s) \\ \dot{w}_{s2} &= [((s + \frac{1}{\tau_+^s})\mathbf{I} - A)^{-1} Q]_{5,1} - \lambda[Q ((s + \frac{1}{\tau_-^s})\mathbf{I} - A^T)^{-1}]_{5,1} + r_e^* r_s^* (\tau_+^s - \lambda\tau_-^s) \end{aligned} \quad (15)$$

In above, τ_+^x and τ_-^x represent the exponent for the positive and negative domains of the stdp curves for the neuron type $x \in \{e, i, s\}$.

Dimensionality reduction for the SOM-PV network

In order to reduce the dimensionality of the system such that a phase portrait plot in $w_{11} - w_{12}$ plane is possible, we rewrite the dynamics of the rates in the Laplace domain:

$$\begin{aligned} \begin{pmatrix} \dot{E}_1 \\ \dot{E}_2 \\ \dot{I} \\ \dot{S}_1 \\ \dot{S}_2 \end{pmatrix} &= -\frac{1}{\tau} \begin{pmatrix} E_1 \\ E_2 \\ I \\ S_1 \\ S_2 \end{pmatrix} + \begin{pmatrix} \alpha & \beta & -\gamma & -k_1 & -k_2 \\ \beta & \alpha & -\gamma & -k_2 & -k_1 \\ J_1 & J_1 & -g_1 & J_1 & -qJ_1 \\ 4x & x & 0 & 0 & 0 \\ x & 4x & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} E_1 \\ E_2 \\ I \\ S_1 \\ S_2 \end{pmatrix} \\ &+ \begin{pmatrix} \sqrt{c} & 0 & 0 & 0 & 0 & k_e(\mu, \sigma) \sqrt{1-c} \\ 0 & \sqrt{c} & 0 & 0 & 0 & k_e(\mu, \sigma) \sqrt{1-c} \\ 0 & 0 & 0 & 0 & 0 & k_i(\mu, \sigma) \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ 0 \\ 0 \\ 0 \\ X_3 \end{pmatrix} \end{aligned} \quad (16)$$

where $X_1 = J k_e(\mu, \sigma) \sigma \xi_{c_1}$, $X_2 = J k_e(\mu, \sigma) \sigma \xi_{c_2}$, and $X_3 = J \sigma \xi$.
 $\alpha = k_e(\mu, \sigma) p N_e w_{11}$,
 $\beta = k_e(\mu, \sigma) p N_e w_{12}$,
 $\gamma = -4 p k_e N_i w_i$,

$$\begin{aligned}k_1 &= -4 p k_e N_s w s_1, \\k_2 &= -4 p k_e N_s w s_2, \\J_1 &= 4 p k_i N_e J, \\g_1 &= N_i/N_e g, \\q &= N_s/N_e g, \\x &= (0.8/3) J N_e k_s(\mu, \sigma).\end{aligned}$$

In the Laplace domain, the equations can be written in the following 3 dimensional form:

$$\begin{aligned}(s + \frac{1}{\tau} - \alpha) E_1 - \beta E_2 + \gamma I + k_1 S_1 + k_2 S_2 &= \sqrt{c} X_1 + k_e(\mu, \sigma) \sqrt{1-c} X_3 \\-\beta E_1 + (s + \frac{1}{\tau} - \alpha) E_2 + \gamma I + k_2 S_1 + k_1 S_2 &= \sqrt{c} X_2 + k_e(\mu, \sigma) \sqrt{1-c} X_3 \\-J_1 E_1 - J_1 E_2 + (s + \frac{1}{\tau} + g_1 J_1) I + q J_1 S_1 + q J_1 S_2 &= k_i(\mu, \sigma) X_3\end{aligned}\tag{17}$$

$$\begin{aligned}S_1 &= \frac{1}{s + \frac{1}{\tau}} (4x E_1 + x E_2) \\S_2 &= \frac{1}{s + \frac{1}{\tau}} (4x E_2 + x E_1)\end{aligned}$$

which yields

$$\begin{pmatrix} E_1 \\ E_2 \\ I \end{pmatrix} = \begin{pmatrix} s + \frac{1}{\tau} - \alpha + \frac{4x}{s + \frac{1}{\tau}} k_1 + \frac{x}{s + \frac{1}{\tau}} k_2 & -\beta + \frac{4x}{s + \frac{1}{\tau}} k_2 + \frac{x}{s + \frac{1}{\tau}} k_1 & \gamma \\ -\beta + \frac{4x}{s + \frac{1}{\tau}} k_2 + \frac{x}{s + \frac{1}{\tau}} k_1 & s + \frac{1}{\tau} - \alpha + \frac{4x}{s + \frac{1}{\tau}} k_1 + \frac{x}{s + \frac{1}{\tau}} k_2 & \gamma \\ -J_1 + \frac{5xqJ_1}{s + \frac{1}{\tau}} & -J_1 + \frac{5xqJ_1}{s + \frac{1}{\tau}} & s + \frac{1}{\tau} + g_1 J_1 \end{pmatrix}^{-1} \begin{pmatrix} \sqrt{c} X_1 + \sqrt{1-c} k_e(\mu, \sigma) X_3 \\ \sqrt{c} X_2 + \sqrt{1-c} k_e(\mu, \sigma) X_3 \\ k_i(\mu, \sigma) X_3 \end{pmatrix}\tag{18}$$

The inverse of the matrix in equation (18) can be represented as

$$\frac{1}{\det \times (s + \frac{1}{\tau})^2} \begin{pmatrix} f_1(s) & f_2(s) & f_3(s) \\ f_2(s) & f_1(s) & f_3(s) \\ f_4(s) & f_4(s) & f_5(s) \end{pmatrix}$$

where

$$\begin{aligned}f_1(s) &= (s + \frac{1}{\tau}) \{ (s + \frac{1}{\tau} + g_1 J_1) ((s + \frac{1}{\tau} - \alpha)(s + \frac{1}{\tau}) + x k_2 + 4x k_1) - \gamma (-J_1(s + \frac{1}{\tau}) + 5x q J_1) \} \\f_2(s) &= (s + \frac{1}{\tau}) \{ \gamma (-J_1(s + \frac{1}{\tau}) + 5x q J_1) - (s + \frac{1}{\tau} + g_1 J_1) (-\beta(s + \frac{1}{\tau}) + x k_1 + 4x k_2) \} \\f_3(s) &= \gamma (s + \frac{1}{\tau}) \{ (-\beta(s + \frac{1}{\tau}) + x k_1 + 4x k_2) - ((s + \frac{1}{\tau})(s + \frac{1}{\tau} - \alpha) + x k_2 + 4x k_1) \}\end{aligned}\tag{19}$$

Therefore, the solutions for $E_1(s)$ and $E_2(s)$ are

$$\begin{aligned}E_1(s) &= \frac{1}{\det} ((X_1 + \sqrt{1-c} k_e(\mu, \sigma) X_3) f_1(s) + (\sqrt{c} X_2 + \sqrt{1-c} k_e(\mu, \sigma) X_3) f_2(s) + k_i(\mu, \sigma) X_3 f_3(s)) \\E_2(s) &= \frac{1}{\det} ((X_2 + \sqrt{1-c} k_e(\mu, \sigma) X_3) f_1(s) + (\sqrt{c} X_1 + \sqrt{1-c} k_e(\mu, \sigma) X_3) f_2(s) + k_i(\mu, \sigma) X_3 f_3(s))\end{aligned}\tag{20}$$

where $\det$ is the determinant of the characteristic matrix in equation (18) (or the product of $(s - \text{eigenvalues})$ of the matrix A in the 5 dimensional system)

$$\det = \prod_{i=1}^5 (s - e_i) \quad (21)$$

The correlation functions in the Laplace domain are

$$\begin{aligned} E_1 \star E_1 &= E_1(-s).E_1(s) \\ &= (J \sigma k_e(\mu, \sigma))^2 \{term_1 + term_2 + term_3\} / (\det(s) \det(-s)) \\ term_1 &= f_1(s).f_1(-s) + f_2(s).f_2(-s) + f_3(s).f_3(-s). \left(\frac{k_i(\mu, \sigma)}{k_e(\mu, \sigma)}\right)^2 \\ term_2 &= (1 - c)(f_2(s).f_1(-s) + f_2(-s).f_1(s)) \\ term_3 &= \frac{k_i(\mu, \sigma)}{k_e(\mu, \sigma)} \sqrt{1 - c} (f_1(-s)f_3(s) + f_2(-s)f_3(s) + f_1(s)f_3(-s) + f_2(s)f_3(-s)) \end{aligned} \quad (22)$$

and

$$\begin{aligned} E_1 \star E_2 &= E_1(-s).E_2(s) \\ &= (J \sigma k_e(\mu, \sigma))^2 \{term_1 + term_2 + term_3\} / (\det(s) \det(-s)) \\ term_1 &= f_2(s).f_1(-s) + f_1(s).f_2(-s) + f_3(s).f_3(-s). \left(\frac{k_i(\mu, \sigma)}{k_e(\mu, \sigma)}\right)^2 \\ term_2 &= (1 - c)(f_1(s).f_1(-s) + f_2(-s).f_2(s)) \\ term_3 &= \frac{k_i(\mu, \sigma)}{k_e(\mu, \sigma)} \sqrt{1 - c} (f_1(-s)f_3(s) + f_2(-s)f_3(s) + f_1(s)f_3(-s) + f_2(s)f_3(-s)) \end{aligned} \quad (23)$$

However, these functions include the correlation functions for both $t > 0$ and $t < 0$. Due to the symmetry of these functions around $t = 0$, we calculate the functions for $t > 0$. To do so, we need to use the technique of partial fractions and keep the terms with components of $\det(s)$ only, and neglect those with components of $\det(-s)$. This results in

$$\begin{aligned} (E_1 \star E_1)^{t+} &= \sum_{i=1}^5 \frac{D_{11i}}{s - e_i} \\ (E_1 \star E_2)^{t+} &= \sum_{i=1}^5 \frac{D_{12i}}{s - e_i} \end{aligned} \quad (24)$$

in which D_{11i} and D_{12i} are obtained by replacing s in $E_1 \star E_1$ and $E_1 \star E_2$ by e_i , respectively. According to equation (13), the dynamics of the weights in the Laplace domain are

$$\begin{aligned} \dot{w}_{11} &= (E_1 \star E_1)^{t+}|_{s=\frac{1}{\tau_+}} - \lambda (E_1 \star E_1)^{t+}|_{s=\frac{1}{\tau_-}} + r_e^{*2} (\tau_+^e - \lambda \tau_-^e) \\ \dot{w}_{12} &= (E_1 \star E_2)^{t+}|_{s=\frac{1}{\tau_+}} - \lambda (E_1 \star E_2)^{t+}|_{s=\frac{1}{\tau_-}} + r_e^{*2} (\tau_+^e - \lambda \tau_-^e) \end{aligned} \quad (25)$$

In the equations above, the term γ contains the plastic inhibitory weight from PV to E neurons. Because homeostasis on the firing rate of E cells guarantees a constant rate, we can assume that the mean membrane potential of the E neurons are constant, and we call this term *const*. From the dynamics of the LIF neuron (6), it is straightforward to replace γ by a function of other variables in the network:

$$\gamma = k_e(\mu, \sigma) \frac{(\alpha + \beta) \frac{r_e^*}{k_e(\mu, \sigma)} - (k_1 + k_2) \frac{r_s^*}{k_e(\mu, \sigma)} + \frac{J \sigma \tau - \text{const}}{\tau}}{r_i^*} \quad (26)$$

Moreover, the dynamics of the inhibitory weights from SOM to E neurons can be written as a function of w_{11} and w_{12} (or in other words, α and β). To get this relation, we use both equations (17) and (15). It is clear that

$$\begin{aligned}\dot{w}_{s1} &= \frac{4xE_1(-s) + xE_2(-s)}{-s + 1/\tau} E_1(s)|_{s=\frac{1}{\tau_-}} - \lambda \frac{4xE_1(s) + xE_2(s)}{s + 1/\tau} E_1(-s)|_{s=\frac{1}{\tau_-}} + r_e^* r_s^* \tau_- (1 - \lambda) \\ \dot{w}_{s2} &= \frac{x E_1(-s) + 4xE_2(-s)}{-s + 1/\tau} E_1(s)|_{s=\frac{1}{\tau_-}} - \lambda \frac{x E_1(s) + 4xE_2(s)}{s + 1/\tau} E_1(-s)|_{s=\frac{1}{\tau_-}} + r_e^* r_s^* \tau_- (1 - \lambda)\end{aligned}\quad (27)$$

Since in the equation above the terms $E_1(-s)E_1(s)$ and $E_2(-s)E_1(s)$ appear, it is conceivable to assume that w_{s1} and w_{s2} are linear functions of w_{11} and w_{12} .

$$\begin{aligned}w_{s1} &= \alpha_1 w_{11} + \alpha_2 w_{12} + \alpha_3 \\ w_{s2} &= \beta_1 w_{11} + \beta_2 w_{12} + \beta_3\end{aligned}\quad (28)$$

Following this idea, we replace k_1 and k_2 with a linear combination of α and β :

$$\begin{aligned}k_1 &= 4(\alpha_1 \frac{\alpha N_s}{N_e} + \alpha_2 \frac{\beta N_s}{N_e}) + 0.4k_e(\mu, \sigma) N_s \alpha_3 \\ k_2 &= 4(\beta_1 \frac{\alpha N_s}{N_e} + \beta_2 \frac{\beta N_s}{N_e}) + 0.4k_e(\mu, \sigma) N_s \beta_3\end{aligned}\quad (29)$$

Using equations (29) and (26) in equation (25), we obtain a system that can describe the weights within and between assemblies entirely as a function of w_{11} and w_{12} , and other constants of the system. The same approach can be used to reduce the dimensionality in the 1pv network. In that case, there is no k_1 and k_2 , and only γ has to be replaced by α and β .

Tuned PV to E connections

In order to study the emergence of tuned inhibitory weights from *PV* to E_1 and E_2 , we assume that in equation (16), S_1 projects slightly more strongly to E_2 than to E_1 . To model this symmetry breaking, we assume that ϵ represents the amount of asymmetry in projections from *SOM* to *PV* populations. The *PV* population, therefore, needs to be split into two different sub-populations, called I_1 and I_2 . Since in this case we have a 6 dimensional system, we represent the new variables by a small symbol on top of the old variables. The new firing rate dynamics in the time domain would be

$$\begin{aligned}\begin{pmatrix} \dot{\tilde{E}}_1 \\ \dot{\tilde{E}}_2 \\ \dot{I}_1 \\ \dot{I}_2 \\ \dot{\tilde{S}}_1 \\ \dot{\tilde{S}}_2 \end{pmatrix} &= -\frac{1}{\tau} \begin{pmatrix} \tilde{E}_1 \\ \tilde{E}_2 \\ I_1 \\ I_2 \\ \tilde{S}_1 \\ \tilde{S}_2 \end{pmatrix} + \begin{pmatrix} \alpha & \beta & -\gamma_1 & -\gamma_2 & -k_1 & -k_2 \\ \beta & \alpha & -\gamma_2 & -\gamma_1 & -k_2 & -k_1 \\ J_1 & J_1 & -0.5 * g_1 & J_1 & -0.5 * g_1 & J_1 \\ J_1 & J_1 & -0.5 * g_1 & J_1 & -0.5 * g_1 & J_1 \\ 4x & x & 0 & 0 & 0 & 0 \\ x & 4x & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \tilde{E}_1 \\ \tilde{E}_2 \\ I_1 \\ I_2 \\ \tilde{S}_1 \\ \tilde{S}_2 \end{pmatrix} \\ &+ \begin{pmatrix} \sqrt{c} & 0 & 0 & 0 & 0 & k_e(\mu, \sigma) \sqrt{1-c} \\ 0 & \sqrt{c} & 0 & 0 & 0 & k_e(\mu, \sigma) \sqrt{1-c} \\ 0 & 0 & 0 & 0 & 0 & k_i(\mu, \sigma) \\ 0 & 0 & 0 & 0 & 0 & k_i(\mu, \sigma) \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ 0 \\ 0 \\ 0 \\ X_3 \end{pmatrix}\end{aligned}\quad (30)$$

Note that the factor 0.5 on the third column and the third row (and similarly other 0.5 factors) are due to the fact that in these new equations, I_1 and I_2 have half of the population size for I in equation (16). More precisely, the dynamics of I_1 and I_2 are

$$\begin{aligned}\dot{I}_1 &= -\frac{1}{\tau} I_1 + J_1 \tilde{E}_1 + J_1 \tilde{E}_2 - 0.5 g J_1 I_1 - 0.5 g J_1 I_2 - (q - \epsilon) J_1 \tilde{S}_1 - (q + \epsilon) J_1 \tilde{S}_2 + k_i(\mu, \sigma) X_3(t) \\ \dot{I}_2 &= -\frac{1}{\tau} I_2 + J_1 \tilde{E}_1 + J_1 \tilde{E}_2 - 0.5 g J_1 I_1 - 0.5 g J_1 I_2 - (q + \epsilon) J_1 \tilde{S}_1 - (q - \epsilon) J_1 \tilde{S}_2 + k_i(\mu, \sigma) X_3(t)\end{aligned}\quad (31)$$

Summing up and subtracting the equations in (31), one obtains

$$\begin{aligned} \dot{I}_1 + \dot{I}_2 &= -\frac{1}{\tau} (I_1 + I_2) + 2J_1 \tilde{E}_1 + 2J_1 \tilde{E}_2 - gJ_1 (I_1 + I_2) - 2qJ_1 \tilde{S}_1 - 2qJ_1 \tilde{S}_2 + 2k_i(\mu, \sigma) X_3(t) \\ \dot{I}_1 - \dot{I}_2 &= -\frac{1}{\tau} (I_1 - I_2) + 2\epsilon J_1 \tilde{S}_1 - 2\epsilon J_1 \tilde{S}_2 \end{aligned} \quad (32)$$

Comparing the first equation in (32) with the third equation in (16), one concludes $I_1 + I_2 = 2\tilde{I}$. Here, $\tilde{I}$ is a variable similar, but not equal to I in (16). Since this variable does not play role in the final result, we are not interested in getting its precise temporal dynamics, and we will leave it as an abstract variable.

The new equations for the firing rates of the excitatory neurons are

$$\begin{aligned} \dot{\tilde{E}}_1 &= -\frac{1}{\tau} \tilde{E}_1 + \alpha \tilde{E}_1 + \beta \tilde{E}_2 - \gamma_1 I_1 - \gamma_2 I_2 - k_1 \tilde{S}_1 - k_2 \tilde{S}_2 + k_e(\mu, \sigma)(\sqrt{c} X_1(t) + \sqrt{1-c} X_3(t)) \\ \dot{\tilde{E}}_2 &= -\frac{1}{\tau} \tilde{E}_2 + \beta \tilde{E}_1 + \alpha \tilde{E}_2 - \gamma_2 I_1 - \gamma_1 I_2 - k_2 \tilde{S}_1 - k_1 \tilde{S}_2 + k_e(\mu, \sigma)(\sqrt{c} X_1(t) + \sqrt{1-c} X_3(t)) \end{aligned} \quad (33)$$

The last two equations in (17) hold in this new system. Using the formula which relate $\tilde{S}_1$ and $\tilde{S}_2$ to $\tilde{E}_1$ and $\tilde{E}_2$, and also including our previous conclusion, we will get the following set of relations between I_1 , I_2 , $\tilde{I}$, $\tilde{E}_1$ and $\tilde{E}_2$ in the Laplace domain

$$\begin{aligned} s(I_1 - I_2) + \frac{1}{\tau}(I_1 - I_2) &= \frac{2\epsilon J_1(3x)}{s + \frac{1}{\tau}}(\tilde{E}_1 - \tilde{E}_2) \\ I_1 + I_2 &= 2\tilde{I} \end{aligned} \quad (34)$$

From the equation above, the solutions for I_1 and I_2 are

$$\begin{aligned} I_1 &= \frac{3\epsilon J_1 x}{(s + \frac{1}{\tau})^2}(\tilde{E}_1 - \tilde{E}_2) + \tilde{I} \\ I_2 &= \frac{-3\epsilon J_1 x}{(s + \frac{1}{\tau})^2}(\tilde{E}_1 - \tilde{E}_2) + \tilde{I} \end{aligned} \quad (35)$$

We are interested in the dynamics of γ_1 and γ_2 , the inhibitory weights from I_1 to E_1 , and E_2 , respectively. Using the formula in (13), the dynamics of γ_1 and γ_2 are

$$\begin{aligned} \dot{\gamma}_1 &= 4pk_e(\mu, \sigma)N_i \int_{-\infty}^{+\infty} pv_{stdp}(\hat{t}) \{C(I_1(t), \tilde{E}_1(t+\hat{t})) + (r_e^* - r_e^{target})r_i^*\} d\hat{t} \\ \dot{\gamma}_2 &= 4pk_e(\mu, \sigma)N_i \int_{-\infty}^{+\infty} pv_{stdp}(\hat{t}) \{C(I_1(t), \tilde{E}_2(t+\hat{t})) + (r_e^* - r_e^{target})r_i^*\} d\hat{t} \end{aligned} \quad (36)$$

Since the second terms in the equation above are similar, we ignore their precise terms for the rest of our analysis, and will show them by dc . In the Laplace domain, the dynamic part of equation (36) are

$$\begin{aligned} \dot{\gamma}_1 &= 2(I_1 \star \tilde{E}_1)^{t+}|_{s=\frac{1}{\tau_+}} + \frac{dc}{s}|_{s=\frac{1}{\tau_+}} \\ &= \left(2\tilde{I} \star \tilde{E}_1 + 2\left(\frac{3\epsilon J_1 x}{(s + \frac{1}{\tau})^2}(\tilde{E}_1 - \tilde{E}_2)\right) \star \tilde{E}_1 + \frac{dc}{s}\right)^{t+}|_{s=\frac{1}{\tau_+}} \\ \dot{\gamma}_2 &= 2(I_1 \star \tilde{E}_2)^{t+}|_{s=\frac{1}{\tau_+}} + \frac{dc}{s}|_{s=\frac{1}{\tau_+}} \\ &= \left(2I \star \tilde{E}_1 + 2\left(\frac{-3\epsilon J_1 x}{(s + \frac{1}{\tau})^2}(\tilde{E}_1 - \tilde{E}_2)\right) \star \tilde{E}_1 + \frac{dc}{s}\right)^{t+}|_{s=\frac{1}{\tau_+}} \end{aligned} \quad (37)$$

which results in

$$\begin{aligned}\dot{\gamma}_1 - \dot{\gamma}_2 &= \left(\left(2 \frac{6\epsilon J_1 x}{(s + \frac{1}{\tau})^2} (\tilde{E}_1 - \tilde{E}_2) \right) \star \tilde{E}_1 \right)^{t+} \Big|_{s=\frac{1}{\tau_+}} \\ \dot{\gamma}_1 - \dot{\gamma}_2 &\sim z_1 (\dot{w}_{11} - \dot{w}_{12}),\end{aligned}\tag{38}$$

in which z_1 is a positive scalar. Equation (38) indicates that as long as $\dot{w}_{11} - \dot{w}_{12}$ is positive, or in other words, as long as assemblies are being formed, $\dot{\gamma}_1 - \dot{\gamma}_2$ would be positive. This means that γ_1 increases faster than γ_2 , and if the initial conditions for those variables were identical at the beginning of the simulation, after some time, $\gamma_1 > \gamma_2$. This shows the emergence of tuned I_1 to $\tilde{E}_1$ connections.

Dimensionality reduction for the 1pv network

The dynamics of the 1pv network is simpler than the SOM-PV network. In this system, 3 dynamical variables E_1 , E_2 , and I , represent the dynamics of the firing rates for the two excitatory assemblies, and the shared inhibitory assembly. Moreover, there are 3 variables representing the within assembly weight dynamics (w_{11}), between assembly weight dynamics (w_{12}), and the plastic inhibitory weights from the PV neurons to the excitatory neurons (w_i).

The dynamics of the firing rates can be written in the following form

$$\begin{aligned}\begin{pmatrix} \dot{E}_1 \\ \dot{E}_2 \\ \dot{I} \end{pmatrix} &= -\frac{1}{\tau} \begin{pmatrix} E_1 \\ E_2 \\ I \end{pmatrix} + \begin{pmatrix} \alpha & \beta & -\gamma \\ \beta & \alpha & -\gamma \\ J_1 & J_1 & -g_1 J_1 \end{pmatrix} \begin{pmatrix} E_1 \\ E_2 \\ I \end{pmatrix} \\ &+ \begin{pmatrix} \sqrt{c} & 0 & 0 & k_e(\mu, \sigma) \sqrt{1-c} \\ 0 & \sqrt{c} & 0 & k_e(\mu, \sigma) \sqrt{1-c} \\ 0 & 0 & 0 & k_i(\mu, \sigma) \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ 0 \\ X_3 \end{pmatrix}\end{aligned}\tag{39}$$

where $X_1 = J k_e(\mu, \sigma) \sigma \xi_{c_1}$, $X_2 = J k_e(\mu, \sigma) \sigma \xi_{c_2}$, and $X_3 = J \sigma \xi$.

$$\alpha = k_e(\mu, \sigma) p N_e w_{11},$$

$$\beta = k_e(\mu, \sigma) p N_e w_{12},$$

$$\gamma = -4 p k_e N_i w_i,$$

$$J_1 = 4 p k_i N_e J,$$

$$g_1 = N_i / N_e g.$$

In the Laplace domain, the equations can be written in the following 3 dimensional form:

$$\begin{aligned}(s + \frac{1}{\tau} - \alpha) E_1 - \beta E_2 + \gamma I &= \sqrt{c} X_1 + k_e(\mu, \sigma) \sqrt{1-c} X_3 \\ -\beta E_1 + (s + \frac{1}{\tau} - \alpha) E_2 + \gamma I &= \sqrt{c} X_2 + k_e(\mu, \sigma) \sqrt{1-c} X_3 \\ -J_1 E_1 - J_1 E_2 + (s + \frac{1}{\tau} + g_1 J_1) I &= k_i(\mu, \sigma) X_3\end{aligned}\tag{40}$$

which yields

$$\begin{pmatrix} E_1 \\ E_2 \\ I \end{pmatrix} = \begin{pmatrix} s + \frac{1}{\tau} - \alpha & -\beta & \gamma \\ -\beta & s + \frac{1}{\tau} - \alpha & \gamma \\ -J_1 & -J_1 & s + \frac{1}{\tau} + g_1 J_1 \end{pmatrix}^{-1} \begin{pmatrix} \sqrt{c} X_1 + \sqrt{1-c} k_e(\mu, \sigma) X_3 \\ \sqrt{c} X_2 + \sqrt{1-c} k_e(\mu, \sigma) X_3 \\ k_i(\mu, \sigma) X_3 \end{pmatrix}\tag{41}$$

The characteristic equation of the system in equation (41) is:

$$\det = (s + \frac{1}{\tau} - \alpha + \beta) [(s + \frac{1}{\tau} + g_1 J_1)(s + \frac{1}{\tau} - \alpha - \beta) + 2\gamma J_1] = (s + \frac{1}{\tau} - \alpha + \beta)(s + \frac{1}{\tau} - p_1)(s + \frac{1}{\tau} - p_2)\tag{42}$$

where p_1 and p_2 are $\frac{\alpha+\beta \pm \sqrt{(-\alpha-\beta+g_1 J_1)^2+4(\alpha+\beta)g_1 J_1-8\gamma J_1}}{2}$
The inverse of the matrix in equation (41) can be represented as

$$\frac{1}{\det} \begin{pmatrix} f_1(s) & f_2(s) & f_3(s) \\ f_2(s) & f_1(s) & f_3(s) \\ f_4(s) & f_4(s) & f_5(s) \end{pmatrix}$$

where

$$\begin{aligned} f_1(s) &= (s + \frac{1}{\tau} - \alpha)(s + \frac{1}{\tau} + g_1 J_1) + J_1 \gamma \\ f_2(s) &= \beta(s + \frac{1}{\tau} + g_1 J_1) - J_1 \gamma \\ f_3(s) &= -\gamma(s + \frac{1}{\tau} - \alpha + \beta) \end{aligned} \quad (43)$$

Therefore, the solutions for $E_1(s)$ and $E_2(s)$ are

$$\begin{aligned} E_1(s) &= \frac{1}{\det} ((X_1 + \sqrt{1-c} k_e(\mu, \sigma) X_3) f_1(s) + (\sqrt{c} X_2 + \sqrt{1-c} k_e(\mu, \sigma) X_3) f_2(s) + k_i(\mu, \sigma) X_3 f_3(s)) \\ E_2(s) &= \frac{1}{\det} ((X_2 + \sqrt{1-c} k_e(\mu, \sigma) X_3) f_1(s) + (\sqrt{c} X_1 + \sqrt{1-c} k_e(\mu, \sigma) X_3) f_2(s) + k_i(\mu, \sigma) X_3 f_3(s)) \end{aligned} \quad (44)$$

The correlation functions in the Laplace domain are

$$\begin{aligned} E_1 \star E_1 &= E_1(-s).E_1(s) \\ &= (J \sigma k_e(\mu, \sigma))^2 \{term_1 + term_2 + term_3\} / (\det(s) \det(-s)) \\ term_1 &= f_1(s).f_1(-s) + f_2(s).f_2(-s) + f_3(s).f_3(-s).(\frac{k_i(\mu, \sigma)}{k_e(\mu, \sigma)})^2 \\ term_2 &= (1-c)(f_2(s).f_1(-s) + f_2(-s).f_1(s)) \\ term_3 &= \frac{k_i(\mu, \sigma)}{k_e(\mu, \sigma)} \sqrt{1-c} (f_1(-s)f_3(s) + f_2(-s)f_3(s) + f_1(s)f_3(-s) + f_2(s)f_3(-s)) \end{aligned} \quad (45)$$

and

$$\begin{aligned} E_1 \star E_2 &= E_1(-s).E_2(s) \\ &= (J \sigma k_e(\mu, \sigma))^2 \{term_1 + term_2 + term_3\} / (\det(s) \det(-s)) \\ term_1 &= f_2(s).f_1(-s) + f_1(s).f_2(-s) + f_3(s).f_3(-s).(\frac{k_i(\mu, \sigma)}{k_e(\mu, \sigma)})^2 \\ term_2 &= (1-c)(f_1(s).f_1(-s) + f_2(-s).f_2(s)) \\ term_3 &= \frac{k_i(\mu, \sigma)}{k_e(\mu, \sigma)} \sqrt{1-c} (f_1(-s)f_3(s) + f_2(-s)f_3(s) + f_1(s)f_3(-s) + f_2(s)f_3(-s)) \end{aligned} \quad (46)$$

The solutions for E_1 and E_2 , and hence the correlation functions are similar to those for the SOM-PV network; however, the functions $f_i(\cdot)$, $i = 1...5$ are different in this system.

Similar to the solution for the SOM-PV network, the result in equation (46) includes the correlation functions for both $t > 0$ and $t < 0$. Due to the symmetry of these functions around $t = 0$, we calculate the functions for $t > 0$. To do so, we need to use the technique of partial fractions and keep the terms with components of $\det(s)$ only, and neglect those with components of $\det(-s)$. This results in

$$\begin{aligned}(E_1 \star E_1)^{t+} &= \sum_{i=1}^3 \frac{D_{11i}}{s - e_i} \\ (E_1 \star E_2)^{t+} &= \sum_{i=1}^3 \frac{D_{12i}}{s - e_i}\end{aligned}\tag{47}$$

in which D_{11i} and D_{12i} are obtained by replacing s in $E_1 \star E_1$ and $E_1 \star E_2$ by the solutions of the characteristic equation (42), respectively. According to equation (13), the dynamics of the weights in the Laplace domain are

$$\begin{aligned}\dot{w}_{11} &= (E_1 \star E_1)^{t+}|_{s=\frac{1}{\tau_+}} - \lambda (E_1 \star E_1)^{t+}|_{s=\frac{1}{\tau_-}} + r_e^{*2} (\tau_+^e - \lambda \tau_-^e) \\ \dot{w}_{12} &= (E_1 \star E_2)^{t+}|_{s=\frac{1}{\tau_+}} - \lambda (E_1 \star E_2)^{t+}|_{s=\frac{1}{\tau_-}} + r_e^{*2} (\tau_+^e - \lambda \tau_-^e)\end{aligned}\tag{48}$$

In the equations above, the term γ contains the plastic inhibitory weight from PV to E neurons. Because homeostasis on the firing rate of E cells guarantees a constant rate, we can assume that the mean membrane potential of the E neurons are constant, and we call this term *const*. From the dynamics of the LIF neuron (6), it is straightforward to replace γ by a function of other variables in the network:

$$\gamma = k_e(\mu, \sigma) \frac{(\alpha + \beta) \frac{r_e^*}{k_e(\mu, \sigma)} + \frac{J \sigma \tau - \text{const}}{\tau}}{r_i^*}\tag{49}$$

After this replacement, equation (48) describes the dynamics of w_{11} and w_{12} entirely based on these variables in a 2-dimensional system.

Network parameters

Network parameters are represented in the following table.

Variable	Value	Explanation
N_e	400	Number of excitatory neurons in each assembly
N_i	80	Number of inhibitory neurons in the SOM-PV network
N_i	200	Number of inhibitory neurons in the 1pv and 2PV network
N_s	60	Number of inhibitory neurons in each SOM population
J	0.06 mV	Epsp amplitude for fixed connections
g	10	Scalar for the fixed IPSP amplitudes
I_e	200 pA	External DC current to the SOM neurons
θ	20 mV	Threshold for the membrane potential
τ	20 ms	membrane potential time constant
C	250 pF	Capacitance of the membrane potential
p	0.1	probability of connection between excitatory neurons
p_t	0.8/3	probability of tuned E to SOM connections in the SOM-PV network
p_t	0.8/3	probability of tuned E to PV connections in the 2PV network
τ_-	30 ms	time constant of the depression side of SOM stdp
τ_-	30 ms	time constant of the depression side of the excitatory stdp
τ_+	30 ms	time constant of the potentiation side of the SOM stdp
τ_+	15 ms	time constant of the potentiation side of the excitatory stdp
τ_+	20 ms	time constant of the symmetric PV to E stdp
η_e	0.0025 ms^{-1}	time constant of $E \rightarrow E$ connections
η_s	0.01 ms^{-1}	time constant of $SOM \rightarrow E$ connections
η_p	0.01 ms^{-1}	time constant of $PV \rightarrow E$ connections
α_1	-3.593	appears in equation (29)
α_2	-14.16799	appears in equation (29)
α_3	1.7141	appears in equation (29)
β_1	6.6686	appears in equation (29)
β_2	-2.9763	appears in equation (29)
β_3	0.4271	appears in equation (29)
c	0.3	percentage of shared input from the total input
σ	$\sqrt{25000}$	standard deviation of the external Poisson process

References

- [1] Kenneth D Harris and Thomas D Mrsic-Flogel. Cortical connectivity and sensory coding. *Nature*, 503(7474):51–58, 2013.
- [2] Anthony Holtmaat and Pico Caroni. Functional and structural underpinnings of neuronal assembly formation in learning. *Nature neuroscience*, 19(12):1553, 2016.
- [3] Ho Ko, Lee Cossell, Chiara Baragli, Jan Antolik, Claudia Clopath, Sonja B. Hofer, and Thomas D. Mrsic-Flogel. The emergence of functional microcircuits in visual cortex. *Nature*, 496(7443):96–100, 2013.
- [4] Wei-Chung Allen Lee, Vincent Bonin, Michael Reed, Brett J Graham, Greg Hood, Katie Glattfelder, and R Clay Reid. Anatomy and function of an excitatory network in the visual cortex. *Nature*, 532(7599):370–374, 2016.
- [5] Lee Cossell, Maria Florencia Iacaruso, Dylan R Muir, Rachael Houlton, Elie N Sader, Ho Ko, Sonja B Hofer, and Thomas D Mrsic-Flogel. Functional organization of excitatory synaptic strength in primary visual cortex. *Nature*, 518(7539):399–403, 2015.

- [6] Jae Eun Kang Miller, Inbal Ayzenshtat, Luis Carrillo-Reid, and Rafael Yuste. Visual stimuli recruit intrinsically generated cortical ensembles. *Proceedings of the National Academy of Sciences of the United States of America*, 111(38):E4053–E4061, 2014.
- [7] Arnaud Malvache, Susanne Reichinnek, Vincent Villette, Caroline Haimerl, and Rosa Cossart. Awake hippocampal reactivations project onto orthogonal neuronal assemblies. *Science*, 353(6305):1280–1283, 2016.
- [8] John Lisman, Katherine Cooper, Megha Sehgal, and Alcino J. Silva. Memory formation depends on both synapse-specific modifications of synaptic strength and cell-specific increases in excitability. *Nature Neuroscience*, 21(3):309–314, 2018.
- [9] Kenneth D Harris. Neural signatures of cell assembly organization. *Nature Reviews Neuroscience*, 6(5):399–407, 2005.
- [10] Luis Carrillo-Reid, Weijian Yang, Yuki Bando, Darcy S Peterka, and Rafael Yuste. Imprinting Cortical Ensembles. *Science*, 353(6300):691–694, 2016.
- [11] Gabriel Koch Ocker and Brent Doiron. Training and spontaneous reinforcement of neuronal assemblies by spike timing plasticity. *Cerebral Cortex*, 29(3):937–951, 2019.
- [12] Ashok Litwin-Kumar and Brent Doiron. Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nature Communications*, 5(May):5319, 2014.
- [13] Friedemann Zenke, Everton J. Agnes, and Wulfram Gerstner. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature Communications*, 6, 2015.
- [14] Gianluigi Mongillo, Daniel J Amit, and Nicolas Brunel. Retrospective and prospective persistent activity induced by hebbian learning in a recurrent cortical network. *European Journal of Neuroscience*, 18(7):2011–2024, 2003.
- [15] Carl Van Vreeswijk and Haim Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726, 1996.
- [16] Bilal Haider, Alvaro Duque, Andrea R Hasenstaub, and David A McCormick. Neocortical Network Activity In Vivo Is Generated through a Dynamic Balance of Excitation and Inhibition. *The Journal of Neuroscience*, 26(17):4535–4545, 2006.
- [17] Michael Okun and Ilan Lampl. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature Neuroscience*, 11(5):535–537, 2008.
- [18] Suixin Deng, Junlong Li, Quansheng He, Xiaoxue Zhang, Jie Zhu, Liang Li, Zhen Mi, Xiufeng Yang, Man Jiang, Qiang Dong, Ying Mao, and Yousheng Shu. Regulation of Recurrent Inhibition by Asynchronous Glutamate Release in Neocortex. *Neuron*, 105(3):522–533.e4, 2020.
- [19] Daniel J Amit and Nicolas Brunel. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, 7(3):237–252, 1997.
- [20] Hirofumi Ozeki, Ian M Finn, Evan S Schaffer, Kenneth D Miller, and David Ferster. Inhibitory stabilization of the cortical network underlies visual surround suppression. *Neuron*, 62(4):578–592, 2009.
- [21] Nicolas Brunel. Phase diagrams of sparsely connected networks of excitatory and inhibitory spiking neurons. *Neurocomputing*, 32-33:307–312, 2000.
- [22] Alexander S. Ecker, Philipp Berens, Georgios A. Keliris, Matthias Bethge, Nikos K. Logothetis, and Andreas S. Tolias. Decorrelated neuronal firing in cortical microcircuits. *Science*, 327(5965):584–587, 2010.
- [23] Alfonso Renart, Jaime De La Rocha, Peter Bartho, Liad Hollender, Néstor Parga, Alex Reyes, and Kenneth D. Harris. The asynchronous state in cortical circuits. *Science*, 327(5965):587–590, 2010.
- [24] Tom Tetzlaff, Moritz Helias, Gaute T Einevoll, and Markus Diesmann. Decorrelation of neural-network activity by inhibitory feedback. *PLoS computational biology*, 8(8):e1002596, aug 2012.
- [25] Robert Rosenbaum, Matthew A. Smith, Adam Kohn, Jonathan E. Rubin, and Brent Doiron. The spatial structure of correlated neuronal variability. *Nature Neuroscience*, 20(1):107–114, 2017.

- [26] Lisa Roux and Gyrgy Buzski. Tasks for inhibitory interneurons in intact brain circuits. *Neuropharmacology*, 88:10–23, 2015. GABAergic Signaling in Health and Disease.
- [27] Thomas Rost, Moritz Deger, and Martin P. Nawrot. Winnerless competition in clustered balanced networks: inhibitory assemblies do the trick. *Biological Cybernetics*, 112(1-2):81–98, 2018.
- [28] Fereshteh Lagzi and Stefan Rotter. Dynamics of competition between sub-networks of spiking neuronal networks in the balanced state. *PLoS One*, September:1–29, 2015.
- [29] Ran Darshan, Carl Van Vreeswijk, and David Hansel. Strength of correlations in strongly recurrent neuronal networks. *Physical Review X*, 8(3):031072, 2018.
- [30] Fereshteh Lagzi, Fatihcan M. Atay, and Stefan Rotter. Bifurcation analysis of the dynamics of interacting subnetworks of a spiking network. *Scientific Reports*, 9(1):1–17, 2019.
- [31] Feng Han, Natalia Caporale, and Yang Dan. Reverberation of recent visual experience in spontaneous cortical waves. *Neuron*, 60(2):321–327, 2008.
- [32] Misha Tsodyks, Tal Kenet, Amiram Grinvald, and A Arieli. Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science*, 286(5446):1943–1946, 1999.
- [33] Robin Tremblay, Soohyun Lee, and Bernardo Rudy. Gabaergic interneurons in the neocortex: from cellular properties to circuits. *Neuron*, 91(2):260–292, 2016.
- [34] Adam Kepecs and Gordon Fishell. Interneuron cell types are fit to function. *Nature*, 505(7483):318–326, 2014.
- [35] Jessica A Cardin. Inhibitory interneurons regulate temporal precision and correlations in cortical circuits. *Trends in neurosciences*, 41(10):689–700, 2018.
- [36] Ryoma Hattori, Kishore V. Kuchibhotla, Robert C. Froemke, and Takaki Komiyama. Functions and dysfunctions of neocortical inhibitory neuron subtypes. *Nature Neuroscience*, 20(9):1199–1208, 2017.
- [37] Katherine C Wood, Jennifer M Blackwell, and Maria Neimark Geffen. Cortical inhibitory interneurons control sensory processing. *Current opinion in neurobiology*, 46:200–207, 2017.
- [38] Melanie A Woodin, Karunesh Ganguly, and Mu-Ming Poo. Coincident Pre-and Postsynaptic Activity Modifies GABAergic Synapses by Postsynaptic Changes in Cl Transporter Activity At many excitatory synapses, coincident pre-and postsynaptic spiking of different temporal patterns has been shown to result in either . *Neuron*, 39:807–820, 2003.
- [39] Julie S. Haas, Thomas Nowotny, and H. D.I. Abarbanel. Spike-timing-dependent plasticity of inhibitory synapses in the entorhinal cortex. *Journal of Neurophysiology*, 96(6):3305–3313, 2006.
- [40] Alexandre W. Moreau and Dimitri M. Kullmann. NMDA receptor-dependent function and plasticity in inhibitory circuits. *Neuropharmacology*, 74:23–31, 2013.
- [41] Dimitri M. Kullmann, Alexandre W. Moreau, Yamina Bakiri, and Elizabeth Nicholson. Plasticity of Inhibition. *Neuron*, 75(6):951–962, 2012.
- [42] Simon X. Chen, An Na Kim, Andrew J. Peters, and Takaki Komiyama. Subtype-specific plasticity of inhibitory circuits in motor cortex during motor learning. *Nature Neuroscience*, 18(8):1109–1115, 2015.
- [43] James A. D’amour and Robert C. Froemke. Inhibitory and excitatory spike-timing-dependent plasticity in the auditory cortex. *Neuron*, 86(2):514–528, 2015.
- [44] Guillaume Hennequin, Everton J Agnes, and Tim P Vogels. Inhibitory plasticity: balance, control, and codependence. *Annual review of neuroscience*, 40:557–579, 2017.
- [45] Tara Keck, Taro Toyozumi, Lu Chen, Brent Doiron, Daniel E Feldman, Kevin Fox, Wulfram Gerstner, Philip G Haydon, Mark Hübener, Hey-Kyoung Lee, et al. Integrating hebbian and homeostatic plasticity: the current state of the field and future research directions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1715):20160158, 2017.

- [46] Natalia Caporale and Yang Dan. Spike timing-dependent plasticity: a hebbian learning rule. *Annu. Rev. Neurosci.*, 31:25–46, 2008.
- [47] Carl D Holmgren and Yuri Zilberter. Coincident spiking activity induces long-term changes in inhibition of neocortical pyramidal cells. *Journal of Neuroscience*, 21(20):8270–8277, 2001.
- [48] Ashok Litwin-Kumar and Brent Doiron. Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nature communications*, 5(1):1–12, 2014.
- [49] Alex C. Kwan and Yang Dan. Dissection of cortical microcircuits by single-neuron stimulation in vivo. *Current Biology*, 22(16):1459–1467, 2012.
- [50] Richard Kempter, Wulfram Gerstner, and J. Leo van Hemmen. Hebbian learning and spiking neurons. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 59(4):4498–4514, 1999.
- [51] Gabriel Koch Ocker, Ashok Litwin-Kumar, and Brent Doiron. Self-Organization of Microcircuits in Networks of Spiking Neurons with Plastic Synapses. *PLoS Computational Biology*, 11(8):1–40, 2015.
- [52] Tzitzitlini Alejandre-García, Samuel Kim, Jesús Pérez-Ortega, and Rafael Yuste. Intrinsic excitability mechanisms of neuronal ensemble formation. *bioRxiv*, page 1, 2020.
- [53] Avital Adler, Ruohe Zhao, Myung Eun Shin, Ryohei Yasuda, and Wen Biao Gan. Somatostatin-Expressing Interneurons Enable and Maintain Learning-Dependent Sequential Activation of Pyramidal Neurons. *Neuron*, 102(1):202–216.e7, 2019.
- [54] Melissa Serrano and Pico Caroni. Managing Neuronal Ensembles: Somatostatin Interneuron Subpopulations Shape and Protect Cortical Neuronal Ensembles for Learning. *Neuron*, 102(1):6–8, 2019.
- [55] Matt Udakis, Victor Pedrosa, Sophie E.L. Chamberlain, Claudia Clopath, and Jack R. Mellor. Interneuron-specific plasticity at parvalbumin and somatostatin inhibitory synapses onto CA1 pyramidal neurons shapes hippocampal output. *Nature Communications*, 11(1), 2020.
- [56] Gilad Silberberg and Henry Markram. Disynaptic Inhibition between Neocortical Pyramidal Cells Mediated by Martinotti Cells. *Neuron*, 53(5):735–746, 2007.
- [57] Gilad Silberberg. Polysynaptic subcircuits in the neocortex: spatial and temporal diversity. *Current Opinion in Neurobiology*, 18(3):332–337, 2008.
- [58] Petr Znamenskiy, Mean Hwan Kim, Dylan R. Muir, Maria Florencia Iacaruso, Sonja B. Hofer, and Thomas D. Mrsic-Flogel. Functional selectivity and specific connectivity of inhibitory neurons in primary visual cortex. *bioRxiv*, 2018.
- [59] Wen Pei Ma, Bao Hua Liu, Ya Tang Li, Z. Josh Huang, Li I. Zhang, and Huizhong W. Tao. Visual representations by cortical somatostatin inhibitory neurons - Selective but with weak and delayed responses. *Journal of Neuroscience*, 30(43):14371–14379, 2010.
- [60] James C.H. Cottam, Spencer L. Smith, and Michael Häusser. Target-specific effects of somatostatin-expressing interneurons on neocortical visual processing. *Journal of Neuroscience*, 33(50):19567–19578, 2013.
- [61] Aaron M. Kerlin, Mark L. Andermann, Vladimir K. Berezovskii, and R. Clay Reid. Broadly Tuned Response Properties of Diverse Inhibitory Neuron Subtypes in Mouse Visual Cortex. *Neuron*, 67(5):858–871, 2010.
- [62] Bassam V. Atallah, William Bruns, Matteo Carandini, and Massimo Scanziani. Parvalbumin-Expressing Interneurons Linearly Transform Cortical Responses to Visual Stimuli. *Neuron*, 73(1):159–170, 2012.
- [63] Elodie Fino and Rafael Yuste. Dense inhibitory connectivity in neocortex. *Neuron*, 69(6):1188–1203, 2011.
- [64] Adil G. Khan, Jasper Poort, Angus Chadwick, Antonin Blot, Maneesh Sahani, Thomas D. Mrsic-Flogel, and Sonja B. Hofer. Distinct learning-induced changes in stimulus selectivity and interactions of GABAergic interneuron classes in visual cortex. *Nature Neuroscience*, 21(6):851–859, 2018.
- [65] Mingshan Xue, Bassam V. Atallah, and Massimo Scanziani. Equalizing excitation-inhibition ratios across visual cortical neurons. *Nature*, 511(7511):596–600, 2014.

- [66] Hiroki Taniguchi, Miao He, Priscilla Wu, Sangyong Kim, Raehum Paik, Ken Sugino, Duda Kvitsani, Yu Fu, Jiangteng Lu, Ying Lin, et al. A resource of cre driver lines for genetic targeting of gabaergic neurons in cerebral cortex. *Neuron*, 71(6):995–1013, 2011.
- [67] Linda Madisen, Tianyi Mao, Henner Koch, Jia-min Zhuo, Antal Berenyi, Shigeyoshi Fujisawa, Yun-Wei A Hsu, Alfredo J Garcia, Xuan Gu, Sebastien Zanella, et al. A toolbox of cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nature neuroscience*, 15(5):793–802, 2012.