

# Inhibition stabilization and paradoxical effects in recurrent neural networks with short-term plasticity

Yue Kris Wu \* and Julijana Gjorgjieva 

*School of Life Sciences, Technical University of Munich, 85354 Freising, Germany  
and Max Planck Institute for Brain Research, 60438 Frankfurt, Germany*



(Received 22 December 2022; accepted 5 June 2023; published 12 July 2023)

*This article is part of the Physical Review Research collection titled [Physics of Neuroscience](#).*

Inhibition stabilization is considered a ubiquitous property of cortical networks, whereby inhibition controls network activity in the presence of strong recurrent excitation. In networks with fixed connectivity, an identifying characteristic of inhibition stabilization is that increasing (decreasing) excitatory input to the inhibitory population leads to a decrease (increase) in inhibitory firing, known as the paradoxical effect. However, population responses to stimulation are highly nonlinear, and drastic changes in synaptic strengths induced by short-term plasticity (STP) can occur on the timescale of perception. How neuronal nonlinearities and STP affect inhibition stabilization and the paradoxical effect is unclear. Using analytical calculations, we demonstrate that in networks with STP the paradoxical effect implies inhibition stabilization, but inhibition stabilization does not imply the paradoxical effect. Interestingly, networks with neuronal nonlinearities and STP can transition nonmonotonically between inhibition-stabilization and noninhibition-stabilization, and between paradoxically- and nonparadoxically-responding regimes with increasing excitatory activity. Furthermore, we generalize our results to more complex scenarios including networks with multiple interneuron subtypes and any monotonically increasing neuronal nonlinearities. In summary, our work reveals the relationship between inhibition stabilization and the paradoxical effect in the presence of neuronal nonlinearity and STP, yielding several testable predictions.

DOI: [10.1103/PhysRevResearch.5.033023](https://doi.org/10.1103/PhysRevResearch.5.033023)

## I. INTRODUCTION

Cortical networks are typically characterized by inhibition stabilization, where inhibition is needed to keep network activity levels in biologically realistic ranges despite the presence of strong recurrent excitation [1]. Networks operating in the inhibition-stabilized regime are capable of performing various computations, including input amplification, response normalization, and network multistability [2–6]. In networks with fixed connectivity, a hallmark of inhibition stabilization is the paradoxical effect: An increase or a decrease of excitatory input to the inhibitory population respectively decreases or increases the inhibitory firing [7]. Over the past decade, much effort has been made to identify the operating regime of cortical networks based on the paradoxical effect [1,8,9].

Yet, various aspects ranging from the network to the synaptic level can considerably affect network dynamics and the operating regime. First, if individual neurons in the network receive large excitatory and inhibitory currents which precisely cancel each other, the network operates in a balanced

state characterized by a linear population response [10–12]. Recent work has argued that neuronal input-output functions are better characterized by supralinear functions, and networks with this type of nonlinearity can exhibit various nonlinear phenomena as observed in biology [13–15]. Second, synapses in the brain are highly dynamic as a result of different STP mechanisms, operating on a timescale of milliseconds to seconds [16,17]. Upon presynaptic stimulation, postsynaptic responses can either get depressed subject to short-term depression (STD) or facilitated subject to short-term facilitation (STF). While short-term synaptic dynamics are widely observed in biological circuits, it is unclear how they interact with the neuronal nonlinearity to jointly determine the network operating regime. Here we ask how the neuronal nonlinearity and STP affect inhibition stabilization and the paradoxical effect.

To address this question, we determine the conditions for inhibition stabilization and the paradoxical effect in networks of excitatory and inhibitory neurons in the presence of STP with linear and supralinear population response functions. We find that, irrespective of the neuronal nonlinearity, in networks with excitatory-to-excitatory (E-to-E) STD, inhibition stabilization does not necessarily imply the paradoxical effect, but the paradoxical effect implies inhibition stabilization. In contrast, in networks with static connectivity or networks with other STP mechanisms, inhibition stabilization and the paradoxical effect imply each other. Interestingly, neuronal nonlinearities and STP endow the network with unconventional behaviors. More specifically, in the presence

\*kris.wu@tum.de

*Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](#) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.*

of a neuronal nonlinearity and E-to-E STD, monotonically increasing excitatory activity can lead to nonmonotonic transitions between noninhibition-stabilization and inhibition-stabilization, as well as between nonparadoxically-responding and paradoxically-responding regimes. Furthermore, we generalize our results to more complex scenarios including networks with multiple interneuron subtypes and any monotonically increasing neuronal nonlinearities. In conclusion, our work reveals the impact of neuronal nonlinearities and STP on inhibition stabilization and the paradoxical effect, and makes several predictions for future experiments.

## II. RESULTS

To understand the relationship between inhibition stabilization and the paradoxical effect in recurrent neural networks with STP, we studied rate-based population models consisting of an excitatory (E) and an inhibitory (I) population. The dynamics of the network are given by

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

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

where  $r_E$  and  $r_I$  denote the firing rates of the excitatory and inhibitory population;  $\tau_E$  and  $\tau_I$  are the corresponding time constants;  $J_{AB}$  represents the synaptic strength from population  $B$  to population  $A$ , where  $A, B \in \{E, I\}$ ;  $g_E$  and  $g_I$  are the external inputs to the respective populations; and  $\alpha_E$  and  $\alpha_I$  are the exponents of the respective input-output functions. Finally,  $p_{AB}$  represents the short-term plasticity variable from population  $B$  to population  $A$ . We implemented short-term plasticity mechanisms based on the Tsodyks and Markram model [16]. For STD, we replaced  $p_{AB}$  with  $x_{AB}$  and described the STD dynamics as follows:

$$\frac{dx_{AB}}{dt} = \frac{1 - x_{AB}}{\tau_x} - U_d x_{AB} r_B, \quad (3)$$

where  $x_{AB}$  is a short-term depression variable that is limited to the interval  $(0, 1]$  for the synaptic connection from population  $B$  to population  $A$ . Biophysically, the short-term depression variable  $x$  represents the fraction of vesicles available for release,  $\tau_x$  is the time constant of STD, and  $U_d$  is the depression factor controlling the degree of depression induced by the presynaptic activity.

For STF, we replaced  $p_{AB}$  by  $u_{AB}$  and expressed the STF dynamics as follows:

$$\frac{du_{AB}}{dt} = \frac{1 - u_{AB}}{\tau_u} + U_f (U_{\max} - u_{AB}) r_B, \quad (4)$$

where  $u_{AB}$  is a short-term facilitation variable that is constrained to the interval  $[1, U_{\max})$  for the synaptic connection from population  $B$  to population  $A$ . Biophysically, the short-term facilitation variable  $u$  represents the ability of releasing neurotransmitter,  $\tau_u$  is the time constant of STF,  $U_f$  is the facilitation factor controlling the degree of facilitation induced by the presynaptic activity, and  $U_{\max}$  is the maximal facilitation value.

To investigate the impact of neuronal nonlinearities on inhibition stabilization and the paradoxical effect, we

considered both threshold-linear networks ( $\alpha_E = \alpha_I = 1$ ) as well as supralinear networks ( $\alpha_E = \alpha_I > 1$ ). In the regime of positive  $r_E$  and  $r_I$ , threshold-linear networks behave as linear networks. In the following, we thus call them linear networks. Furthermore, while we keep our analysis for supralinear networks in a general form, we use  $\alpha_E = \alpha_I = 2$  for the numerical simulations. Note that the neuronal nonlinearity in our study refers to the nonlinearity of population-averaged responses to input when no STP mechanisms are taken into account, which is fully determined by  $\alpha_E$  and  $\alpha_I$ .

In addition, for the sake of analytical tractability, we included one STP mechanism at a time. To investigate how inhibition stabilization is affected by the neuronal nonlinearity and STP, we computed the real part of the leading eigenvalue of the Jacobian matrix of the excitatory-to-excitatory subnetwork incorporating STP, and refer to it as the “Inhibition Stabilization index” (IS index) (see Supplemental Material [18]). A positive (negative) IS index implies that the network is in the IS (non-IS) regime. To reveal how inhibition stabilization changes with network activity and network connectivity, we investigated how the IS index changes with the excitatory activity  $r_E$  and the excitatory to excitatory connection strength  $J_{EE}$ . These two quantities,  $r_E$  and  $J_{EE}$ , are directly involved in the definition of the IS index (see Supplemental Material [18]).

### A. Inhibition stabilization in recurrent neural networks with short-term depression at E-to-E synapses

We first examined inhibition stabilization for networks with E-to-E STD, evaluated at the fixed point of the system [Fig. 1(a)]. The distinction between non-IS and IS is reflected in network responses to perturbations induced by injecting additional excitatory currents into excitatory neurons while inhibition is fixed. Networks initially in the non-IS regime return back to their initial activity level after a small transient perturbation to the excitatory activity when inhibition is fixed, whereas networks initially in the IS regime deviate from their initial activity (Fig. S1). For linear networks with E-to-E STD, if  $J_{EE}$  is less than one, the network is always in the non-IS regime regardless of  $r_E$  [Fig. 1(b)]. If  $J_{EE}$  is greater than one, the network transitions from IS to non-IS with increasing  $r_E$  [Fig. 1(b)]. In contrast, supralinear networks with E-to-E STD manifest different behaviors. When  $J_{EE}$  is large, the network first transitions from non-IS to IS, and then back to non-IS with increasing  $r_E$  [Figs. 1(c) and S1]. When  $J_{EE}$  is small, the supralinear network stays in the non-IS regime for all values of  $r_E$  [Fig. 1(c)].

To better understand the transition between non-IS and IS in the presence of neuronal nonlinearities and E-to-E STD, we investigated how the boundary between non-IS and IS, defined as “IS boundary,” changes with  $r_E$  (see Fig. S2; Supplemental Material [18]). Mathematically, the IS boundary is determined by the recurrent excitatory-to-excitatory connection strength for different  $r_E$  at which the IS index is zero, denoted by  $J_{EE}^{IS}$ . By computing the derivative of  $J_{EE}^{IS}$  with respect to  $r_E$  (see Supplemental Material [18]), we found that the derivative is always positive for linear networks with E-to-E STD, suggesting that the IS boundary increases with increasing  $r_E$  [Fig. 1(d)]. Therefore, for networks with large

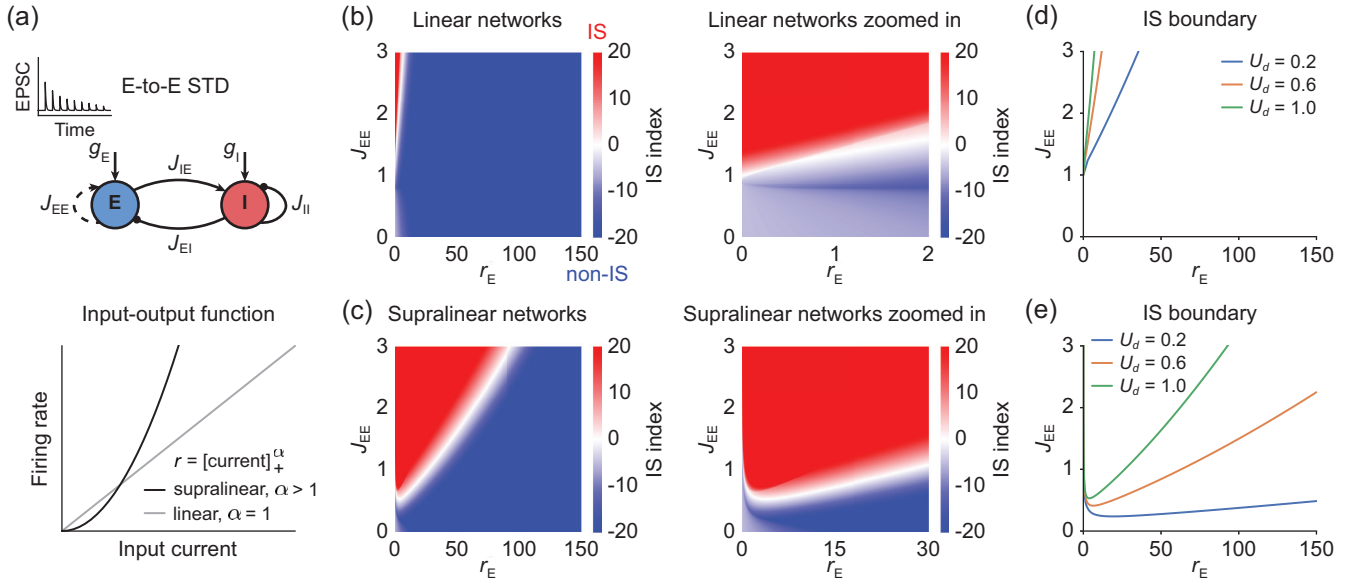


FIG. 1. Inhibition stabilization in recurrent neural networks with E-to-E short-term depression. (a) Top: Schematic of the recurrent network model consisting of an excitatory (blue) and an inhibitory population (red) with E-to-E STD. Bottom: Different nonlinearities of population-averaged responses to input. Linear (gray) and supralinear (black) input-output function given by a rectified power law with exponent  $\alpha = 1$  and  $\alpha = 2$  [cf. Eqs. (1) and (2)], respectively. (b) Left: IS index as a function of excitatory firing rate  $r_E$  and excitatory-to-excitatory connection strength  $J_{EE}$  for linear networks with E-to-E STD. The non-IS and IS regime are marked in blue and red, respectively. Right: Zoomed-in version of (b) left. Here, the depression factor  $U_d$  is 1.0, the firing rate  $r_E$  is plotted in a biologically realistic range from 0 to 150 Hz. (c) Same as (b) but for supralinear networks with E-to-E STD. Here, the depression factor  $U_d$  is 1.0. (d) IS boundary for linear networks with E-to-E STD, defined as the corresponding excitatory-to-excitatory connection strength  $J_{EE}^{IS}$  for different  $r_E$  at which the IS index is zero. Different colors represent the IS boundary for different values of depression factor  $U_d$ . (e) Same as (d) but for supralinear networks with E-to-E STD. Here,  $\tau_x$  is 200 ms in (b)–(e).

$J_{EE}$ , as  $r_E$  increases, only the transition from IS to non-IS is possible [Figs. 1(b) and 1(d)]. In contrast, for supralinear networks with E-to-E STD, the derivative changes from negative to positive with increasing  $r_E$  (see Supplemental Material [18]), implying that the IS boundary first decreases and then increases as  $r_E$  increases [Fig. 1(e)]. As a result, networks can undergo nonmonotonic transitions between non-IS and IS with increasing  $r_E$ . More specifically, networks can switch from non-IS to IS, and then back to non-IS with increasing  $r_E$  [Figs. 1(c) and 1(e); Fig. S1]. The nonmonotonic transitions arise from the competition between the increasing neuronal gain due to the supralinear neuronal input-output function and the decreasing synaptic strength due to STD. Despite high firing rates in the large  $r_E$  limit, E-to-E synaptic strengths are significantly depressed and STD effectively decouples excitatory neurons, rendering the network non-inhibition stabilized. Furthermore, the turning point of the IS boundary appears when  $U_d x_{EE} r_E$  is of order one (see Supplemental Material [18]). Increasing the depression factor  $U_d$  or the STD time constant  $\tau_x$  shifts the turning point to the upper left in the  $(r_E, J_{EE})$  coordinate system (see Fig. S3; Supplemental Material [18]). We also found that these nonmonotonic transitions cannot be observed in networks with static connectivity (see Fig. S4; Supplemental Material [18]). Taken together, our results suggest that E-to-E STD can nontrivially affect the inhibition stabilization property, especially in the presence of neuronal nonlinearities.

## B. Inhibition stabilization in recurrent neural networks with short-term facilitation at E-to-E synapses

To determine if the observed effects are specific to the type of STP at E-to-E synapses, we next examined networks with E-to-E STF [Fig. 2(a)]. Unlike the scenario with STD, for both linear networks or supralinear networks, only a monotonic transition from non-IS to IS is possible with increasing  $r_E$  in the presence of E-to-E STF [Figs. 2(b) and 2(c)]. In contrast to supralinear networks, linear networks with  $J_{EE}$  larger than one are always in the IS regime regardless of  $r_E$ . In both cases, the parameter regime of  $J_{EE}$  and  $r_E$  which supports IS is much larger than in the corresponding network with STD (Fig. 2). Furthermore, independent of the neuronal nonlinearity, the derivative of  $J_{EE}^{IS}$  with respect to  $r_E$  is always negative (see Supplemental Material [18]), indicating that the IS boundary decreases as  $r_E$  increases [Figs. 2(d) and 2(e)]. These results indicate that E-to-E STF exerts a more intuitive influence on the inhibition stabilization property than E-to-E STD even in the presence of neuronal nonlinearities.

## C. Inhibition stabilization in recurrent neural networks with short-term plasticity at other synapses

Finally, we performed the same analyses for networks with different types of STP at all synapses other than E-to-E, including E-to-I STD/STF, I-to-E STD/STF, and I-to-I STD/STF, respectively [see Fig. 3(a); Supplemental

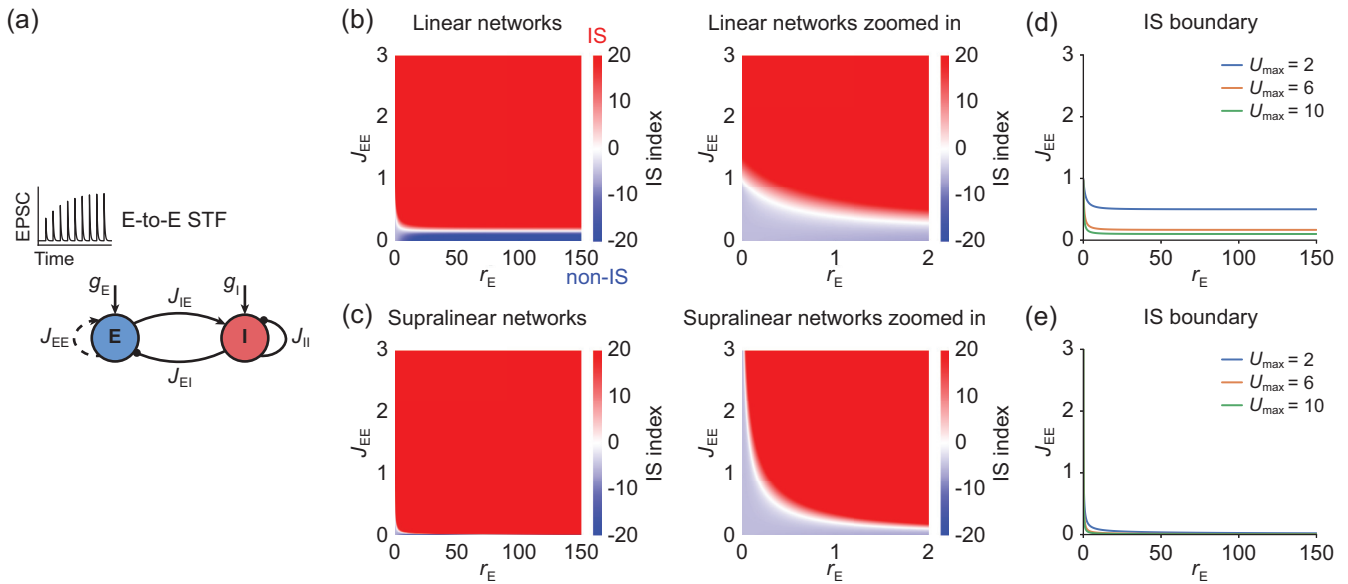


FIG. 2. Inhibition stabilization in recurrent neural networks with E-to-E short-term facilitation. (a) Schematic of the recurrent network model consisting of an excitatory (blue) and an inhibitory population (red) with E-to-E STF. (b) Left: IS index as a function of excitatory firing rate  $r_E$  and excitatory-to-excitatory connection strength  $J_{EE}$  for linear networks with E-to-E STF. The non-IS and IS regime are marked in blue and red, respectively. Right: Zoomed-in version of (b) left. Here, the facilitation factor  $U_f$  is 1.0, and the maximal facilitation value  $U_{\max}$  is 6.0. (c) Same as (b) but for supralinear networks with E-to-E STF. Here, the facilitation factor  $U_f$  is 1.0, and  $U_{\max}$  is 6.0. (d) IS boundary for linear networks with E-to-E STF, defined as the corresponding excitatory-to-excitatory connection strength  $J_{EE}^{IS}$  for different  $r_E$  at which the IS index is zero. Different colors represent the IS boundary for different values of  $U_{\max}$ . (e) Same as (d) but for supralinear networks with E-to-E STF. Here,  $\tau_H$  is 200 ms in (b)–(e).

Material [18]]. Including these STP mechanisms does not change the IS condition relative to networks with fixed connectivity (see Supplemental Material [18]). For networks with a linear input-output function, the IS boundary does not change with  $r_E$  [Figs. 3(b) and 3(d); Supplemental Material [18]], and  $J_{EE}$  completely determines whether the network is non-IS or IS. In contrast, for networks with a supralinear input-output function, the derivative of  $J_{EE}^{IS}$  with respect to  $r_E$  is always negative, suggesting that the IS boundary decreases with increasing  $r_E$  [see Figs. 3(c) and 3(e); Supplemental Material [18]]. Therefore, the transition from non-IS to IS with increasing  $r_E$  in static supralinear networks or supralinear networks with STP at all synapses other than E-to-E can only happen for large  $J_{EE}$  [Figs. 3(c) and 3(e); Fig. S4]. No transition between non-IS and IS can occur in the biological realistic firing regime from 0 to 150 Hz for small  $J_{EE}$  [Figs. 3(c) and 3(e)].

In summary, by considering all possible STP mechanisms, our results demonstrate a nontrivial influence of the neuronal nonlinearity and STP on inhibition stabilization. Specifically, we revealed how inhibition stabilization changes with excitatory activity and network connectivity when considering neuronal nonlinearities and STP.

#### D. Paradoxical effects in recurrent neural networks with short-term plasticity

Previous theoretical studies have suggested that in excitatory and inhibitory networks with static connectivity, one identifying characteristic of inhibition stabilization is that injecting excitatory (inhibitory) current into inhibitory neurons

decreases (increases) inhibitory firing rates, known as the paradoxical effect [4,7]. Here, we sought to identify the conditions under which a paradoxical effect can arise in recurrent neural networks with STP. We assumed that the system is stable locally around the fixed point, in other words, a small transient perturbation to the system will not lead to deviation from the initial fixed point over time. Furthermore, the perturbation used to probe the paradoxical effect (e.g., the excitatory current injected to the inhibitory population) is small enough that it will not lead to a transition between non-IS and IS. To determine the conditions for the presence of the paradoxical effect under these assumptions, we considered the phase plane of the excitatory (abscissa) and inhibitory (ordinate) firing rate dynamics. The first condition involves a larger slope of the inhibitory nullcline than of the excitatory nullcline locally around the fixed point in the phase plane, while the second condition involves a positive slope of the excitatory nullcline around the fixed point (see Ref. [7]; Fig. S5; Supplemental Material [18]). We compared the above two conditions for the presence of the paradoxical effect with the conditions to be in the IS regime. We found that irrespective of the shape of the neuronal nonlinearity, in networks with E-to-E STD, the paradoxical effect implies inhibition stabilization, whereas inhibition stabilization does not imply the paradoxical effect [see Fig. 4(a); Supplemental Material [18]]. In contrast, for networks with E-to-E STF, E-to-I STD/STF, I-to-I STD/STF, and I-to-I STD/STF, inhibition stabilization and the paradoxical effect imply each other [see Fig. 4(a); Supplemental Material [18]]. To highlight the difference between inhibition stabilization and the paradoxical effect in networks with E-to-E STD [Fig. 4(b)], we plotted the paradoxical effect

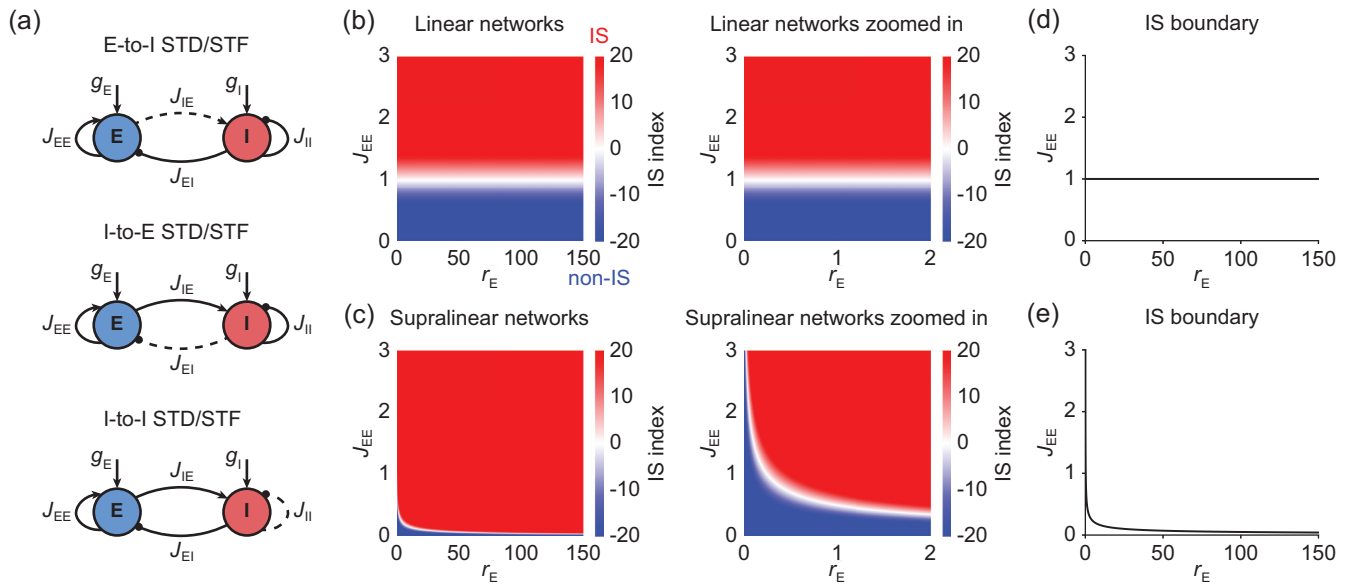


FIG. 3. Inhibition stabilization in recurrent neural networks with E-to-I short-term depression/facilitation, I-to-E short-term depression/facilitation, or I-to-I short-term depression/facilitation. (a) Schematic of the recurrent network model consisting of an excitatory (blue) and an inhibitory population (red) with E-to-I STD/STF (top), with I-to-E STD/STF (middle), or with I-to-I STD/STF (bottom). (b) Left: IS index as a function of excitatory firing rate  $r_E$  and excitatory-to-excitatory connection strength  $J_{EE}$  for linear networks with E-to-I STD/STF, with I-to-E STD/STF, or with I-to-I STD/STF. The non-IS and IS regime are marked in blue and red, respectively. Right: Zoomed-in version of (b) left. (c) Same as (b) but for supralinear networks with E-to-I STD/STF, with I-to-E STD/STF, or with I-to-I STD/STF. (d) IS boundary for linear networks with E-to-I STD/STF, with I-to-E STD/STF, or with I-to-I STD/STF, defined as the corresponding excitatory-to-excitatory connection strength  $J_{EE}^{IS}$  for different  $r_E$  at which the IS index is zero. (e) Same as (d) but for supralinear networks with E-to-I STD/STF, with I-to-E STD/STF, or with I-to-I STD/STF.

boundary that separates the paradoxically-responding and the nonparadoxically-responding regime together with the IS boundary for both linear networks and supralinear networks with E-to-E STD [Figs. 4(c)–4(f)]. In the two dimensional  $r_E$ - $J_{EE}$  plane, the parameter regime for the paradoxical effect is much narrower than the IS regime, suggesting that there is a large parameter space, in which inhibition-stabilized networks do not exhibit the paradoxical effect [Figs. 4(c)–4(f)]. It is noteworthy that in the presence of E-to-E STD, networks in which dynamic STD regulation is required to ensure stability, as studied in [19,20], are a subset of networks which do not exhibit paradoxical effects (see Supplemental Material [18]). For inhibition stabilized networks which do not exhibit paradoxical effects, the corresponding excitatory subnetwork with dynamical short-term plasticity variables is unstable.

Furthermore, by analyzing how the paradoxical effect boundary changes with  $r_E$ , we found that it exhibits a similar trend to the IS boundary (see Supplemental Material [18]). In particular, the paradoxical effect boundary of supralinear networks with E-to-E STD is also a nonmonotonic function of  $r_E$ . Therefore, in this case, networks can undergo non-monotonic transitions between the paradoxically-responding regime and nonparadoxically-responding regime with monotonically changing excitatory activity  $r_E$ .

#### E. Generalization to networks with a mixture of STP mechanisms and multiple interneuron subtypes

So far, we only considered individual STP mechanisms one at a time. Here, we generalized our findings to four more complex and biologically realistic scenarios. First, our results can

be generalized to networks with a mixture of STP mechanisms at other types of synapses except E-to-E synapses where we assume only STD or STF (see Supplemental Material [18]). More specifically, the conditions and results derived for networks with either E-to-E STD or STF alone are the same for networks with either E-to-E STD or STF and a mixture of STP mechanisms at other types of synapses.

Second, by incorporating both STD and STF at E-to-E synapses, we found that the paradoxical effect implies inhibition stabilization, whereas inhibition stabilization does not necessarily imply the paradoxical effect (see Supplemental Material [18]).

Third, inhibitory neurons in the cortex are highly diverse in terms of anatomy, electrophysiology, and function [21–23]. Recent studies have investigated the relationship between inhibition stabilization and the paradoxical effect in networks with multiple interneuron subtypes in the absence of STP [24–28]. Yet, synapses between different interneuron subtypes exhibit considerable short-term dynamics [29]. We then extended the analysis to networks with multiple interneuron subtypes. Theoretical studies have shown that in networks with static connectivity, if the excitatory subnetwork is non-IS (IS), then the sign of the change in the firing rate of the excitatory population and of the change in the total inhibitory current to the excitatory population are opposite (the same) [24]. We found that in the presence of E-to-E STD, if the network is IS, the sign of the change in the firing rate of the excitatory population and of the change in the total inhibitory current to the excitatory population are the same (see Supplemental Material [18]). However, the same sign

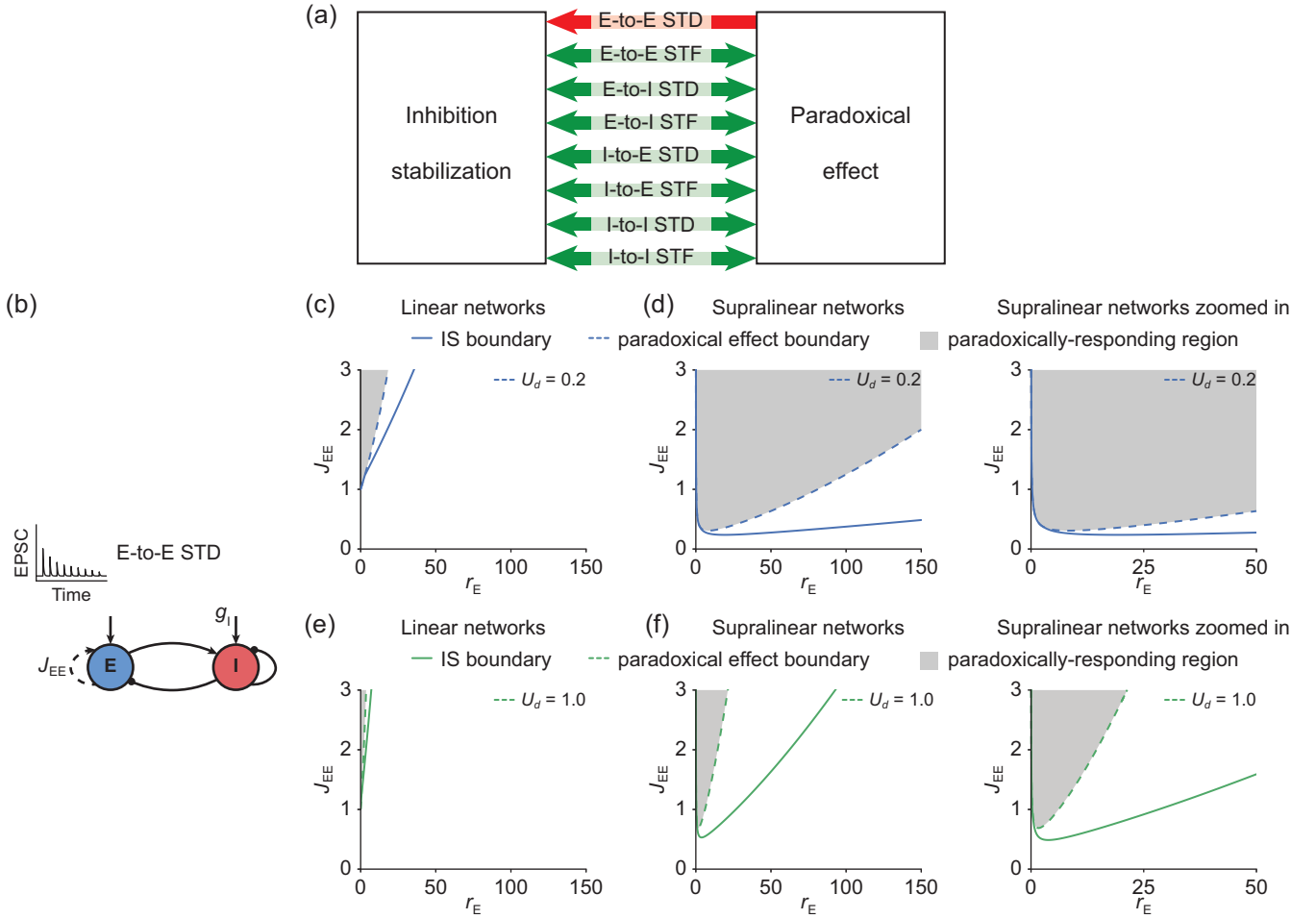


FIG. 4. Relationship between inhibition stabilization and the paradoxical effect in recurrent neural networks with short-term plasticity. (a) In networks with E-to-E STD, the paradoxical effect implies inhibition stabilization, whereas inhibition stabilization does not imply the paradoxical effect indicated by the unidirectional red arrow. In contrast, for networks with E-to-E STF, E-to-I STD/STF, I-to-I STD/STF, and I-to-E STD/STF, inhibition stabilization and the paradoxical effect imply each other indicated by the bidirectional green arrows. (b) Schematic of the recurrent network model consisting of an excitatory (blue) and an inhibitory population (red) with E-to-E STD. (c) An example of the paradoxical effect boundary (dashed line) and the inhibition stabilization boundary (solid line) as a function of excitatory firing rate  $r_E$  for linear networks with E-to-E STD. The paradoxically-responding region is marked in gray. Here, the depression factor  $U_d$  is 0.2 (Supplemental Methods). (d) Same as (c) but for supralinear networks with E-to-E STD. Right: Zoomed-in version of (d) left. (e), (f) Same as (c), (d) but the depression factor  $U_d$  is 1.0. Here,  $\tau_x$  is 200 ms in (c)–(f).

of the change in the firing rate of the excitatory population and of the change in the total inhibitory current to the excitatory population does not imply that the network is IS (see Supplemental Material [18]). In the presence of E-to-E STF and STP at other types of synapses, inhibition stabilization and the same sign change in the firing rate of the excitatory population and of the change in the total inhibitory current to the excitatory population imply each other (see Supplemental Material [18]).

Finally, the derived relationship between inhibition stabilization and the paradoxical effect is independent from the power of the power-law input-output function. Therefore, the derived relationship holds for any monotonically increasing neuronal nonlinearities including sublinear (see Fig. S6; Supplemental Material [18]), despite the fact that sublinear networks have different IS transitions from supralinear networks (see Fig. S7; Supplemental Material [18]). Taken together, our

results indicate that the relationship between inhibition stabilization and paradoxical effect in networks with STP becomes nontrivial in the presence of short-term plasticity.

### III. DISCUSSION

In this study, we combined analytical and numerical methods to reveal the effects of neuronal nonlinearities and STP on inhibition stabilization, the paradoxical effect, and the relationship between them. Including STD at E-to-E synapses, in contrast to other types of STP and other synapse types, generates the most surprising results. Under these conditions, the paradoxical effect implies inhibition stabilization, whereas inhibition stabilization does not imply the paradoxical effect. For networks with other STP mechanisms and networks with static connectivity, inhibition stabilization and the paradoxical effect imply each other. Additionally, in the presence of a neu-

ronal nonlinearity and E-to-E STD, a nonmonotonic transition between different regimes can occur when excitatory activity changes monotonically.

Network models applied to neural circuit development have previously investigated inhibition stabilization in the presence of STP [30]. Recent studies have examined inhibition stabilization and the paradoxical effect in threshold-linear networks with E-to-E STP to demonstrate that inhibition stabilization can be probed by the paradoxical effect [1]. Recent work has conducted similar analysis for supralinear networks with short-term plasticity on specific synapses [31]. Here, we systematically analyzed networks with all forms of STP mechanisms, for both linear networks and supralinear networks. By mathematically defining the IS boundary and the paradoxical effect boundary, we further demonstrated how network activity and connectivity affect the inhibition stabilization property and the paradoxical effect. Importantly, we generalized our results to several more complex scenarios including networks with a mixture of STP mechanisms, networks with both E-to-E STD and STF, networks with multiple interneuron subtypes, and any monotonically increasing neuronal nonlinearities.

In this work, we assumed that the network activity has reached a fixed point, and we did not consider scenarios like multistability or oscillations that could arise from neuronal nonlinearities or STP [15,32,33]. While multistability and oscillations have been observed in the brain [34,35], the single stable fixed point assumed in our study is believed to be a reasonable approximation of awake sensory cortex [36].

Our model makes several predictions that can be tested in optogenetic experiments. Across cortical layers and across brain regions, synaptic strengths can differ by an order of magnitude [37]. Furthermore, the degrees of balance between excitation and inhibition may also vary [38,39], resulting in different neuronal nonlinearities [11,13,39]. Therefore, different behaviors predicted by our analysis may be observable in different neural circuits. For example, in the presence of E-to-E STD, our model shows that networks with weak excitatory to excitatory connection strength  $J_{EE}$  are always non-IS in the biologically realistic activity regime and therefore exhibit no paradoxical effects. In contrast, with E-to-E STD and strong  $J_{EE}$ , network models with a linear population-averaged response function can undergo the transition from IS to non-IS with increasing excitatory activity  $r_E$ . Different from linear networks, our model predicts that nonmonotonic transitions between non-IS and IS can be found in supralinear networks. More specifically, supralinear networks can switch from non-IS to IS, and then from IS to non-IS with increasing  $r_E$ . Although inhibition stabilization does not imply the paradoxical effect in the presence of E-to-E STD, the transition between paradoxically-responding and nonparadoxically-responding regime is also nonmonotonic with increasing  $r_E$  in supralinear networks, whereas in linear networks, only transitions from the paradoxically-responding

to the nonparadoxically-responding regime with increasing  $r_E$  is possible. Therefore, depending on the excitation-inhibition balance, and the specific short term plasticity mechanisms operating in different brain regions, our work proposes that the circuits will exhibit different properties when interrogated with common experimental techniques.

Second, in the presence of STF on E-to-E synapses or STP on other synapses, our results demonstrate that inhibition stabilization and the paradoxical effect imply each other. Linear network models with  $J_{EE}$  larger than one that have E-to-E STF are always IS and thus exhibit the paradoxical effect. In linear network models with STP on other synapses, activity does not affect inhibition stabilization and the paradoxical effect. In contrast, regardless of the strength of  $J_{EE}$ , supralinear networks with E-to-E STF or STP on other synapses can switch from non-IS to IS with increasing  $r_E$ . This transition from non-IS and IS can be directly tested experimentally by probing the paradoxical effect, because of the equivalence of inhibition stabilization and the paradoxical effect found in network models with E-to-E STF or STP on other synapses.

Last, our analysis shows that in most cases substantially altering either  $J_{EE}$  or  $r_E$  can switch the network operating regime. Multiple factors can modify  $J_{EE}$  and  $r_E$  experimentally. On a short timescale, the strength of sensory stimulation, and behavioral states like arousal [40], attention [41], and locomotion [42] can dramatically change activity levels  $r_E$ . Regime switching may therefore be experimentally observable across different stimulation conditions and different behavioral states. On a long timescale,  $J_{EE}$  or  $r_E$  can be modified by long term plasticity mechanisms [43,44]. In this case, regime switching could be experimentally detectable across different developmental stages.

Taken together, our theoretical framework provides a systematic analysis of how short-term synaptic plasticity and response nonlinearities interact to determine the network operating regime, revealing unexpected relationships and their signatures as a guide for future experimental studies.

The code used for model simulations is available at GitHub [45]. All simulation parameters are listed in Supplemental Material.

## ACKNOWLEDGMENTS

We thank Elizabeth Herbert, Felix Waitzmann, Leonidas M. A. Richter, and Dylan Festa for comments on the manuscript. This work was supported by the Max Planck Society and has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 804824 to J.G.) and from the Deutsche Forschungsgemeinschaft in the Collaborative Research Centre 1080 (project C7 to J.G.). Y.K.W. is supported by the Add-on Fellowship of the Joachim Herz Foundation.

[1] A. Sanzeni, B. Akitake, H. C. Goldbach, C. E. Leedy, N. Brunel, and M. H. Histed, Inhibition stabilization is a

widespread property of cortical networks, *eLife* **9**, 54875 (2020).

- [2] D. J. Amit and N. Brunel, Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex, *Cereb. Cortex* **7**, 237 (1997).
- [3] B. K. Murphy and K. D. Miller, Balanced amplification: A new mechanism of selective amplification of neural activity patterns, *Neuron* **61**, 635 (2009).
- [4] H. Ozeki, I. M. Finn, E. S. Schaffer, K. D. Miller, and D. Ferster, Inhibitory stabilization of the cortical network underlies visual surround suppression, *Neuron* **62**, 578 (2009).
- [5] D. B. Rubin, S. D. VanHooser, and K. D. Miller, The stabilized supralinear network: A unifying circuit motif underlying multi-input integration in sensory cortex, *Neuron* **85**, 402 (2015).
- [6] S. Sadeh and C. Clopath, Inhibitory stabilization and cortical computation, *Nat. Rev. Neurosci.* **22**, 21 (2021).
- [7] M. V. Tsodyks, W. E. Skaggs, T. J. Sejnowski, and B. L. McNaughton, Paradoxical effects of external modulation of inhibitory interneurons, *J. Neurosci.* **17**, 4382 (1997).
- [8] S. Sadeh, R. A. Silver, T. D. Mrsic-Flogel, and D. R. Muir, Assessing the role of inhibition in stabilizing neocortical networks requires large-scale perturbation of the inhibitory population, *J. Neurosci.* **37**, 12050 (2017).
- [9] N. Li, S. Chen, Z. V. Guo, H. Chen, Y. Huo, H. K. Inagaki, G. Chen, C. Davis, D. Hansel, C. Guo, and K. Svoboda, Spatiotemporal constraints on optogenetic inactivation in cortical circuits, *eLife* **8**, 48622 (2019).
- [10] C. van Vreeswijk and H. Sompolinsky, Chaos in neuronal networks with balanced excitatory and inhibitory activity, *Science* **274**, 1724 (1996).
- [11] C. van Vreeswijk and H. Sompolinsky, Chaotic balanced state in a model of cortical circuits, *Neural Comput.* **10**, 1321 (1998).
- [12] A. Renart, J. D. Rocha, P. Bartho, L. Hollender, N. Parga, A. Reyes, and K. D. Harris, The asynchronous state in cortical circuits, *Science* **327**, 587 (2010).
- [13] Y. Ahmadian, D. B. Rubin, and K. D. Miller, Analysis of the stabilized supralinear network, *Neural Comput.* **25**, 1994 (2013).
- [14] G. Hennequin, Y. Ahmadian, D. B. Rubin, M. Lengyel, and K. D. Miller, The dynamical regime of sensory cortex: Stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability, *Neuron* **98**, 846 (2018).
- [15] N. Kraynyukova and T. Tchumatchenko, Stabilized supralinear network can give rise to bistable, oscillatory, and persistent activity, *Proc. Natl. Acad. Sci. USA* **115**, 3464 (2018).
- [16] M. V. Tsodyks and H. Markram, The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability, *Proc. Natl. Acad. Sci. USA* **94**, 719 (1997).
- [17] H. Markram, E. Muller, S. Ramaswamy, M. W. Reimann, M. Abdellah, C. A. Sanchez, A. Ailamaki, L. Alonso-Nanclares, N. Antille, S. Arsever *et al.*, Reconstruction and simulation of neocortical microcircuitry, *Cell* **163**, 456 (2015).
- [18] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevResearch.5.033023> for detailed mathematical derivations.
- [19] A. Loebel and M. Tsodyks, Computation by ensemble synchronization in recurrent networks with synaptic depression, *J. Comput. Neurosci.* **13**, 111 (2002).
- [20] A. Loebel, I. Nelken, and M. Tsodyks, Processing of sounds by population spikes in a model of primary auditory cortex, *Front. Neurosci.* **1**, 197 (2007).
- [21] C. K. Pfeffer, M. Xue, M. He, Z. J. Huang, and M. Scanziani, Inhibition of inhibition in visual cortex: The logic of connections between molecularly distinct interneurons, *Nat. Neurosci.* **16**, 1068 (2013).
- [22] X. Jiang, S. Shen, C. R. Cadwell, P. Berens, F. Sinz, A. S. Ecker, S. Patel, and A. S. Tolias, Principles of connectivity among morphologically defined cell types in adult neocortex, *Science* **350**, aac9462 (2015).
- [23] R. Tremblay, S. Lee, and B. Rudy, GABAergic interneurons in the neocortex: From cellular properties to circuits, *Neuron* **91**, 260 (2016).
- [24] A. Litwin-Kumar, R. Rosenbaum, and B. Doiron, Inhibitory stabilization and visual coding in cortical circuits with multiple interneuron subtypes, *J. Neurophysiol.* **115**, 1399 (2016).
- [25] A. Mahrach, G. Chen, N. Li, C. van Vreeswijk, and D. Hansel, Mechanisms underlying the response of mouse cortical networks to optogenetic manipulation, *eLife* **9**, e49967 (2020).
- [26] K. D. Miller and A. Palmigiano, Generalized paradoxical effects in excitatory/inhibitory networks, *bioRxiv*, <https://www.biorxiv.org/content/10.1101/2020.10.13.336727v1>.
- [27] A. Palmigiano, F. Fumarola, D. P. Mossing, N. Kraynyukova, H. Adesnik, and K. D. Miller, Common rules underlying optogenetic and behavioral modulation of responses in multi-cell-type V1 circuits, *bioRxiv*, <https://www.biorxiv.org/content/10.1101/2020.11.11.378729v3>.
- [28] L. M. A. Richter and J. Gjorgjieva, A circuit mechanism for independent modulation of excitatory and inhibitory firing rates after sensory deprivation, *Proc. Nat. Acad. Sci. USA* **119**, e2116895119 (2022).
- [29] L. Campagnola, S. C. Seeman, T. Chartrand, L. Kim, A. Hoggarth, C. Gamlin, S. Ito, J. Trinh, P. Davoudian, C. Radaelli *et al.*, Local connectivity and synaptic dynamics in mouse and human neocortex, *Science* **375**, eabj5861 (2022).
- [30] V. Rahmati, K. Kirmse, K. Holthoff, L. Schwabe, and S. J. Kiebel, Developmental emergence of sparse coding: A dynamic systems approach, *Sci. Rep.* **7**, 13015 (2017).
- [31] Y. K. Wu and F. Zenke, Nonlinear transient amplification in recurrent neural networks with short-term plasticity, *eLife* **10**, 71263 (2021).
- [32] G. Mongillo, O. Barak, and M. Tsodyks, Synaptic theory of working memory, *Science* **319**, 1543 (2008).
- [33] G. Mongillo, D. Hansel, and C. van Vreeswijk, Bistability and Spatiotemporal Irregularity in Neuronal Networks with Nonlinear Synaptic Transmission, *Phys. Rev. Lett.* **108**, 158101 (2012).
- [34] X.-J. Wang, Synaptic reverberation underlying mnemonic persistent activity, *Trends Neurosci.* **24**, 455 (2001).
- [35] G. Buzsáki and A. Draguhn, Neuronal oscillations in cortical networks, *Science* **304**, 1926 (2004).
- [36] K. D. Miller, Canonical computations of cerebral cortex, *Current Opinion in Neurobiology* **37**, 75 (2016).
- [37] Allen Institute for Brain Science, Synaptic physiology coarse matrix dataset, 2019.
- [38] J. Barral and A. D'Reyes, Synaptic scaling rule preserves excitatory-inhibitory balance and salient neuronal network dynamics, *Nat. Neurosci.* **19**, 1690 (2016).
- [39] Y. Ahmadian and K. D. Miller, What is the dynamical regime of cerebral cortex? *Neuron* **109**, 3373 (2021).
- [40] M. Vinck, R. Batista-Brito, U. Knoblich, and J. A. Cardin, Arousal and locomotion make distinct contributions to cortical activity patterns and visual encoding, *Neuron* **86**, 740 (2015).

- [41] J. H. Reynolds and L. Chelazzi, Attentional modulation of visual processing, *Annu. Rev. Neurosci.* **27**, 611 (2004).
- [42] Y. Fu, J. M. Tucciarone, J. S. Espinosa, N. Sheng, D. P. Darcy, R. A. Nicoll, Z. J. Huang, and M. P. Stryker, A cortical circuit for gain control by behavioral state, *Cell* **156**, 1139 (2014).
- [43] N. L. Rochefort, O. Garaschuk, R. I. Milos, M. Narushima, N. Marandi, B. Pichler, Y. Kovalchuk, and A. Konnerth, Sparsification of neuronal activity in the visual cortex at eye-opening, *Proc. Natl. Acad. Sci. USA* **106**, 15049 (2009).
- [44] H. Ko, L. Cossell, C. Baragli, J. Antolik, C. Clopath, S. B. Hofer, and T. D. Mrsic-Flogel, The emergence of functional microcircuits in visual cortex, *Nature (London)* **496**, 96 (2013).
- [45] Code repository, <https://github.com/comp-neural-circuits/inhibition-stabilization-paradoxical-effect-STP>.