Recurrent connectivity structure controls the emergence of co-tuned excitation and inhibition

Emmanouil Giannakakis, Oleg Vinogradov, Victor Buendía, Anna Levina
(Dated: February 27, 2023)

Experimental studies have shown that in cortical neurons, excitatory and inhibitory incoming currents are strongly correlated, which is hypothesized to be essential for efficient computations. Additionally, cortical neurons exhibit strong preference to particular stimuli, which combined with the co-variability of excitatory and inhibitory inputs indicates a detailed co-tuning of the corresponding populations. Such co-tuning is hypothesized to emerge during development in a self-organized manner. Indeed, theoretical studies have demonstrated that a combination of plasticity rules could lead to the emergence of E/I co-tuning in neurons driven by low noise signals from feedforward connections. However, cortical signals are very noisy and originate in highly recurrent networks, which raises a question on the ability of known plasticity mechanisms to self-organize co-tuned connectivity. We demonstrate that high noise levels combined with random recurrence destroy co-tuning. However, we demonstrate that introducing structure in the connectivity patterns of the recurrent E/I network, the recurrence does not hinder but enhances the formation of the co-tuned selectivity. We employ a combination of analytical methods and simulation-based inference to uncover constraints on the recurrent connectivity that allow E/I co-tuning to emerge. We find that stronger excitatory connectivity within similarly tuned neurons, combined with more homogeneous inhibitory connectivity enhances the ability of plasticity to produce co-tuning in an upstream population. Our results suggest that structured recurrent connectivity controls information propagation and can enhance the ability of synaptic plasticity to learn input-output relationships in higher brain areas.

I. INTRODUCTION

Input selectivity, the ability of neurons to respond differently to distinct stimuli, is a prevalent mechanism for encoding information in the nervous system. This selectivity can range from simple orientation selectivity in lower sensory areas to more complex spatiotemporal pattern selectivity in higher areas [1]. Such selectivity is shown to be self-organized under the influence of structured input, enabling, for example, the emergence of visual orientation preference in non-visual sensory areas upon rewiring [2] or changing the whiskers representation in the barrel cortex of rats depending on the level of sensory input [3]. The mechanisms underlying the emergence of input selectivity have been the subject of extensive investigation, both through experimental and computational modeling studies [4–7].

Although initially, the stimulus-selectivity was primarily attributed to the excitatory neurons and their network structure, we now know that inhibitory neurons are also tuned to stimuli, and the coordination of the E/I currents is a central component of efficient neural computation [8, 9]. In particular, it has been shown that excitatory and inhibitory inputs are often correlated [10], with preferred stimuli eliciting stronger excitatory and inhibitory responses within relatively small time windows. [11, 12]. This co-tuning of excitation and inhibition is theorized to be beneficial for a variety of computations such as gain control [13, 14], visual surround suppression [15, 16], novelty detection [17] and optimal spike-timing [9, 18].

Although it is still unclear how E/I co-tuning emerges, the dominant view is that it arises via the interaction of several synaptic plasticity mechanisms [19], a hypothesis that has been reinforced by the findings of multiple theoretical studies over the last decade. First, it has been demonstrated that different inhibitory plasticity rules can match static excitatory connectivity [20–23]. More recently, it was also shown that various combinations of plasticity and diverse normalisation mechanisms allow for simultaneous development of matching excitatory and inhibitory connectivity in feedforward settings [6, 24], as well as the simultaneous learning of excitatory and inhibitory connectivity in recurrent settings [25–31].

Most of these plasticity studies use Hebbian-like rules that learn the statistical dependencies of different inputs. The statistical structure of these inputs can be assumed to have specific features in areas that receive direct sensory signals, but this assumption becomes less viable in higher areas where neurons receive inputs from highly recurrent and noisy networks. Thus a question arises on whether the outputs of realistically structured recurrent networks can have the necessary statistical structure for E/I co-tuning to emerge via synaptic plasticity.

Here, we investigate how the development of co-tuned excitation and inhibition via neuronal plasticity is affected by biologically plausible levels of noise and recurrence. We combine excitatory and inhibitory plasticity rules [20, 24, 32, 33] in a spiking network to develop detailed co-tuning of excitatory and inhibitory connectivity. We demonstrate that the ability of these plasticity mechanisms to create co-tuning is significantly reduced in the presence of noise and random recurrent connectivity. We subsequently build a simple neuronal mass model exhibiting the same dependence but allowing for analytical understanding of the underlying phenomena. We show that the effects of recurrence and

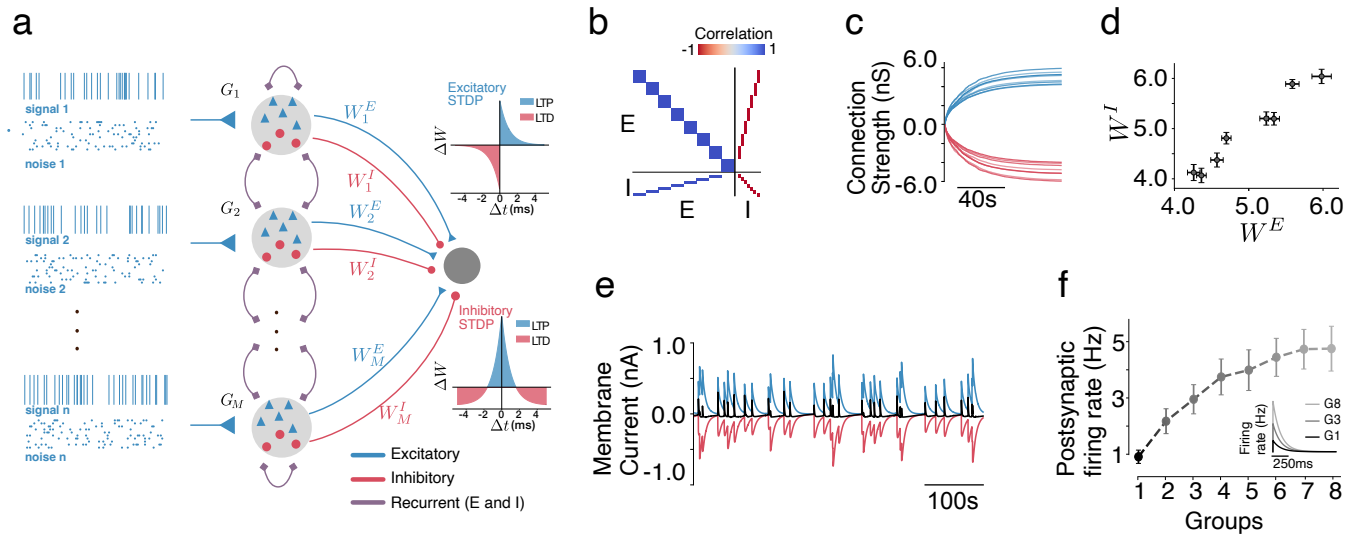

FIG. 1. **Emergence of tuning in a feedforward network a.** A diagram of the network. The purple recurrent connections are absent in the feedforward version of the model (panels b-f), but will be included in later sections. **b.** The modified (the sign of inhibitory activities is inverted) covariance matrix determines the convergence point of the plasticity protocol. Here we see a near-optimal matrix that leads to very clear co-tuning. **c.** The development of E and I weights in a feedforward network with very low noise. **d.** The relation between E/I weights after 100s in C demonstrates co-tuning: most groups have clearly distinct weights, and the E and I weights of each group match each other. **e.** Time-course of the currents incoming onto the post-synaptic neuron after the convergence of plasticity. The blue E current is canceled after a brief delay by the red I current; the black trace depicts the sum of the currents. **f.** The post-synaptic neuron's response to a brief pulse of input current given to different groups differs in the resulting firing rates due to the input selectivity.

noise on excitatory/inhibitory co-tuning can be ameliorated by the formation of synapse-type specific assemblies of neurons. The near-optimal connectivity configurations involve strong excitatory assemblies and weaker inhibitory assemblies. Finally, we demonstrate that assemblies allow co-tuning to emerge even in sparsely connected networks if their relative strengths are adjusted for the sparsity level.

II. RESULTS

A. Co-tuning and its self-organization by synaptic plasticity in a low-noise feedforward setting

We use previously studied plasticity mechanisms to generate diverse co-tuned excitatory/inhibitory weights in a feedforward, low-noise setting. We model a single read-out postsynaptic unit driven by a population of $N = 1000$ neurons. The pre-synaptic population is divided into M groups G_i, $i \in \{1, \ldots, M\}$. Each group is comprised of 80% excitatory and 20% inhibitory neurons, which are driven by an identical, group-specific Poisson spike train — a shared external input. Additionally, each neuron receives low-intensity independent external noise [20]. This setting leads to highly correlated firing among neurons of the same input group; it is a commonly used simplified setting for studying the effect of different plasticity rules Fig. 1a.

The post-synaptic neuron can discriminate the inputs from different groups (by responding with a different firing rate) if the feedforward projections from each group are sufficiently diverse. Specifically, groups with stronger connections will elicit stronger post-synaptic responses upon activation, while groups with weaker connections will elicit weak or no response upon activation Fig. 1f. Moreover, connections from neurons with highly correlated firing (i.e., from the same group) should have a similar strength. To quantify this feature of the network, we define a diversity metric:

$$D = 1 - \frac{1}{M \cdot \mathrm{Std}\,(W^E)} \sum_{i=1}^{M} \mathrm{Std}(W^E_{G_i}), \tag{1}$$

where W^k, $k \in \{E, I\}$ is the set of (E or I) feedforward connection weights and $W^k_{G_i}$, $k \in \{E, I\}$ is the subset of

(E or I) feedforward connection weights from input group i. Diversity $D \in [0, 1]$ equals unity when the feedforward connections from each group are similar within a group, but different across groups; D is close to zero when there is no difference between groups.

An important network feature that helps optimize the network's coding capabilities is the balance between incoming excitatory and inhibitory input currents [8, 34]. Balanced excitatory/inhibitory inputs emerge both structurally and dynamically in recurrent networks [12, 35, 36]. For a post-synaptic neuron to be balanced, the average inhibitory current must be equal to the average excitatory current. In the simple setting we are studying, this can be achieved in two ways. First, by setting all the inhibitory connections to a constant such that their sum equals the sum of all excitatory connections. Alternatively, one can set the inhibitory connections of each group to match the group's excitatory connections. The latter setting creates a detailed balance between excitation and inhibition [13, 20] which results in the canceling of the excitatory and inhibitory inputs following a small delay generated by the difference in the synaptic timescales, Fig. 1e. This specific type of E/I balance has been linked to efficient coding [8, 9] and the processing of multiple signals [13]. To quantify detailed balance, we use the Pearson correlation coefficient between the mean excitatory and inhibitory weights of each group,

$$B = \frac{\mathrm{Cov}\left(\langle W_G^E \rangle, \langle W_G^I \rangle, \right)}{\mathrm{Std}\left(\langle W_G^E \rangle\right) \cdot \mathrm{Std}\left(\langle W_G^I \rangle\right)}, \tag{2}$$

where $\langle W_G^k \rangle = \left(\langle W_{G_1}^k \rangle, \langle W_{G_2}^k \rangle, \ldots, \langle W_{G_m}^k \rangle\right)$, $k \in \{I, E\}$ and $\langle W_{G_i}^k \rangle$ is the average projection weight from the excitatory ($k = E$) or inhibitory ($k = I$) neurons in group i. In networks with high balance B the strength of incoming E and I currents is highly correlated.

We then verify that high diversity ($D \approx 1$) and detailed balance ($B \approx 1$) can organically emerge via a combination of plasticity mechanisms in the feedforward connections. Specifically, the excitatory connections follow a triplet STDP rule [33] that implements Hebbian learning using triplets of spikes. The inhibitory connections follow a homeostatic learning rule that adjusts inhibitory weights in order to maintain a constant post-synaptic firing rate [20]. We additionally use a competitive normalization mechanism in both inhibitory and excitatory connections that has been applied previously to networks of rate units [24]. This normalization mechanism, in addition to preventing runaway plasticity, also leads to the emergence of detailed E/I co-tuning by amplifying small transient differences between the firing rates of different input groups which leads to the development of distinct connections. This plasticity protocol consistently generates near-perfect detailed E-I balance and creates strong input selectivity, Fig. 1c-d.

The point at which the feedforward weights converge is fully determined by the covariance matrix of the presynaptic neurons' activities. Specifically, it has been demonstrated [6, 20, 24] that the fixed points of the weight dynamics are eigenvectors of the modified covariance matrix:

$$\overline{C} = \left\langle \begin{pmatrix} EE^T & -IE^T \\ EI^T & -II^T \end{pmatrix} \right\rangle \tag{3}$$

where E and I the firing rates of the presynaptic excitatory and inhibitory neurons, respectively, and $\langle \dot{\rangle}$ stands for the average over time, Fig. 1b. We verify that in our model the feedforward weights indeed converge to an eigenvector of this matrix (see Appendix B).

B. Noise and recurrent connectivity compromise the ability of STDP to produce E/I co-tuning

Strictly feed-forward and low noise circuits are unrealistic approximations whose dynamics might deviate critically from the ones observed in real brain networks. Thus, we introduce either noise, or recurrent connectivity, both of which are ubiquitously present in biological networks [37, 38], in our network and investigate how they affect the emerging E/I co-tuning by changing the structure of the activity covariance matrix.

Noise decorrelates the activity of neurons from the same input group, which reduces weight diversity. We control noise level by changing the fraction of signal spikes that all neurons of the same group receive and the noise spikes coming from an independent Poisson process. As the ratio of noise to signal increases, the cross-correlations within each input group decrease, while the cross-correlations between neurons of different input groups remain very low, Fig. 2b. The effect of this in-group decorrelation is an increased variability in the learned projections to the postsynaptic neuron from neurons of the same input group and thus a decrease in the resulting diversity. At the same time, increased noise has only a small effect on the correlation between E and I populations as measured by the balance metric, which visibly declines only once the noise becomes overwhelmingly stronger than the input (more than 80% incoming spikes are not shared between neurons of the same group), Fig. 2a.

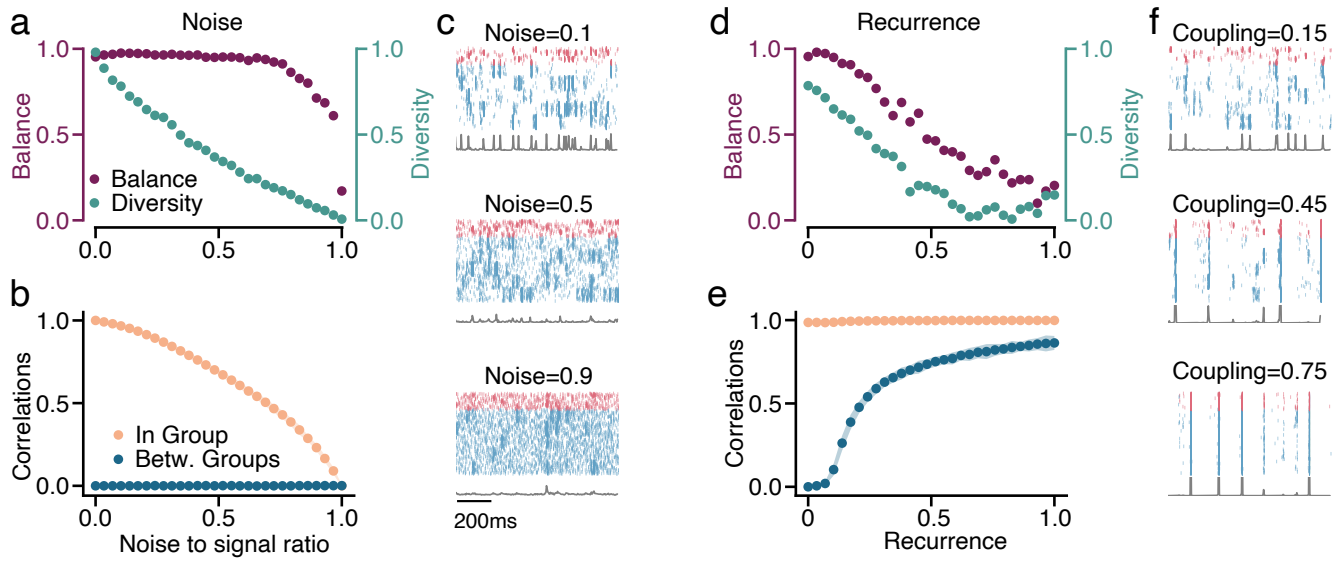

FIG. 2. **Noise and Recurrence Destroy E/I Co-Tuning.** **a.** An increase in the noise the network receives leads to a reduction in diversity (green) and after some point also balance (purple). **b.** This decrease is caused by increase in in-group correlation. **c.** As the noise increases (indicated above the panel) spiking activity becomes more asynchronous. **d.** An increase in the recurrent coupling strength leads to a rapid decrease in balance and diversity, which is caused by (**e.**) an increase of the between-group correlation. **f.** The spiking activity becomes more synchronous as the coupling strength increases.

Recurrent connectivity in the pre-synaptic network introduces cross-correlations between the neurons from different input groups, which compromises both the diversity and balance. To test the extent of this impact, we connect all-to-all N presynaptic neurons(creating a fully connected recurrent network) and use the coupling strength W as a control parameter. Changing W, we can control the ratio between the input received from the feedforward connections (whose rate and connection strength is fixed) and the other neurons in the network via recurrent connections. The recurrent connectivity increases cross-correlations between groups while maintaining correlation within each input group, Fig. 2e. The effect of these cross-correlations is stronger than the effect of the noise since they affect both the diversity and the E/I balance, both of which decline as the recurrent connections become stronger Fig. 2d.

The combination of noise and recurrent connectivity usually fuses the two previous effects with in-group and between-group correlations, converging at some intermediate value as the noise and recurrent connection strength increase, which severely impacts both tuning metrics Fig. 3c-d.

We develop a formal description of the effect of noise and recurrence on the covariance structure in a simplified linear neural mass model. To this end, we consider $M = 8$ mesoscopic units instead of the previously studied M inter-connected groups, represented by continuous rate variables $x_j(t)$, $j = 1, \ldots, M$. These units evolve in time, subject to stochastic white noise. The linear approximation is justified for any system at a stationary state with a constant average firing rate, and it serves as a simplified model for a wide range of parameters of the spiking network (for details on the linear model and its relation to the spiking network, see Appendix C).

In this simplified case, it is possible to derive analytical equations for all the relevant in- and between-group covariances, which yield the correlation coefficients. These covariances are the solution to a linear system of equations, which can be solved exactly using numerical methods. Furthermore, one can find close-form solutions in some simple scenarios. For example, in the case of a completely homogeneous network, where all coupling weights are the same, correlation coefficients can be written explicitly (see Appendix C). If the coupling strength increases $W \to +\infty$ all correlations grow to 1 as $1 - \mathcal{O}(1/W^2)$, while they decrease to $MW/(M-1)^2 + \mathcal{O}(r^{-2})$ when increasing the noise to signal ratio $r \to 0_+$. Both cases eliminate any possible differentiation between the groups, thus compromising the ability of the plasticity mechanisms to create diversity D. Another observation is that in the linear network, increasing noise affects the correlation coefficient quadratically, while coupling increases it linearly. Therefore, increasing the coupling has a larger impact on the co-tuning, a consequence that is recovered in the spiking network, Fig. 2.

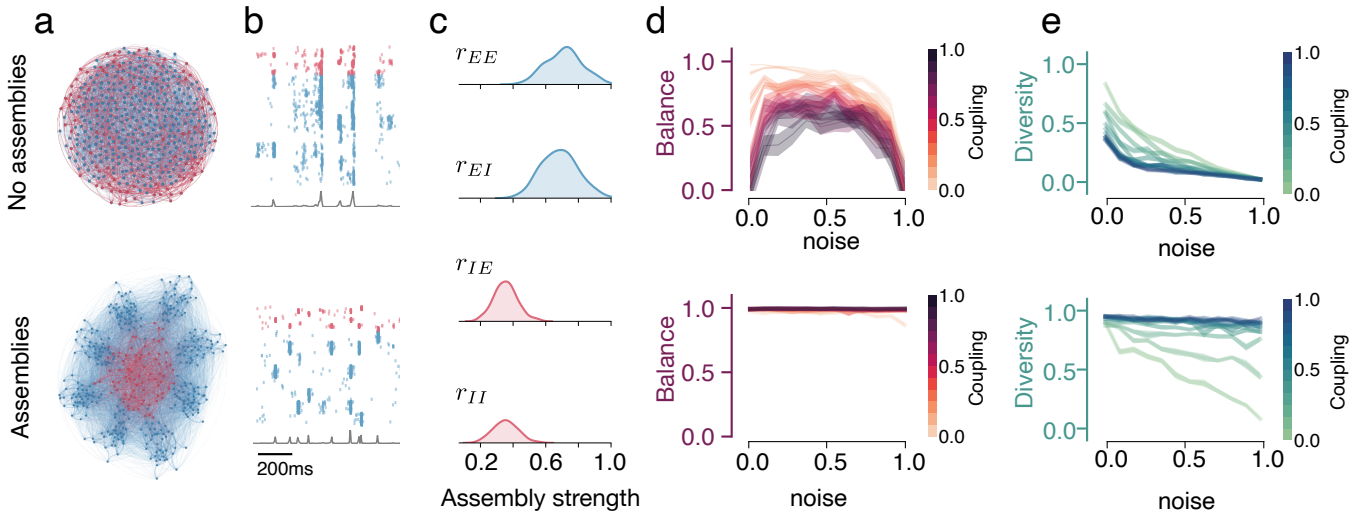

FIG. 3. **Optimized assemblies of neurons restore the co-tuning in recurrent noisy networks.** Top row — networks with uniform connectivity, bottom — networks with inferred optimal assembly strengths. **a.** Diagram of the network, with uniform connectivity and inferred optimal assembly strengths. **b.** Stimulus-driven activity of the networks with (bottom) and without (top) assemblies. **c.** Approximate posterior distributions of excitatory and inhibitory assemblies strength. **d.** Balance in homogeneous (top) and optimized assemblies (bottom) networks with different strengths of noise and coupling. **e.** Same for diversity. Assemblies restore co-tuning even with strong noise if given sufficient coupling strength, but for homogeneous networks, the strong coupling is detrimental to the development of co-tuning.

C. Neuronal type-specific assemblies restore the ability of STDP to produce co-tuning

A homogeneous all-to-all connectivity is not a realistic assumption that could be particularly detrimental for the self-organization of co-tuning. Next, we examine the impact of different types of inhomogeneous connectivity. In particular, we study whether stronger recurrent connectivity between neurons of the same input group would lead to beneficial correlations in the activity of the network (we treat each input group as a neuronal assembly). We define a metric of assembly strengths as a ratio of the average input from the same group and type (E/I) neurons to the total average input from the given type of neurons:

$$r_{ab} = \frac{C^{ab}_{\text{in}}}{C^{ab}_{\text{in}} + C^{ab}_{\text{out}}} = \frac{W^{ab}_{\text{in}}}{W^{ab}_{\text{in}} + (M-1)\cdot W^{ab}_{\text{out}}}, \quad a, b \in \{E, I\}, \tag{4}$$

where C^{ab}_{in} in the total input a neuron of type b receives from neurons of type a, for $a, b \in \{E, I\}$, of its own input group and C^{ab}_{out} is the total input a neuron of type b receives from neurons of type a, for $a, b \in \{E, I\}$, from all other input groups. In our network, we keep all connections of the same type equal, thus, $C^{ab}_{\text{in}} = \frac{p\cdot N}{M} \cdot W^{ab}_{\text{in}}$ and $C^{ab}_{\text{out}} = \frac{p\cdot(M-1)\cdot N}{M} \cdot W^{ab}_{\text{out}}$, where p is the probability of connection between two neurons, W^{ab}_{in} the connection strength between neurons of the same group and W^{ab}_{out} the connection strength between neurons of different groups. We vary assembly strengths for each type of connection r_{EE}, r_{EI}, r_{IE}, and r_{II} while keeping the total input to a neuron $C^{ab}_{\text{in}} + C^{ab}_{\text{out}} = \frac{p\cdot N}{M} \cdot (W^{ab}_{\text{in}} + (M-1)\cdot W^{ab}_{\text{out}}) =: p \cdot N \cdot W$ constant. Here W is a coupling strength, same as in the network without assemblies. Thus, we can vary the fraction of input coming to a neuron from its own input group without changing the average recurrent E or I input it receives.

For the reduced linear neural mass model, we compute analytically the optimal assembly strengths, (Appendix C). We find that for all combinations of noise and sufficiently strong recurrent connectivity, the optimal connectivity is to have very strong excitatory assemblies (high r_{EE} and r_{EI}) and uniform inhibitory connectivity (low r_{IE} and r_{II}). This connectivity allows the correlating excitatory currents to remain mostly within the input group/assembly and maintain high in-group correlation while inhibitory currents are diffused and reduce correlations between groups.

The reduced model does not account for many essential features of the spiking network, like sparsity of connections, in-group interactions between neurons of the same type, and non-stationary dynamical states of the groups. Therefore, although the analytic solution obtained for the linear neural mass model can serve to develop intuition, the results need to also be validated for the spiking recurrent network. Particularly, we seek to estimate the effect of various assembly strengths on the covariance structure of the spiking network's activity and, thus, on balance and diversity.

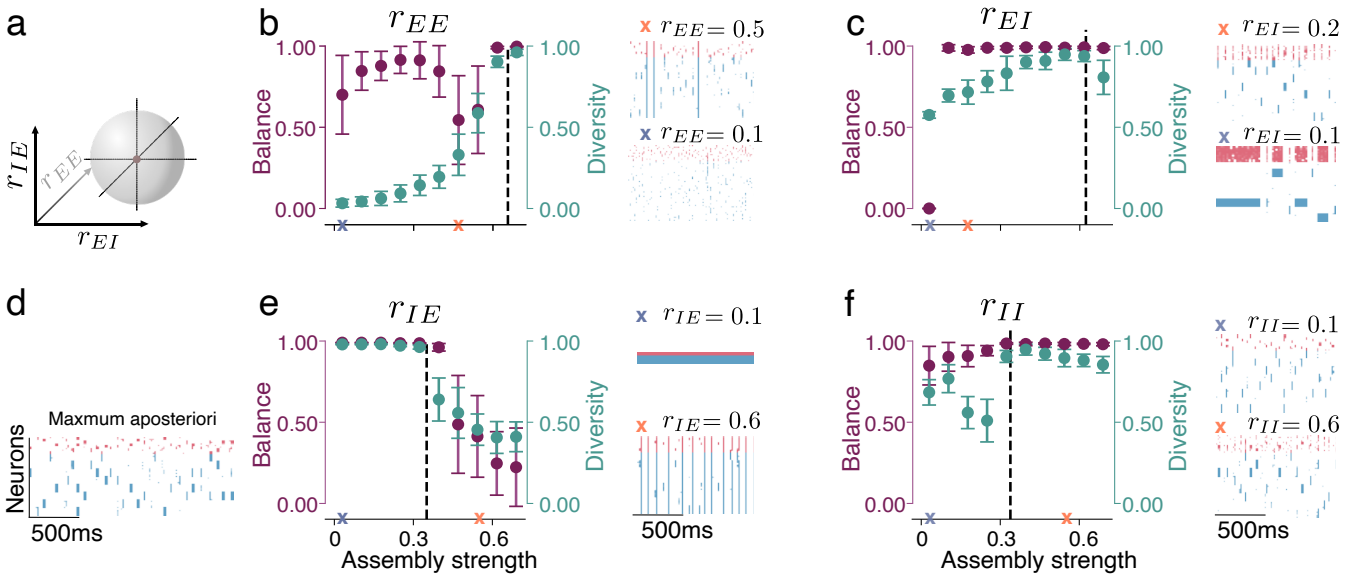

FIG. 4. **Changes of the various assemblies strengths differently affect balance and diversity. a.** We sequentially vary the value of each assembly strength, while keeping the rest of the parameters fixed at the maximum aposteriori (MAP) solution (**d**). **b.** A decrease in the $E \to E$ assembly strength (R_{EE}) introduces synchronous burst-like events that jeopardize co-tuning and weight diversity. Further reduction of the $E \to E$ assembly strength results in sparse, asynchronous spiking which significantly reduces tuning quality. **c.** Reduction of $E \to I$ assembly strength first leads to synchronous inhibitory firing across groups and further reduction leads to persistent activity of the whole inhibitory population combined with bursts of excitatory activity that prevent the development of diversity. **e.** Decreasing the $I \to E$ assembly strength leads to persistent activation of a single group of neurons (which affects the tuning only marginally). While the increase of the $I \to E$ assembly strength leads to oscillatory behavior of the whole network. **f.** Weakening the $I \to I$ assemblies decreases the weight diversity by introducing occasional synchronous bursts in the network, while strengthening them leads to asynchronous inhibitory activity.

We search for all combinations of assembly strength r_{EE}, r_{EI}, r_{IE}, r_{II} that lead to the detailed E/I balance ($B \approx 1$) and maximum weight diversity ($D \approx 1$), mathematically formalized as the posterior destitution of parameters. To this end, we use the sequential Approximate Bayesian Computation (ABC) [39] to minimize the loss function defined to be zero when the in-group correlations are equal one and all between-group correlations vanish (for details, see Methods). This method allows us to find the approximate posterior distribution of network parameters. This distribution can later be tested on whether it leads to the self-organization of co-tuning in the post-synaptic neuron.

Networks with optimized assemblies regain the ability to develop E/I co-tuning. Assembly strengths that are drawn from an approximate posterior result in a covariance structure very similar to the one observed in a feedforward/low noise network, which allows the plasticity to produce near-optimal co-tuning of the feedforward connections, Fig. 3d-e. We find that the optimal assembly structure involves very strong $E \to E$ and $E \to I$ assemblies and medium-strength $I \to E$ and $I \to I$, Fig. 3 c. This is in contrast to the neural mass model result, which predicted optimal performance for uniform inhibitory weights.

Changes in inhibitory and excitatory connectivity are known to affect network dynamics differently [40]. To investigate the relative importance of the different connection types, we perturb various assemblies away from the optimal solutions. We study how fast the balance and diversity deteriorate as the parameters are shifted away from the optimal solution inferred with ABC Fig. 4a. We find that $E \to E$ and $I \to E$ assemblies have a stronger impact on co-tuning and weight diversity compared to the $E \to I$ and $I \to I$ assemblies Fig. 4b-f.

D. The sparsity of a network's recurrent connectivity shifts the optimal assembly structure

Biological neural networks are usually very sparsely connected [41–43] and the sparsity of connections is associated with distinct dynamics [44]. We observed that the impact of noise and recurrence on the deterioration of balance and weight diversity in sparse networks without assemblies is qualitatively similar to fully connected networks. Thus, we examined the ability of neuronal assemblies to produce activity that restores balance and weight diversity in sparsely connected recurrent networks that receive noisy input.

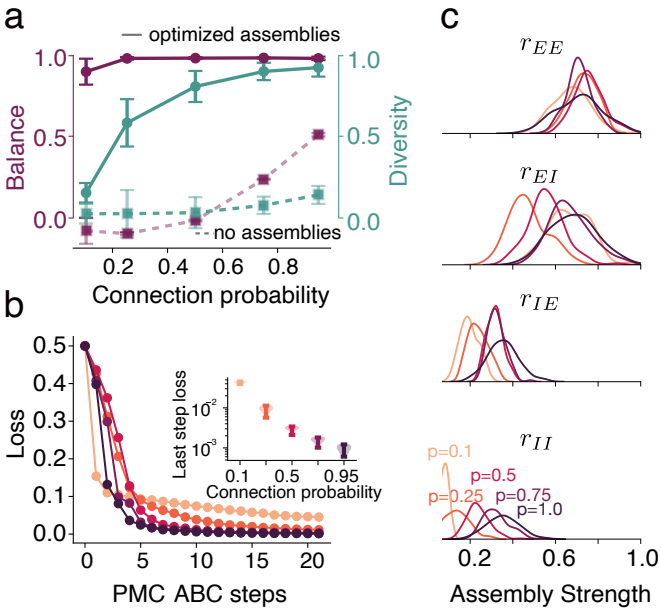

FIG. 5. Assemblies improve co-tuning and allow for co-tuning in sparse networks. **a.** E/I balance (purple) and weights diversity (teal) in the networks with assemblies compared to non-structured networks (dashed lines), error bars — standard deviation. The noise level is 0.1 for all sparsities and the coupling is 0.85 (scaled by $1/p$). **b.** Loss for the sparser networks is higher, which results in the overall worse performance, inset shows the loss for 50 accepted samples at the last ABC step. **c.** Posterior distributions of all assembly strengths change with sparsity. Sparser networks require weaker inhibitory assemblies (more uniform connections) to produce co-tuning.

The optimal assembly strength values are shifted for different sparsity levels. We use ABC to discover the approximate posterior distribution of assembly strengths for 5 different levels of sparsity, corresponding to the probability of connection $p = 1.0$, $p = 0.75$, $p = 0.5$, $p = 0.25$, and $p = 0.1$. We preserve the total input per neuron across different sparsities by scaling the coupling strength inversely proportional to p. The optimal strength of most assemblies is reduced as the connection probability is decreased, Fig. 5c. Specifically, we find that all but $E \to E$ assemblies should be weaker in sparser networks, with the greatest decrease observed in the $I \to I$ assemblies, which completely disappear for very sparse networks.

As sparsity increases, the ability of assemblies to improve the tuning diminishes. The overall loss after 21 ABC steps is larger for the sparse networks than for fully-connected networks and increases with sparseness, Fig. 5b. Therefore, despite an improvement in the tuning metrics for most sparse networks (compare dashed and solid lines in Fig. 5a), particularly diversity is strongly affected by sparseness and cannot be recovered by assemblies at the same extent as for the fully connected networks, Fig. 5a. This reduced effectiveness is expected given the smaller number of connections and the greater variance in the network's connectivity.

III. DISCUSSION

Input selectivity is a universal attribute of brain networks that is maintained across brain hierarchies, including in brain areas that only receive input from highly recurrent networks. Here, we demonstrated that two ubiquitous features of biological networks, namely internal noise and recurrent connectivity between different sub-networks, can impact the statistics of inputs coming from a population in ways that completely prevent known plasticity mechanisms [6, 20, 21, 24] from forming any kind of input selectivity in neurons found in higher areas.

We hypothesized that the strongly negative effects of the recurrent connectivity on the ability of STDP to produce co-tuning are to a great extent due to the biologically unrealistic, homogeneous connectivity we used as a baseline. Cortical networks consist of hundreds of interacting neuron types, each characterized by distinct dynamics and highly detailed local connectivity patterns. It is natural to assume that the neural activity that is produced by such networks can hardly be approximated by random networks without any spatial structure or any distinction between the connectivity patterns of E and I units. Our findings indicate that non-uniform network connectivities, even of a much

simpler nature than the intricate patterns of actual brain networks, can have a very significant impact on the ability of known STDP rules to produce E/I co-tuning in higher areas.

Cortical networks are known to be highly clustered [45] and the clustering seems to have functional as well as spatial criteria. For example, neurons that share common input [46] or targets [47] are more likely to form recurrent connections amongst themselves. Additionally, there is strong evidence that groups of highly interconnected neurons (neuronal assemblies) share common functions within recurrent networks [31, 48, 49]. Moreover, evidence has accumulated [50, 51] that different neurons type (excitatory and inhibitory subtypes) follow distinct spatial connectivity patterns, which have implications for neural computation.

The ubiquitous presence of neuronal assemblies and the fact that different neuron types seem to follow different connectivity patterns, along with the evidence for the creation of distinct dynamics in networks [29] with clustered excitatory connectivity, led us to explore the possibility that different neuron types may form overlapping functional assemblies of varying strengths. Our results suggest that such overlapping, synaptic type-specific assemblies can have a very strong effect on the dynamics of a recurrent network and can effectively control the ability of STDP to produce E/I co-tuning in upstream areas.

Despite the general biological plausibility of the inferred network structure, the extent to which it is realized in the *in vivo* networks remains unclear. Although we identify different assembly strengths for networks of different sparsity, a general finding is that excitatory assemblies should ubiquitously be stronger than the corresponding inhibitory ones. The formation of E/I assemblies via synaptic plasticity has been extensively studied [20, 24, 25, 28] and it has been demonstrated that a variety of plasticity mechanisms can lead to stable, matching E/I assemblies. Still, the question of different assembly strengths for different connection types has not been fully explored. Potential mechanisms that can control the formation of assemblies of varying strengths is variation in the learning rates of different synaptic types or the presence of multiple plasticity mechanisms that regulate assembly formation on various connection types. Additionally, the presence of different regulatory interneurons, which has been linked with assembly formation [52], could potentially play a role in modulating the relative assembly strengths of different connections. Finally, the varied clustering levels observed in cortical networks may act as a driving force of assembly formation, favoring stronger excitatory assemblies.

For the purposes of our study, we parameterized the network topology by adopting a quantitative metric for the strength of different types of neuronal assemblies. This approach allowed the efficient simulation-based inference [53] of the optimal assembly strengths as well as the analytical treatment of a linear model with analogous connectivity constraints. However, direct optimization of each individual recurrent weight to find optimal recurrent connectivity for E/I co-tuning to emerge would almost certainly deliver an even better solution than our simplified parametrization. Such optimization could potentially be achieved by one of several gradient-based methods for training spiking networks [54, 55] or alternatively via an evolutionary algorithm [56]. It is an open question whether this optimal recurrent connectivity would maintain any features of the topology we identified via our parameter inference method, although the analytic solution of the simplified linear model suggests the existence of a general pattern that is not limited to the specific network parametrization we chose.

IV. MATERIALS AND METHODS

A. Neuron Model

We modelled all neurons of our networks as leaky integrate-and-fire (LIF) neurons with leaky synapses. The evolution of their membrane potential is given by the ODE

$$C_m \cdot \frac{dV(t)}{dt} = g_{\text{leak}} \cdot (V_{\text{rest}} - V(t)) + g_I(t) \cdot (V_I - V(t)) - g_E(t) \cdot (V_E - V(t)), \tag{5}$$

where V_{rest} is the neuron's resting potential, V_E, V_I are the excitatory and inhibitory reversal potentials and g_{leak} the leak conductance. Additionally, the excitatory and inhibitory conductances g_E, g_I decay exponentially over time and get boosts upon excitatory or inhibitory pre-synaptic spiking respectively, as

$$\begin{aligned} \tau_E \cdot \frac{dg_E(t)}{dt} &= -g_E(t) + \overline{g_E} \cdot \sum_j W_j^E \cdot \sum_f \delta(t - t_j^f), \\ \tau_I \cdot \frac{dg_I(t)}{dt} &= -g_I(t) + \overline{g_I} \cdot \sum_j W_j^I \cdot \sum_f \delta(t - t_j^f). \end{aligned} \tag{6}$$

Here t_j^f denotes the time at which the f-th spike of the $j-$th neuron happened. When the membrane potential reaches the spiking theshold V_{th}, a spike is emitted, and the potential is changed to a reset potential V_{reset}. Values of the constants used in simulations can be found in Appendix D.

B. Network Input

The external input to each of the 1000 pre-synaptic neurons is the mixture of two Poisson spike trains. The first Poisson spike train is shared with all the other neurons of the same group, while the second Poisson spike train is the individual noise of the neuron,

$$C_{\text{total}} = C_{\text{signal}} + C_{\text{noise}}, \tag{7}$$

where $C_{\text{signal}} \sim \text{Poisson}((1-c)\cdot r)$ and $C_{\text{noise}} \sim \text{Poisson}(c \cdot r)$. Here, r is the total firing rate of the input, and c is the strength of the noise. C_{signal} is the same for all neurons of the same input group, while C_{noise} is individual to each neuron

Plasticity

1. *Triplet excitatory STDP*

The feedforward excitatory connections are modified according to a simplified form of the triplet STDP rule [32], which has been shown to generalize the Bienenstock–Cooper–Munro (BCM) rule [4] for higher-order correlations [33]. The firing rates of the pre-synaptic excitatory neurons and the post-synaptic neuron are approximated by traces with two different timescales,

$$\begin{aligned}
\tau_1^{estdp} \cdot \frac{dy_k^E}{dt} &= -y_k^E + \sum_f \delta(t - t_k^f), \\
\tau_2^{estdp} \cdot \frac{dz_k^E}{dt} &= -z_k^E + \sum_f \delta(t - t_k^f), \\
\tau_1^{estdp} \cdot \frac{dx_1}{dt} &= -x_1 + \sum_f \delta(t - t_x^f), \\
\tau_2^{estdp} \cdot \frac{dx_2}{dt} &= -x_2 + \sum_f \delta(t - t_x^f),
\end{aligned} \tag{8}$$

where $\tau_1^{estdp} < \tau_2^{estdp}$ are the two timescales of the plasticity rule, y_k^E, z_k^E and x_1, x_2 represent the slow and fast traces of the k-th excitatory pre-synaptic and the single post-synaptic neuron respectively. t_k^f and t_x^f are the firing times of the pre and post-synaptic neurons. The connection weights are updated upon pre and post-synaptic spiking according to

$$\Delta W_k^E = \eta_E \cdot A_{LTP} \cdot x_1 \cdot z_k^E \cdot \sum_f \delta(t - t_k^f) - \eta_E \cdot A_{LTD} \cdot x_2 \cdot y_k^E \cdot \sum_f \delta(t - t_x^f), \tag{9}$$

where η_E is the excitatory learning rate and A_{LTP}, A_{LTD} the amplitudes of long term depression and potentiation respectively.

2. *Inhibitory STDP*

We used the homeostatic STDP rule first proposed in [20] for the inhibitory feedforward connections. Approximations of the firing rates are kept via a trace for each of the pre-synaptic inhibitory neurons as well as the post-synaptic

neuron,

$$\tau^{istdp} \cdot \frac{dy_k^I}{dt} = -y_k^I + \sum_f \delta(t - t_k^f),$$
$$\tau^{istdp} \cdot \frac{dx}{dt} = -x + \sum_f \delta(t - t_x^f), \tag{10}$$

where τ^{istdp} is the single timesclae of the plasticity rule, y_k^I and x are the traces of the the k_{th} inhibitory pre-synaptic and the single post-synaptic neuron, and t_k^f, t_x^f are the spike times of the k_{th} inhibitory pre-synaptic and the post-synaptic neuron respectively. The connection weights are updated upon pre and post-synaptic spiking as

$$\Delta W_k^I = \eta_I \cdot (x - 2\rho_0 \tau^{stdp}) \cdot \sum_f \delta(t - t_k^f) + \eta_I \cdot y_k^I \cdot \sum_f \delta(t - t_x^f). \tag{11}$$

Here, η_I is the inhibitory learning rate, and ρ_0 is the target rate of the post-synaptic neuron.

3. *Synaptic Scaling*

Due to the instability of the triplet STDP rule, some sort of normalization mechanism is necessary to constrain weight development. We use the novel competitive normalization protocol first proposed in [24], which we adapt for spiking neurons. The normalization is separately applied to both excitatory and inhibitory incoming connections,

$$W_k^A \leftarrow W_k^A \left(1 - \eta_N + \eta_N \cdot \frac{W_{target}^A}{\sum_{i=1}^{N_A} W_i^A}\right), \quad A \in \{E, I\}. \tag{12}$$

Where W_{target}^A is the target total weight of each connection type and η_N is the normalization learning rate. The normalization maintains the sum of the excitatory and the sum of the inhibitory feedforward connections weights close to the set target total weights W_{target}^E and W_{target}^I.

C. Approximating the posterior distribution of the model parameters

To estimate the set of parameters that lead to high in-group correlations and low out-group correlations, we used simulation-based inference [53]. The basic idea is to use simulation with known parameters to approximate the full posterior distributions for the model given the required output, i.e., the distribution of parameters and samples from which produce the required correlation structure. To approximate the posterior distribution we use sequential Approximate Bayesian Computation (ABC) [39]. We define a loss function that maximizes in-group correlations and minimizes between-group correlations,

$$\mathcal{L} = -\alpha C_{in}^2 - \beta \left[(1 - C_{out}^{E\to E})^2 + (1 - C_{out}^{E\to I})^2 + (1 - C_{out}^{E\to I})^2\right].$$

We define a uniform prior $p(\theta)$. A set of parameters $\theta = [r_{ee}, r_{ei}, r_{ie}, r_{ii}]$ is sampled from it and used to run the simulations for 3 seconds. From the simulation results, correlations are computed, which allows us to obtain the loss. We accept a parameter set if the loss is below the error ϵ, and keep sampling until the number of accepted samples is 60. We use the kernel density estimate on the accepted samples to obtain an approximate posterior. Next, we rescale this approximate posterior with the original prior to obtain a proposal distribution that we use as a prior in the next step of the ABC. In each step, we reduce ϵ by setting it to the 75th percentile of the losses for the accepted samples (see [39] for more details). As a rule, we run 20 to 30 steps of the sequential ABC until the loss converges. We run separate fits for networks with different levels of sparsity with connection probabilities $p = 0.1, 0.25, 0.5, 0.75, 1.0$. The fitting was done using a modified version of the simple-abc toolbox.

D. Reduced model

The dynamics of the system can be studied analytically using a simplified, reduced linear model. Here, each pair of variables (x_i, y_i) represents the excitatory and inhibitory mean firing rate of a neuron group. In theory, these variables

display complicated non-linear interactions that arise from the microscopic details of the LIF spiking network and synapse dynamics. However, in the stationary state –and away from any critical point– a linearised model can capture the essential features of the correlations between different populations.

Internal noise, modeled as independent Poisson trains to each individual neuron, becomes Gaussian white noise in the large-population limit, characterized by zero mean and variance σ_{int}. Each population is affected by different internal fluctuations. For simplicity, external noise, which is applied as the same train of Poisson spikes to all the neurons inside an input group, will also be approximated as a Gaussian white noise of mean η_0 and variance σ_{ext}.

Therefore, the simplified linear model reads

$$\dot{x}_i = a x_i + b y_i + \frac{1}{M-1} \sum_{j \neq i} \left(W^{EE} x_j + W^{EI} y_j \right) + \sigma_{\text{int}} \xi_i^x(t) + \sigma_{\text{ext}} \eta_i(t) + \eta_0, \tag{13a}$$

$$\dot{y}_i = c x_i + d y_i + \frac{1}{M-1} \sum_{j \neq i} \left(W^{IE} x_j + W^{II} y_j \right) + \sigma_{\text{int}} \xi_i^y(t) + \sigma_{\text{ext}} \eta_i(t) + \eta_0, \tag{13b}$$

where M is the number of populations, a, b, c, d are parameters controlling in-group recurrent coupling, and W^{EE}, W^{EI}, W^{IE}, W^{II} are couplings between different clusters. Internal noise for each population is represented by $\xi_i^{x,y}(t)$, while external noise is notated as $\eta_i(t)$. All noises are uncorrelated, meaning that

$$\left\langle \xi_i^c \xi_j^{c'} \right\rangle = \delta_{cc'} \delta_{ij} \delta(t - t'), \tag{14a}$$

$$\langle \xi_i^c(t) \eta_j(t') \rangle = 0 \quad \forall i, j, t, t', \tag{14b}$$

$$\langle \eta_i(t) \eta_j(t') \rangle = \delta_{ij} \delta(t - t'), \tag{14c}$$

with $c, c' = \{x, y\}$, and where $\langle \ldots \rangle$ represents an ensemble average, i.e., an average over noise realizations. From this model, it is possible to obtain closed equations for Pearson correlation coefficients (see Appendix C for details). Notice that stochastic differential equations are never complete without an interpretation, and we choose to interpret these in the Itô sense, which will be relevant for computations.

V. ACKNOWLEDGEMENTS

This work was supported by a Sofja Kovalevskaja Award from the Alexander von Humboldt Foundation. EG thank the International Max Planck Research School for Intelligent Systems (IMPRS-IS) for support. We acknowledge the support from the BMBF through the Tübingen AI Center (FKZ: 01IS18039B). AL is a member of the Machine Learning Cluster of Excellence, EXC number 2064/1 – Project number 39072764.

[1] M. Riesenhuber and T. Poggio, Hierarchical models of object recognition in cortex, Nature neuroscience **2**, 1019 (1999).
[2] J. Sharma, A. Angelucci, and M. Sur, Induction of visual orientation modules in auditory cortex, Nature **404**, 841 (2000).
[3] D. E. Feldman and M. Brecht, Map plasticity in somatosensory cortex, Science **310**, 810 (2005).
[4] E. Bienenstock, L. Cooper, and P. Munro, Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex, Journal of Neuroscience **2**, 32 (1982), https://www.jneurosci.org/content/2/1/32.full.pdf.
[5] B. S. Blais, N. Intrator, H. Shouval, and L. N. Cooper, Receptive Field Formation in Natural Scene Environments: Comparison of Single-Cell Learning Rules, Neural Computation **10**, 1797 (1998), https://direct.mit.edu/neco/article-pdf/10/7/1797/813973/089976698300017142.pdf.
[6] C. Clopath, T. P. Vogels, R. C. Froemke, and H. Sprekeler, Receptive field formation by interacting excitatory and inhibitory synaptic plasticity, bioRxiv 10.1101/066589 (2016), https://www.biorxiv.org/content/early/2016/07/29/066589.full.pdf.
[7] C. S. N. Brito and W. Gerstner, Nonlinear hebbian learning as a unifying principle in receptive field formation, PLOS Computational Biology **12**, 1 (2016).
[8] S. Deneve and C. Machens, Efficient codes and balanced networks, Nature Neuroscience **19**, 375 (2016).
[9] S. Zhou and Y. Yu, Synaptic e-i balance underlies efficient neural coding, Frontiers in Neuroscience **12**, 10.3389/fnins.2018.00046 (2018).
[10] W. Singer, Recurrent dynamics in the cerebral cortex: Integration of sensory evidence with stored knowledge, Proceedings of the National Academy of Sciences **118**, e2101043118 (2021), https://www.pnas.org/doi/pdf/10.1073/pnas.2101043118.
[11] J. Isaacson and M. Scanziani, How Inhibition Shapes Cortical Activity, Neuron **72**, 231 (2011).

[12] M. Okun and I. Lampl, Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities, Nature Neuroscience **11**, 535 (2008).

[13] T. Vogels and L. Abbott, Gating multiple signals through detailed balance of excitation and inhibition in spiking networks, Nature neuroscience **12**, 483 (2009).

[14] A. Bhatia, S. Moza, and U. S. Bhalla, Precise excitation-inhibition balance controls gain and timing in the hippocampus, eLife **8**, e43415 (2019).

[15] H. Ozeki, I. Finn, E. Schaffer, K. Miller, and D. Ferster, Inhibitory stabilization of the cortical network underlies visual surround suppression, Neuron **62**, 578 (2009).

[16] D. Rubin, S. Van Hooser, and K. Miller, The stabilized supralinear network: A unifying circuit motif underlying multi-input integration in sensory cortex, Neuron **85**, 402 (2015).

[17] A. Schulz, C. Miehl, I. Berry, Michael J, and J. Gjorgjieva, The generation of cortical novelty responses through inhibitory plasticity, eLife **10**, e65309 (2021).

[18] M. Wehr and A. M. Zador, Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex, Nature **426**, 442 (2003).

[19] Y. K. Wu, C. Miehl, and J. Gjorgjieva, Regulation of circuit organization and function through inhibitory synaptic plasticity, Trends in Neurosciences **45**, 884 (2022).

[20] T. P. Vogels, H. Sprekeler, F. Zenke, C. Clopath, and W. Gerstner, Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks, Science **334**, 1569 (2011), https://science.sciencemag.org/content/334/6062/1569.full.pdf.

[21] Y. Luz and M. Shamir, Balancing feed-forward excitation and inhibition via hebbian inhibitory synaptic plasticity, PLOS Computational Biology **8**, 1 (2012).

[22] P. J. Hellyer, B. Jachs, C. Clopath, and R. Leech, Local inhibitory plasticity tunes macroscopic brain dynamics and allows the emergence of functional brain networks, NeuroImage **124**, 85 (2016).

[23] *Reservoir computing with self-organizing neural oscillators*, ALIFE 2021: The 2021 Conference on Artificial Life, Vol. ALIFE 2021: The 2021 Conference on Artificial Life (2021) 78, https://direct.mit.edu/isal/proceedings-pdf/isal/33/78/1929805/isal_a_00409.pdf.

[24] S. Eckmann and J. Gjorgjieva, Synapse-type-specific competitive hebbian learning forms functional recurrent networks, bioRxiv 10.1101/2022.03.11.483899 (2022), https://www.biorxiv.org/content/early/2022/03/14/2022.03.11.483899.full.pdf.

[25] O. Mackwood, L. B. Naumann, and H. Sprekeler, Learning excitatory-inhibitory neuronal assemblies in recurrent networks, eLife **10**, e59715 (2021).

[26] R. Larisch, L. Gönner, M. Teichmann, and F. H. Hamker, Sensory coding and contrast invariance emerge from the control of plastic inhibition over emergent selectivity, bioRxiv 10.1101/2020.04.07.029157 (2021), https://www.biorxiv.org/content/early/2021/09/22/2020.04.07.029157.full.pdf.

[27] F. Effenberger, J. Jost, and A. Levina, Self-organization in balanced state networks by stdp and homeostatic plasticity, PLoS computational biology **11**, e1004420 (2015).

[28] F. Zenke, E. Agnes, and W. Gerstner, Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks, Nature Communications **6**, 6922 (2015).

[29] A. Litwin-Kumar and B. Doiron, Formation and maintenance of neuronal assemblies through synaptic plasticity, Nature communications **5**, 5319 (2014).

[30] E. Giannakakis, F. Hutchings, C. A. Papasavvas, C. E. Han, B. Weber, C. Zhang, and M. Kaiser, Computational modelling of the long-term effects of brain stimulation on the local and global structural connectivity of epileptic patients, PLOS ONE **15**, 1 (2020).

[31] C. Miehl and J. Gjorgjieva, Stability and learning in excitatory synapses by nonlinear inhibitory plasticity, bioRxiv 10.1101/2022.03.28.486052 (2022), https://www.biorxiv.org/content/early/2022/03/29/2022.03.28.486052.full.pdf.

[32] J.-P. Pfister and W. Gerstner, Triplets of spikes in a model of spike timing-dependent plasticity, Journal of Neuroscience **26**, 9673 (2006), https://www.jneurosci.org/content/26/38/9673.full.pdf.

[33] J. Gjorgjieva, C. Clopath, J. Audet, and J.-P. Pfister, A triplet spike-timing–dependent plasticity model generalizes the bienenstock–cooper–munro rule to higher-order spatiotemporal correlations, Proceedings of the National Academy of Sciences **108**, 19383 (2011), https://www.pnas.org/content/108/48/19383.full.pdf.

[34] S. Zhou and Y. Yu, Synaptic e-i balance underlies efficient neural coding, Frontiers in Neuroscience **12**, 10.3389/fnins.2018.00046 (2018).

[35] G. Liu, Local structural balance and functional interaction of excitatory and inhibitory synapses in hippocampal dendrites, Nature Neuroscience **7**, 373 (2004).

[36] N. Sukenik, O. Vinogradov, E. Weinreb, M. Segal, A. Levina, and E. Moses, Neuronal circuits overcome imbalance in excitation and inhibition by adjusting connection numbers, Proceedings of the National Academy of Sciences **118**, 10.1073/pnas.2018459118 (2021), publisher: National Academy of Sciences Section: Biological Sciences.

[37] A. S. Ecker and A. S. Tolias, Is there signal in the noise?, Nature neuroscience **17**, 750 (2014).

[38] D. A. Aponte, G. Handy, A. M. Kline, H. Tsukano, B. Doiron, and H. K. Kato, Recurrent network dynamics shape direction selectivity in primary auditory cortex, Nature Communications **12**, 314 (2021), number: 1 Publisher: Nature Publishing Group.

[39] M. A. Beaumont, J.-M. Cornuet, J.-M. Marin, and C. P. Robert, Adaptive approximate Bayesian computation, Biometrika **96**, 983 (2009).

[40] G. Mongillo, S. Rumpel, and Y. Loewenstein, Inhibitory connectivity defines the realm of excitatory plasticity, Nature neuroscience **21**, 1463 (2018).

[41] S. C. Seeman, L. Campagnola, P. A. Davoudian, A. Hoggarth, T. A. Hage, A. Bosma-Moody, C. A. Baker, J. H. Lee, S. Mihalas, C. Teeter, A. L. Ko, J. G. Ojemann, R. P. Gwinn, D. L. Silbergeld, C. Cobbs, J. Phillips, E. Lein, G. Murphy, C. Koch, H. Zeng, and T. Jarsky, Sparse recurrent excitatory connectivity in the microcircuit of the adult mouse and human cortex, eLife **7**, e37349 (2018).

[42] G. A. Wildenberg, M. R. Rosen, J. Lundell, D. Paukner, D. J. Freedman, and N. Kasthuri, Primate neuronal connections are sparse in cortex as compared to mouse, Cell Reports **36**, 109709 (2021).

[43] J. Barral and A. D Reyes, Synaptic scaling rule preserves excitatory–inhibitory balance and salient neuronal network dynamics, Nature Neuroscience **19**, 1690 (2016).

[44] Brunel, Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons, Journal of Computational Neuroscience **8** (2000).

[45] C. Hilgetag and M. Kaiser, Clustered organization of cortical connectivity, Neuroinformatics **2**, 353 (2004).

[46] Y. Yoshimura, J. Dantzker, and E. Callaway, Excitatory cortical neurons form fine-scale functional networks, Nature **433**, 868 (2005).

[47] S. Brown and S. Hestrin, Intracortical circuits of pyramidal neurons reflect their long-range axonal targets, Nature **457**, 1133 (2009).

[48] A.-S. Badin, F. Fermani, and S. A. Greenfield, The features and functions of neuronal assemblies: Possible dependency on mechanisms beyond synaptic transmission, Frontiers in Neural Circuits **10**, 10.3389/fncir.2016.00114 (2017).

[49] G. S. Umbach, R. J. Tan, J. Jacobs, B. E. Pfeiffer, and B. C. Lega, Flexibility of functional neuronal assemblies supports human memory, Nature Communications **13** (2021).

[50] S. Hofer, H. Ko, B. Pichler, J. Vogelstein, H. Ros, H. Zeng, E. Lein, N. Lesica, and T. Mrsic-Flogel, Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex, Nature neuroscience **14**, 1045 (2011).

[51] R. B. Levy and A. D. Reyes, Spatial profile of excitatory and inhibitory synaptic connectivity in mouse primary auditory cortex, Journal of Neuroscience **32**, 5609 (2012), https://www.jneurosci.org/content/32/16/5609.full.pdf.

[52] F. Lagzi, M. C. Bustos, A.-M. Oswald, and B. Doiron, Assembly formation is stabilized by parvalbumin neurons and accelerated by somatostatin neurons, bioRxiv 10.1101/2021.09.06.459211 (2021), https://www.biorxiv.org/content/early/2021/09/07/2021.09.06.459211.full.pdf.

[53] K. Cranmer, J. Brehmer, and G. Louppe, The frontier of simulation-based inference, Proceedings of the National Academy of Sciences , 201912789 (2020).

[54] G. Bellec, F. Scherr, A. Subramoney, E. Hajek, D. Salaj, R. Legenstein, and W. Maass, A solution to the learning dilemma for recurrent networks of spiking neurons, Nature Communications **11** (2020).

[55] F. Zenke and T. P. Vogels, The Remarkable Robustness of Surrogate Gradient Learning for Instilling Complex Function in Spiking Neural Networks, Neural Computation **33**, 899 (2021), https://direct.mit.edu/neco/article-pdf/33/4/899/1902294/neco_a_01367.pdf.

[56] K. Stanley, J. Clune, J. Lehman, and R. Miikkulainen, Designing neural networks through neuroevolution, Nature Machine Intelligence **1** (2019).

[57] N. Caporale and Y. Dan, Spike timing–dependent plasticity: A hebbian learning rule, Annual review of neuroscience **31**, 25 (2008).

[58] One could argue that external noise should be interpreted as Stratonovich and internal as Itô. Since both noises are additive, this difference is not so relevant and we treat all as Itô for simplicity.

[59] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, Springer Series in Synergetics (Springer, 2009).

Appendix A: Alternative plasticity protocol

We examined whether our results are dependent on the particular plasticity protocol we used [24] and we verified that they hold for alternative plasticity mechanisms. At first, we examine our result's robustness with respect to the form of the excitatory and inhibitory learning rules used, while maintaining the same competitive normalization protocol [24]. We find that replacing the triplet rule [33] with a classic spike pair Hebbian rule [57] leads to no discernible effect in the quality of the tuning. Furthermore, we examine a variety of different LTD and LTP amplitudes in the excitatory plasticity as well as several terget rates ρ_0 for the inhibitory plasticity without observing any noticeable changes on our main findings

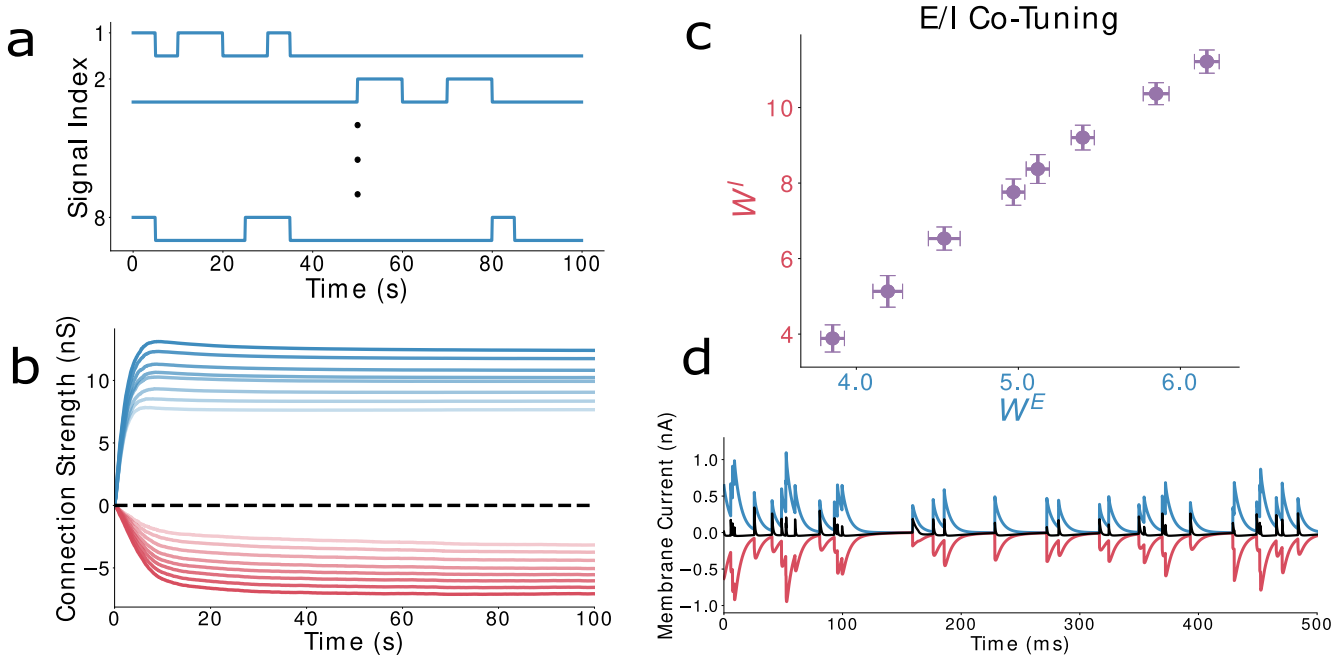

FIG. 6. **Alternative plasticity protocol. a.** : For diversity to emerge, there needs to be a mechanism that enforces inhomogeneity in the pre0synaptic firing rates, such as pulses of input. **b.** Given such a mechanism, the weights converge rapidly, in a state of near-perfect E/I co-tuning (**c**), leading to very tight balance of incoming E/I currents (**d**).

Moreover, we studied both the triplet [33] and pair Hebbian rules [57] combined with a subtractive normalization mechanism that has been previously used in plasticity studies [17] :

$$\Delta W_k^E = \frac{\sum_{i=1}^{N_A} W_{ik}^E - W^{target}}{N_E} \tag{A1}$$

In 2016 [6] it was demonstrated that subtractive normalization only on the excitatory connections will lead to all the weights converging on the same point due to the inhibitory plasticity creating a moving threshold. In order to prevent this collapse of the receptive field, enformed inhomogeneity on the firing rates of different groups is needed. We solved this problem by giving the network's input as pulses of 500 mS, enforcing inhomogeneous firing rates which result in the emergence of diverse and balanced receptive fields.

We verified that the co-tuning achieved this setting, similarly to the mechanism presented in the main text, suffers from the intoduction of noise and recurrent connectivity. Furthermore, the assembling principles that we derived for the original network, seem to have a similarly beneficial effect on this setting, restoring the original covariance structure of the network's activity and leading to detailed co-tuning between the E and I feedforward connections.

Appendix B: Convergence of weights to an eigenvector of the covariance matrix

We verify that our model agrees with the analytics of previous studies [6, 20, 24] that the convergence point of the weights is an eigenvector of the modified covariance matrix:

$$\overline{C} = \left\langle \begin{pmatrix} EE^T & -IE^T \\ EI^T & -II^T \end{pmatrix} \right\rangle \tag{B1}$$

where E, I are the activities of the excitatory and inhibitory populations respectively.

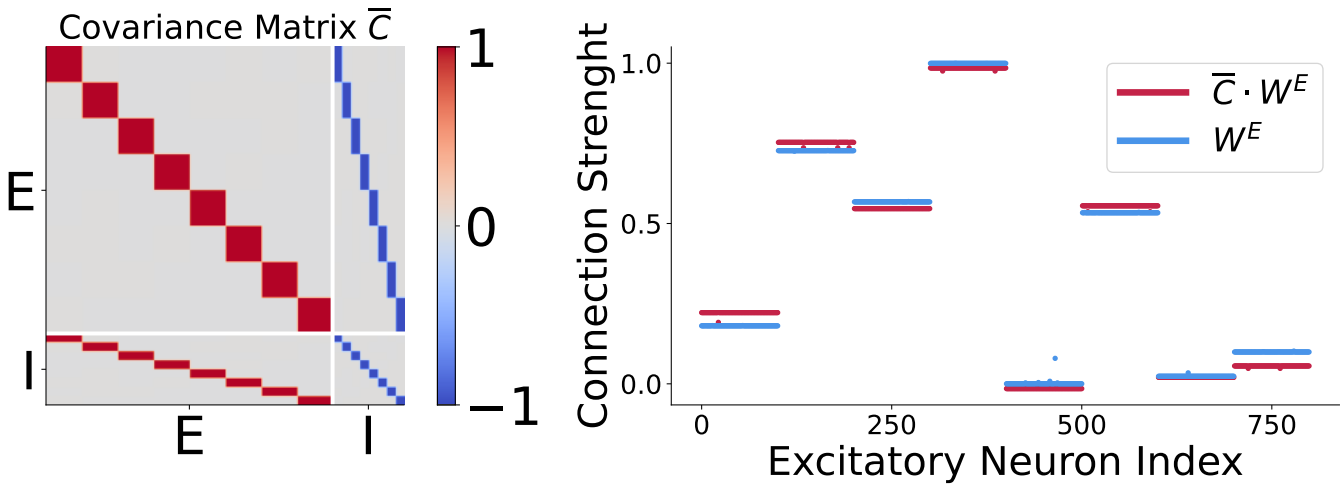

FIG. 7. **Weights converge to an eigenvector of the covariance matrix.** A: The estimated covariance matrix for a feedforward network. B. We verify that the convergence point of the weights is an eigenvector of the covariance matrix.

Specifically, we multiply the converged weight vector with a numerical calculation of the covariance (estimated via binning of the spike trains with a bin size of 1 mS) and verify that the resulting product is approximately equal to a multiple of the original weight matrix.

Appendix C: Reduced model calculations

1. Derivation of the equations

In this Appendix, we start from main text Eqs. (IV D),

$$\dot{x}_i = ax_i + by_i + \frac{1}{M-1}\sum_{j\neq i}\left(W^{EE}x_j + W^{EI}y_j\right) + \sigma_{\text{int}}\xi_i^x(t) + \sigma_{\text{ext}}\eta_i(t) + \eta_0, \tag{C1a}$$

$$\dot{y}_i = cx_i + dy_i + \frac{1}{M-1}\sum_{j\neq i}\left(W^{IE}x_j + W^{II}y_j\right) + \sigma_{\text{int}}\xi_i^y(t) + \sigma_{\text{ext}}\eta_i(t) + \eta_0. \tag{C1b}$$

where we consider N groups composed by excitatory $x_i(t)$ and inhibitory $y_i(t)$ populations ($i = 1, \ldots, N$) coupled linearly (see also Methods). We will derive closed expression for the correlation coefficients. We would like to remark that $\langle\rangle$ means an ensemble average over noise realizations. All stochastic equations are to be interpreted in the Itô convention[58].

First of all, we redefine the noise terms, which will prove convenient later to simplify the algebra. For this reason, we define

$$\xi_i^1(t) = \sigma_{\text{int}}\xi_i^x(t) + \sigma_{\text{ext}}\eta_i(t), \tag{C2a}$$

$$\xi_i^2(t) = \sigma_{\text{int}}\xi_i^y(t) + \sigma_{\text{ext}}\eta_i(t), \tag{C2b}$$

which are Gaussian white noises with zero mean and correlation matrix

$$\left\langle \xi_i^1(t)\,\xi_j^1(t')\right\rangle = \left\langle \xi_i^2(t)\,\xi_j^2(t')\right\rangle = \delta_{ij}\delta(t-t')\left(\sigma_{int}^2+\sigma_{ext}^2\right), \tag{C3a}$$

$$\left\langle \xi_i^1(t)\,\xi_j^2(t')\right\rangle = \delta_{ij}\delta(t-t')\,\sigma_{ext}^2. \tag{C3b}$$

To start with, one can obtain the average values for the stationary rates by applying averages to both sides of equations C1 and imposing the stationary state condition, $\langle \dot{x}_i\rangle = \langle \dot{y}_i\rangle = 0$. Once this is done, it is immediate to solve the resulting linear system and check that $\langle x_i^*\rangle, \langle y_i^*\rangle \propto \eta_0$, where the star (*) indicates that these values correspond to the stationary state. Hence, making $\eta_0 = 0$ the mean values vanish. One can demonstrate that correlations do not depend on η_0, and hence we can make $\eta_0 = 0$ without loss of generality. Conceptually, this means just shifting up or down the baseline of fluctuations of the firing rate, which does not affect the fluctuations themselves.

In order to compute correlations, we need to evaluate the second order moments between different populations, as $\langle x_i y_j\rangle$ or $\langle x_i^2\rangle$. A possible way of doing this is realising that we have a linear system, thus starting from the analytical solution of the multidimensional Orstein-Uhlenbeck process [59]. However, this approach will yield a linear system with $N(N+1)/2$ variables to solve for, which are all the elements of the (symmetric) correlation matrix. But all the groups are identical (or *indistinguishable*), so we expect correlations not to depend on the particular population chosen. Therefore, all the equations will be reduced to just 6 covariances: $\langle x_i x_j\rangle$, $\langle x_i y_j\rangle$, $\langle y_i y_j\rangle$, $\langle x_i^2\rangle$, $\langle x_i y_i\rangle$ and $\langle y_i^2\rangle$.

In this context, it is conceptually simpler to obtain equations for the evolution of the second moments, and then evaluate them in the stationary state. We report here in detail the computation of two of these moments as an example, giving just the final answer for the other four, which is performed in an analogous way.

First, we define $X_{ij} = x_i x_j$, and then look for the time evolution of X_{ij}, i.e., $\dot{X}_{ij}$. Notice that this a non-linear change of variables, and thus Itô's lemma is required. The lemma tells us that if we have a change of variables $z = z(\vec{x})$ then one has to include the second-order terms in the expansion,

$$dz = \underbrace{\sum_{i=1}^{N}\partial_{x_i} z\,dx_i}_{\text{Chain rule}} + \underbrace{\frac{1}{2}\sum_{i=1}^{N}\partial_{x_i}\partial_{x_j} z\,dx_i dx_j}_{\text{Itô's lemma}}. \tag{C4}$$

The terms dx_i can be obtained as $\dot{x}_i dt$. It is important to remark that in this procedure noise terms are rewritten as the differential of the Wiener processes, i.e., $\eta_i(t)dt = dW_i$. After applying the Itô lemma above, only terms up to order dt should be taken into account. Recall that $dW_i(t) \propto \sqrt{dt}$. Finally, one just divides again by dt to recover the stochastic differential equation and applies the ensemble average.

For X_{ij}, this reads as

$$\begin{aligned}\frac{d\langle x_i x_j\rangle}{dt} = \langle \dot{x}_i x_j\rangle + \langle x_i \dot{x}_j\rangle + \frac{1}{2}\langle \dot{x}_i \dot{x}_j\rangle = \\ a\langle x_i x_j\rangle + b\langle y_i x_j\rangle + \frac{W^{EE}}{M-1}\sum_{k\neq i}\langle x_k x_j\rangle + \frac{W^{EI}}{M-1}\sum_{k\neq i}\langle y_k x_j\rangle + \\ + a\langle x_i x_j\rangle + b\langle x_i y_j\rangle + \frac{W^{EE}}{M-1}\sum_{k\neq j}\langle x_i x_k\rangle + \frac{W^{EI}}{M-1}\sum_{k\neq j}\langle x_i y_k\rangle + \\ \left\langle \xi_i^1 x_j\right\rangle + \left\langle x_i \xi_j^1\right\rangle + \left\langle \xi_i^1 \xi_j^1\right\rangle + \mathcal{O}\left(dt^2\right),\end{aligned} \tag{C5}$$

where all the averages between the noise and the variable yield 0, due to Itô's prescription. The next step is to simplify the sums involving correlations. As it was discussed above, since clusters are indistinguishable all the terms are exactly the same. However, notice that an index k running from 1 to N will inevitably hit $k = j$, and this has to be taken into account separately, since the correlation with a population inside of my group is different to that of a population outside of it. Then,

$$\sum_{k\neq i}\langle x_k x_j\rangle = (M-2)\langle x_i x_j\rangle + \left\langle x_i^2\right\rangle, \tag{C6}$$

allowing us to simplify the equation. At this step we simplify the notation by letting $X_{ij} = \langle x_i x_j\rangle$, $Z_{ij} = \langle x_i y_j\rangle$, $Y_i = \langle y_i^2\rangle$, etc., leading to

$$\frac{1}{2}\dot{X}_{ij} = \left(a + \frac{M-2}{M-1}W^{EE}\right) X_{ij} + \left(b + \frac{M-2}{M-1}W^{EI}\right) Z_{ij} + \frac{1}{M-1}\left[W^{EE}X_i + W^{EI}Z_i\right]. \tag{C7}$$

The same procedure can be repeated for all the other correlations, such as

$$\begin{aligned}\frac{d\left\langle y_i^2\right\rangle}{dt} &= \langle 2y_i\dot{y}_i\rangle + \frac{1}{2}\left\langle 2\dot{y}_i^2\right\rangle = \\ &2c\left\langle x_i y_i\right\rangle + 2d\left\langle y_i^2\right\rangle + \frac{2W^{IE}}{M-1}\sum_{k\neq i}\left\langle y_i x_k\right\rangle + \frac{2W^{II}}{M-1}\sum_{k\neq i}\left\langle y_i y_k\right\rangle + 2\left\langle y_i\xi_i^2\right\rangle + \left\langle \xi_i^2\xi_i^2\right\rangle = \\ &2c\left\langle Z_i\right\rangle + 2d\left\langle Y_i\right\rangle + 2W^{IE}\left\langle Z_{ij}\right\rangle + 2W^{II}\left\langle Y_{ij}\right\rangle + \sigma_{\text{int}}^2 + \sigma_{\text{ext}}^2,\end{aligned} \tag{C8}$$

where now the correlation between noises yields a non vanishing value. This operation is repeated with all the remaining terms, in order to find a linear system of differential equations with 6 variables and 6 equations,

$$\frac{1}{2}\dot{X}_i = aX_i + bZ_i + \left[W^{EE}X_{ij} + W^{EI}Z_{ij}\right] + \frac{1}{2}\left(\sigma_{\text{int}}^2 + \sigma_{\text{ext}}^2\right), \tag{C9a}$$

$$\frac{1}{2}\dot{Y}_i = cZ_i + dY_i + \left[W^{IE}Z_{ij} + W^{II}Y_{ij}\right] + \frac{1}{2}\left(\sigma_{\text{int}}^2 + \sigma_{\text{ext}}^2\right), \tag{C9b}$$

$$\dot{Z}_i = cX_i + (a+d)\,Z_i + bY_i + \left[W^{IE}X_{ij} + \left(W^{EE} + W^{II}\right)Z_{ij} + W^{EI}Y_{ij}\right] + \sigma_{\text{ext}}^2, \tag{C9c}$$

$$\frac{1}{2}\dot{X}_{ij} = \left(a + \frac{M-2}{M-1}W^{EE}\right) X_{ij} + \left(b + \frac{M-2}{M-1}W^{EI}\right) Z_{ij} + \frac{1}{M-1}\left[W^{EE}X_i + W^{EI}Z_i\right], \tag{C9d}$$

$$\frac{1}{2}\dot{Y}_{ij} = \left(d + \frac{M-2}{M-1}W^{II}\right) Y_{ij} + \left(c + \frac{M-2}{M-1}W^{IE}\right) Z_{ij} + \frac{1}{M-1}\left[W^{II}Y_i + W^{IE}Z_i\right], \tag{C9e}$$

$$\dot{Z}_{ij} = \left(c + \frac{M-2}{M-1}W^{IE}\right) X_{ij} + \left(b + \frac{M-2}{M-1}W^{EI}\right) Y_{ij} + \left(a + d + \frac{M-2}{M-1}\left(W^{EE} + W^{II}\right)\right) Z_{ij} + \tag{C9f}$$

$$+ \frac{1}{M-1}\left[W^{IE}X_i + W^{EI}Y_i + \left(W^{EE} + W^{II}\right)Z_i\right]. \tag{C9g}$$

This system can be solved in the stationary limit, when all the derivatives of the left-hand side are zero. From these, one is able to obtain the Pearson correlation coefficients. Notice that since correlation with itself is always unity, there is only four coefficients remaining: the correlation between excitation and inhibition inside a group $C_{EI}^{\text{int}} = Z_i^*/\sqrt{X_i^* Y_i^*}$, and all three between-group correlations, $C_{EE}^{\text{ext}} = X_{ij}^*/X_i^*$, $C_{II}^{\text{ext}} = Y_{ij}^*/Y_i^*$, and $C_{EI}^{\text{ext}} = Z_{ij}^*/\sqrt{X_i^* Y_i^*}$.

2. Solutions for the homogeneous network

In some special cases it is possible to give a simple solution in closed form for the correlation coefficients. One example is the homogeneous network: when all weights are identical, and an intrinsic decay is added to both the excitatory and inhibitory populations (i.e., with $c = W^{EE} = W^{IE} = +W$, $b = W^{IE} = W^{II} = -W$ and $a = W - 1$, $d = -W - 1$) the solution reads

$$C_{EI}^{\text{int}} = \frac{r^2(M-1)^2 + W^2(1-r)^2 M}{\sqrt{M^2(1-r)^4 W^4 + (M-1)(1-r)^2 W^2\left[M(2-4(1-r)r) - (1-r)^2\right] + (M-1)^4\left[2(r-1)r+1\right]^2}}, \tag{C10a}$$

$$C_{EI}^{\text{ext}} = \frac{W^2(1-r)^2 M}{\sqrt{M^2(1-r)^4 W^4 + (M-1)(1-r)^2 W^2\left(M(4(r-1)r+2) - (r-1)^2\right) + (M-1)^4(2(r-1)r+1)^2}}, \tag{C10b}$$

$$C_{EE}^{\text{ext}} = \frac{(1-r)^2 W\left((W+1)M - 1\right)}{M(1-r)^2W^2 + (M-1)(1-r)^2W + (1-M)^2\left[1-2(1-r)r\right]}, \tag{C10c}$$

$$C_{II}^{\text{ext}} = \frac{(1-r)^2 W\left((W-1)M + 1\right)}{M(1-r)^2W^2 - (M-1)(1-r)^2W + (1-M)^2\left[1-2(1-r)r\right]} \tag{C10d}$$

where we additionally defined r as the signal-to-noise ratio, $\sigma_{\text{int}} = r\sigma$, $\sigma_{\text{ext}} = (1-r)\sigma$. This analytical solution has some interesting features. First, notice it does not depend on the total amount of noise σ that the system receives, but only on the ratio between external and internal noise. Second, if $W \to \infty$ all correlations go to 1. Expanding in series around $\epsilon = 1/W = 0$, one can see that all coefficients are $C = 1 - \mathcal{O}(1/W^2)$ for large coupling. It is also possible to study the limiting values of the noise. $r = 1$ makes all the between-group correlations equal to zero, while coupling determines the in-group value. On the other hand, when $r \ll 1$, one gets

$$C_{EI}^{\text{int}} \simeq C_{EI}^{\text{ext}} \simeq \frac{MW^2}{(M-1)^2} + \mathcal{O}(r^2), \tag{C11a}$$

$$C_{EE}^{\text{ext}} \simeq -C_{II}^{\text{ext}} \simeq \frac{W}{M-1} + \mathcal{O}(r^2). \tag{C11b}$$

meaning that the external correlations grow linearly with the coupling, but quadratically with the signal to noise ratio: a small increase in coupling needs to be followed by a larger increase in signal intensity in order to recover the previous tuning. As a result, coupling has a larger impact in tuning than the signal to noise ratio, a effect that can be measured in the full spiking network.

Finally, we see that between-group correlations also tend to zero as the limit $M \to +\infty$ is taken, since in that case, the input that a module receives from all others becomes just white noise. A finite number of clusters (or a finite connectivity among them) is thus required for tuning.

3. Clustering optimisation

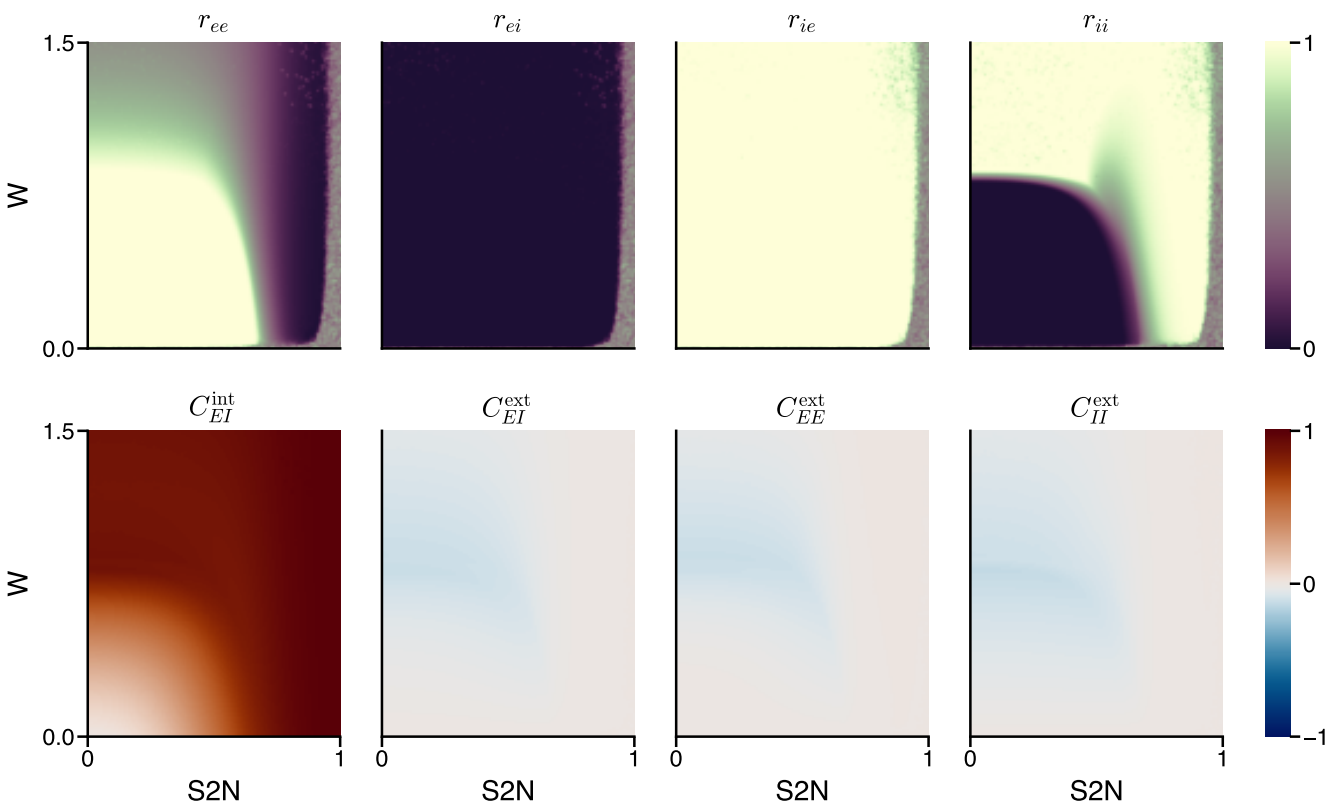

FIG. 8. Analytical results for optimal clustering

Optimizasion of the clustering for the fully-connected network can be done by minimising a loss function which depends on the correlations. A simple possibility is to employ minimum squares,

$$\mathcal{L}^{an}[C;p] = \left(1 - C_{EI}^{\text{int}}\right)^2 + \left(C_{EE}^{\text{ext}}\right)^2 + \left(C_{EI}^{\text{ext}}\right)^2 + \left(C_{II}^{\text{ext}}\right)^2 . \tag{C12}$$

The solution and associated optimal correlations are shown in 8. There are several key remarks following from this figure:

1. Even for very large clustering and extremely low signal to noise ratio, the clustering is able to provide an in-group correlation close to unity combined with low between-group correlations, thus ensuring co-tuning.

2. Inhibition to excitation never clusters. Inhibitory neurons act over excitatory individuals regardless of their cluster.

3. Excitatory connections are clustered. In particular, excitatory connections always project to inhibitory neurons in their own cluster, but not to other ones. Excitatory-to-excitatory connectivity is also strongly clustered, except for large coupling.

4. Inhibition controls excitation for large W. If one keeps highly clustered excitation and increases the coupling, the dynamics of single modules becomes unstable at a critical value $W_c(r)$. However, the network can remain stable if the excitatory clustering is reduced and the amount of inhibition in the group increases, which can be accomplished by increasing r_{II}.

5. When the signal to noise ratio is close to one, clustering becomes mostly irrelevant, since system is driven by the external input which allows co-tuning easily.

Notice that the optimization algorithm automatically finds solutions where the equations are well-defined –i.e., where the system reaches a stationary state– thus selecting to increase the inhibitory clustering when W goes over the instability threshold.

Therefore, the analytical approach is able to find a good candidate for the optimal clustering depending on the network dynamics. Although it cannot be directly applied to the spiking network, which is able to display richer dynamics, tells us that as rule of thumb excitatory clustering should be as high as possible while avoid crossing an instability threshold. If this happens, inhibition needs to be increased.

Appendix D: Tables of parameters

We mostly used the neuron model parameters from the original inhibitory STDP paper [20]

Network Model		
Symbol	Description	Value
N	Number of neurons	1000
N_E	Number of E neurons	800
N_I	Number of I neurons	200
M	Number of input groups	8
g_{leak}	Leak conductance	10 nS
V_{rest}	Resting potential	-60 mV
V_{reset}	Reset potential	-60 mV
V_{th}	Spiking threshold	-50 mV
V_E	Excitatory reversal potential	0 mV
V_I	Inhibitory reversal potential	-80 mV
C_m	Membrane capacitance	200 pF
τ_{ref}	Absolute refractory period	5 ms
τ_E	Decay time constant of E conductance	5 ms
τ_I	Decay time constant of I conductance	10 ms
$\overline{g_E}$	E weight scaling constant	1.4 nS
$\overline{g_I}$	I weight scaling constant	3.5 nS

Plasticity Rules		
Symbol	Description	Value
τ_1^{estdp}	Slow eSTDP timescale	50 ms
τ_2^{estdp}	Fast eSTDP timescale	10 ms
η_E	eSTDP learning rate	0.0025
A_{LTP}	Long term potentiation amplitude	1.0
A_{LTD}	Long term depression amplitude	0.1
τ^{istdp}	iSTDP timescale	10 ms
η_I	iSTDP learning rate	0.01
ρ_0	iSTDP target rate	3 Hz
η_N	Normalization learning rate	0.003
W^E_{target}	Excitatory normalization target	5.0
W^I_{target}	Inhibitory normalization target	5.0

ABC Optimization		
Symbol	Description	Value
α	weight of in-group correlation	0.1
β	weight of between-group correlations	0.3