Nonlinear transient amplification in recurrent neural networks with short-term plasticity

Yue Kris Wu^{1,2,3,4} and Friedemann Zenke^{1,2*}

*For correspondence:
friedemann.zenke@fmi.ch (FZ)

¹Friedrich Miescher Institute for Biomedical Research, Basel, Switzerland; ²Faculty of Natural Sciences, University of Basel, Basel, Switzerland; ³Max Planck Institute for Brain Research, Frankfurt, Germany; ⁴School of Life Sciences, Technical University of Munich, Freising, Germany

Abstract To rapidly process information, neural circuits have to amplify specific activity patterns transiently. How the brain performs this nonlinear operation remains elusive. Hebbian assemblies are one possibility whereby strong recurrent excitatory connections boost neuronal activity. However, such Hebbian amplification is often associated with dynamical slowing of network dynamics, non-transient attractor states, and pathological run-away activity. Feedback inhibition can alleviate these effects but typically linearizes responses and reduces amplification gain. Here we study nonlinear transient amplification (NTA), a plausible alternative mechanism that reconciles strong recurrent excitation with rapid amplification while avoiding the above issues. NTA has two distinct temporal phases. Initially, positive feedback excitation selectively amplifies inputs that exceed a critical threshold. Subsequently, short-term plasticity quenches the run-away dynamics into an inhibition-stabilized network state. By characterizing NTA in supralinear network models, we establish that the resulting onset transients are stimulus selective and well-suited for speedy information processing. Further, we find that excitatory-inhibitory co-tuning widens the parameter regime in which NTA is possible in the absence of persistent activity. In summary, NTA provides a parsimonious explanation for how excitatory-inhibitory co-tuning and short-term plasticity collaborate in recurrent networks to achieve transient amplification.

Introduction

Perception in the brain is reliable and strikingly fast. Recognizing a familiar face or locating an animal in a picture only takes a split second (*Thorpe et al., 1996*). This pace of processing is truly remarkable since it involves several recurrently connected brain areas each of which has to selectively amplify or suppress specific signals before propagating them further. This processing is mediated through circuits with several intriguing properties. First, excitatory-inhibitory (EI) currents into individual neurons are commonly correlated in time and co-tuned in stimulus space (*Wehr and Zador, 2003; Froemke et al., 2007; Okun and Lampl, 2008; Hennequin et al., 2017; Rupprecht and Friedrich, 2018; Znamenskiy et al., 2018*). Second, neural responses to stimulation are shaped through diverse forms of short-term plasticity (STP) (*Tsodyks and Markram, 1997; Markram et al., 1998; Zucker and Regehr, 2002; Pala and Petersen, 2015*). Finally, mounting evidence suggests that amplification rests on neuronal ensembles with strong recurrent excitation (*Marshall et al., 2019; Peron et al., 2020*), whereby excitatory neurons with similar tuning preferentially form reciprocal

connections (Ko et al., 2011; Cossell et al., 2015). Such predominantly symmetric connectivity between excitatory cells is consistent with the notion of Hebbian cell assemblies (Hebb, 1949), which are considered an essential component of neural circuits and the putative basis of associative memory (Harris, 2005; Josselyn and Tonegawa, 2020). Computationally, Hebbian cell assemblies can amplify specific activity patterns through positive feedback, also referred to as Hebbian amplification. Based on these principles, several studies have shown that Hebbian amplification can drive persistent activity that outlasts a preceding stimulus (Hopfield, 1982; Amit and Brunel, 1997; Yakovlev et al., 1998; Wong and Wang, 2006; Zenke et al., 2015; Gillary et al., 2017), comparable to selective delay activity observed in the prefrontal cortex when animals are engaged in working memory tasks (Funahashi et al., 1989; Romo et al., 1999).

However, in most brain areas, evoked responses are transient and sensory neurons typically exhibit pronounced stimulus onset responses, after which the circuit dynamics settle into a low-activity steady-state even when the stimulus is still present (DeWeese et al., 2003; Mazor and Laurent, 2005; Bolding and Franks, 2018). Preventing run-away excitation and multi-stable attractor dynamics in recurrent networks requires powerful and often finely tuned feedback inhibition resulting in EI balance (Amit and Brunel, 1997; Compte et al., 2000; Litwin-Kumar and Doiron, 2012; Ponce-Alvarez et al., 2013; Mazzucato et al., 2019). However, strong feedback inhibition tends to linearize steady-state activity (Van Vreeswijk and Sompolinsky, 1996; Baker et al., 2020). Murphy and Miller (2009) proposed *balanced amplification* which reconciles transient amplification with strong recurrent excitation by tightly balancing recurrent excitation with strong feedback inhibition (Goldman, 2009; Hennequin et al., 2012, 2014; Bondanelli and Ostojic, 2020; Gillett et al., 2020). Importantly, balanced amplification was formulated for linear network models of excitatory and inhibitory neurons. Due to linearity, it intrinsically lacks the ability to nonlinearly amplify stimuli which limits its capabilities for pattern completion and pattern separation. Further, how balanced amplification relates to nonlinear neuronal activation functions and nonlinear synaptic transmission as, for instance, mediated by STP (Tsodyks and Markram, 1997; Markram et al., 1998; Zucker and Regehr, 2002; Pala and Petersen, 2015), remains elusive. This begs the question of whether there are alternative nonlinear amplification mechanisms and how they relate to existing theories of recurrent neural network processing.

Here we address this question by studying an alternative mechanism for the emergence of transient dynamics that relies on recurrent excitation, supralinear neuronal activation functions, and STP. Specifically, we build on the notion of ensemble synchronization in recurrent networks with STP (Loebel and Tsodyks, 2002; Loebel et al., 2007) and study this phenomenon in analytically tractable network models with rectified quadratic activation functions (Ahmadian et al., 2013; Rubin et al., 2015; Hennequin et al., 2018; Kraynyukova and Tchumatchenko, 2018) and STP. We first characterize the conditions under which individual neuronal ensembles with supralinear activation functions and recurrent excitatory connectivity succumb to explosive run-away activity in response to external stimulation. We then show how STP effectively mitigates this instability by re-stabilizing ensemble dynamics in an inhibition-stabilized network (ISN) state, but only after generating a pronounced stimulus-triggered onset transient. We call this mechanism NTA and show that it yields selective onset responses that carry more relevant stimulus information than the subsequent steady-state. Finally, we characterize the functional benefits of inhibitory co-tuning, a feature that is widely observed in the brain (Wehr and Zador, 2003; Froemke et al., 2007; Okun and Lampl, 2008; Rupprecht and Friedrich, 2018) and readily emerges in computational models endowed with activity-dependent plasticity of inhibitory synapses (Vogels et al., 2011). We find that co-tuning prevents persistent attractor states but does not preclude NTA from occurring. Importantly, NTA purports that, following transient amplification, neuronal ensembles settle into a stable ISN state, consistent with recent studies suggesting that inhibition stabilization is a ubiquitous feature of cortical networks (Sanzeni et al., 2020). In summary, our work indicates that NTA is ideally suited to amplify stimuli rapidly through the interaction of strong recurrent excitation with STP.

Results

To understand the emergence of transient responses in recurrent neural networks, we studied rate-based population models with a supralinear, power law input-output function (Figure 1A, B; Ahmadian et al., 2013; Hennequin et al., 2018), which captures essential aspects of neuronal activation (Priebe et al., 2004), while also being analytically tractable. We first considered an isolated neuronal ensemble consisting of one excitatory (E) and one inhibitory (I) population (Figure 1A).

The dynamics of this network are given by

$$\tau_E \frac{dr_E}{dt} = -r_E + \left[J_{EE}r_E - J_{EI}r_I + g_E \right]_+^{\alpha_E}, \quad (1)$$

$$\tau_I \frac{dr_I}{dt} = -r_I + \left[J_{IE}r_E - J_{II}r_I + g_I \right]_+^{\alpha_I}, \quad (2)$$

where r_E and r_I are the firing rates of the excitatory and inhibitory population, τ_E and τ_I represent the corresponding time constants, J_{XY} denotes the synaptic strength from the population Y to the population X , where $X, Y \in \{E, I\}$, g_E and g_I are the external inputs to the respective populations. Finally, α_E and α_I , the exponents of the respective input-output functions, are fixed at two unless mentioned otherwise. For ease of notation, we further define the weight matrix $\mathbf{J}$ of the compound system as follows:

$$\mathbf{J} = \begin{bmatrix} J_{EE} & -J_{EI} \\ J_{IE} & -J_{II} \end{bmatrix}. \quad (3)$$

We were specifically interested in networks with strong recurrent excitation that can generate positive feedback dynamics in response to external inputs g_E . Therefore, we studied networks with

$$\det(\mathbf{J}) = -J_{EE}J_{II} + J_{IE}J_{EI} < 0. \quad (4)$$

In contrast, networks in which recurrent excitation is met by strong feedback inhibition such that $\det(\mathbf{J}) > 0$ are unable to generate positive feedback dynamics provided that inhibition is fast enough (Ahmadian et al., 2013). Importantly, we assumed that most inhibition originates from recurrent connections (Franks et al., 2011; Large et al., 2016) and, hence, we kept the input to the inhibitory population g_I fixed unless mentioned otherwise.

Nonlinear amplification of inputs above a critical threshold

We initialized the network in a stable low-activity state in the absence of external stimulation, consistent with spontaneous activity in cortical networks (Figure 1C). However, an input g_E of sufficient strength, destabilized the network (Figure 1C). Importantly, this behavior is distinct from linear network models in which the network stability is independent of inputs (Methods). The transition from stable to unstable dynamics can be understood by examining the phase portrait of the system (Figure 1D). Before stimulation, the system has a stable and an unstable fixed point (Figure 1D, left). However, both fixed points disappear for an input g_E above a critical stimulus strength (Figure 1D, right).

To further understand the system's bifurcation structure, we consider the characteristic function

$$F(z) = J_{EE} \left[z \right]_+^{\alpha_E} - J_{EI} \left[\det(\mathbf{J}) \cdot J_{EI}^{-1} \left[z \right]_+^{\alpha_E} + J_{EI}^{-1} J_{II} z - J_{EI}^{-1} J_{II} g_E + g_I \right]_+^{\alpha_I} - z + g_E, \quad (5)$$

where z denotes the total current into the excitatory population and $\det(\mathbf{J})$ represents the determinant of the weight matrix (Kraynyukova and Tchumatchenko, 2018; Methods). The characteristic function reduces the original two-dimensional system to one dimension, whereby the zero crossings of the characteristic function correspond to the fixed points of the original system (cf. Eq. (1)-(2)). We use this correspondence to visualize how the fixed points of the system change with the input g_E . Increasing g_E shifts $F(z)$ upwards, which eventually leads to all zero crossings disappearing and the ensuing unstable dynamics (Figure 1E; Methods). Importantly, for any weight matrix $\mathbf{J}$ with negative determinant, there exists a critical input g_E at which all fixed points disappear (Methods). While for weak recurrent E-to-E connection strength J_{EE} , the transition from stable dynamics

Figure 1. Neuronal ensembles nonlinearly amplify inputs above a critical threshold. **(A)** Schematic of the recurrent ensemble model consisting of an excitatory (blue) and an inhibitory population (red). **(B)** Supralinear input-output function given by a rectified power law with exponent $\alpha = 2$. **(C)** Firing rates of the excitatory (blue) and inhibitory population (red) in response to external stimulation during the interval from 2–4 s (gray bar). The stimulation was implemented by temporarily increasing the input g_E . **(D)** Phase portrait of the system before stimulation (left; cf. C orange) and during stimulation (right; cf. C green). **(E)** Characteristic function $F(z)$ for varying input strength g_E . Note that the function loses its zero crossings, which correspond to fixed points of the system for increasing external input. **(F)** Heat map showing the evoked firing rate of the excitatory population for different parameter combinations J_{EE} and g_E . The gray region corresponds to the parameter regime with unstable dynamics.

Figure 1–Figure supplement 1. Unstable ensemble dynamics can be triggered by additional stimulation in supralinear networks with negative determinant even in the presence of substantial feedforward inhibition.

to unstable is gradual, in that it happens at higher firing rates (**Figure 1F**), it becomes more abrupt for stronger J_{EE} . Thus, our analysis demonstrates that individual neuronal ensembles with negative determinant $\det(\mathbf{J})$ nonlinearly amplify inputs above a critical threshold by switching from initially stable to unstable dynamics.

Short-term plasticity, but not spike-frequency adaptation, can re-stabilize ensemble dynamics

Since unstable dynamics are not observed in neurobiology, we wondered whether neuronal spike-frequency adaptation (SFA) or STP could re-stabilize the ensemble dynamics while keeping the non-linear amplification character of the system. Specifically, we considered SFA of excitatory neurons, E-to-E short-term depression (STD), and E-to-I short-term facilitation (STF). We focused on these particular mechanisms because they are ubiquitously observed in the brain. Most pyramidal cells exhibit SFA (*Barkai and Hasselmo, 1994*) and most synapses show some form of STP (*Markram et al., 1998; Zucker and Regehr, 2002; Pala and Petersen, 2015*). Moreover, the time scales of these mechanisms are well-matched to typical timescales of perception, i.e., from milliseconds to seconds (*Tsodyks and Markram, 1997; Fairhall et al., 2001; Pozzorini et al., 2013*).

When we simulated our model with SFA (Eqs. (21)–(23)), we observed different network behaviors depending on the adaptation strength. When adaptation strength was weak, SFA was unable to stabilize run-away excitation (**Figure 2A**; Methods). Increasing the adaptation strength eventually prevented run-away excitation, but to give way to oscillatory ensemble activity (**Figure 2–Figure Supplement 1**). Finally, we confirmed analytically that SFA cannot stabilize excitatory run-away dynamics at a stable fixed point (Methods). In particular, while the input is present, strong SFA creates a stable limit cycle with associated oscillatory ensemble activity (**Figure 2–Figure Supplement 1**; Methods), which was also shown in previous modeling studies (*Van Vreeswijk and Hansel, 2001*),

156 but is not typically observed in sensory systems (DeWeese et al., 2003; Rupprecht and Friedrich,
157 2018).

158 Next, we considered STP, which is capable of saturating the effective neuronal input-output
159 function (Mongillo et al., 2012; Zenke et al., 2015) (Eqs. (37)–(39), Eqs. (41)–(43)). We first analyzed
160 the stimulus-evoked network dynamics when we added STD to the recurrent E-to-E connections.
161 Strong depression of synaptic efficacy resulted in a brief onset transient after which the ensemble
162 dynamics quickly settled into a stimulus-evoked steady-state with slightly higher activity than the
163 baseline (Figure 2B, left). After stimulus removal, the ensemble activity returned back to its base-
164 line level (Figure 2B, left; Figure 2C). Notably, the ensemble dynamics settled at a stable steady
165 state with a much higher firing rate, when inhibition was inactivated during stimulus presentation
166 (Figure 2B, right). This shows that STP is capable of creating a stable high-activity fixed point, which
167 is fundamentally different from the SFA dynamics discussed above. This difference in ensemble
168 dynamics can be readily understood by analyzing the stability of the three-dimensional dynamical
169 system (Methods; Appendix 1). We can gain a more intuitive understanding by considering self-
170 consistent solutions of the characteristic function $F(z)$. Initially, the ensemble is at the stable low
171 activity fixed point. But the stimulus causes this fixed point to disappear, thus giving way to positive
172 feedback which creates the leading edge of the onset transient (cf. Figure 2B). However, because
173 E-to-E synaptic transmission is rapidly reduced by STD, the curvature of $F(z)$ changes and a stable
174 fixed point is created, thereby allowing excitatory run-away dynamics to terminate and the ensem-
175 ble dynamics settle into a steady-state at low activity levels (Figure 2D). We found that E-to-I STF
176 leads to similar dynamics (Figure 2E, left) with the only difference that this configuration requires
177 inhibition for network stability (Figure 2E, right), whereas E-to-E STD stabilizes activity even without
178 inhibition, albeit at physiologically implausibly high activity levels. Importantly, the re-stabilization
179 through either form of STP did not impair an ensemble’s ability to amplify stimuli during the initial
180 onset phase.

181 Crucially, transient amplification in supralinear networks with STP occurs above a critical thresh-
182 old (Figure 2–Figure Supplement 2), and requires recurrent excitation J_{EE} to be sufficiently strong
183 (Figure 2–Figure Supplement 2C, D). To quantify the amplification ability of these networks, we cal-
184 culated the ratio of the evoked peak firing rate to the input strength, henceforth called the “Ampli-
185 fication index”. We found that amplification in STP-stabilized supralinear networks can be orders
186 of magnitude larger than in linear networks with equivalent weights and comparable stabilized
187 supralinear networks (SSNs) without STP (Figure 2–Figure Supplement 3). We stress that the re-
188 sulting firing rates are parameter-dependent (Figure 2–Figure Supplement 4) and their absolute
189 value can be high due to the high temporal precision of the onset peak and its short duration.
190 In experiments, such high rates manifest themselves as precisely time-locked spikes with millisec-
191 ond resolution (DeWeese et al., 2003; Wehr and Zador, 2003; Bolding and Franks, 2018; Gjoni et al.,
192 2018).

193 Recent studies suggest that cortical networks operate as inhibition-stabilized networks (ISNs)
194 (Sanzeni et al., 2020; Sadeh and Clopath, 2021), in which the excitatory network is unstable in the
195 absence of feedback inhibition (Tsodyks et al., 1997; Ozeki et al., 2009). To that end, we investigated
196 how ensemble re-stabilization relates to the network operating regime at baseline and during stim-
197 ulation. Whether a network is an ISN or not is mathematically determined by the real part of the
198 leading eigenvalue of the Jacobian of the excitatory-to-excitatory subnetwork (Tsodyks et al., 1997).
199 We computed the leading eigenvalue in our model incorporating STP and referred to it as “ISN in-
200 dex” (Methods; Appendix 2). We found that in networks with STP the ISN index can switch sign
201 from negative to positive during external stimulation, indicating that the ensemble can transition
202 from a non-ISN to an ISN (Figure 2F). Notably, this behavior is distinct from linear network models
203 in which the network operating regime is independent of the input (Methods). Whether this switch
204 between non-ISN to ISN occurred, however, was parameter dependent and we also found network
205 configurations that were already in the ISN regime at baseline and remained ISNs during stimula-
206 tion (Figure 2–Figure Supplement 5). Thus, re-stabilization was largely unaffected by the network

Figure 2. Short-term plasticity, but not spike-frequency adaptation, re-stabilizes ensemble dynamics. **(A)** Firing rates of the excitatory (blue) and inhibitory population (red) in the presence of spike-frequency adaptation (SFA). During stimulation (gray bar) additional input is injected into the excitatory population. The inset shows a cartoon of how SFA affects spiking neuronal dynamics in response to a step current input. **(B)** Left: Same as (A) but in the presence of E-to-E short-term depression (STD). Right: Same as left but inactivating inhibition in the period marked in purple. **(C)** 3D plot of the excitatory activity r_E , inhibitory activity r_I , and the STD variable x of the network in B left. The orange and green points mark the fixed points before/after and during stimulation. **(D)** Characteristic function $F(z)$ in networks with E-to-E STD. Different brightness levels correspond to different time points in B left. **(E)** Same as (B) but in the presence of E-to-I short-term facilitation (STF). **(F)** Inhibition-stabilized network (ISN) index, which corresponds to the largest real part of the eigenvalues of the Jacobian matrix of the E-E subnetwork with STD, as a function of time for the network with E-to-E STD in B left. For values above zero (dashed line), the ensemble is an ISN. **(G)** Analytical solution of non-ISN (magenta), ISN (green), paradoxical, and non-paradoxical regions for different parameter combinations J_{EE} and the STD variable x . The solid line separates the non-ISN and ISN regions, whereas the dashed line separates the non-paradoxical and paradoxical regions. **(H)** The normalized firing rates of the excitatory (blue) and inhibitory population (red) when injecting additional excitatory current into the inhibitory population before stimulation (left; cf. orange bar in B), and during stimulation (right; cf. green bar in B). Initially, the ensemble is in the non-ISN regime and injecting excitatory current into the inhibitory population increases its firing rate. During stimulation, however, the ensemble is an ISN. In this case, excitatory current injection into the inhibitory population results in a reduction of its firing rate, also known as the *paradoxical effect*.

Figure 2–Figure supplement 1. Ensemble dynamics in supralinear networks with strong SFA.

Figure 2–Figure supplement 2. Dependence of peak amplitude and fixed point activity on input g_E and E-to-E connection strength J_{EE} .

Figure 2–Figure supplement 3. Comparisons of amplification ability between NTA and linear networks, and between NTA and SSNs.

Figure 2–Figure supplement 4. Dependence of peak amplitude and fixed point activity on STP parameters.

Figure 2–Figure supplement 5. Networks initially in the ISN regime can exhibit strong NTA.

Figure 2–Figure supplement 6. ISN index and paradoxical effect test for networks with E-to-I STF.

Figure 2–Figure supplement 7. Inhibition stabilization does not imply paradoxical response in networks with E-to-E STD.

Figure 2–Figure supplement 8. Transition from non-ISN to ISN indicating by frozen inhibition test.

Figure 2–Figure supplement 9. Similar qualitative behavior in rate-based models with maximal firing rate capped at 300 Hz.

Figure 2–Figure supplement 10. Similar qualitative behavior in spiking neural networks.

Figure 2–Figure supplement 11. Unstable dynamics can emerge in supralinear networks with positive determinant and slow inhibition.

Figure 2–Figure supplement 12. Networks with substantial feedforward inhibition can exhibit strong NTA.

state and consistent with experimentally observed ISN states (*Sanzeni et al., 2020*).

Theoretical studies have shown that one defining characteristic of ISNs in static excitatory and inhibitory networks is that injecting excitatory (inhibitory) current into inhibitory neurons decreases (increases) inhibitory firing rates, which is also known as the paradoxical effect (*Tsodyks et al., 1997; Miller and Palmigiano, 2020*). Yet, it is unclear whether in networks with STP, inhibitory stabilization implies paradoxical response and vice versa. We therefore analyzed the condition of being an ISN and the condition of having paradoxical response in networks with STP (Methods; *Appendix 2; Appendix 3*). Interestingly, we found that in networks with E-to-E STD, the paradoxical effect implies inhibitory stabilization, whereas inhibitory stabilization does not necessarily imply paradoxical response (*Figure 2G*; Methods), suggesting that having paradoxical effect is a sufficient but not necessary condition for being an ISN. In contrast, in networks with E-to-I STF, inhibitory stabilization and paradoxical effect imply each other (*Appendix 2; Appendix 3*). Therefore, paradoxical effect can be exploited as a proxy for inhibition stabilization for networks with STP we considered here. By injecting excitatory current into the inhibitory population, we found that the network did not exhibit the paradoxical effect before stimulation (*Figure 2H*, left; *Figure 2–Figure Supplement 6*). In contrast, injecting excitatory inputs into the inhibitory population during stimulation reduced their activity (*Figure 2H*, right; *Figure 2–Figure Supplement 6*). As demonstrated in our analysis, non-paradoxical response does not imply non-ISN (*Figure 2–Figure Supplement 7*; Methods). We therefore examined the inhibition stabilization property of the ensemble by probing the ensemble behavior when a small transient perturbation to excitatory population activity is introduced while inhibition is frozen before stimulation and during stimulation. Before stimulation, the firing rate of the excitatory population slightly increases and then returns to its baseline after the transient perturbation (*Figure 2–Figure Supplement 8*). During stimulation, however, the transient perturbation leads to a transient explosion of the excitatory firing rate (*Figure 2–Figure Supplement 8*). These results further confirm that the ensemble shown in our example is initially a non-ISN before stimulation and can transition to an ISN with stimulation. By elevating the input level at the baseline in the model, the ensemble can be initially an ISN (*Figure 2–Figure Supplement 5*), resembling recent studies revealing that cortical circuits in the mouse V1 operate as ISNs in the absence of sensory stimulation (*Sanzeni et al., 2020*).

Despite the fact that the supralinear input-output function of our framework captures some aspects of intracellular recordings (*Priebe et al., 2004*), it is unbounded and thus allows infinitely

high firing rates. This is in contrast to neurobiology where firing rates are bounded due to neuronal refractory effects. While this assumption permitted us to analytically study the system and therefore to gain a deeper understanding of the underlying ensemble dynamics, we wondered whether our main conclusions were also valid when we limited the maximum firing rates. To that end, we carried out the same simulations while capping the firing rate at 300 Hz. In the absence of additional SFA or STP mechanisms, the firing rate saturation introduced a stable high-activity state in the ensemble dynamics which replaced the unstable dynamics in the uncapped model. As above, the ensemble entered this high-activity steady-state when stimulated with an external input above a critical threshold and exhibited persistent activity after stimulus removal (*Figure 2–Figure Supplement 9*). While weak SFA did not change this behavior, strong SFA resulted in oscillatory behavior during stimulation consistent with previous analytical work (*Figure 2–Figure Supplement 9, Van Vreeswijk and Hansel, 2001*), but did not in stable steady-states commonly observed in biological circuits. In the presence of E-to-E STD or E-to-I STF, however, the ensemble exhibited transient evoked activity at stimulation onset that was comparable to the uncapped case. Importantly, the ensemble did not show persistent activity after the stimulation (*Figure 2–Figure Supplement 9*). Finally, we confirmed that all of these findings were qualitatively similar in a realistic spiking neural network model (*Figure 2–Figure Supplement 10; Methods*).

In summary, we found that neuronal ensembles can rapidly, nonlinearly, and transiently amplify inputs by briefly switching from stable to unstable dynamics before being re-stabilized through STP mechanisms. We call this mechanism nonlinear transient amplification (NTA) which, in contrast to balanced amplification (*Murphy and Miller, 2009; Hennequin et al., 2012*), arises from population dynamics with supralinear neuronal activation functions interacting with STP. While we acknowledge that there may be other nonlinear transient amplification mechanisms, in this article we restrict our analysis to the definition above. NTA is characterized by a large onset response, a subsequent ISN steady-state while the stimulus persists, and a return to a unique baseline activity state after the stimulus is removed. Thus, NTA is ideally suited to rapidly and nonlinearly amplify sensory inputs through recurrent excitation, like reported experimentally (*Ko et al., 2011; Cossell et al., 2015*), while avoiding persistent activity.

Co-tuned inhibition broadens the parameter regime of NTA in the absence of persistent activity

Up to now, we have focused on a single neuronal ensemble. However, to process information in the brain, several ensembles with different stimulus selectivity presumably coexist and interact in the same circuit. This coexistence creates potential problems. It can lead to multi-stable persistent attractor dynamics, which are not commonly observed and could have adverse effects on the processing of subsequent stimuli. One solution to this issue could be EI co-tuning, which arises in network models with plastic inhibitory synapses (*Vogels et al., 2011*) and has been observed experimentally in several sensory systems (*Wehr and Zador, 2003; Froemke et al., 2007; Okun and Lampl, 2008; Rupperecht and Friedrich, 2018*).

To characterize the conditions under which neuronal ensembles nonlinearly amplify stimuli without persistent activity, we analyzed the case of two interacting ensembles. More specifically, we considered networks with two excitatory ensembles and distinguished between global and co-tuned inhibition (*Figure 3A*). In the case of global inhibition, one inhibitory population non-specifically inhibits both excitatory populations (*Figure 3A, left*). In contrast, in networks with co-tuned inhibition, each ensemble is formed by a dedicated pair of an excitatory and an inhibitory population which can have cross-over connections, for instance, due to overlapping ensembles (*Figure 3A, right*).

Global inhibition supports winner-take-all competition and is therefore often associated with multi-stable attractor dynamics (*Wong and Wang, 2006; Mongillo et al., 2008*). We first illustrated this effect in a network model with global inhibition. When the recurrent excitatory connections within each ensemble were sufficiently strong, small amounts of noise in the initial condition led to

Figure 3. Co-tuned inhibition broadens the parameter regime of NTA in the absence of persistent activity. **(A)** Schematic of two neuronal ensembles with global inhibition (left) and with co-tuned inhibition (right). **(B)** Firing rate dynamics of bi/multi-stable ensemble dynamics (left) and uni-stable (right). In both cases, additional excitatory inputs are injected into excitatory ensemble *E1* during the period marked in gray. **(C)** Analytical solution of uni- and bi/multi-stability regions for global inhibition (left) and co-tuned inhibition (right). Co-tuning results in a larger parameter regime of uni-stability. The triangles correspond to the two examples in B.

Figure 3—Figure supplement 1. Ensembles with co-tuned inhibition exhibit weaker — but still strong — NTA in comparison to ensembles with global inhibition.

one of the ensembles spontaneously activating at elevated firing rates, while the other ensemble's activity remained low (**Figure 3B**, left). A specific external stimulation could trigger a switch from one state to the other in which the other ensemble was active at a high firing rate. Importantly, this change persisted even after the stimulus had been removed, a hallmark of multi-stable dynamics. In contrast, uni-stable systems have a global symmetric state in which both ensembles have the same activity in the absence of stimulation. While the stimulated ensemble showed elevated firing rates in response to the stimulus, its activity returned to the baseline level after the stimulus is removed (**Figure 3B**, right), consistent with experimental observations (*DeWeese et al., 2003; Rupprecht and Friedrich, 2018; Bolding and Franks, 2018*). Note that the only difference between these two models is that J_{EE} is larger in the multi-stable example than in the uni-stable one.

Symmetric baseline activity is most consistent with activity observed in sensory areas. Hence, we sought to understand which inhibitory connectivity would be most conducive to maintain it. To that end, we analytically identified the uni-stability conditions, which are determined by the leading eigenvalue of the Jacobian matrix of the system, for networks with varying degrees of EI co-tuning (Methods). We found that a broader parameter regime underlies uni-stability in networks with co-tuned inhibition than global inhibition (**Figure 3C**). Notably, this conclusion is general and extends

to networks with an arbitrary number of ensembles (Methods). In comparison to the ensemble with global inhibition, the ensemble with co-tuned inhibition exhibits weaker — but still strong — NTA (*Figure 3–Figure Supplement 1*). Thus, co-tuned inhibition broadens the parameter regime in which NTA is possible while simultaneously avoiding persistent attractor dynamics.

NTA provides better pattern completion than fixed points while retaining stimulus selectivity

Neural circuits are capable of generating stereotypical activity patterns in response to partial cues and forming distinct representations in response to different stimuli (*Carrillo-Reid et al., 2016; Marshel et al., 2019; Bolding et al., 2020; Vinje and Gallant, 2000; Cayco-Gajic and Silver, 2019*). To test whether NTA achieves pattern completion while retaining stimulus selectivity, we analyzed the transient onset activity in our models and compared it to the fixed point activity.

To investigate pattern completion and stimulus selectivity in our model, we considered a co-tuned network with E-to-E STD and two distinct excitatory ensembles $E1$ and $E2$. We gave additional input g_{E1} to a Subset 1, consisting of 75% of the neurons in ensemble $E1$ (*Figure 4A*). We then measured the evoked activity in the remaining 25% of the excitatory neurons in $E1$ to quantify pattern completion. To assess stimulus selectivity, we injected additional input g_{E1} into the entire $E1$ ensemble during the second stimulation phase (*Figure 4A*) while measuring the activity of $E2$. We found that neurons in Subset 2, which did not receive additional input, showed large onset responses, their steady-state activity was largely suppressed (*Figure 4B*). Despite the fact that inputs to $E1$ caused increased transient onset responses in $E2$, the amount of increase was orders of magnitude smaller than in $E1$ (*Figure 4B*). To quantify pattern completion, we defined the

$$\text{Association index} = 1 + \frac{r_{E1_2} - r_{E1_1}}{r_{E1_2} + r_{E1_1}} . \quad (6)$$

Here, r_{E1_1} and r_{E1_2} correspond to the subpopulation activities of $E1$, respectively. By definition, the Association index ranges from zero to one, with larger values indicating stronger associativity. In addition, to quantify the selectivity between $E1$ and $E2$, we considered a symmetric binary classifier (*Figure 4A*, inset) and measured the distance to the decision boundary (Methods). Note that the Association index was computed during Phase 1 and the distance to the decision boundary during Phase 2 in this simulation paradigm (*Figure 4B*).

With these definitions, we ran simulations with different input strengths g_{E1} . We found that the onset peaks showed stronger association than the fixed point activity (*Figure 4C*). Note that the Association index at the fixed point remained zero, a direct consequence of r_{E1_2} being suppressed to zero. Furthermore, we found that the distance between the transient onset response and the decision boundary was always greater than for the fixed point activity (*Figure 4D*) showing that onset responses retain stimulus selectivity. While the fixed point activity of the unstimulated co-tuned neurons is zero in the given example, stimulating a subset of neurons in one ensemble can lead to an increase in the fixed point activity of the unstimulated neurons in the same ensemble under certain conditions (*Figure 4–Figure Supplement 1; Appendix 4*), which is consistent with pattern completion experiments (*Carrillo-Reid et al., 2016; Marshel et al., 2019*) showing that unstimulated neurons from the same ensemble can remain active throughout the whole stimulation period.

To investigate how the recurrent excitatory connectivity affects both pattern completion and stimulus selectivity, we introduced the parameter β which controls recurrent excitatory tuning by trading off within-ensemble E-to-E strength J_{EE} relative to the inter-ensemble strength J'_{EE} (*Figure 4A*) such that $J_{EE} = \beta J_{\text{tot}}$ and $J'_{EE} = (1 - \beta) J_{\text{tot}}$. These definitions ensure that the total weight $J_{\text{tot}} = J_{EE} + J'_{EE}$ remains constant for any choice of β . Notably, the overall recurrent excitation strength within an ensemble J_{EE} increases with increasing β . When β is larger than 0.5, the excitatory connection strength within the ensemble J_{EE} exceeds the one between ensembles J'_{EE} .

Figure 4. NTA yields stronger pattern completion than fixed points while retaining stimulus selectivity. **(A)** Schematic of the network setup used to probe pattern completion and stimulus selectivity. To assess the effect on pattern completion, 75% of the neurons (Subset 1) in ensemble $E1$ received additional input g_{E1} during Phase 1 (2–4 s), while we recorded the firing rate of the remaining 25% (Subset 2) in the excitatory ensemble $E1$. To evaluate the impact on stimulus selectivity, all neurons in $E1$ received additional inputs g_{E1} in Phase 2 (6–8 s) while the firing rate of $E2$ was measured. A downstream neuron's ability to discriminate between $E1$ or $E2$ being active depends on whether their activity is well separated by a symmetric decision boundary (inset). **(B)** Examples of firing rates of Subset 2 of $E1$ (left, blue) and $E2$ (right, green) with E-to-E STD. **(C)** Association index as a function of input g_{E1} for the onset peak amplitude (magenta solid line) and fixed point activity (gray dashed line) for E-to-E STD. **(D)** Distance to the decision boundary (see panel A, inset) as a function of input g_{E1} for the onset peak amplitude (magenta solid line) and fixed point activity (gray dashed line) for E-to-E STD. **(E and F)** Same as C and D but as a function of β , which controls the inner- and inter-ensemble connection strength.

Figure 4–Figure supplement 1. Change in steady state activity for unstimulated co-tuned neurons in the rate-based model.

Figure 4–Figure supplement 2. Quantification of pattern completion and stimulus selectivity in networks with E-to-I STF.

We found that pattern completion ability monotonically increases with β with a pronounced onset for $\beta > 0.6$ where NTA takes hold (Figure 4E). Moreover, in this regime the two stimulus representations are well separated (Figure 4F) which ensures stimulus selectivity also during onset transients. Together, these findings recapitulate the point that recurrent excitatory tuning is a key determinant of network dynamics. Finally, we confirmed that our findings were also valid in networks with E-to-I STF (Figure 4–Figure Supplement 2), which is commonly observed in the brain (Markram et al., 1998; Zucker and Regehr, 2002; Pala and Petersen, 2015). In summary, NTA's transient onset responses maintain stimulus selectivity and result in overall better pattern completion than fixed point activity.

NTA provides higher amplification and pattern separation in morphing experiments

So far, we only considered input to one ensemble. To examine how representations in our model are affected by ambiguous inputs to several ensembles, we performed additional morphing experiments (Freedman et al., 2001; Niessing and Friedrich, 2010). To that end, we introduced the parameter p which interpolates between two input stimuli which target $E1$ and $E2$ respectively. When p is zero, all additional input is injected into $E1$. For p equal to one, all additional input is injected into $E2$. Finally, p equal to 0.5 corresponds to the symmetric case in which $E1$ and $E2$ receive the same amount of additional input (Figure 5A).

First, we investigated how the recurrent excitatory connection strength within each ensemble

Figure 5. NTA provides stronger amplification and pattern separation in morphing experiments than fixed point activity. **(A)** Schematic of the morphing stimulation paradigm. The fraction of the additional inputs into the two excitatory ensembles is controlled by the parameter p . **(B)** Peak amplitude of $E1$ (blue) and $E2$ (green) as a function of p for E-to-E STD. Brightness levels represent different recurrent E-to-E connection strengths J_{EE} . **(C)** Same as in B but for fixed point activity. **(D)** Distance to the decision boundary as a function of p for the peak onset response (magenta solid line) and fixed point activity (gray dashed line) for E-to-E STD in a network with $J'_{IE} = 0.4$. **(E)** Same as D but with different E-to-I connection strengths J'_{IE} across ensembles for a network with $J_{EE} = 1.2$.

Figure 5–Figure supplement 1. Quantification of pattern separation in morphing experiments using a normalized measure.

Figure 5–Figure supplement 2. Quantification of pattern separation in morphing experiments for networks with E-to-I STF.

J_{EE} affects the onset peak amplitude and fixed point activity. We found that the peak amplitudes depend strongly on J_{EE} , whereas the fixed point activity was only weakly dependent on J_{EE} (Figure 5B, C). When we disconnected the ensembles by completely eliminating all recurrent excitatory connections, activity was noticeably decreased (Figure 5B, C). This illustrates, that recurrent excitation does play an important role in selectively amplifying specific stimuli similar to experimental observations (Marshall et al., 2019; Peron et al., 2020), but that amplification is highest at the onset. Further, we examined the impact of competition through lateral inhibition as a function of the E-to-I inter-ensemble strength J'_{IE} (Methods). As above, we quantified its impact by measuring the representational distance to the decision boundary for the transient onset responses and fixed point activity. We found that regardless of the specific STP mechanism, the distance was larger for the onset responses than for the fixed point activity, consistent with the notion that the onset dynamics separate stimulus identity reliably (Figure 5D, E). Since the absolute activity levels between onset and fixed point differed substantially, we further computed the relative pattern Separation index $(r_{E2} - r_{E1}) / (r_{E1} + r_{E2})$ and found that the onset transient provides better pattern separation ability for ambiguous stimuli with p close to 0.5 (Figure 5–Figure Supplement 1) provided that the E-to-I connection strength across ensembles J'_{IE} is strong enough. All the while separability for the onset transient was slightly decreased for distinct inputs with $p \in \{0, 1\}$ in comparison to the fixed point. In contrast, fixed points clearly separated such pure stimuli while providing weaker pattern separation for ambiguous input combinations. Importantly, these findings qualitatively held for networks with NTA mediated by E-to-I STF (Figure 5–Figure Supplement 2). Thus, NTA provides stronger amplification and pattern separation than fixed point activity in response to ambiguous stimuli.

Figure 6. Spiking neural network simulations qualitatively reproduce NTA dynamics of rate models. **(A)** Spiking activity of excitatory (blue) and inhibitory (red) neurons in a spiking neural network. From 2–20 s, Ensembles 1–5 individually received additional input for 2 s each (colored bars). From 22–24 s, 75% of Ensemble 5 neurons (Subset 1) received additional input, whereas the rest 25% of Ensemble 5 neurons (Subset 2) did not receive additional input. The symbols at the top designate the different simulation phases of baseline activity, the onset transients, and the fixed point activity. Different colors correspond to the distinct stimulation periods. **(B)** Ensemble activity (Hz). **(C)** Difference in ensemble activity between the stimulated ensemble with the remaining ensembles for the transient onset peak and the fixed point. Points correspond to the different stimulation periods. **(D)** Spiking activity during the interval 0–10 s represented in the PCA basis spanned by the first two principal components which captured approximately 40% of the total variance. The colored lines represent the PC trajectories of the first two stimuli shown in A and B. Triangles, points and crosses correspond to the onset peak, fixed point, and baseline activity, respectively. **(E)** Ensemble activity of Subset 1 (purple) and Subset 2 (gray) of Ensemble 5 from 16–26 s. Onset peaks are marked by triangles.

Figure 6–Figure supplement 1. Change in steady state activity for unstimulated co-tuned neurons in spiking neural networks.

NTA in spiking neural networks

Thus far, our analysis relied on power law neuronal input-output functions in the interest of analytical tractability. To test whether our findings also qualitatively apply to more realistic network models, we built a spiking neural network consisting of randomly connected 800 excitatory and 200 inhibitory neurons, in which the E-to-E synaptic connections were subject to STD (Methods). Here, we defined five overlapping ensembles, each corresponding to 200 randomly selected excitatory neurons. During an initial simulation phase (0–22s), we consecutively stimulated each ensemble by giving additional input to their excitatory neurons, whereas the input to other neurons remained unchanged (**Figure 6A**). In addition, we also tested pattern completion by stimulating only 75% (Subset 1) of the neurons belonging to Ensemble 5 (22–24 s; **Figure 6A**). We quantified each ensemble’s activity by calculating the population firing rate of the ensemble (Methods). As in the case of the rate-based model, the neuronal ensembles in the spiking model generated pronounced transient onset responses. We then measured the difference of peak ensemble activity and fixed point activity between the stimulated ensemble and the remaining unstimulated ensembles (Methods). As for the rate-based networks, this difference was consistently larger for the onset peak than for the fixed point (**Figure 6B, C**). Thus, transient onset responses allow better stimulus separation than fixed points also in spiking neural network models.

Finally, to visualize the neural activity, we projected the binned spiking activity during the first 10 s of our simulation onto its first two principal components. Notably, the PC trajectory does not exhibit a pronounced rotational component (*Figure 6D*) as activity is confined to one specific ensemble, consistent with experiments (*Marshall et al., 2019*). Furthermore, we computed the fifth ensemble's activity for Subset 1 and 2 during the time interval 16–26 s. In agreement with our rate models, neurons in Subset 2 which did not receive additional inputs showed a strong response at the onset (*Figure 6E*), but not at the fixed point, suggesting that the strongest pattern completion occurs during the initial amplification phase. Finally, we also observed higher-than-baseline fixed point activity in unstimulated neurons of Subset 2 in spiking neural networks (*Figure 6–Figure Supplement 1*). Thus, the key characteristics of NTA are preserved across rate-based and more realistic spiking neural network models.

Discussion

In this study, we demonstrated that neuronal ensemble models with recurrent excitation and suitable forms of STP exhibit nonlinear transient amplification (NTA), a putative mechanism underlying selective amplification in recurrent circuits. NTA combines a supralinear neuronal transfer function, recurrent excitation between neurons with similar tuning, and pronounced STP. Using analytical and numerical methods, we showed that NTA generates rapid transient onset responses during which optimal stimulus separation occurs rather than at steady-states. Additionally, we showed that co-tuned inhibition is conducive to prevent the emergence of persistent activity, which could otherwise interfere with processing subsequent stimuli. In contrast to balanced amplification (*Murphy and Miller, 2009*), NTA is an intrinsically nonlinear mechanism for which only stimuli above a critical threshold are amplified effectively. While the precise threshold value is parameter-dependent, it can be arbitrarily low provided the excitatory recurrent connections are sufficiently strong (cf. *Figure 1F*). Importantly, such a critical activation threshold offers a possible explanation for sensory perception experiments which show similar threshold behavior (*Marshall et al., 2019; Peron et al., 2020*). Following transient amplification, ensemble dynamics are inhibition-stabilized, which renders our model compatible with existing work on SSNs (*Ahmadian et al., 2013; Rubin et al., 2015; Hennequin et al., 2018; Kraynyukova and Tchumatchenko, 2018; Echeveste et al., 2020*). Thus, NTA provides a parsimonious explanation for why sensory systems may rely upon neuronal ensembles with recurrent excitation in combination with EI co-tuning, and pronounced STP dynamics.

Several theoretical studies approached the problem of transient amplification in recurrent neural network models. *Loebel and Tsodyks (2002)* have described an NTA-like mechanism as a driver for powerful ensemble synchronization in rate-based networks and in spiking neural network models of auditory cortex (*Loebel et al., 2007*). Here we generalized this work to both E-to-E STD and E-to-I STF and provide an in-depth characterization of its amplification capabilities, pattern completion properties, and the resulting network states with regard to their inhibition-stabilization properties. Moreover, we showed that SFA cannot provide similar network stabilization and explored how EI co-tuning interacts with NTA. Finally, we contrasted NTA to alternative transient amplification mechanisms. Balanced amplification is a particularly well-studied transient amplification mechanism (*Murphy and Miller, 2009; Goldman, 2009; Hennequin et al., 2014; Bondanelli and Ostojic, 2020; Gillett et al., 2020; Christodoulou et al., 2021*) that relies on non-normality of the connectivity matrix to selectively and rapidly amplify stimuli. Importantly, balanced amplification occurs in networks in which strong recurrent excitation is appropriately balanced by strong recurrent inhibition. It is capable of generating rich transient activity in linear network models (*Hennequin et al., 2014*), and selectively amplifies specific activity patterns, but without a specific activation threshold. In addition, in spiking neural networks, strong input can induce synchronous firing at the population level which is subsequently stabilized by strong feedback inhibition without the requirement for STP mechanisms (*Stern et al., 2018*). These properties contrast with NTA, which has a nonlinear activation threshold and intrinsically relies on STP to stabilize otherwise unstable run-away

dynamics. Due to the switch of the network's dynamical state, NTA's amplification can be orders of magnitudes larger than balanced amplification (cf. **Figure 2–Figure Supplement 3**). Interestingly, after the transient amplification phase, ensemble dynamics settle in an inhibitory-stabilized state, which renders NTA compatible with previous work on SSNs but in the presence of STP. Finally, although NTA and balanced amplification rely on different amplification mechanisms, they are not mutually exclusive and could, in principle, co-exist in biological networks.

NTA's requirement to generate positive feedback dynamics through recurrent excitation, motivated our focus on networks with $\det(\mathbf{J}) < 0$. As demonstrated in previous work (**Ahmadian et al., 2013**), supralinear networks with $\det(\mathbf{J}) > 0$ and instantaneous inhibition ($\tau_I/\tau_E \rightarrow 0$) are always stable for any given input, they are thus unable to generate positive feedback dynamics. In addition, networks with $\det(\mathbf{J}) > 0$ can exhibit a range of interesting behaviors, e.g., oscillatory dynamics and persistent activity (**Kranyukova and Tchumatchenko, 2018**). It is worth noting, however, that for delayed or slow inhibition, stimulation can still lead to unstable network dynamics in networks with $\det(\mathbf{J}) > 0$. Nevertheless, our simulations suggest that our main conclusions about the stabilization mechanisms still hold (**Figure 2–Figure Supplement 11**).

NTA shares some properties with the notion of network criticality in the brain, like synchronous activation of cell ensembles (**Plenz and Thiagarajan, 2007**) and STP which can tune networks to a critical state (**Levina et al., 2007**). However, in contrast to most models of criticality, in NTA an ensemble briefly transitions to supercritical dynamics in a controlled, stimulus-dependent manner rather than spontaneously. Yet, how the two paradigms are connected at a more fundamental level, is an intriguing question left for future work. Furthermore, recurrent co-tuned inhibition is essential for NTA to ensure uni-stability and selectivity through the suppression of ensembles with different tuning. This requirement is similar in flavor to semi-balanced networks characterized by excess inhibition to some excitatory ensembles while others are balanced (**Baker et al., 2020**). However, the theory of semi-balanced networks has, so far, only been applied to steady-state dynamics while ignoring transients and STP. EI co-tuning prominently features in several models and was shown to support network stability (**Vogels et al., 2011; Hennequin et al., 2017; Znamenskiy et al., 2018**), efficient coding (**Denève and Machens, 2016**), novelty detection (**Schulz et al., 2021**), changes in neuronal variability (**Hennequin et al., 2018; Rost et al., 2018**), and correlation structure (**Wu et al., 2020**). Moreover, some studies have argued that EI balance and co-tuning could increase robustness to noise in the brain (**Rubin et al., 2017**). The present work mainly highlights its importance for preventing multi-stability and delay activity in circuits not requiring such long-timescale dynamics.

NTA is consistent with several experimental findings. First, our model recapitulates the key findings of **Shew et al. (2015)** who showed *ex vivo* that strong sensory inputs cause a transient shift to a supercritical state, after which adaptive changes rapidly tune the network to criticality. Second, NTA requires strong recurrent excitatory connectivity between neurons with similar tuning, which has been reported in experiments (**Ko et al., 2011; Cossell et al., 2015; Peron et al., 2020**). Third, ensemble activation in our model depends on a critical stimulus strength in line with recent all-optical experiments in the visual cortex, which further link ensemble activation with a perceptual threshold (**Marshall et al., 2019**). Fourth, sensory networks are uni-stable in that they return to a non-selective activity state after the removal of the stimulus and usually do not show persistent activity (**DeWeese et al., 2003; Mazor and Laurent, 2005; Rupprecht and Friedrich, 2018**). Fifth, our work shows that NTA's onset responses encode stimulus identity better than the fixed point activity, consistent with experiments in the locust antennal lobe (**Mazor and Laurent, 2005**) and research supporting that the brain relies on coactivity on short timescales to represent information (**Stopfer et al., 1997; Engel et al., 2001; Harris et al., 2003; El-Gaby et al., 2021**). Yet, it remains to be seen whether these findings are also coherent with data on the temporal evolution in other sensory systems. Finally, EI co-tuning, which is conducive for NTA, has been found ubiquitously in different sensory circuits (**Wehr and Zador, 2003; Froemke et al., 2007; Okun and Lampl, 2008; Rupprecht and Friedrich, 2018; Znamenskiy et al., 2018**).

508 In our model, we made several simplifying assumptions. For instance, we kept the input to
 509 inhibitory neurons fixed and only varied the input to the excitatory population. This step was mo-
 510 tivated by experiments in the piriform cortex where the total inhibition is dominated by feedback
 511 inhibition (*Franks et al., 2011*). Nevertheless, significant feedforward inhibition was observed in
 512 other areas (*Bissière et al., 2003; Cruikshank et al., 2007; Ji et al., 2016; Miska et al., 2018*). While
 513 an in-depth comparison for different origins of inhibition was beyond the scope of the present
 514 study, we found that increasing the inputs to the excitatory population and inhibitory population
 515 by the same amount can still lead to NTA (*Figure 2–Figure Supplement 1; Figure 2–Figure Supple-*
 516 *ment 12*; Methods), suggesting that our main findings can remain unaffected in the presence of
 517 substantial feedforward inhibition. In addition, we limited our analysis to only a few overlapping
 518 ensembles. It will be interesting future work to study NTA in the case of many interacting and
 519 potentially overlapping ensembles and to determine the maximum storage capacity above which
 520 performance degrades. Finally, we anticipate that temporal differences in excitatory and inhibitory
 521 synaptic transmission may be important to preserve NTA's stimuli selectivity.

522 Our model makes several predictions. In contrast to balanced amplification, in which the net-
 523 work operating regime depends solely on the connectivity, an ensemble involved in NTA can transi-
 524 tion from a non-ISCN to an ISCN state. Such a transition is consistent with noise variability observed in
 525 sensory cortices (*Hennequin et al., 2018*) and could be tested experimentally by probing the para-
 526 doxical effect under different stimulation conditions (*Figure 2G–H; Figure 2–Figure Supplement 6*).
 527 Moreover, NTA predicts that onset activity provides a better stimulus encoding and its activity is
 528 correlated with the fixed point activity. This signature is different from purely non-normal ampli-
 529 fication mechanisms which would involve a wave of neuronal activity across several distinct ensem-
 530 bles similar to a synfire chain (*Abeles, 1991*). The difference should be clearly discernible in data.
 531 Since NTA relies on recurrent excitation between ensemble neurons, it suggests normal dynamics
 532 in which distinct ensembles first activate and then inactivate. The resulting dynamics have weak
 533 rotational components (cf. *Figure 6D*) as seen in some experiments (*Marshall et al., 2019*). Strong
 534 non-normal amplification, on the other hand, relies on sequential activation associated with pro-
 535 nounced rotational dynamics (*Hennequin et al., 2014; Gillett et al., 2020*), as for instance observed
 536 in motor areas (*Churchland et al., 2012*). Although both non-normal mechanisms and NTA are
 537 likely to co-exist in the brain, we speculate that strong NTA is best suited for, and thus most like to
 538 be found in, sensory systems.

539 In summary, we introduced a general theoretical framework of selective transient signal ampli-
 540 fication in recurrent networks. Our approach derives from the minimal assumptions of a nonlinear
 541 neuronal transfer function, recurrent excitation within neuronal ensembles, and STP. Importantly,
 542 our analysis revealed the functional benefits of STP and EI co-tuning, both pervasively found in
 543 sensory circuits. Finally, our work suggests that transient onset responses rather than steady-state
 544 activity are ideally suited for coactivity-based stimulus encoding and provides several testable pre-
 545 dictions.

546 Methods and Materials

547 Stability conditions for supralinear networks

548 The dynamics of a neuronal ensemble consisting of one excitatory and one inhibitory population
 549 with a supralinear, power law input-output function can be described as follows:

$$\tau_E \frac{dr_E}{dt} = -r_E + \left[J_{EE}r_E - J_{EI}r_I + g_E \right]_+^{\alpha_E} \quad (7)$$

$$\tau_I \frac{dr_I}{dt} = -r_I + \left[J_{IE}r_E - J_{II}r_I + g_I \right]_+^{\alpha_I} \quad (8)$$

550 The Jacobian $\mathbf{M}$ of the system is given by

$$\mathbf{M} = \begin{bmatrix} \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) & -\tau_E^{-1}J_{EI}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} \\ \tau_I^{-1}J_{IE}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}} & -\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) \end{bmatrix} \quad (9)$$

551 To ensure that the system is stable, the product of $\mathbf{M}$'s eigenvalues $\lambda_1 \lambda_2$, which is equivalent to
 552 the determinant of $\mathbf{M}$, has to be positive. In addition, the sum of the two eigenvalues $\lambda_1 + \lambda_2$,
 553 which corresponds to $\text{Tr}(\mathbf{M})$, has to be negative. We therefore obtained the following two stability
 554 conditions

$$\lambda_1 \lambda_2 = -\tau_E^{-1}\tau_I^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1)(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) + \tau_E^{-1}\tau_I^{-1}J_{EI}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} J_{IE}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}} > 0 \quad (10)$$

$$\lambda_1 + \lambda_2 = \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) - \tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) < 0 \quad (11)$$

555 Notably, the stability conditions depend on the firing rate of the excitatory population r_E and
 556 the inhibitory population r_I . Since firing rates are input-dependent, the stability of supralinear
 557 networks is input-dependent. In contrast, in linear networks in which $\alpha_E = \alpha_I = 1$, the conditions
 558 can be simplified to

$$\lambda_1 \lambda_2 = -\tau_E^{-1}\tau_I^{-1}(J_{EE} - 1)(1 + J_{II}) + \tau_E^{-1}\tau_I^{-1}J_{EI}J_{IE} > 0 \quad (12)$$

$$\lambda_1 + \lambda_2 = \tau_E^{-1}(J_{EE} - 1) - \tau_I^{-1}(1 + J_{II}) < 0 \quad (13)$$

559 and are thus input-independent.

560 ISN index for supralinear networks

561 If an ensemble is unstable without feedback inhibition, then the ensemble is an ISN (*Tsodyks et al.,*
 562 *1997*). To determine whether a given system is an ISN, we analyzed the stability of the E-E sub-
 563 network, which is determined by the real part of the leading eigenvalue of the Jacobian of the E-E
 564 subnetwork. In the following we call this leading eigenvalue the "ISN index", which is defined as
 565 follows:

$$\text{ISN index} = \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) \quad (14)$$

566 A positive ISN index indicates the system is an ISN. Otherwise, the system is non-ISN. For supra-
 567 linear networks in which $\alpha_E > 1$, the ISN index depends on the firing rates, inputs can therefore
 568 switch the network from non-ISN to ISN. In contrast, $\alpha_E = 1$ for linear networks which renders the
 569 ISN index firing rate independent.

570 Characteristic function

571 To investigate how network stability changes with input, we trace the steps of *Krainyukova and*
 572 *Tchumatchenko (2018)* and define the characteristic function $F(z)$ as follows:

$$F(z) = J_{EE} \left[z \right]_+^{\alpha_E} - J_{EI} \left[\det(\mathbf{J}) \cdot J_{EI}^{-1} \left[z \right]_+^{\alpha_E} + J_{EI}^{-1} J_{II} z - J_{EI}^{-1} J_{II} g_E + g_I \right]_+^{\alpha_I} - z + g_E \quad (15)$$

573 where

$$z = J_{EE} r_E - J_{EI} r_I + g_E \quad (16)$$

574 is the current into the excitatory population. The characteristic function simplifies the original two-
 575 dimensional system to a one-dimensional system, and the zero crossings of $F(z)$ correspond to
 576 the fixed points of the original system. For $z \geq 0$, we note:

$$\frac{dF(z)}{dz} = J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - J_{EI}\alpha_I \left(\det(\mathbf{J}) \cdot J_{EI}^{-1}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} + J_{EI}^{-1}J_{II} \right) r_I^{\frac{\alpha_I-1}{\alpha_I}} - 1 = -\tau_E\tau_I\lambda_1\lambda_2 \quad (17)$$

577 Therefore, if the derivative of $F(z)$ evaluated at one of its roots is positive, the corresponding fixed
 578 point is a saddle point. Note that as r_E and r_I increase, the term in parenthesis becomes dominant.

579 To ensure that $\lambda_1 \lambda_2$ is negative also for large r_E and r_I , the determinant of the weight matrix $\det(\mathbf{J})$
 580 has to be positive. Therefore, $\det(\mathbf{J})$ has a decisive impact on the curvature of $F(z)$. In systems
 581 with negative determinant, $F(z)$ bends upwards for large z . In contrast, $F(z)$ asymptotically bends
 582 downwards in systems with positive determinant. Hence, the high-activity steady-state of systems
 583 with negative determinant is unstable. In addition, we can simplify the above condition to the
 584 determinant of the weight matrix which is a necessary condition for network stability at any firing
 585 rate:

$$\det(\mathbf{J}) = -J_{EE}J_{II} + J_{IE}J_{EI} > 0 \quad (18)$$

586 To investigate how the network stability changes with input g_E , we examined how $F(z)$ varies
 587 with changing input g_E by calculating the derivative of $F(z)$ with respect to g_E ,

$$\frac{dF(z)}{dg_E} = \alpha_I J_{II} \left[\det(\mathbf{J}) \cdot J_{EI}^{-1} \left[z \right]_+^{\alpha_E} + J_{EI}^{-1} J_{II} z - J_{EI}^{-1} J_{II} g_E + g_I \right]_+^{\alpha_I - 1} + 1 \quad (19)$$

588 Since $\frac{dF(z)}{dg_E}$ is positive, increasing g_E always shifts $F(z)$ upwards, eventually leading to the vanishing
 589 of all roots and, thus, unstable dynamics in supralinear networks with negative $\det(\mathbf{J})$. In scenarios
 590 in which feedforward input to the inhibitory population also changes, we have

$$\begin{aligned} \frac{dF(z)}{dt} &= \frac{\partial F(z)}{\partial g_E} \frac{dg_E}{dt} + \frac{\partial F(z)}{\partial g_I} \frac{dg_I}{dt} \\ &= \left(\alpha_I J_{II} \left[\det(\mathbf{J}) \cdot J_{EI}^{-1} \left[z \right]_+^{\alpha_E} + J_{EI}^{-1} J_{II} z - J_{EI}^{-1} J_{II} g_E + g_I \right]_+^{\alpha_I - 1} + 1 \right) \Delta g_E \\ &\quad - \alpha_I J_{EI} \left[\det(\mathbf{J}) \cdot J_{EI}^{-1} \left[z \right]_+^{\alpha_E} + J_{EI}^{-1} J_{II} z - J_{EI}^{-1} J_{II} g_E + g_I \right]_+^{\alpha_I - 1} \Delta g_I \end{aligned} \quad (20)$$

591 When the change in stimulation strength into the excitatory (Δg_E) and the inhibitory population
 592 (Δg_I) are the same, $\frac{dF(z)}{dt}$ is always positive provided J_{II} is greater than J_{EI} . Hence, depending on
 593 the value of $\frac{J_{II}}{J_{EI}}$, stimulation can lead to unstable network dynamics even when the input to the
 594 inhibitory population increases more than to the excitatory population.

595 Spike-frequency adaptation (SFA)

596 We modeled SFA of excitatory neurons as an activity-dependent negative feedback current (**Benda**
 597 **and Herz, 2003; Brette and Gerstner, 2005**):

$$\tau_E \frac{dr_E}{dt} = -r_E + \left[J_{EE}r_E - J_{EI}r_I + g_E \right]_+^{\alpha_E} - a \quad (21)$$

$$\tau_I \frac{dr_I}{dt} = -r_I + \left[J_{IE}r_E - J_{II}r_I + g_I \right]_+^{\alpha_I} \quad (22)$$

$$\tau_a \frac{da}{dt} = -a + br_E \quad (23)$$

598 where a is the adaptation variable, τ_a is the adaptation time constant, and b is the adaptation
 599 strength.

600 Stability conditions in networks with SFA

601 The Jacobian $\mathbf{M}_{\text{SFA}}$ of the system with SFA is given by

$$\mathbf{M}_{\text{SFA}} = \begin{bmatrix} \tau_E^{-1} (J_{EE} \alpha_E r_E^{\alpha_E - 1} - 1) & -\tau_E^{-1} J_{EI} \alpha_E r_E^{\alpha_E - 1} & -\tau_E^{-1} \\ \tau_I^{-1} J_{IE} \alpha_I r_I^{\alpha_I - 1} & -\tau_I^{-1} (1 + J_{II} \alpha_I r_I^{\alpha_I - 1}) & 0 \\ \tau_a^{-1} b & 0 & -\tau_a^{-1} \end{bmatrix} \quad (24)$$

602 The characteristic polynomial of the system with SFA can be written as follows (**Horn and Johnson,**
 603 **1985**):

$$\lambda^3 - \text{tr}(\mathbf{M}_{\text{SFA}}) \lambda^2 + (A_{11} + A_{22} + A_{33}) \lambda - \det(\mathbf{M}_{\text{SFA}}) = 0 \quad (25)$$

where $\text{tr}(\mathbf{M}_{\text{SFA}})$ and $\det(\mathbf{M}_{\text{SFA}})$ are the trace and the determinant of the Jacobian matrix $\mathbf{M}_{\text{SFA}}$, A_{11} , A_{22} , and A_{33} are the matrix cofactors. More specifically,

$$\text{tr}(\mathbf{M}_{\text{SFA}}) = \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) - \tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) - \tau_a^{-1} \quad (26)$$

$$A_{11} = \begin{vmatrix} -\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) & 0 \\ 0 & -\tau_a^{-1} \end{vmatrix} = \tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}})\tau_a^{-1} \quad (27)$$

$$A_{22} = \begin{vmatrix} \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) & -\tau_E^{-1} \\ \tau_a^{-1}b & -\tau_a^{-1} \end{vmatrix} = -\tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1)\tau_a^{-1} + \tau_a^{-1}b\tau_E^{-1} \quad (28)$$

$$A_{33} = \begin{vmatrix} \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) & -\tau_E^{-1}J_{EI}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} \\ \tau_I^{-1}J_{IE}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}} & -\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) \end{vmatrix} = -\tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1)\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) + \tau_E^{-1}J_{EI}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}}\tau_I^{-1}J_{IE}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}} \quad (29)$$

$$A_{11} + A_{22} + A_{33} = \tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}})\tau_a^{-1} - \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1)\tau_a^{-1} + \tau_a^{-1}b\tau_E^{-1} \quad (30)$$

$$\det(\mathbf{M}_{\text{SFA}}) = \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1)\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}})\tau_a^{-1} - \tau_E^{-1}J_{EI}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}}\tau_I^{-1}J_{IE}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}} - \tau_a^{-1}b\tau_E^{-1}\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) \quad (31)$$

To ensure that the dynamics of the system are stable, the real parts of the eigenvalues of the Jacobian at the fixed point, and thus all roots of the characteristic polynomial have to be negative. Since the product of the roots is equal to $\det(\mathbf{M}_{\text{SFA}})$, $-\det(\mathbf{M}_{\text{SFA}})$ has to be positive. We then have

$$b > \frac{\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} (J_{EE} - \det(\mathbf{J}) \cdot \alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}})}{1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}} - 1 \quad (32)$$

Since SFA does not modify the synaptic connections, the term $J_{EE} - \det(\mathbf{J}) \cdot \alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}$ is positive for networks with $\det(\mathbf{J}) < 0$.

In the large r_E limit, if b is small such that the above condition cannot be fulfilled, $\det(\mathbf{M}_{\text{SFA}})$ is then positive, suggesting that the Jacobian of the system has always at least one positive eigenvalue. Therefore, the dynamics of the system cannot be stabilized in the presence of small b .

In addition, $A_{11} + A_{22} + A_{33}$ is equal to $\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_1\lambda_3$, with the roots of the characteristic polynomial λ_1 , λ_2 , and λ_3 . If all roots are real and negative, $A_{11} + A_{22} + A_{33}$ has to be positive. If one root is real and negative and two other roots are complex conjugates, to ensure that all roots have negative real parts, one necessary condition is $A_{11} + A_{22} + A_{33} > 0$. From the $\text{tr}(\mathbf{M}_{\text{SFA}})$ and $\det(\mathbf{M}_{\text{SFA}})$ conditions, we have

$$A_{11} + A_{22} + A_{33} > \tau_a^{-1}(-\tau_a^{-1} + b\tau_E^{-1}) - b\tau_E^{-1}\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) \quad (33)$$

As a result, if $\tau_a^{-1}(-\tau_a^{-1} + b\tau_E^{-1}) - b\tau_E^{-1}\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) > 0$, $A_{11} + A_{22} + A_{33}$ is guaranteed to be positive. We therefore have

$$b[\tau_a^{-1}\tau_E^{-1} - \tau_E^{-1}\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}})] > \tau_a^{-2} \quad (34)$$

621 Note that τ_a has to be small, in other words, SFA has to be fast, so that $\tau_a^{-1}\tau_E^{-1} - \tau_E^{-1}\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}})$
 622 is positive for arbitrary r_I . For positive $\tau_a^{-1}\tau_E^{-1} - \tau_E^{-1}\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}})$, we have

$$b > \frac{\tau_a^{-2}}{\tau_a^{-1}\tau_E^{-1} - \tau_E^{-1}\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}})} \quad (35)$$

623 Since τ_a has to be small, the above condition cannot be satisfied for small b .

624 Next, we consider the system with large b . Suppose that the firing rate r_E and r_I in the ini-
 625 tial network are of order 1, and b is of order K , where K is a large number. We therefore have
 626 $-\text{tr}(\mathbf{M}_{\text{SFA}}) \sim O(1)$, $A_{11} + A_{22} + A_{33} \sim O(K)$, and $-\det(\mathbf{M}_{\text{SFA}}) \sim O(K)$. The discriminant of the character-
 627 istic polynomial is

$$\begin{aligned} & (-\text{tr}(\mathbf{M}_{\text{SFA}}))^2(A_{11} + A_{22} + A_{33})^2 - 4(A_{11} + A_{22} + A_{33})^3 - 4(-\text{tr}(\mathbf{M}_{\text{SFA}}))^3(-\det(\mathbf{M}_{\text{SFA}})) \\ & - 27(-\det(\mathbf{M}_{\text{SFA}}))^2 + 18(-\text{tr}(\mathbf{M}_{\text{SFA}}))(A_{11} + A_{22} + A_{33})(-\det(\mathbf{M}_{\text{SFA}})) \\ & = (A_{11} + A_{22} + A_{33})^3 \left[\frac{(-\text{tr}(\mathbf{M}_{\text{SFA}}))^2}{A_{11} + A_{22} + A_{33}} - 4 - \frac{4(-\text{tr}(\mathbf{M}_{\text{SFA}}))^3(-\det(\mathbf{M}_{\text{SFA}}))}{(A_{11} + A_{22} + A_{33})^3} \right. \\ & \quad \left. - \frac{27(-\det(\mathbf{M}_{\text{SFA}}))^2}{(A_{11} + A_{22} + A_{33})^3} + \frac{18(-\text{tr}(\mathbf{M}_{\text{SFA}}))(-\det(\mathbf{M}_{\text{SFA}}))}{(A_{11} + A_{22} + A_{33})^2} \right] \end{aligned} \quad (36)$$

628 Clearly, in the large b limit, the discriminant is negative, suggesting that the characteristic polyno-
 629 mial has one real root and two complex conjugate roots (*Irving, 2004*).

630 As the input g_E increases, the complex conjugate eigenvalues cross the imaginary axis when
 631 $\text{tr}(\mathbf{M}_{\text{SFA}})(A_{11} + A_{22} + A_{33})$ equals $\det(\mathbf{M}_{\text{SFA}})$. As a result, the system undergoes a supercritical Hopf
 632 bifurcation. We numerically confirmed that the resulting limit cycle is stable (*Figure 2-Figure Sup-*
 633 *plement 1*), consistent with previous work (*Van Vreeswijk and Hansel, 2001*). Thus, the system
 634 shows oscillatory behavior instead of stable steady state.

635 Short-term plasticity (STP)

636 We modeled E-to-E STD following previous work (*Tsodyks and Markram, 1997; Varela et al., 1997*):

$$\tau_E \frac{dr_E}{dt} = -r_E + \left[x J_{EE} r_E - J_{EI} r_I + g_E \right]_+^{\alpha_E} \quad (37)$$

$$\tau_I \frac{dr_I}{dt} = -r_I + \left[J_{IE} r_E - J_{II} r_I + g_I \right]_+^{\alpha_I} \quad (38)$$

$$\frac{dx}{dt} = \frac{1-x}{\tau_x} - U_d x r_E \quad (39)$$

637 where x is the depression variable, which is limited to the interval $(0, 1]$, τ_x is the depression time
 638 constant, and U_d is the depression rate. The steady-state solution x^* is given by

$$x^* = \frac{1}{1 + U_d r_E \tau_x} \quad (40)$$

639 Similarly, we modeled E-to-I STF as

$$\tau_E \frac{dr_E}{dt} = -r_E + \left[J_{EE} r_E - J_{EI} r_I + g_E \right]_+^{\alpha_E} \quad (41)$$

$$\tau_I \frac{dr_I}{dt} = -r_I + \left[u J_{IE} r_E - J_{II} r_I + g_I \right]_+^{\alpha_I} \quad (42)$$

$$\frac{du}{dt} = \frac{1-u}{\tau_u} + U_f (U_{\max} - u) r_E \quad (43)$$

640 where u is the facilitation variable constrained to the interval $[1, U_{\max})$, $U_{\max}$ is the maximal facilita-
 641 tion value, τ_u is the time constant of STF, and U_f is the facilitation rate. The steady-state solution
 642 u^* is given by

$$u^* = \frac{1 + U_f U_{\max} r_E \tau_u}{1 + U_f r_E \tau_u} \quad (44)$$

643 Stability conditions for networks with E-to-E STD

644 The Jacobian $\mathbf{M}_{\text{STD}}$ of the system with E-to-E STD is given by

$$\mathbf{M}_{\text{STD}} = \begin{bmatrix} \tau_E^{-1}(xJ_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) & -\tau_E^{-1}J_{EI}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} & \tau_E^{-1}J_{EE}\alpha_E r_E^{\frac{2\alpha_E-1}{\alpha_E}} \\ \tau_I^{-1}J_{IE}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}} & -\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) & 0 \\ -U_d x & 0 & -\tau_x^{-1} - U_d r_E \end{bmatrix} \quad (45)$$

645 and the characteristic polynomial can be written as follows:

$$\lambda^3 - \text{tr}(\mathbf{M}_{\text{STD}})\lambda^2 + (A_{11} + A_{22} + A_{33})\lambda - \det(\mathbf{M}_{\text{STD}}) = 0 \quad (46)$$

646 where $\text{tr}(\mathbf{M}_{\text{STD}})$ and $\det(\mathbf{M}_{\text{STD}})$ are the trace and the determinant of the Jacobian matrix $\mathbf{M}_{\text{STD}}$. A_{11} ,

647 A_{22} , and A_{33} are the matrix cofactors. More specifically,

$$\text{tr}(\mathbf{M}_{\text{STD}}) = \tau_E^{-1}(xJ_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) - \tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) - \tau_x^{-1} - U_d r_E \quad (47)$$

648 In the case of unstable dynamics, r_E goes to infinity due to run-away excitation. However, the

649 depression variable x approaches zero in this limit, as $\lim_{r_E \rightarrow \infty} x = \lim_{r_E \rightarrow \infty} \frac{1}{1 + U_d r_E \tau_x} = 0$. Therefore,

650 in the large r_E limit, $-\text{tr}(\mathbf{M}_{\text{STD}})$ is positive.

$$\begin{aligned} A_{11} + A_{22} + A_{33} &= \tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}})(\tau_x^{-1} + U_d r_E) \\ &+ \tau_E^{-1}(xJ_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1)(-\tau_x^{-1} - U_d r_E) - \tau_E^{-1}J_{EE}\alpha_E r_E^{\frac{2\alpha_E-1}{\alpha_E}}(-U_d x) \\ &- \tau_E^{-1}(xJ_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1)\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) + \tau_E^{-1}J_{EI}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}}\tau_I^{-1}J_{IE}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}} \end{aligned} \quad (48)$$

651 Similarly, in the large r_E limit, $A_{11} + A_{22} + A_{33}$ is positive.

$$\begin{aligned} \det(\mathbf{M}_{\text{STD}}) &= \tau_E^{-1}(xJ_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1)\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}})(\tau_x^{-1} + U_d r_E) \\ &- \tau_E^{-1}J_{EI}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}}\tau_I^{-1}J_{IE}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}(\tau_x^{-1} + U_d r_E) \\ &- \tau_E^{-1}J_{EE}\alpha_E r_E^{\frac{2\alpha_E-1}{\alpha_E}}U_d x \tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) \end{aligned} \quad (49)$$

652 Similarly, in the large r_E limit, $-\det(\mathbf{M}_{\text{STD}})$ is positive.

653 According to the Descartes' rule of signs, the number of positive roots is at most the number of
654 sign changes in the sequences of polynomial's coefficients. Therefore, there are no positive roots
655 for the above characteristic polynomial and the network dynamics can be stabilized by E-to-E STD.

656 Characteristic function approximation for networks with E-to-E STD

657 As demonstrated above, E-to-E STD is able to restabilize the system, there exists a stable steady
658 state for which the STD variable x is constant $x = x^*$. Because x changes slowly compared to the
659 neuronal dynamics, we can approximate it as constant which results in a natural reduction to a 2D
660 system in which the weights with STD are modified. The stability of this 2D system can be readily
661 characterized by the characteristic function $F(z)$ (*Kraynyukova and Tchumatchenko, 2018*), which
662 depends on the previous steady state value of x . The characteristic function approximation with
663 E-to-E STD can therefore be written as follows:

$$F(z) = xJ_{EE} \left[z \right]_+^{\alpha_E} - J_{EI} \left[\det(\mathbf{J}_{\text{STD}}) \cdot J_{EI}^{-1} \left[z \right]_+^{\alpha_E} + J_{EI}^{-1} J_{II} z - J_{EI}^{-1} J_{II} g_E + g_I \right]_+^{\alpha_I} - z + g_E \quad (50)$$

664 where

$$\det(\mathbf{J}_{\text{STD}}) = \begin{vmatrix} xJ_{EE} & -J_{EI} \\ J_{IE} & -J_{II} \end{vmatrix} = -xJ_{EE}J_{II} + J_{IE}J_{EI} \quad (51)$$

665 Note that $\det(\mathbf{J}_{\text{STD}})$ can now change its sign due to E-to-E STD, the characteristic function can there-
666 fore change its bending shape. We used this relation to visualize how E-to-E STD effectively changes
667 the network stability of the reduced system in **Figure 2D**.

668 Conditions for ISN in networks with E-to-E STD

669 Here, we identify the condition of being in the ISN regime in supralinear networks with E-to-E STD.

670 When the level of inhibition is frozen, the Jacobian of the system reduces to the following:

$$\mathbf{M}_1 = \begin{bmatrix} \tau_E^{-1}(x J_{EE} \alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) & \tau_E^{-1} J_{EE} \alpha_E r_E^{\frac{2\alpha_E-1}{\alpha_E}} \\ -U_d x & -\tau_x^{-1} - U_d r_E \end{bmatrix} \quad (52)$$

671 For the system with frozen inhibition, the dynamics are stable if

$$\text{tr}(\mathbf{M}_1) = \tau_E^{-1}(x J_{EE} \alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) - \tau_x^{-1} - U_d r_E < 0 \quad (53)$$

672 and

$$\det(\mathbf{M}_1) = \tau_E^{-1}(x J_{EE} \alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1)(-\tau_x^{-1} - U_d r_E) + \tau_E^{-1} J_{EE} \alpha_E r_E^{\frac{2\alpha_E-1}{\alpha_E}} U_d x > 0 \quad (54)$$

673 Therefore, if the network is an ISN at the fixed point, the following condition has to be satisfied:

$$x > \min \left(\sqrt{\frac{1}{J_{EE} \alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}}}}, \frac{\tau_x + \tau_E + \tau_E \tau_x U_d r_E}{\tau_x J_{EE} \alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}}} \right) \quad (55)$$

674 Furthermore, we define the largest real part of the eigenvalues of $\mathbf{M}_1$ as the ISN index for networks

675 with E-to-E STD. More specifically,

$$\begin{aligned} \text{ISN index} = \text{Re} \left[\frac{\tau_E^{-1}(x J_{EE} \alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) - \tau_x^{-1} - U_d r_E}{2} \right. \\ \left. + \sqrt{\frac{1}{4}(\tau_E^{-1}(x J_{EE} \alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) + \tau_x^{-1} + U_d r_E)^2 - \tau_E^{-1} J_{EE} \alpha_E r_E^{\frac{2\alpha_E-1}{\alpha_E}} U_d x} \right] \end{aligned} \quad (56)$$

676 Conditions for paradoxical response in networks with E-to-E STD

677 Next, we identify the condition of having the paradoxical effect in supralinear networks with E-to-E

678 STD. To that end, we exploit a separation of timescales between the fast neural activity and the slow

679 STP variable. Therefore, set the depression variable to its value at the fixed point corresponding

680 to the fixed point value of r_E . The excitatory nullcline is defined as follows

$$\tau_E \frac{dr_E}{dt} = -r_E + \left[\frac{1}{1 + \tau_x U_d r_E} J_{EE} r_E - J_{EI} r_I + g_E \right]_+^{\alpha_E} = 0 \quad (57)$$

681 For $r_{E,I} > 0$, we have

$$r_I = \frac{\frac{1}{1 + \tau_x U_d r_E} J_{EE} r_E - r_E^{\frac{1}{\alpha_E}} + g_E}{J_{EI}} \quad (58)$$

682 The slope of the excitatory nullcline in the r_E/r_I plane where x axis is r_E and y axis is r_I can be

683 written as follows

$$k_{STD}^E = \frac{1}{J_{EI}} \left(-\frac{J_{EE}}{(1 + \tau_x U_d r_E)^2} \tau_x U_d r_E + \frac{J_{EE}}{1 + \tau_x U_d r_E} - \frac{1}{\alpha_E} r_E^{\frac{1}{\alpha_E}-1} \right) \quad (59)$$

684 Note that the slope of the excitatory nullcline is nonlinear. To have paradoxical effect, the slope

685 of the excitatory nullcline at the fixed point of the system has to be positive. Therefore, the STD

686 variable x at the fixed point has to satisfy the following condition

$$x > \sqrt{\frac{1}{J_{EE} \alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}}}} \quad (60)$$

687 The inhibitory nullcline can be written as follows

$$\tau_I \frac{dr_I}{dt} = -r_I + \left[J_{IE}r_E - J_{II}r_I + g_I \right]_+^{\alpha_I} = 0 \quad (61)$$

688 In the region of rates $r_{E,I} > 0$, we have

$$r_I = \frac{J_{IE}r_E - r_I^{\frac{1}{\alpha_I}} + g_I}{J_{II}} \quad (62)$$

689 The slope of the inhibitory nullcline can be written as follows

$$k_{STD}^I = \frac{J_{IE}}{J_{II} + \frac{1}{\alpha_I} r_I^{\alpha_I - 1}} \quad (63)$$

690 In addition to the positive slope of the excitatory nullcline, the slope of the inhibitory nullcline at the
691 fixed point of the system has to be larger than the slope of the excitatory nullcline. We therefore
692 have

$$J_{EI}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} J_{IE}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}} (\tau_x^{-1} + U_d r_E) > \left(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}} \right) \left(-\frac{J_{EE}U_d r_E}{1 + \tau_x U_d r_E} \alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} \right. \\ \left. + \frac{J_{EE}}{1 + \tau_x U_d r_E} \alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} (\tau_x^{-1} + U_d r_E) - (\tau_x^{-1} + U_d r_E) \right) \quad (64)$$

693 The above condition is the same as the stability condition of the determinant of the Jacobian of the
694 system with E-to-E STD (Eq. (50)). Therefore, the condition is always satisfied when the system with
695 E-to-E STD is stable.

696 Based on the condition of being ISN shown in Eq. (55) and the condition of having paradoxical
697 effect shown in Eq. (60), we therefore can conclude that in supralinear networks with E-to-E STD,
698 the paradoxical effect implies inhibitory stabilization, whereas inhibitory stabilization does not nec-
699 essarily imply paradoxical responses. This is consistent with recent work by *Sanzeni et al. (2020)*,
700 in which threshold-linear networks with STP have been studied. Here, we showed analytically that
701 the conclusion holds for any rectified power-law activation function with positive α .

702 To visualize the conditions in a two-dimensional plane, we reduced the conditions into a func-
703 tion of J_{EE} and x . For *Figure 2G*, $r_E = 1$. In *Figure 2-Figure Supplement 5C* and *Figure 2-Figure*
704 *Supplement 8A*, the depression variable thresholds above which the network exhibits the paradox-
705 ical effect were calculated based on Eq. (60).

706 Uni-stability conditions

707 The system is said to be “uni-stable”, when it has a single stable fixed point. We first identified the
708 uni-stability condition for networks with global inhibition. To that end, we considered a general
709 network with N excitatory populations and N inhibitory populations. To treat this problem analyt-
710 ically, we did not take STP into account in our analysis. The Jacobian matrix of networks with global
711 inhibition $\mathbf{Q}$, can be written as follows,

$$\mathbf{Q} = \begin{bmatrix} \mathbf{J}_{E \leftarrow E} & \mathbf{J}_{E \leftarrow I} \\ \mathbf{J}_{I \leftarrow E} & \mathbf{J}_{I \leftarrow I} \end{bmatrix} \quad (65)$$

712 where $\mathbf{J}_{E \leftarrow E}$, $\mathbf{J}_{E \leftarrow I}$, $\mathbf{J}_{I \leftarrow E}$, and $\mathbf{J}_{I \leftarrow I}$ are N by N block matrices defined below.

$$\mathbf{J}_{E \leftarrow E} = \begin{bmatrix} a - e & ka & \cdots & ka \\ ka & a - e & \cdots & ka \\ \vdots & \vdots & \ddots & \vdots \\ ka & ka & \cdots & a - e \end{bmatrix} \quad (66)$$

713

$$\mathbf{J}_{E \leftarrow I} = -b\mathbf{J}_{N,N} \quad (67)$$

714

$$\mathbf{J}_{I \leftarrow E} = c \mathbf{J}_{N,N} \quad (68)$$

715

$$\mathbf{J}_{I \leftarrow I} = \begin{bmatrix} -d-f & -d & \cdots & -d \\ -d & -d-f & \cdots & -d \\ \vdots & \vdots & \ddots & \vdots \\ -d & -d & \cdots & -d-f \end{bmatrix} \quad (69)$$

716 where $a = \tau_E^{-1} J_{EE} \alpha_E [z_E]_+^{\alpha_E-1}$, $b = \tau_E^{-1} J_{EI} \alpha_E [z_E]_+^{\alpha_E-1}$, $c = \tau_I^{-1} J_{IE} \alpha_I [z_I]_+^{\alpha_I-1}$, $d = \tau_I^{-1} J_{II} \alpha_I [z_I]_+^{\alpha_I-1}$, $e = \tau_E^{-1}$,
 717 and $f = \tau_I^{-1}$. Here, z_E and z_I denote the total current into the excitatory and inhibitory population,
 718 respectively. Note that all these parameters are non-negative. Parameter k controls the excitatory
 719 connection strength across different populations. $\mathbf{J}_{N,N}$ is a N by N matrix of ones.

720 The eigenvalues of the Jacobian $\mathbf{Q}$ are roots of its characteristic polynomial,

$$\det((\mathbf{J}_{E \leftarrow E} - \lambda \mathbb{1})(\mathbf{J}_{I \leftarrow I} - \lambda \mathbb{1}) - \mathbf{J}_{E \leftarrow I} \mathbf{J}_{I \leftarrow E}) = 0 \quad (70)$$

721 where $\mathbb{1}$ represents the identity matrix of size N . The characteristic polynomial can be expanded
 722 to:

$$\left[(a - e - ka - \lambda)(-f - \lambda) \right]^{N-1} \left[(a - e + (N-1)ka - \lambda)(-Nd - f - \lambda) + N^2 bc \right] = 0 \quad (71)$$

723 We therefore had four distinct eigenvalues:

$$\lambda_1 = a - e - ka \quad (72)$$

$$\lambda_2 = -f \quad (73)$$

and

$$\lambda_{3/4} = \frac{1}{2} \left[(a - e - f - Nd + (N-1)ka) \pm \sqrt{(a - e - f - Nd + (N-1)ka)^2 - 4((-af + ef + k af) - N(a - e)d - N k af - N(N-1)kad + N^2 bc)} \right] \quad (74)$$

724 Note that the eigenvalues λ_1 and λ_2 have an algebraic and geometric multiplicity of $(N-1)$, whereas
 725 the eigenvalues λ_3 and λ_4 have an algebraic and geometric multiplicity of 1.

726 In analogy to networks with global inhibition, the Jacobian matrix of networks with co-tuned
 727 inhibition $\mathbf{R}$, can be written as

$$\mathbf{R} = \begin{bmatrix} \mathbf{J}_{E \leftarrow E} & \mathbf{J}_{E \leftarrow I} \\ \mathbf{J}_{I \leftarrow E} & \mathbf{J}_{I \leftarrow I} \end{bmatrix} \quad (75)$$

728 where $\mathbf{J}_{E \leftarrow E}$, $\mathbf{J}_{E \leftarrow I}$, $\mathbf{J}_{I \leftarrow E}$, and $\mathbf{J}_{I \leftarrow I}$ are N by N block matrices defined as follows:

$$\mathbf{J}_{E \leftarrow E} = \begin{bmatrix} a - e & ka & \cdots & ka \\ ka & a - e & \cdots & ka \\ \vdots & \vdots & \ddots & \vdots \\ ka & ka & \cdots & a - e \end{bmatrix} \quad (76)$$

$$\mathbf{J}_{E \leftarrow I} = \begin{bmatrix} -Nb + (N-1)mb & -mb & \cdots & -mb \\ -mb & -Nb + (N-1)mb & \cdots & -mb \\ \vdots & \vdots & \ddots & \vdots \\ -mb & -mb & \cdots & -Nb + (N-1)mb \end{bmatrix} \quad (77)$$

$$\mathbf{J}_{I \leftarrow E} = \begin{bmatrix} Nc - (N-1)mc & mc & \cdots & mc \\ mc & Nc - (N-1)mc & \cdots & mc \\ \vdots & \vdots & \ddots & \vdots \\ mc & mc & \cdots & Nc - (N-1)mc \end{bmatrix} \quad (78)$$

$$\mathbf{J}_{I \leftarrow I} = \begin{bmatrix} -Nd + (N-1)md - f & -md & \cdots & -md \\ -md & -Nd + (N-1)md - f & \cdots & -md \\ \vdots & \vdots & \ddots & \vdots \\ -md & -md & \cdots & -Nd + (N-1)md - f \end{bmatrix} \quad (79)$$

where m controls the degree of co-tuning in the network. If $m = 0$, the network decouples into N independent ensembles and inhibition is perfectly co-tuned with excitation. In the case $m = 1$, inhibition is global and the block matrices become identical to the above case of global inhibition. The eigenvalues of the matrix $\mathbf{R}$ are given as the roots of the characteristic polynomial defined by:

$$\det((\mathbf{J}_{E \leftarrow E} - \lambda \mathbf{1})(\mathbf{J}_{I \leftarrow I} - \lambda \mathbf{1}) - \mathbf{J}_{E \leftarrow I} \mathbf{J}_{I \leftarrow E}) = 0 \quad (80)$$

which yields the following expression:

$$\left[\lambda^2 - (a - e - ka - Nd + Nmd - f)\lambda - (a - e - ka)(Nd - Nmd - f) + N^2 bc(1 - m)^2 \right]^{N-1} \left[(a - e + (N - 1)ka - \lambda)(-Nd - f - \lambda) + N^2 bc \right] = 0 \quad (81)$$

We therefore had four distinct eigenvalues:

$$\lambda'_{1/2} = \frac{1}{2} \left[(a - e - ka - Nd + Nmd - f) \pm \sqrt{(a - e - ka + Nd - Nmd + f)^2 - 4N^2 bc(1 - m)^2} \right] \quad (82)$$

736

$$\lambda'_{3/4} = \frac{1}{2} \left[(a - e - f - Nd + (N - 1)ka) \pm \sqrt{(a - e - f - Nd + (N - 1)ka)^2 - 4((-af + ef + kaf) - N(a - e)d - Nkaf - N(N - 1)kad + N^2 bc)} \right] \quad (83)$$

The eigenvalues λ'_1 and λ'_2 have an algebraic and geometric multiplicity of $(N - 1)$, whereas the eigenvalues λ'_3 and λ'_4 have an algebraic and geometric multiplicity of 1. We noted that $\lambda_3 = \lambda'_3$, $\lambda_4 = \lambda'_4$.

To compare under which conditions networks with different structures are uni-stable, we examined the different eigenvalues derived above. As $\lambda_2 < 0$, and $\lambda'_1 > \lambda'_2$, we only had to compare λ'_1 to λ_1 . For networks with co-tuned inhibition, we have $m < 1$,

$$\begin{aligned} \lambda'_1 &= \frac{1}{2} \left[(a - e - ka - Nd + Nmd - f) + \sqrt{(a - e - ka + Nd - Nmd + f)^2 - 4N^2 bc(1 - m)^2} \right] \\ &< \frac{1}{2} \left[(a - e - ka - Nd + Nmd - f) + \sqrt{(a - e - ka + Nd - Nmd + f)^2} \right] = a - e - ka = \lambda_1 \end{aligned} \quad (84)$$

The inequality, $\lambda'_1 < \lambda_1$, indicates that networks with co-tuned inhibition have a broad parameter regime in which they are uni-stable than networks with global inhibition. Note that in the absence of a saturating nonlinearity of the input-output function and in the absence of any additional stabilization mechanisms, systems with positive eigenvalues of the Jacobian are unstable. In this case, networks with co-tuned inhibition have a broad parameter regime of being stable than networks with global inhibition.

To visualize the conditions in a two-dimensional plane, we reduced the conditions into a function of a and d . For **Figure 3C**, $k = 0.1$, $m = 0.5$ and $bc = 0.9ad$.

Distance to the decision boundary

To calculate the distance to the decision boundary in **Figure 4**, **Figure 5**, **Figure 4–Figure Supplement 2** and **Figure 5–Figure Supplement 2**, we first projected the excitatory activity in Phase 2 onto a two-dimensional Cartesian coordinate system in which the horizontal axis is the activity of the first excitatory ensemble r_{E1} and the vertical axis is the activity of the second excitatory ensemble r_{E2} . We denote the location of the projected data point in the Cartesian coordinate system by (x, y) , where x and y equal r_{E1} and r_{E2} , respectively. The distance L between the projected data and the decision boundary which corresponds to the diagonal line in the coordinate system can be expressed as follows:

$$L = \sqrt{x^2 + y^2} \sin(|45^\circ - \arcsin(\frac{x}{\sqrt{x^2 + y^2}})|) \quad (85)$$

Note that the inverse trigonometric function $\arcsin$ gives the value of the angle in degrees.

760 Inhibitory feedback pathways for suppressing unwanted neural activation

761 To identify the important neural pathways for the suppression of unwanted neural activation, we
 762 analyzed how the activity of the second excitatory ensemble r_{E2} changes with the input to the first
 763 excitatory ensemble g_{E1} . To that end, we considered a general weight matrix for networks with
 764 two interacting ensembles

$$\mathbf{J} = \begin{bmatrix} J_{E1E1} & J_{E1E2} & -J_{E1I1} & -J_{E1I2} \\ J_{E2E1} & J_{E2E2} & -J_{E2I1} & -J_{E2I2} \\ J_{I1E1} & J_{I1E2} & -J_{I1I1} & -J_{I1I2} \\ J_{I2E1} & J_{I2E2} & -J_{I2I1} & -J_{I2I2} \end{bmatrix} \quad (86)$$

765 We can write the change in firing rate of the excitatory population in the second ensemble δr_{E2}
 766 as a function of the change in the input to the other δg_{E1} :

$$\begin{aligned} \delta r_{E2} = \frac{1}{\det(\mathbb{1} - \mathbf{FJ})} & \left[(-f'_{E2} J_{E2E1}) f'_{I1} J_{I1I2} f'_{I2} J_{I2I1} + f'_{E2} J_{E2I1} (-f'_{I1} J_{I1E1}) (1 + f'_{I2} J_{I2I2}) \right. \\ & + f'_{E2} J_{E2I2} (1 + f'_{I1} J_{I1I1}) (-f'_{I2} J_{I2E1}) - (-f'_{E2} J_{E2E1}) (1 + f'_{I1} J_{I1I1}) (1 + f'_{I2} J_{I2I2}) \\ & \left. - f'_{E2} J_{E2I1} f'_{I1} J_{I1I2} (-f'_{I2} J_{I2E1}) - f'_{E2} J_{E2I2} (-f'_{I1} J_{I1E1}) f'_{I2} J_{I2I1} \right] f'_{E1} \delta g_{E1} \end{aligned} \quad (87)$$

767 where $\mathbb{1}$ is the identity matrix. And $\mathbf{F}$ is given by

$$\mathbf{F} = \begin{bmatrix} f'_{E1} & 0 & 0 & 0 \\ 0 & f'_{E2} & 0 & 0 \\ 0 & 0 & f'_{I1} & 0 \\ 0 & 0 & 0 & f'_{I2} \end{bmatrix} \quad (88)$$

768 where f'_{E1} , f'_{E2} , f'_{I1} and f'_{I2} are the derivatives of the input-output functions evaluated at the fixed
 769 point.

770 Assuming that $J_{E1E1} = J_{E2E2} = J_{EE}$, $J_{I1E1} = J_{I2E2} = J_{IE}$, $J_{E1I1} = J_{E2I2} = J_{EI}$, $J_{I1I1} = J_{I2I2} = J_{II}$,
 771 $J_{E1E2} = J_{E2E1} = J'_{EE}$, $J_{I1E2} = J_{I2E1} = J'_{IE}$, $J_{E1I2} = J_{E2I1} = J'_{EI}$ and $J_{I1I2} = J_{I2I1} = J'_{II}$, we find

$$\begin{aligned} \delta r_{E2} = \frac{1}{\det(\mathbb{1} - \mathbf{FJ})} & \left[(-f'_{E2} J'_{EE}) f'_{I1} J'_{II} f'_{I2} J'_{II} + f'_{E2} J'_{EI} (-f'_{I1} J_{IE}) (1 + f'_{I2} J_{II}) \right. \\ & + f'_{E2} J_{EI} (1 + f'_{I1} J_{II}) (-f'_{I2} J'_{IE}) - (-f'_{E2} J'_{EE}) (1 + f'_{I1} J_{II}) (1 + f'_{I2} J_{II}) \\ & \left. - f'_{E2} J'_{EI} f'_{I1} J'_{II} (-f'_{I2} J'_{IE}) - f'_{E2} J_{EI} (-f'_{I1} J_{IE}) f'_{I2} J'_{II} \right] f'_{E1} \delta g_{E1} \end{aligned} \quad (89)$$

772 By further assuming that the weight strengths across ensembles are weak and ignoring the corre-
 773 sponding higher-order terms, we get

$$\begin{aligned} \delta r_{E2} & \approx \frac{1}{\det(\mathbb{1} - \mathbf{FJ})} \left[f'_{E2} J'_{EI} (-f'_{I1} J_{IE}) (1 + f'_{I2} J_{II}) + f'_{E2} J_{EI} (1 + f'_{I1} J_{II}) (-f'_{I2} J'_{IE}) \right. \\ & \left. - (-f'_{E2} J'_{EE}) (1 + f'_{I1} J_{II}) (1 + f'_{I2} J_{II}) - f'_{E2} J_{EI} (-f'_{I1} J_{IE}) f'_{I2} J'_{II} \right] f'_{E1} \delta g_{E1} \\ & = \frac{1}{\det(\mathbb{1} - \mathbf{FJ})} \left[\left(\frac{J'_{II}}{J'_{EI}} f'_{I2} - \left(\frac{1}{J_{EI}} + f'_{I2} \frac{J_{II}}{J_{EI}} \right) \right) J'_{EI} J_{EI} J_{IE} f'_{E2} f'_{I1} \right. \\ & \left. + \left(\frac{J'_{EE}}{J'_{IE}} (1 + J_{II} f'_{I2}) - J_{EI} f'_{I2} \right) J'_{IE} f'_{E2} (1 + f'_{I1} J_{II}) \right] f'_{E1} \delta g_{E1} \end{aligned} \quad (90)$$

774 Note that $\frac{J'_{EE}}{J'_{IE}}$ and $\frac{J'_{II}}{J'_{EI}}$ are terms regulating the respective excitatory and inhibitory input from one
 775 ensemble to the excitatory and inhibitory population in another ensemble. The term $\det(\mathbb{1} - \mathbf{FJ})$ is
 776 positive to ensure the stability of the system.

777 To suppress the activity of the excitatory population in the second ensemble r_{E2} , in other words,
 778 to ensure that $\delta r_{E2} < 0$, J'_{IE} or/and J'_{EI} have to be large. Therefore, we identified J'_{IE} and J'_{EI}
 779 as important synaptic connections which lead to suppression of the unwanted neural activation,

780 suggesting that inhibition can be provided via J'_{IE} through the $E1-I2-E2$ pathway or via J'_{EI} through
 781 the $E1-I1-E2$ pathway.

782 For **Figure 4-5**, the rate-based model consists of two ensembles, each of which is composed of
 783 100 excitatory and 25 inhibitory neurons with all-to-all connectivity.

784 Spiking neural network model

785 The spiking neural network model was composed of N_E excitatory and N_I inhibitory leaky integrate-
 786 and-fire neurons. Neurons were randomly connected with probability of 20%. The dynamics of
 787 membrane potential of neuron i , U_i , are given by (Zenke et al., 2015):

$$\tau^m \frac{dU_i}{dt} = (U^{\text{rest}} - U_i) + g_i^{\text{ext}}(t)(U^{\text{exc}} - U_i) + g_i^{\text{inh}}(t)(U^{\text{inh}} - U_i) \quad (91)$$

788 Here, τ^m is the membrane time constant and U^{rest} is the resting potential. Spikes are triggered when
 789 the membrane potential reaches the spiking threshold U^{thr} . After a spike is emitted, the membrane
 790 potential is reset to U^{rest} and the neuron enters a refractory period of τ^{ref} . Inhibitory neurons
 791 obeyed the same integrate-and-fire formalism but with a shorter membrane time constant.

792 Excitatory synapses contain a fast AMPA component and a slow NMDA component. The dy-
 793 namics of the excitatory conductance are described by:

$$\tau^{\text{ampa}} \frac{dg_i^{\text{ampa}}}{dt} = -g_i^{\text{ampa}} + \sum_{j \in \text{exc}} J_{ij} S_j(t) \quad (92)$$

$$\tau^{\text{nmda}} \frac{dg_i^{\text{nmda}}}{dt} = -g_i^{\text{nmda}} + g_i^{\text{ampa}} \quad (93)$$

$$g_i^{\text{exc}}(t) = \xi g_i^{\text{ampa}}(t) + (1 - \xi) g_i^{\text{nmda}}(t) \quad (94)$$

794 Here, J_{ij} denotes the synaptic strength from neuron j to neuron i . If the connection does not exist,
 795 J_{ij} was set to 0. $S_j(t)$ is the spike train of neuron j , which is defined as $S_j(t) = \sum_k \delta(t - t_j^k)$, where δ
 796 is the Dirac delta function and t_j^k the spikes times k of neuron j . ξ is a weighting parameter. The
 797 dynamics of inhibitory conductances are governed by:

$$\tau^{\text{gaba}} \frac{dg_i^{\text{inh}}}{dt} = -g_i^{\text{inh}} + \sum_{j \in \text{inh}} J_{ij} S_j(t) \quad (95)$$

798 In the spiking neural network models, SFA of excitatory neurons is modeled as follows,

$$\tau^m \frac{dU_i}{dt} = (U^{\text{rest}} - U_i) + g_i^{\text{ext}}(t)(U^{\text{exc}} - U_i) + (g_i^{\text{inh}}(t) + a_i(t))(U^{\text{inh}} - U_i) \quad (96)$$

$$\frac{da_i}{dt} = -\frac{a_i}{\tau_a} + b S_i(t) \quad (97)$$

799 where i is the index of excitatory neurons.

800 The dynamics of E-to-E STD are given by

$$\frac{dx_{ij}}{dt} = \frac{1 - x_{ij}}{\tau_x} - U_d x_{ij} S_j(t) \quad (98)$$

$$\tau^{\text{ampa}} \frac{dg_i^{\text{ampa}}}{dt} = -g_i^{\text{ampa}} + \sum_{j \in \text{exc}} x_{ij} J_{ij} S_j(t) \quad (99)$$

801 where i represents the index of excitatory neurons.

802 The dynamics of E-to-I STF are governed by

$$\frac{du_{ij}}{dt} = \frac{1 - u_{ij}}{\tau_u} + U_f (U_{\text{max}} - u_{ij}) S_j(t) \quad (100)$$

$$\tau^{\text{ampa}} \frac{dg_i^{\text{ampa}}}{dt} = -g_i^{\text{ampa}} + \sum_{j \in \text{exc}} u_{ij} J_{ij} S_j(t) \quad (101)$$

803 where i denotes the index of inhibitory neurons.

804 For **Figure 6**, each excitatory and inhibitory neuron received external excitatory input from 300
805 neurons firing with Poisson statistics at an average firing rate of 0.1 Hz at baseline. During stimula-
806 tion, the excitatory neurons corresponding to the activated ensemble received external excitatory
807 input from 300 neurons firing with Poisson statistics at an average firing rate of 0.5 Hz. The ensem-
808 ble activity is computed from the instantaneous firing rates of the respective ensembles with 10 ms
809 bin size. The difference in ensemble activity for the peak amplitude is calculated by subtracting the
810 average maximal ensemble activity of the unstimulated ensembles from the maximal ensemble
811 activity of the activated ensemble. Similarly, the difference in ensemble activity for the fixed point
812 is calculated by subtracting the average ensemble activity of the unstimulated ensembles at the
813 fixed point from the ensemble activity of the activated ensemble at the fixed point. Fixed point ac-
814 tivity is computed by averaging the activity of the middle 1 second within the 2-second stimulation
815 period.

816 For **Figure 2–Figure Supplement 10**, each excitatory and inhibitory neuron received external
817 excitatory input from 300 neurons firing with Poisson statistics at an average firing rate of 0.1 Hz
818 at the baseline. During stimulation, each excitatory neuron received external excitatory input from
819 300 neurons firing with Poisson statistics at an average firing rate of 0.3 Hz.

820 For **Figure 6–Figure Supplement 1**, the firing rates of 300 neurons are varying from 4/15 Hz to
821 7/15 Hz.

822 Simulations

823 Simulations were performed in Python and Mathematica. All differential equations were imple-
824 mented by Euler integration with a time step of 0.1 ms. All simulation parameters are listed in
825 Tables 1–5 and **Appendix 5** Tables 1–10. The simulation source code to reproduce the figures is
826 publicly available at <https://github.com/fmi-basel/gzenke-nonlinear-transient-amplification>.

Table 1. Parameters for Figure 1C-E

Symbol	Value	Unit	Description
J_{EE}	1.8	-	E-to-E connection strength
J_{IE}	1.0	-	E-to-I connection strength
J_{EI}	1.0	-	I-to-E connection strength
J_{II}	0.6	-	I-to-I connection strength
α_E	2	-	power of excitatory input-output function
α_I	2	-	power of inhibitory input-output function
τ_E	20	ms	time constant of excitatory firing dynamics
τ_I	10	ms	time constant of inhibitory firing dynamics
g_E^{bs}	1.55	-	input to the E population at baseline
g_E^{stim}	3.0	-	input to the E population during stimulation
g_I	2.0	-	input to the I population
Parameters for Figure 1F			
J_{IE}	0.45	-	E-to-I connection strength
J_{EI}	1.0	-	I-to-E connection strength
J_{II}	1.5	-	I-to-I connection strength

Table 2. Parameters for Figure 2

Symbol	Value	Unit	Description
τ_a	200	ms	time constant of SFA
b	1.0	-	strength of SFA
τ_x	200	ms	time constant of STD
U_d	1.0	-	depression rate
τ_u	200	ms	time constant of STF
U_f	1.0	-	facilitation rate
U_{max}	6.0	-	maximal facilitation value
Note that these values are also applied elsewhere unless mentioned otherwise.			

Table 3. Parameters for Figure 3 bi/multi-stable example

Symbol	Value	Unit	Description
J_{EE}	1.4	-	within-ensemble E-to-E connection strength
J_{IE}	0.6	-	within-ensemble E-to-I connection strength
J_{EI}	1.0	-	within-ensemble I-to-E connection strength
J_{II}	0.6	-	within-ensemble I-to-I connection strength
J'_{EE}	0.14	-	inter-ensemble E-to-E connection strength
J'_{IE}	0.6	-	inter-ensemble E-to-I connection strength
J'_{EI}	1.0	-	inter-ensemble I-to-E connection strength
J'_{II}	0.6	-	inter-ensemble I-to-I connection strength
g_{E1}^{bs}	2.2	-	input to the $E1$ population at baseline
g_{E1}^{stim}	3.0	-	input to the $E1$ population during stimulation
g_{E2}	2.2	-	input to the $E2$ population
g_I	2.0	-	input to the I population
Parameters for Figure 3 uni-stable example			
J_{EE}	1.3	-	within-ensemble E-to-E connection strength
J'_{EE}	0.13	-	inter-ensemble E-to-E connection strength

Table 4. Parameters for Figure 4-5

Symbol	Value	Unit	Description
N_E	200	-	number of excitatory neurons
N_I	50	-	number of inhibitory neurons
N	2	-	number of ensembles
J_{EE}	$1.2/(N_E/2 - 1)$	-	within-ensemble E-to-E connection strength
J_{IE}	$1.0/(N_E/2)$	-	within-ensemble E-to-I connection strength
J_{EI}	$1.0/(N_I/2)$	-	within-ensemble I-to-E connection strength
J_{II}	$1.0/(N_I/2 - 1)$	-	within-ensemble I-to-I connection strength
J'_{EE}	$0.36/(N_E/2 - 1)$	-	inter-ensemble E-to-E connection strength
J'_{IE}	$0.4/(N_E/2)$	-	inter-ensemble E-to-I connection strength
J'_{EI}	$0.1/(N_I/2)$	-	inter-ensemble I-to-E connection strength
J'_{II}	$0.1/(N_I/2)$	-	inter-ensemble I-to-I connection strength
g_I	2.0	-	input to the I population
Parameters for Figure 4			
g_{E1}^{bs}	1.35	-	input to the $E1$ population at baseline
g_{E1}^{stim}	4.0	-	input to the $E1$ population during stimulation
g_{E2}	1.35	-	input to the $E2$ population
Parameters for Figure 5			
g_{E1}^{bs}	1.35	-	input to the $E1$ population at baseline
g_{E1}^{stim}	$1.35 + (4.0 - 1.35)(1-p)$	-	input to the $E1$ population during stimulation
g_{E2}^{bs}	1.35	-	input to the $E2$ population at baseline
g_{E2}^{stim}	$1.35 + (4.0 - 1.35)p$	-	input to the $E2$ population during stimulation
Here, p is a parameter between 0 and 1 controlling the additional inputs to $E1$ and $E2$.			

Table 5. Parameters for Figure 6

Symbol	Value	Unit	Description
N_E	400	-	number of excitatory neurons
N_I	100	-	number of inhibitory neurons
U^{rest}	-70	mV	resting membrane potential
U^{exc}	0	mV	excitatory reversal potential
U^{inh}	-80	mV	inhibitory reversal potential
τ^{ref}	3	ms	duration of refractory period
τ_{exc}^m	20	ms	membrane time constant of excitatory neurons
τ_{inh}^m	10	ms	membrane time constant of inhibitory neurons
τ^{ampa}	5	ms	time constant of AMPA receptor
τ^{gaba}	10	ms	time constant of GABA receptor
τ^{nmda}	100	ms	time constant of NMDA receptor
ξ	0.5	-	receptor weighting factor
J_{EE}	0.19	-	within-ensemble E-to-E connection strength
J_{IE}	0.10	-	within-ensemble E-to-I connection strength
J_{EI}	0.10	-	within-ensemble I-to-E connection strength
J_{II}	0.06	-	within-ensemble I-to-I connection strength
J_{EE}^I	0.019	-	inter-ensemble E-to-E connection strength
J_{IE}^I	0.05	-	inter-ensemble E-to-I connection strength
J_{EI}^I	0.04	-	inter-ensemble I-to-E connection strength
J_{II}^I	0.006	-	inter-ensemble I-to-I connection strength

Acknowledgments

We thank Rainer W. Friedrich, Claire Meissner-Bernard, William F. Podlaski, and members of the Zenke Group for comments and discussions. This work was supported by the Novartis Research Foundation.

Competing interests

The authors declare no competing interests.

References

- Abeles M. Corticonics. Cambridge: Cambridge University Press; 1991. doi: <https://doi.org/10.1017/CBO9780511574566>.
- Ahmadian Y, Rubin DB, Miller KD. Analysis of the stabilized supralinear network. *Neural Computation* 2013;25(8):1994–2037. doi: https://doi.org/10.1162/NECO_a_00472.
- Amit DJ, Brunel N. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex* 1997;7(3):237–252. doi: <https://doi.org/10.1093/cercor/7.3.237>.
- Baker C, Zhu V, Rosenbaum R. Nonlinear stimulus representations in neural circuits with approximate excitatory-inhibitory balance. *PLoS Computational Biology* 2020;16(9):1–30. doi: <https://doi.org/10.1371/journal.pcbi.1008192>.
- Barkai E, Hasselmo ME. Modulation of the input/output function of rat piriform cortex pyramidal cells. *Journal of Neurophysiology* 1994;72(2):644–658. doi: <https://doi.org/10.1152/jn.1994.72.2.644>.
- Benda J, Herz AVM. A universal model for spike-frequency adaptation. *Neural Computation* 2003;15(11):2523–2564. doi: <https://doi.org/10.1162/089976603322385063>.
- Bissière S, Humeau Y, Lüthi A. Dopamine gates LTP induction in lateral amygdala by suppressing feedforward inhibition. *Nature Neuroscience* 2003;6(6):587–592. doi: <https://doi.org/10.1038/nn1058>.
- Bolding KA, Franks KM. Recurrent cortical circuits implement concentration-invariant odor coding. *Science* 2018;361(6407). doi: <https://doi.org/10.1126/science.aat6904>.
- Bolding KA, Nagappan S, Han BX, Wang F, Franks KM. Recurrent circuitry is required to stabilize piriform cortex odor representations across brain states. *eLife* 2020;9:1–23. doi: <https://doi.org/10.7554/eLife.53125>.
- Bondanelli G, Ostojic S. Coding with transient trajectories in recurrent neural networks. *PLoS Computational Biology* 2020;16(2):1–36. doi: <https://doi.org/10.1371/journal.pcbi.1007655>.
- Brette R, Gerstner W. Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of Neurophysiology* 2005;94(5):3637–3642. doi: <https://doi.org/10.1152/jn.00686.2005>.
- Carrillo-Reid L, Yang W, Bando Y, Peterka DS, Yuste R. Imprinting and recalling cortical ensembles. *Science* 2016;353(6300):691–694. doi: <https://doi.org/10.1126/science.aaf7560>.
- Cayco-Gajic NA, Silver RA. Re-evaluating Circuit Mechanisms Underlying Pattern Separation. *Neuron* 2019;101(4):584–602. doi: <https://doi.org/10.1016/j.neuron.2019.01.044>.
- Christodoulou G, Vogels TP, Agnes EJ. Regimes and mechanisms of transient amplification in abstract and biological networks. *bioRxiv* 2021; doi: <https://doi.org/10.1101/2021.04.01.437964>.
- Churchland MM, Cunningham JP, Kaufman MT, Foster JD, Nuyujukian P, Ryu SI, et al. Neural population dynamics during reaching. *Nature* 2012;487(7405):51–56. doi: <https://doi.org/10.1038/nature11129>.
- Compte A, Brunel N, Goldman-Rakic PS, Wang XJ. Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex* 2000;10(9):910–923. doi: <https://doi.org/10.1093/cercor/10.9.910>.
- Cossell L, Iacaruso MF, Muir DR, Houlton R, Sader EN, Ko H, et al. Functional organization of excitatory synaptic strength in primary visual cortex. *Nature* 2015;518(7539):399–403. doi: <https://doi.org/10.1038/nature14182>.
- Cruikshank SJ, Lewis TJ, Connors BW. Synaptic basis for intense thalamocortical activation of feedforward inhibitory cells in neocortex. *Nature Neuroscience* 2007;10(4):462–468. doi: <https://doi.org/10.1038/nn1861>.

- 872 Denève S, Machens CK. Efficient codes and balanced networks. *Nature Neuroscience* 2016;19(3):375–382. doi:
873 <https://doi.org/10.1038/nn.4243>.
- 874 DeWeese MR, Wehr M, Zador AM. Binary spiking in auditory cortex. *Journal of Neuroscience* 2003;23(21):7940–
875 7949. doi: <https://doi.org/10.1523/JNEUROSCI.23-21-07940.2003>.
- 876 Echeveste R, Aitchison L, Hennequin G, Lengyel M. Cortical-like dynamics in recurrent circuits optimized for
877 sampling-based probabilistic inference. *Nature Neuroscience* 2020;23(9):1138–1149. doi: [http://dx.doi.org/](http://dx.doi.org/10.1038/s41593-020-0671-1)
878 [10.1038/s41593-020-0671-1](http://dx.doi.org/10.1038/s41593-020-0671-1).
- 879 El-Gaby M, Reeve HM, Lopes-dos Santos V, Campo-Urriza N, Perestenko PV, Morley A, et al. An emergent neural
880 coactivity code for dynamic memory. *Nature Neuroscience* 2021;24(5):694–704. doi: [https://doi.org/10.1038/](https://doi.org/10.1038/s41593-021-00820-w)
881 [s41593-021-00820-w](https://doi.org/10.1038/s41593-021-00820-w).
- 882 Engel AK, Fries P, Singer W. Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature*
883 *Reviews Neuroscience* 2001;2(10):704–716. doi: <https://doi.org/10.1038/35094565>.
- 884 Fairhall AL, Lewen GD, Bialek W, de Ruyter van Steveninck RR. Efficiency and ambiguity in an adaptive neural
885 code. *Nature* 2001;412(6849):787–792. doi: <https://doi.org/10.1038/35090500>.
- 886 Franks KM, Russo MJ, Sosulski DL, Mulligan AA, Siegelbaum SA, Axel R. Recurrent circuitry dynamically shapes
887 the activation of piriform cortex. *Neuron* 2011;72(1):49–56. doi: <https://doi.org/10.1016/j.neuron.2011.08.020>.
- 888 Freedman DJ, Riesenhuber M, Poggio T, Miller EK. Categorical representation of visual stimuli in the primate
889 prefrontal cortex. *Science* 2001;291(5502):312–316. doi: <https://doi.org/10.1126/science.291.5502.312>.
- 890 Froemke RC, Merzenich MM, Schreiner CE. A synaptic memory trace for cortical receptive field plasticity. *Nature*
891 2007;450(7168):425–429. doi: <https://doi.org/10.1038/nature06289>.
- 892 Funahashi S, Bruce CJ, Goldman-Rakic PS. Mnemonic coding of visual space in the monkey's dorsolateral pre-
893 frontal cortex. *Journal of Neurophysiology* 1989;61(2):331–349. doi: <https://doi.org/10.1152/jn.1989.61.2.331>.
- 894 Gillary G, von der Heydt R, Niebur E. Short-term depression and transient memory in sensory cortex. *Journal*
895 *of Computational Neuroscience* 2017;43(3):273–294. doi: <https://doi.org/10.1007/s10827-017-0662-8>.
- 896 Gillett M, Pereira U, Brunel N. Characteristics of sequential activity in networks with temporally asymmet-
897 ric Hebbian learning. *Proceedings of the National Academy of Sciences of the United States of America*
898 2020;117(47):29948–29958. doi: <https://doi.org/10.1073/pnas.1918674117>.
- 899 Gjoni E, Zenke F, Bouhours B, Schneggenburger R. Specific synaptic input strengths determine the compu-
900 tational properties of excitation–inhibition integration in a sound localization circuit. *Journal of Physiology*
901 2018;596(20):4945–4967. doi: <https://doi.org/10.1113/JP276012>.
- 902 Goldman MS. Memory without feedback in a neural network. *Neuron* 2009;61(4):621–634. doi: [https://doi.org/](https://doi.org/10.1016/j.neuron.2008.12.012)
903 [10.1016/j.neuron.2008.12.012](https://doi.org/10.1016/j.neuron.2008.12.012).
- 904 Harris KD. Neural signatures of cell assembly organization. *Nature Reviews Neuroscience* 2005;6(5):399–407.
905 doi: <https://doi.org/10.1038/nrn1669>.
- 906 Harris KD, Csicsvari J, Hirase H, Dragoi G, Buzsáki G. Organization of cell assemblies in the hippocampus. *Nature*
907 2003;424(6948):552–556. doi: <https://doi.org/10.1038/nature01834>.
- 908 Hebb DO. *The organization of behavior: a neuropsychological theory*. Wiley, New York; 1949. doi: [https:](https://doi.org/10.4324/9781410612403)
909 [//doi.org/10.4324/9781410612403](https://doi.org/10.4324/9781410612403).
- 910 Hennequin G, Agnes EJ, Vogels TP. Inhibitory plasticity: Balance, control, and codependence. *Annual Review of*
911 *Neuroscience* 2017;40:557–579. doi: <https://doi.org/10.1146/annurev-neuro-072116-031005>.
- 912 Hennequin G, Ahmadian Y, Rubin DB, Lengyel M, Miller KD. The dynamical regime of sensory cortex: Sta-
913 ble dynamics around a single stimulus-tuned attractor account for patterns of noise variability. *Neuron*
914 2018;98(4):846–860.e5. doi: <https://doi.org/10.1016/j.neuron.2018.04.017>.
- 915 Hennequin G, Vogels TP, Gerstner W. Non-normal amplification in random balanced neuronal networks. *Phys-*
916 *ical Review E - Statistical, Nonlinear, and Soft Matter Physics* 2012;86(1):1–12. doi: [https://doi.org/10.1103/](https://doi.org/10.1103/PhysRevE.86.011909)
917 [PhysRevE.86.011909](https://doi.org/10.1103/PhysRevE.86.011909).

- 918 Hennequin G, Vogels TP, Gerstner W. Optimal control of transient dynamics in balanced networks supports
919 generation of complex movements. *Neuron* 2014;82(6):1394–1406. doi: [https://doi.org/10.1016/j.neuron.](https://doi.org/10.1016/j.neuron.2014.04.045)
920 [2014.04.045](https://doi.org/10.1016/j.neuron.2014.04.045).
- 921 Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. *Proceed-*
922 *ings of the national academy of sciences* 1982;79(8):2554–2558. doi: <https://doi.org/10.1073/pnas.79.8.2554>.
- 923 Horn RA, Johnson CR. *Matrix Analysis*. Cambridge University Press; 1985. doi: [https://doi.org/10.1017/](https://doi.org/10.1017/cbo9780511810817)
924 [cbo9780511810817](https://doi.org/10.1017/cbo9780511810817).
- 925 Irving RS. *Integers, polynomials, and rings: a course in algebra*. Springer; 2004. doi: [https://doi.org/10.1007/](https://doi.org/10.1007/b97633)
926 [b97633](https://doi.org/10.1007/b97633).
- 927 Ji XY, Zingg B, Mesik L, Xiao Z, Zhang LI, Tao HW. Thalamocortical innervation pattern in mouse auditory and
928 visual cortex: Laminar and cell-type specificity. *Cerebral Cortex* 2016;26(6):2612–2625. doi: [https://doi.org/](https://doi.org/10.1093/cercor/bhv099)
929 [10.1093/cercor/bhv099](https://doi.org/10.1093/cercor/bhv099).
- 930 Josselyn SA, Tonegawa S. Memory engrams: Recalling the past and imagining the future. *Science*
931 2020;367(6473). doi: <https://doi.org/10.1126/science.aaw4325>.
- 932 Ko H, Hofer SB, Pichler B, Buchanan KA, Sjöström PJ, Mrsic-Flogel TD. Functional specificity of local synaptic con-
933 nections in neocortical networks. *Nature* 2011;473(7345):87–91. doi: <https://doi.org/10.1038/nature09880>.
- 934 Kraynyukova N, Tchumatchenko T. Stabilized supralinear network can give rise to bistable, oscillatory, and
935 persistent activity. *Proceedings of the National Academy of Sciences of the United States of America*
936 2018;115(13):3464–3469. doi: <https://doi.org/10.1073/pnas.1700080115>.
- 937 Large AM, Vogler NW, Mielo S, Oswald AMM. Balanced feedforward inhibition and dominant recurrent inhibi-
938 tion in olfactory cortex. *Proceedings of the National Academy of Sciences of the United States of America*
939 2016;113(8):2276–2281. doi: <https://doi.org/10.1073/pnas.1519295113>.
- 940 Levina A, Herrmann JM, Geisel T. Dynamical synapses causing self-organized criticality in neural networks.
941 *Nature Physics* 2007;3(12):857–860. doi: <https://doi.org/10.1038/nphys758>.
- 942 Litwin-Kumar A, Doiron B. Slow dynamics and high variability in balanced cortical networks with clustered
943 connections. *Nature Neuroscience* 2012;15(11):1498–1505. doi: <https://doi.org/10.1038/nn.3220>.
- 944 Loebel A, Nelken I, Tsodyks M. Processing of sounds by population spikes in a model of primary auditory cortex.
945 *Frontiers in Neuroscience* 2007;1:15. doi: <https://doi.org/10.3389/neuro.01.1.1.015.2007>.
- 946 Loebel A, Tsodyks M. Computation by Ensemble Synchronization in Recurrent Networks with Synaptic Depres-
947 sion. *J Comput Neurosci* 2002 Sep;13(2):111–124. doi: <https://doi.org/10.1023/A:1020110223441>.
- 948 Markram H, Wang Y, Tsodyks M. Differential signaling via the same axon of neocortical pyramidal neurons.
949 *Proceedings of the National Academy of Sciences of the United States of America* 1998;95(9):5323–5328.
950 doi: <https://doi.org/10.1073/pnas.95.9.5323>.
- 951 Marshel JH, Kim YS, Machado TA, Quirin S, Benson B, Kadmon J, et al. Cortical layer-specific critical dynamics
952 triggering perception. *Science* 2019;365(6453). doi: <https://doi.org/10.1126/science.aaw5202>.
- 953 Mazor O, Laurent G. Transient dynamics versus fixed points in odor representations by locust antennal lobe
954 projection neurons. *Neuron* 2005;48(4):661–673. doi: <https://doi.org/10.1016/j.neuron.2005.09.032>.
- 955 Mazzucato L, La Camera G, Fontanini A. Expectation-induced modulation of metastable activity underlies
956 faster coding of sensory stimuli. *Nature Neuroscience* 2019;22(5):787–796. doi: [https://doi.org/10.1038/](https://doi.org/10.1038/s41593-019-0364-9)
957 [s41593-019-0364-9](https://doi.org/10.1038/s41593-019-0364-9).
- 958 Miller KD, Palmigiano A. Generalized paradoxical effects in excitatory/inhibitory networks. *bioRxiv* 2020; doi:
959 <https://doi.org/10.1101/2020.10.13.336727>.
- 960 Miska NJ, Richter LMA, Cary BA, Gjorgjieva J, Turrigiano GG. Sensory experience inversely regulates feedforward
961 and feedback excitation-inhibition ratio in rodent visual cortex. *eLife* 2018;7:1–27. doi: [https://doi.org/10.](https://doi.org/10.7554/eLife.38846)
962 [7554/eLife.38846](https://doi.org/10.7554/eLife.38846).
- 963 Mongillo G, Barak O, Tsodyks M. Synaptic theory of working memory. *Science* 2008;319(5869):1543–1546. doi:
964 <https://doi.org/10.1126/science.1150769>.

- 965 Mongillo G, Hansel D, van Vreeswijk C. Bistability and Spatiotemporal Irregularity in Neuronal Networks with
966 Nonlinear Synaptic Transmission. *Physical Review Letters* 2012 Apr;108(15):158101. doi: <https://doi.org/10.1103/PhysRevLett.108.158101>.
- 968 Murphy BK, Miller KD. Balanced amplification: A new mechanism of selective amplification of neural activity
969 patterns. *Neuron* 2009;61(4):635–648. doi: <https://doi.org/10.1016/j.neuron.2009.02.005>.
- 970 Niessing J, Friedrich RW. Olfactory pattern classification by discrete neuronal network states. *Nature*
971 2010;465(7294):47–52. doi: <https://doi.org/10.1038/nature08961>.
- 972 Okun M, Lampl I. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked
973 activities. *Nature Neuroscience* 2008;11(5):535–537. doi: <https://doi.org/10.1038/nn.2105>.
- 974 Ozeki H, Finn IM, Schaffer ES, Miller KD, Ferster D. Inhibitory Stabilization of the Cortical Network Underlies
975 Visual Surround Suppression. *Neuron* 2009;62(4):578–592. doi: <http://doi.org/10.1016/j.neuron.2009.03.028>.
- 976 Pala A, Petersen CCH. In vivo measurement of cell-type-specific synaptic connectivity and synaptic transmission
977 in layer 2/3 mouse barrel cortex. *Neuron* 2015;85(1):68–75. doi: <https://doi.org/10.1016/j.neuron.2014.11.025>.
- 978 Peron S, Pancholi R, Voelcker B, Wittenbach JD, Ólafsdóttir HF, Freeman J, et al. Recurrent interactions in local
979 cortical circuits. *Nature* 2020;579(7798):256–259. doi: <https://doi.org/10.1038/s41586-020-2062-x>.
- 980 Plenz D, Thiagarajan TC. The organizing principles of neuronal avalanches: cell assemblies in the cortex? *Trends*
981 *in Neurosciences* 2007 Mar;30(3):101–110. doi: <https://doi.org/10.1016/j.tins.2007.01.005>.
- 982 Ponce-Alvarez A, Thiele A, Albright TD, Stoner GR, Deco G. Stimulus-dependent variability and noise correlations
983 in cortical MT neurons. *Proceedings of the National Academy of Sciences of the United States of America*
984 2013;110(32):13162–13167. doi: <https://doi.org/10.1073/pnas.1300098110>.
- 985 Pozzorini C, Naud R, Mensi S, Gerstner W. Temporal whitening by power-law adaptation in neocortical neurons.
986 *Nature Neuroscience* 2013;16(7):942–948. doi: <https://doi.org/10.1038/nn.3431>.
- 987 Priebe NJ, Mechler F, Carandini M, Ferster D. The contribution of spike threshold to the dichotomy of cortical
988 simple and complex cells. *Nature Neuroscience* 2004;7(10):1113–1122. doi: <https://doi.org/10.1038/nn1310>.
- 989 Romo R, Brody CD, Hernández A, Lemus L. Neuronal correlates of parametric working memory in the prefrontal
990 cortex. *Nature* 1999;399(6735):470–473. doi: <https://doi.org/10.1038/20939>.
- 991 Rost T, Deger M, Nawrot MP. Winnerless competition in clustered balanced networks: Inhibitory assemblies
992 do the trick. *Biological Cybernetics* 2018;112(1):81–98. doi: <https://doi.org/10.1007/s00422-017-0737-7>.
- 993 Rubin DB, VanHooser SD, Miller KD. The stabilized supralinear network: A unifying circuit motif underlying
994 multi-input integration in sensory cortex. *Neuron* 2015;85(2):402–417. doi: <https://doi.org/10.1016/j.neuron.2014.12.026>.
- 996 Rubin R, Abbott LF, Sompolinsky H. Balanced excitation and inhibition are required for high-capacity, noise-
997 robust neuronal selectivity. *Proceedings of the National Academy of Sciences of the United States of America*
998 2017;114(44):E9366–E9375. doi: <https://doi.org/10.1073/pnas.1705841114>.
- 999 Rupprecht P, Friedrich RW. Precise synaptic balance in the zebrafish homolog of olfactory cortex. *Neuron*
1000 2018;100(3):669–683.e5. doi: <https://doi.org/10.1016/j.neuron.2018.09.013>.
- 1001 Sadeh S, Clopath C. Inhibitory stabilization and cortical computation. *Nature Reviews Neuroscience*
1002 2021;22(1):21–37. doi: <http://doi.org/10.1038/s41583-020-00390-z>.
- 1003 Sanzeni A, Akitake B, Goldbach HC, Leedy CE, Brunel N, Histed MH. Inhibition stabilization is a widespread
1004 property of cortical networks. *eLife* 2020;9:1–39. doi: <https://doi.org/10.7554/eLife.54875>.
- 1005 Schulz A, Miehl C, Berry MJ, Gjorgjieva J. The generation of cortical novelty responses through inhibitory plas-
1006 ticity. *eLife* 2021;10:1–28. doi: <https://doi.org/10.7554/elife.65309>.
- 1007 Shew WL, Clawson WP, Pobst J, Karimipanah Y, Wright NC, Wessel R. Adaptation to sensory input tunes visual
1008 cortex to criticality. *Nature Physics* 2015;11(8):659–663. doi: <https://doi.org/10.1038/nphys3370>.
- 1009 Stern M, Bolding KA, Abbott LF, Franks KM. A transformation from temporal to ensemble coding in a model of
1010 piriform cortex. *eLife* 2018;7:1–26. doi: <https://doi.org/10.7554/eLife.34831>.

- 1011 Stopfer M, Bhagavan S, Smith BH, Laurent G. Normalization for sparse encoding of odours by a wide-filed
1012 interneurons. *Nature* 1997;390(6655):70–74. doi: <https://doi.org/10.1126/science.1201835>.
- 1013 Thorpe S, Fize D, Marlot C. Speed of processing in the human visual system. *Nature* 1996;381(6582):520–522.
1014 doi: <https://doi.org/10.1038/381520a0>.
- 1015 Tsodyks MV, Markram H. The neural code between neocortical pyramidal neurons depends on neurotransmit-
1016 ter release probability. *Proceedings of the National Academy of Sciences of the United States of America*
1017 1997;94(2):719–723. doi: <https://doi.org/10.1073/pnas.94.2.719>.
- 1018 Tsodyks MV, Skaggs WE, Sejnowski TJ, McNaughton BL. Paradoxical effects of external modulation of in-
1019 hibitory interneurons. *Journal of Neuroscience* 1997;17(11):4382–4388. doi: <https://doi.org/10.1523/jneurosci.17-11-04382.1997>.
- 1020
- 1021 Van Vreeswijk C, Hansel D. Patterns of synchrony in neural networks with spike adaptation. *Neural Computa-*
1022 *tion* 2001;13(5):959–992. doi: <https://doi.org/10.1162/08997660151134280>.
- 1023 Van Vreeswijk C, Sompolinsky H. Chaos in neuronal networks with balanced excitatory and inhibitory activity.
1024 *Science* 1996;274(5293):1724–1726. doi: <https://doi.org/10.1126/science.274.5293.1724>.
- 1025 Varela JA, Sen K, Gibson J, Fost J, Abbott LF, Nelson SB. A quantitative description of short-term plasticity at
1026 excitatory synapses in layer 2/3 of rat primary visual cortex. *Journal of Neuroscience* 1997;17(20):7926–7940.
1027 doi: <https://doi.org/10.1523/jneurosci.17-20-07926.1997>.
- 1028 Vinje WE, Gallant JL. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*
1029 2000;287(5456):1273–1276. doi: <http://doi.org/10.1126/science.287.5456.1273>.
- 1030 Vogels TP, Sprekeler H, Zenke F, Clopath C, Gerstner W. Inhibitory plasticity balances excitation and inhibition
1031 in sensory pathways and memory networks. *Science* 2011;334(6062):1569–1573. doi: <https://doi.org/10.1126/science.1211095>.
- 1032
- 1033 Wehr M, Zador AM. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature*
1034 2003;426(6965):442–446. doi: <https://doi.org/10.1038/nature02116>.
- 1035 Wong KF, Wang XJ. A recurrent network mechanism of time integration in perceptual decisions. *Journal of*
1036 *Neuroscience* 2006;26(4):1314–1328. doi: <https://doi.org/10.1523/JNEUROSCI.3733-05.2006>.
- 1037 Wu YK, Hengen KB, Turrigiano GG, Gjorgjieva J. Homeostatic mechanisms regulate distinct aspects of cor-
1038 tical circuit dynamics. *Proceedings of the National Academy of Sciences of the United States of America*
1039 2020;117(39):24514–24525. doi: <https://doi.org/10.1073/pnas.1918368117>.
- 1040 Yakovlev V, Fusi S, Herman E, Zohary E. Inter-trial neuronal activity in inferior temporal cortex: A putative vehicle
1041 to generate long-term visual associations. *Nature Neuroscience* 1998;1(4):310–317. doi: <https://doi.org/10.1038/1131>.
- 1042
- 1043 Zenke F, Agnes EJ, Gerstner W. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve mem-
1044 ories in spiking neural networks. *Nature Communications* 2015;6(1):1–13. doi: <https://doi.org/10.1038/ncomms7922>.
- 1045
- 1046 Znamenskiy P, Kim MH, Muir DR, Iacaruso MF, Hofer SB, Mrsic-Flogel TD. Functional selectivity and specific
1047 connectivity of inhibitory neurons in primary visual cortex. *bioRxiv* 2018; doi: <https://doi.org/10.1101/294835>.
- 1048 Zucker RS, Regehr WG. Short-term synaptic plasticity. *Annual Review of Physiology* 2002;64:355–405. doi:
1049 <https://doi.org/10.1146/annurev.physiol.64.092501.114547>.

Stability conditions in networks with E-to-I STF

The dynamics of supralinear networks with E-to-I STF can be described as follows:

$$\tau_E \frac{dr_E}{dt} = -r_E + \left[J_{EE}r_E - J_{EI}r_I + g_E \right]_+^{\alpha_E} \quad (102)$$

$$\tau_I \frac{dr_I}{dt} = -r_I + \left[uJ_{IE}r_E - J_{II}r_I + g_I \right]_+^{\alpha_I} \quad (103)$$

$$\frac{du}{dt} = \frac{1-u}{\tau_u} + U_f(U_{max} - u)r_E \quad (104)$$

The Jacobian $\mathbf{M}_{STF}$ of the system with E-to-I STF is given by:

$$\mathbf{M}_{STF} = \begin{bmatrix} \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) & -\tau_E^{-1}J_{EI}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} & 0 \\ \tau_I^{-1}uJ_{IE}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}} & -\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) & \tau_I^{-1}J_{IE}r_E\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}} \\ U_f(U_{max} - u) & 0 & -\tau_u^{-1} - U_f r_E \end{bmatrix} \quad (105)$$

The characteristic polynomial for the system with E-to-I STF can be written as follows:

$$\lambda^3 - \text{tr}(\mathbf{M}_{STF})\lambda^2 + (A_{11} + A_{22} + A_{33})\lambda - \det(\mathbf{M}_{STF}) = 0 \quad (106)$$

where $\text{tr}(\mathbf{M}_{STF})$ and $\det(\mathbf{M}_{STF})$ are the trace and the determinant of the Jacobian matrix $\mathbf{M}_{STF}$, A_{11} , A_{22} , and A_{33} are the matrix cofactors. More specifically,

$$\begin{aligned} \text{tr}(\mathbf{M}_{STF}) &= \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) - \tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) - \tau_u^{-1} - U_f r_E \\ &\propto \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} / r_I^{\frac{\alpha_I-1}{\alpha_I}} - r_I^{\frac{1-\alpha_I}{\alpha_I}}) - \tau_I^{-1}(r_I^{\frac{1-\alpha_I}{\alpha_I}} + J_{II}\alpha_I) - \tau_u^{-1} r_I^{\frac{1-\alpha_I}{\alpha_I}} - U_f r_E r_I^{\frac{1-\alpha_I}{\alpha_I}} \end{aligned} \quad (107)$$

Assuming that $\alpha_E = \alpha_I = \alpha$, we then have

$$\text{tr}(\mathbf{M}_{STF}) \propto \tau_E^{-1} \left[J_{EE}\alpha \left(\frac{r_E}{r_I} \right)^{\frac{\alpha-1}{\alpha}} - r_I^{\frac{1-\alpha}{\alpha}} \right] - \tau_I^{-1} \left(r_I^{\frac{1-\alpha}{\alpha}} + J_{II}\alpha \right) - \tau_u^{-1} r_I^{\frac{1-\alpha}{\alpha}} - U_f r_E r_I^{\frac{1-\alpha}{\alpha}} \quad (108)$$

Substituting the firing rates with the current into excitatory population z , we then have

$$\begin{aligned} \text{tr}(\mathbf{M}_{STF}) &\propto \tau_E^{-1} \left[J_{EE}\alpha \left(\frac{z}{\det(\mathbf{J}_{STF}) \cdot J_{EI}^{-1}[z]_+^\alpha + J_{EI}^{-1}J_{II}z - J_{EI}^{-1}J_{II}g_E + g_I} \right)^{\alpha-1} - r_I^{\frac{1-\alpha}{\alpha}} \right] \\ &\quad - \tau_I^{-1} \left(r_I^{\frac{1-\alpha}{\alpha}} + J_{II}\alpha \right) - \tau_u^{-1} r_I^{\frac{1-\alpha}{\alpha}} - U_f r_E r_I^{\frac{1-\alpha}{\alpha}} \end{aligned} \quad (109)$$

where

$$\det(\mathbf{J}_{STF}) = \begin{vmatrix} J_{EE} & -J_{EI} \\ uJ_{IE} & -J_{II} \end{vmatrix} = -J_{EE}J_{II} + uJ_{IE}J_{EI} \quad (110)$$

In the large r_E limit, z is large, $\lim_{r_E \rightarrow \infty} u = \lim_{r_E \rightarrow \infty} \frac{1+U_f U_{max} r_E \tau_u}{1+U_f r_E \tau_u} \approx U_{max}$. Therefore, we can guarantee that $\det(\mathbf{J}_{STF})$ becomes positive for sufficiently large U_{max} . Since the denominator $\det(\mathbf{J}_{STF}) \cdot J_{EI}^{-1}[z]_+^\alpha + J_{EI}^{-1}J_{II}z - J_{EI}^{-1}J_{II}g_E + g_I$ grows faster than the numerator for $z \gg 1$, $\text{tr}(\mathbf{M}_{STF})$ becomes negative for large r_E .

$$\begin{aligned} A_{11} + A_{22} + A_{33} &= \tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}})(\tau_u^{-1} + U_f r_E) \\ &\quad + \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1)(-\tau_u^{-1} - U_f r_E) \\ &\quad - \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1)\tau_I^{-1}(1 + J_{II}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}}) + \tau_E^{-1}J_{EI}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} \tau_I^{-1}uJ_{IE}\alpha_I r_I^{\frac{\alpha_I-1}{\alpha_I}} \end{aligned} \quad (111)$$

1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093

Similarly, in the large r_E limit, $A_{11} + A_{22} + A_{33}$ is positive.

$$\begin{aligned} \det(\mathbf{M}_{\text{STF}}) = & \tau_E^{-1} (J_{EE} \alpha_E r_E^{\frac{a_E-1}{a_E}} - 1) \tau_I^{-1} (1 + J_{II} \alpha_I r_I^{\frac{a_I-1}{a_I}}) (\tau_u^{-1} + U_f r_E) \\ & + \tau_E^{-1} J_{EI} \alpha_E r_E^{\frac{a_E-1}{a_E}} (\tau_I^{-1} u J_{IE} \alpha_I r_I^{\frac{a_I-1}{a_I}} (-\tau_u^{-1} - U_f r_E) - \tau_I^{-1} J_{IE} r_E \alpha_I r_I^{\frac{a_I-1}{a_I}} U_f (U_{\max} - 1)) \end{aligned} \quad (12)$$

Similarly, in the large r_E limit, $\det(\mathbf{M}_{\text{STF}})$ is negative.

Therefore, similar to E-to-E STD, networks dynamics can also be stabilized by E-to-I STF.

Appendix 2

Conditions for ISN in networks with E-to-I STF

Here, we identify the condition of being ISN in supralinear networks with E-to-I STF. If inhibition is frozen, in other words, if feedback inhibition is absent, the Jacobian of the system becomes as follows:

$$\mathbf{M}_2 = \begin{bmatrix} \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) & 0 \\ U_f(U_{max} - u) & -\tau_u^{-1} - U_f r_E \end{bmatrix} \quad (113)$$

For the system with frozen inhibition, the dynamics are stable if

$$\text{tr}(\mathbf{M}_2) = \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) - \tau_u^{-1} - U_f r_E < 0 \quad (114)$$

and

$$\det(\mathbf{M}_2) = \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1)(-\tau_u^{-1} - U_f r_E) > 0 \quad (115)$$

Therefore, if the network is an ISN at the fixed point, the following condition has to be satisfied:

$$\tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) > 0 \quad (116)$$

Note that this condition is independent of the facilitation variable u of E-to-I STF. We further define the ISN index for the system with E-to-I STF as follows:

$$\text{ISN index} = \tau_E^{-1}(J_{EE}\alpha_E r_E^{\frac{\alpha_E-1}{\alpha_E}} - 1) \quad (117)$$

1121 **Conditions for paradoxical response in networks with E-to-I STF**

1122 Next, we identify the condition of having the paradoxical effect in supralinear networks with
1123 E-to-I STF. The excitatory nullcline is defined by

$$1124 \tau_E \frac{dr_E}{dt} = -r_E + \left[J_{EE}r_E - J_{EI}r_I + g_E \right]_+^{\alpha_E} = 0 \quad (118)$$

1125 For $r_{E,I} > 0$, we have

$$1126 r_I = \frac{J_{EE}r_E - r_E^{\frac{1}{\alpha_E}} + g_E}{J_{EI}} \quad (119)$$

1127 The slope of the excitatory nullcline in the r_E/r_I plane where x axis is r_E and y axis is r_I can
1128 be written as follows

$$1129 k_{STF}^E = \frac{J_{EE} - \frac{1}{\alpha_E} r_E^{\frac{1}{\alpha_E}-1}}{J_{EI}} \quad (120)$$

1130 Note that the slope of the excitatory nullcline is nonlinear. To have paradoxical effect, the
1131 slope of the excitatory nullcline at the fixed point of the system has to be positive. We
1132 therefore have

$$1133 J_{EE}\alpha_E r_E^{\alpha_E-1} - 1 > 0 \quad (121)$$

1134 We exploit a separation of timescales between fast neural activity and slow short-term plas-
1135 ticity variable, we therefore set the facilitation variable to the value at its fixed point corre-
1136 sponding to the dynamical value of r_E . Then we can write the inhibitory nullcline as follows

$$1137 \tau_I \frac{dr_I}{dt} = -r_I + \left[\frac{1 + U_f U_{max} r_E \tau_u}{1 + U_f r_E \tau_u} J_{IE} r_E - J_{II} r_I + g_I \right]_+^{\alpha_I} = 0 \quad (122)$$

1138 In the region of rates $r_{E,I} > 0$, we have

$$1139 r_I = \frac{\frac{1 + U_f U_{max} r_E \tau_u}{1 + U_f r_E \tau_u} J_{IE} r_E - r_I^{\frac{1}{\alpha_I}} + g_I}{J_{II}} \quad (123)$$

1140 The slope of the inhibitory nullcline can be written as follows

$$1141 k_{STF}^I = \frac{\frac{1 + U_f U_{max} r_E \tau_u}{1 + U_f r_E \tau_u} J_{IE} + \frac{U_f U_{max} \tau_u - U_f \tau_u}{(1 + U_f r_E \tau_u)^2} J_{IE} r_E}{J_{II} + \frac{1}{\alpha_I} r_I^{\frac{1}{\alpha_I}-1}} \quad (124)$$

1142 In addition to the positive slope of the excitatory nullcline, the slope of the inhibitory null-
1143 cline at the fixed point of the system has to be larger than the slope of the excitatory nullcline.
1144 We therefore have

$$1145 \begin{aligned} 1146 & -(J_{EE}\alpha_E r_E^{\alpha_E-1} - 1)(1 + J_{II}\alpha_I r_I^{\alpha_I-1}) + J_{IE}\alpha_E r_E^{\alpha_E-1} \frac{1 + U_f U_{max} r_E \tau_u}{1 + U_f r_E \tau_u} J_{EI}\alpha_I r_I^{\alpha_I-1} \\ 1147 & + J_{IE}\alpha_E r_E^{\alpha_E-1} \frac{U_f U_{max} \tau_u - U_f \tau_u}{(1 + U_f r_E \tau_u)^2} J_{EI}\alpha_I r_I^{\alpha_I-1} r_E > 0 \end{aligned} \quad (125)$$

1148 The above condition is the same as the stability condition of the determinant of the
1149 Jacobian of the system with E-to-I STF (Eq. (112)). Therefore, the condition is always satisfied
1150 when the system with E-to-I STF is stable.

1165 Note that the condition of being ISN shown in Eq. (116) is identical to the condition of
1166 having paradoxical effect shown in Eq. (121). Therefore, in networks with E-to-I STF alone,
1167 paradoxical effect implies ISN and ISN implies paradoxical effect. We thus use paradoxical
1168 effect as a proxy for inhibitory stabilization.

Change in steady-state activity of unstimulated co-tuned neurons

To analyze the pattern completion in supralinear networks, we considered a network with one excitatory population and one inhibitory population. Neurons in the excitatory population are co-tuned to the same stimulus feature and are separated into two subsets denoting by E_{11} and E_{12} . The dynamics of the system can be described as follows:

$$\tau_E \frac{dr_{E_{11}}}{dt} = -r_{E_{11}} + \left[J_{E_{11}E_{11}} r_{E_{11}} + J_{E_{11}E_{12}} r_{E_{12}} - J_{E_{11}I} r_I + g_{E_{11}} \right]_+^{\alpha_E} \quad (126)$$

$$\tau_E \frac{dr_{E_{12}}}{dt} = -r_{E_{12}} + \left[J_{E_{12}E_{11}} r_{E_{11}} + J_{E_{12}E_{12}} r_{E_{12}} - J_{E_{12}I} r_I + g_{E_{12}} \right]_+^{\alpha_E} \quad (127)$$

$$\tau_I \frac{dr_I}{dt} = -r_I + \left[J_{IE_{11}} r_{E_{11}} + J_{IE_{12}} r_{E_{12}} - J_{II} r_I + g_I \right]_+^{\alpha_I} \quad (128)$$

The change in the firing rate of the Subset 2 in the excitatory population $\delta r_{E_{12}}$ can be written as a function of the change in the input to the Subset 1 $\delta g_{E_{11}}$:

$$\begin{aligned} \delta r_{E_{12}} &= \frac{1}{\det(\mathbb{1} - \mathbf{FJ})} [-f'_{E_{12}} J_{E_{12}I} f'_I J_{IE_{11}} - (-f'_{E_{12}} J_{E_{12}E_{11}})(1 + f'_I J_{II})] f'_{E_{11}} \delta g_{E_{11}} \\ &= \frac{1}{\det(\mathbb{1} - \mathbf{FJ})} [J_{E_{12}E_{11}} + J_{E_{12}E_{11}} J_{II} f'_I - J_{E_{12}I} J_{IE_{11}} f'_I] f'_{E_{11}} \delta g_{E_{11}} \end{aligned} \quad (129)$$

where $\mathbb{1}$ is the identity matrix. And $\mathbf{F}$ is given by

$$\mathbf{F} = \begin{bmatrix} f'_{E_{11}} & 0 & 0 \\ 0 & f'_{E_{12}} & 0 \\ 0 & 0 & f'_I \end{bmatrix} \quad (130)$$

where $f'_{E_{11}}$, $f'_{E_{12}}$, and f'_I are the derivatives of the input-output functions evaluated at the fixed point. The term $\det(\mathbb{1} - \mathbf{FJ})$ is positive to ensure the stability of the system.

Clearly, if the term $J_{E_{12}E_{11}} + J_{E_{12}E_{11}} J_{II} f'_I - J_{E_{12}I} J_{IE_{11}} f'_I$ is positive (negative), increasing the input to the Subset 1 leads to an increase (a decrease) in the activity of neurons in the Subset 2. As the input to the Subset 1 increases, the firing rate of the inhibitory population r_I and also f'_I will increase. In the presence of E-to-E STD or E-to-I STF, $J_{E_{12}E_{11}}$ or $J_{IE_{11}}$ will decrease or increase with the input to the Subset 1. As a result, the sign of $J_{E_{12}E_{11}} + J_{E_{12}E_{11}} J_{II} f'_I - J_{E_{12}I} J_{IE_{11}} f'_I$ can switch from positive to negative as the input to the Subset 1 increases, indicating that the effect on the activity of the co-tuned unstimulated neurons in the same ensemble can switch from potentiation to suppression. Note that this behavior is different from linear networks in which the change is independent of the input or firing rates.

Appendix 5 Table 1. Parameters for Figure 1–Figure Supplement 1

Symbol	Value	Unit	Description
J_{EE}	0.5	-	E-to-E connection strength
J_{IE}	0.45	-	E-to-I connection strength
J_{EI}	1.0	-	I-to-E connection strength
J_{II}	1.5	-	I-to-I connection strength
g_E^{bs}	0.5	-	input to the E population at baseline
g_I^{bs}	1.5	-	input to the I population at baseline

Appendix 5 Table 2. Parameters for Figure 2–Figure Supplement 2C-D

Symbol	Value	Unit	Description
g_E^{stim}	2.0	-	input to the E population during stimulation
Note that values of the unlisted parameters are the same as Tables 1–2.			

Appendix 5 Table 3. Parameters for Figure Figure 2–Figure Supplement 3 SSN example

Symbol	Value	Unit	Description
J_{EE}	1.8	-	E-to-E connection strength
J_{IE}	2.0	-	E-to-I connection strength
J_{EI}	1.0	-	I-to-E connection strength
J_{II}	1.0	-	I-to-I connection strength

Appendix 5 Table 4. Parameters for Figure 2–Figure Supplement 5

Symbol	Value	Unit	Description
g_E^{bs}	1.8	-	input to the E population at baseline
Note that values of the unlisted parameters are the same as Tables 1–2.			

Appendix 5 Table 5. Parameters for Figure 2–Figure Supplement 10

Symbol	Value	Unit	Description
N_E	400	-	number of excitatory neurons
N_I	100	-	number of inhibitory neurons
J_{EE}	0.05	-	E-to-E connection strength
J_{IE}	0.02	-	E-to-I connection strength
J_{EI}	0.05	-	I-to-E connection strength
J_{II}	0.03	-	I-to-I connection strength

Appendix 5 Table 6. Parameters for Figure 2–Figure Supplement 11

Symbol	Value	Unit	Description
J_{EE}	0.9	-	E-to-E connection strength
J_{IE}	1.2	-	E-to-I connection strength
J_{EI}	0.5	-	I-to-E connection strength
J_{II}	0.5	-	I-to-I connection strength
τ_E	20	ms	time constant of excitatory firing dynamics
τ_I	60	ms	time constant of inhibitory firing dynamics
g_E^{bs}	1.0	-	input to the E population at baseline
g_E^{stim}	2.0	-	input to the E population during stimulation
g_I	2.0	-	input to the I population
Note that values of the unlisted parameters are the same as Tables 1–2.			

Appendix 5 Table 7. Parameters for Figure 2–Figure Supplement 12

1219

Symbol	Value	Unit	Description
J_{EE}	1.0	-	E-to-E connection strength
J_{IE}	1.2	-	E-to-I connection strength
J_{EI}	0.5	-	I-to-E connection strength
J_{II}	1.0	-	I-to-I connection strength
g_E^{bs}	0.5	-	input to the E population at baseline
g_I^{bs}	1.0	-	input to the I population at baseline

1220

Appendix 5 Table 8. Parameters for *Figure 3–Figure Supplement 1* global inhibition example

Symbol	Value	Unit	Description
J_{EE}	1.6	-	within-ensemble E-to-E connection strength
J_{IE}	1.0	-	within-ensemble E-to-I connection strength
J_{EI}	1.0	-	within-ensemble I-to-E connection strength
J_{II}	1.2	-	within-ensemble I-to-I connection strength
J_{EE}^I	0.16	-	inter-ensemble E-to-E connection strength
J_{IE}^I	1.0	-	inter-ensemble E-to-I connection strength
J_{EI}^I	1.0	-	inter-ensemble I-to-E connection strength
J_{II}^I	1.2	-	inter-ensemble I-to-I connection strength
g_{E1}^{bs}	1.5	-	input to the $E1$ population at baseline
g_{E2}	1.5	-	input to the $E2$ population
g_{I1}	2.5	-	input to the $I1$ population
g_{I2}	2.5	-	input to the $I2$ population

1222

Parameters for *Figure 3–Figure Supplement 1* co-tuned example

J_{IE}	$1.0 * (4/3)$	-	within-ensemble E-to-I connection strength
J_{EI}	$1.0 * (4/3)$	-	within-ensemble I-to-E connection strength
J_{II}	$1.2 * (4/3)$	-	within-ensemble I-to-I connection strength
J_{IE}^I	$1.0 * (2/3)$	-	inter-ensemble E-to-I connection strength
J_{EI}^I	$1.0 * (2/3)$	-	inter-ensemble I-to-E connection strength
J_{II}^I	$1.2 * (2/3)$	-	inter-ensemble I-to-I connection strength

1223

Appendix 5 Table 9. Parameters for *Figure 4–Figure Supplement 1*

Symbol	Value	Unit	Description
J_{EE}	$1.5/(N_E/2 - 1)$	-	within-ensemble E-to-E connection strength
J_{IE}	$1.0/(N_E/2)$	-	within-ensemble E-to-I connection strength
J_{EI}	$1.0/(N_I/2)$	-	within-ensemble I-to-E connection strength
J_{II}	$1.0/(N_I/2 - 1)$	-	within-ensemble I-to-I connection strength
J_{EE}^I	$0.1/(N_E/2 - 1)$	-	inter-ensemble E-to-E connection strength
J_{IE}^I	$0.3/(N_E/2)$	-	inter-ensemble E-to-I connection strength
J_{EI}^I	$0.3/(N_I/2)$	-	inter-ensemble I-to-E connection strength
J_{II}^I	$0.1/(N_I/2)$	-	inter-ensemble I-to-I connection strength
g_{E1}^{bs}	1.5	-	input to the $E1$ population at baseline
g_{E2}	1.5	-	input to the $E2$ population
g_I	2.0	-	input to the I population

1225

1227

Appendix 5 Table 10. Parameters for *Figure 6–Figure Supplement 1*

Symbol	Value	Unit	Description
J_{EE}	0.20	-	within-ensemble E-to-E connection strength
J_{IE}	0.09	-	within-ensemble E-to-I connection strength
J_{EI}	0.10	-	within-ensemble I-to-E connection strength
J_{II}	0.10	-	within-ensemble I-to-I connection strength
J_{EE}^I	0.02	-	inter-ensemble E-to-E connection strength
J_{IE}^I	0.054	-	inter-ensemble E-to-I connection strength
J_{EI}^I	0.07	-	inter-ensemble I-to-E connection strength
J_{II}^I	0.01	-	inter-ensemble I-to-I connection strength

1228

Note that values of the unlisted parameters are the same as Table 5.

Figure 1-Figure supplement 1. (A) Characteristic function $F(z)$ for different inputs g_E and g_I . (B) Firing rates of the excitatory (blue) and inhibitory population (red) in response to stimulation from 2–4 s (gray bar). During stimulation g_E and g_I are simultaneously changed to the stated values.

Figure 2-Figure supplement 1. (A) Firing rates of the excitatory (blue) and inhibitory population (red) in the presence of strong SFA (left) with $b = 200$. The zoomed-in activity from 3.0 s to 3.2 s (right) corresponding to the green period (left) indicates oscillatory behavior in networks with strong SFA. (B) 3D plot of the excitatory activity r_E , inhibitory activity r_I , and the SFA variable a of the network in the period marked with the green bar in A. Networks with different conditions starting from different gray dots converge to the same stable limit cycle marked with the black line.

Figure 2-Figure supplement 2. (A) The onset peak amplitude (triangles) and fixed point activity (circles) as a function of input g_E in networks with E-to-E STD. Here, $J_{EE} = 1.8$. (B) The ratio of the onset peak amplitude to fixed point activity as a function of input g_E . The vertical dashed line indicates the input level which leads to unstable network dynamics without E-to-E STD. (C and D) Same as A and B but as a function of J_{EE} . Here, $g_E = 2.0$.

Figure 2-Figure supplement 3. (A) Amplification index, which is defined as the ratio of the peak amplitude to input g_E , as a function of input g_E for NTA (blue) and linear network (orange) in the presence of E-to-E STD. **(B)** Same as A but for SSN without STP (green). SSN is constructed by keeping J_{EE} the same while changing J_{IE} and J_{II} such that the determinant of the weight matrix $\det(\mathbf{J})$ becomes positive. **(C)** Example of firing rates of the excitatory (blue) and inhibitory population (red) in linear networks with E-to-E STD. During stimulation (gray bar) additional input is injected into the excitatory population. Same as **Figure 2B** (left) but with $\alpha_E = \alpha_I = 1$ and $g_E = 2$. **(D)** Example of firing rates of the excitatory (blue) and inhibitory population (red) in SSNs with $g_E = 2$.

Figure 2-Figure supplement 4. (A) Firing rates of the excitatory (blue) and inhibitory population (red) in response to stimulation in the presence of E-to-I STF with different time constants τ_u . The stimulation period from 2 s to 4 s is marked with the gray bar. The stimulation is implemented by changing input g_E . **(B)** Peak amplitude (solid line) and fixed point activity (dashed line) of the excitatory population during stimulation with different STF time constants. **(C and D)** Same as A and B, but with different maximum allowed facilitation levels U_{max} .

Figure 2-Figure supplement 5. (A) Firing rates of the excitatory (blue) and inhibitory population (red) in the presence of E-to-E STD in an example of a network initially in the ISN regime. During stimulation (gray bar) additional input is injected into the excitatory population. **(B)** ISN index of the network. **(C)** Depression variable (black) in an example of a network shown in A. The orange dashed line represents the depression variable threshold above which the network exhibits the paradoxical effect. **(D)** The normalized firing rates of the excitatory (blue) and inhibitory population (red) when injecting additional excitatory current into the inhibitory population before stimulation (left; cf. orange bar in A), and during stimulation (right; cf. green bar in A). Both before stimulation and during stimulation, the ensemble is in the ISN regime and injecting excitatory current into inhibitory population decreases its firing rate. **(E)** Firing rates of the excitatory (blue) and inhibitory population (red) when a small transient perturbation to excitatory population activity is introduced marked with arrows while freezing inhibition before stimulation (left) and during stimulation (right). The periods in which inhibition is frozen are marked with black bars. Both before stimulation and during stimulation, the ensemble is in the ISN regime and a small transient excitatory perturbation to the excitatory population results in a transient explosion of the excitatory firing rate.

1235

Figure 2-Figure supplement 6. (A) ISN index of the network with E-to-I STF in **Figure 2E** (left). The horizontal dashed line indicates an ISN index of 0. **(B)** The normalized firing rates of the excitatory (blue) and inhibitory population (red) when injecting additional excitatory current into the inhibitory population before stimulation (left, cf. **Figure 2E**), and during stimulation (right, cf. **Figure 2E**). The horizontal dashed lines indicate a normalized firing rate of 1.0.

1236

Figure 2-Figure supplement 7. (A) ISN index of the network with E-to-E STD as a function of input g_E . The horizontal dashed line indicates an ISN index of 0. For values above zero, the ensemble is an ISN. The vertical dashed line separates the non-paradoxical and paradoxical regions. **(B)** Firing rates of the excitatory (blue) and inhibitory population (red) in an ensemble with the input level marked with the black dot in A when a small transient perturbation to excitatory population activity is introduced marked with arrows while freezing inhibition. The periods in which inhibition is frozen are marked with the black bar. A small transient perturbation to the excitatory population results in a transient explosion of the excitatory firing rate, indicating that the ensemble is an ISN. **(C)** The normalized firing rates of the excitatory (blue) and inhibitory population (red) when injecting excitatory input into the inhibitory population of an ensemble with the input level marked with the black dot in A. The horizontal dashed line indicates a normalized firing rate of 1.0. Despite being an ISN, the ensemble does not exhibit the paradoxical effect.

Figure 2-Figure supplement 8. (A) Depression variable (black) in an example of a network initially in the non-ISN regime shown in **Figure 2B** left. The orange dashed line represents the depression variable threshold above which the network exhibits the paradoxical effect. **(B)** Firing rates of the excitatory (blue) and inhibitory population (red) when a small transient perturbation to excitatory population activity is introduced marked with arrows while freezing inhibition before stimulation (left; cf. **Figure 2B** left) and during stimulation (right; cf. **Figure 2B** left). The periods in which inhibition is frozen are marked with black bars. Initially, the ensemble is in the non-ISN regime and the firing rate of the excitatory slightly increases and then returns to its baseline after introducing a small transient perturbation to the excitatory population. During stimulation, however, the ensemble is an ISN. In this case, a small transient perturbation to the excitatory population results in a transient explosion of the excitatory firing rate.

Figure 2-Figure supplement 9. (A) Firing rates of the excitatory (blue) and inhibitory population (red) in response to stimulation in a rate-based model. Additional excitatory inputs are injected into excitatory neurons in the period marked in gray. Firing rates are capped at 300 Hz. **(B)** Same as A but incorporating SFA. **(C)** Same as A but incorporating E-to-E STD. **(D)** Same as A but incorporating E-to-I STF. **(E)** Same as B but incorporating stronger SFA (left). The zoomed-in activity from 3.0 s to 3.2 s (right) corresponding to the green period (left) indicates oscillatory behavior in networks with strong SFA.

Figure 2-Figure supplement 10. (A) Spike raster plot of the excitatory (top, blue) and inhibitory population (top, red) in response to stimulation in a spiking neural network model, and firing rates calculated with 10 ms time bins (bottom). Additional excitatory inputs are injected into excitatory neurons in the period marked in gray. (B) Same as A but incorporating SFA. (C) Same as A but incorporating E-to-E STD. (D) Same as A but incorporating E-to-I STF.

Figure 2-Figure supplement 11. (A) Firing rates of the excitatory (blue) and inhibitory population (red) in networks with positive determinant in response to external stimulation during the interval from 2–4 s (gray bar). The stimulation was implemented by temporarily increasing the input g_E . (B) Phase portrait of the system before stimulation (left; cf. A orange) and during stimulation (right; cf. A green). (C) Same as (A) but in the presence of weak SFA (left) and strong SFA (right). (D) Same as (A) but in the presence of E-to-E STD. (E) Same as (A) but in the presence of E-to-I STF.

1241

Figure 2-Figure supplement 12. (A) Example of activity when the change in the input of the inhibitory population Δg_I is larger than the the change in the input of the excitatory population Δg_E . Here, $\Delta g_E = 1.8$, and $\Delta g_I = 3$. (B) The ratio of the onset peak amplitude to the baseline activity as a function of the change in the input of the excitatory population Δg_E and the change in the input of the inhibitory population Δg_I . The diagonal dashed line indicates the border at which Δg_E is identical to Δg_I in networks with E-to-E STD. The star symbol represents the example shown in A.

1242

Figure 3-Figure supplement 1. (A) Examples of activity in networks with global inhibition (left) and in networks with co-tuned inhibition (right) in the presence of E-to-E STD. (B) Amplification index for networks with global inhibition (black) and network with co-tuned inhibition (gray) as a function of input g_E in networks with E-to-E STD. Data points at $g_E = 2.5$ correspond to the examples shown in A.

1243

Figure 4-Figure supplement 1. (A) Examples of firing rates of Subset 2 of E1 (blue) and E2 (green) for $g_E = 3$ (left) and for $g_E = 4$ (right) in networks with E-to-E STD. (B) The ratio of the fixed point activity to the baseline activity for unstimulated co-tuned neurons (E1 Subset 2) as a function of input g_E in rate-based models with E-to-E STD. The horizontal dashed line indicates a ratio of 1 at which the fixed point activity is identical to the baseline activity.

Figure 4-Figure supplement 2. (A) Association index as a function of input g_{E1} for the onset peak amplitude (magenta solid line) and fixed point activity (gray dashed line) for E-to-I STF. (B) Distance to the decision boundary as a function of input g_{E1} for the onset peak amplitude (magenta solid line) and fixed point activity (gray dashed line) for E-to-I STF. (C and D) Same as A and B but as a function of β , which controls the inner- and inter-ensemble connection strength.

Figure 5-Figure supplement 1. (A) Separation index, defined as $(r_{E2} - r_{E1}) / (r_{E1} + r_{E2})$, as a function of p for the peak onset response (magenta solid line) and fixed point activity (gray dashed line) for E-to-E STD. Different levels of brightness represent different recurrent E-to-E connection strengths J_{EE} . (B) Same as A but with different E-to-I connection strengths J'_{IE} across ensembles.

Figure 5-Figure supplement 2. (A) Distance to the decision boundary as a function of p for the peak onset response (magenta solid line) and fixed point activity (gray dashed line) for E-to-I STF. (B) Same as A but with different E-to-I connection strengths J'_{IE} across ensembles. (C and D) Same as A and B but quantified by Separation index, defined as $(r_{E2} - r_{E1}) / (r_{E1} + r_{E2})$.

Figure 6–Figure supplement 1. The ratio of the fixed point activity to the baseline activity for unstimulated co-tuned neurons (Ensemble 5, Subset 2) as a function of feedforward input to Subset 1 in spiking neural networks with E-to-E STD. The horizontal dashed line indicates a ratio of one at which the fixed point activity is identical to the baseline activity. Each point corresponds to a measurement from a randomly initialized network.