TUM School of Life Sciences
Technische Universität München

Functional roles of inhibitory plasticity: stability, control of excitatory plasticity, and beyond

Christoph Miehl

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Frank Johannes

Prüfer*innen der Dissertation:

1. Prof. Julijana Gjorgjieva, Ph.D.
2. Prof. Dr. Peter Jedlička
3. Assoc. Prof. Dr. Matthias H. Hennig

Die Dissertation wurde am 08.11.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 09.01.2023 angenommen.

Christoph Miehl. *Functional roles of inhibitory plasticity: stability, control of excitatory plasticity, and beyond.* Dissertation, 2023.

Abstract

Synaptic changes underlying learning and memory are believed to be implemented by synaptic plasticity based on the patterns of neuronal activity. While many studies have focused on synaptic plasticity at excitatory synapses, fewer have investigated synaptic plasticity at inhibitory synapses.

In my dissertation, I investigate the functional role of long-term plasticity at inhibitory synapses and the interaction with plasticity at excitatory synapses. I use computational and mathematical approaches to gain a mechanistic understanding of how plasticity shapes the connectivity and neuronal activity structure in feed-forward and recurrent networks. My work reveals that a novel form of inhibitory plasticity, which depends nonlinear on the postsynaptic firing rate, is a sufficient mechanism to homeostatically stabilize firing rates and excitatory weight dynamics, while simultaneously enabling the learning of new representations. Furthermore, I suggest together with my co-author Auguste Schulz that inhibitory plasticity is the underlying mechanism regulating neuronal responses to familiar versus novel stimuli. Together with my co-author Yue Kris Wu, I summarize the recent literature on the role of inhibitory plasticity and outline future directions and open questions. Based on experimental findings by our collaborators from the laboratory of Prof. Dr. Robert Froemke, I show how the interaction of homosynaptic and heterosynaptic plasticity at excitatory and inhibitory synapses can lead to a set-point for cortical excitatory-inhibitory balance. Finally, I summarize together with my co-author Sebastian Onasch the current literature on the formation of certain recurrent structures, i.e. assemblies, based on excitatory synaptic plasticity and I show as a contributing author with Dr. Lisandro Montangie how such assemblies can emerge spontaneously without structured external input. Taken together, I extend the existing work on the functional role of excitatory and, specifically, inhibitory synaptic plasticity in neuronal circuits.

Zusammenfassung

Es wird allgemein angenommen, dass synaptische Plastizität, also die Änderung der Stärke von synaptischen Verbindungen auf Basis der Aktivität des presynaptischen und des postsynaptischen Neurons, der unterliegende Mechanisms für Lernen und das Formen von Erinnerungen ist. Bisher haben sich die meisten Studien auf synaptische Plastizität an exitatorischen Synapses fokussiert.

In meiner Dissertation untersuche ich die funktionellen Eigenschaften von synaptischer Plastizität an exitatorischen und besonders inhibitorischen Synapsen. Ich verwende Computersimulationen und mathematische Methoden um ein mechanistisches Verständnis davon zu bekommen, wie Plastizität neuronale Verbindungen und Aktivitätsmuster in verschiedenen Netzwerkstrukturen beeinflusst. Meine Arbeit zeigt, dass eine neue Form der inhibitorischen Plastizität ausreichend ist um neuronale Aktivität und exitatorische Plastizität zu stabilisieren, und gleichzeitig das Erlernen von neuen Repräsentationen erlaubt. Weiters schlage ich mit meiner Co-Autorin, Auguste Schulz, vor, dass inhibitorische Plastizität der Mechanismus ist, der die neuronalen Aktivitätsmuster zu "erwarteten" und "unerwarteten" Stimuli kontrolliert. Zusammen mit meinem Co-Autor Yue Kris Wu fasse ich die aktuelle Literatur zur Rolle der inhibitorischen Plastizität zusammen und gebe einen Überblick über zukünftige Forschungsfragen. Ich zeige, basierend auf den experimentellen Studien von unserem Kollaborateur, Prof. Dr. Robert Froemke, dass homosynaptische und heterosynaptische Plastizität zusammen einen sogenannten "Set-Point" an exitatorischer und inhibitorischer Balance erreichen. Zusätzlich fasse ich mit meinem Co-Autor Sebastian Onasch die derzeitige Literatur zur Entstehung von bestimmten neuronalen Verbindungen, sogenannten "Assemblies", zusammen und ich zeige als beitragender Autor mit Dr. Lisandro Montangie wie Assembly-Strukturen spontan aus unstrukturiertem externen Input entstehen können.

Contents

List of Figures

1 Introduction

In the field of neuroscience, the underlying hypothesis is that neurons encode information in their activity patterns. High activity, or firing rates, of neurons in response to a sensory stimulus is interpreted as a representation of this stimulus (Kriegeskorte & Diedrichsen, 2019). The activity of a neuron is determined by the relative amount of excitatory and inhibitory inputs it receives. Changes in the strength of either excitatory or inhibitory synaptic connections can lead to changes in neuronal activity and, therefore, also to changes in stimulus representation. The mechanism based on which synaptic strength is modified is called synaptic plasticity and depends on the activity of the two neurons the synapse is connecting. Hence, a cycle of interaction emerges, where neuronal activity is affected if synaptic connection strength changes, which follows from modifications via synaptic plasticity which, to close the circle, depends on neuronal activity (Fig. 1.1). Many experimental and theoretical studies have investigated certain aspects of this circular interaction between activity and connectivity. In my dissertation, I contribute to a further understanding of how activity and connectivity influence each other via the process of synaptic plasticity, with a specific focus on the role of synaptic plasticity at inhibitory synapses.

Neuronal activity as a representation of sensory stimulus features

Many decades of research have led to the conclusion that neuronal activity encodes information about the environment. A prominent example is the study by Hubel & Wiesel, 1962 in which the authors show that neurons in the cat visual cortex strongly respond to a visual stimulus consisting of a bar with a certain orientation. Following this pioneering work, researchers have found that neurons in many brain

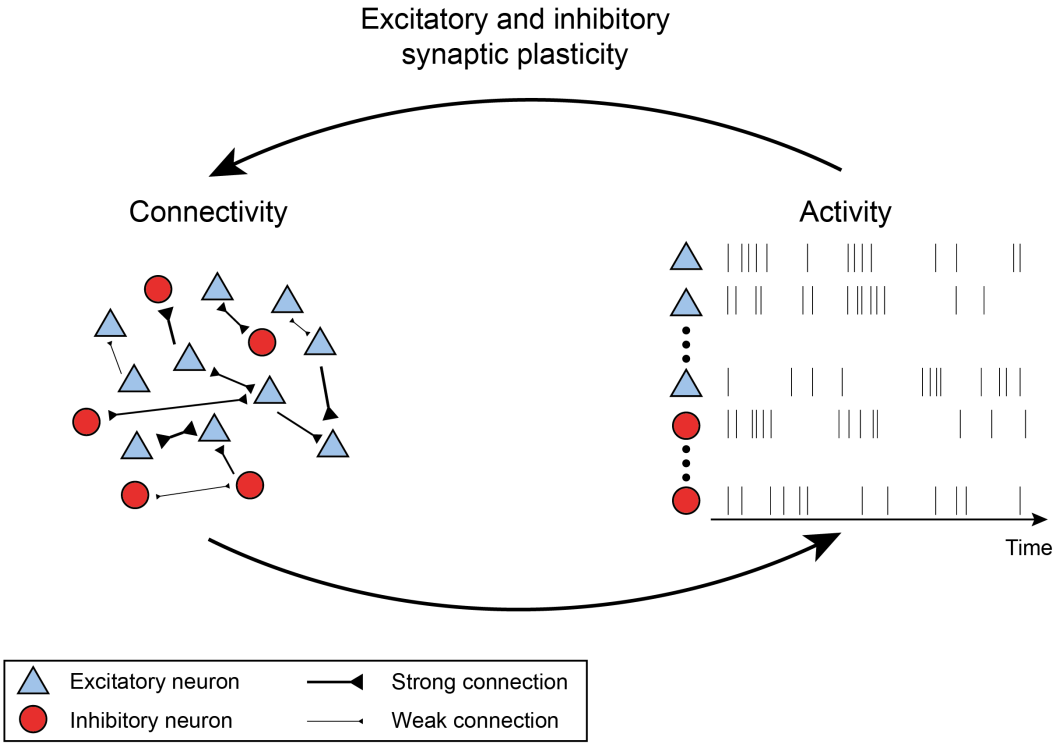


Figure 1.1: Activity – connectivity cycle. Activity of excitatory neurons (blue triangles) and inhibitory neurons (red circles) follows from the connectivity among those neurons, while activity leads to a change in connectivity via the mechanisms of excitatory and inhibitory synaptic plasticity.

regions and different animal models are reliably activated when a specific stimulus is presented or a specific condition is met. For example neurons in the somatosensory cortex in mice respond to whisker deflection (reviewed e.g. in Maravall & Diamond, 2014), neurons in the auditory cortex respond to auditory stimuli (reviewed e.g. in Shamma & Fritz, 2014), space is represented via place cells in Hippocampus (O'Keefe *et al.,* 1971; O'Keefe, 1976) and grid cells in the entorhinal cortex (Fyhn *et al.,* 2004; Hafting *et al.,* 2005), and many more. While much evidence exists that neurons are selectively highly responsive (or 'tuned') to specific stimulus features, it has also been suggested that for neuronal coding also exact spike timings, interspike intervals, and correlations matter (Insanally *et al.,* 2019; Panzeri *et al.,* 2022).

An important concept that determines the activity of a neuron and, therefore, to which stimulus feature the neuron is tuned, is the balance of excitatory and inhibitory inputs (E/I balance) onto the neuron. E/I balance can be broadly defined as the proportionality of total excitatory and inhibitory inputs (Froemke, 2015).

Experimental and theoretical work has found different forms of E/I balance, dependent on its spatial and temporal properties (Hennequin *et al.,* 2017). If excitatory and inhibitory inputs are co-tuned, i.e. E/I balance exists across stimulus space, the balance is termed 'detailed' (Dorrn *et al.,* 2010; Froemke *et al.,* 2007; Wehr & Zador, 2003), and 'global' if excitatory and inhibitory inputs are not co-tuned (Brunel, 2000; van Vreeswijk & Sompolinsky, 1996). In addition, if excitatory and inhibitory inputs are balanced on short timescales (milliseconds), the E/I balance is termed 'tight' (Okun & Lampl, 2008), and 'loose' otherwise (Denève & Machens, 2016). Tight and detailed balance, as found experimentally (Bhatia *et al.,* 2019; Rupprecht & Friedrich, 2018), is called 'precise' E/I balance. Disruption or mismatch of E/I balance has been identified as underlying neurodevelopmental and neurodegenerative disorders in the brain, like autism, schizophrenia, and epilepsy (Lopatina *et al.,* 2019).

The neuronal responses to sensory inputs are not fixed but depend on the history of previous sensory input. A prominent change in neuronal responses on the timescales of seconds to hours is 'adaptation' (Whitmire & Stanley, 2016). Experimental studies find that neuronal responses in many sensory areas are reduced (or adapted) for repeated (or predictable) stimuli, while novel (or unpredictable) stimuli elicit strong neuronal responses (Fairhall, 2014; Homann *et al.,* 2022; Näätänen *et al.,* 1982; Natan *et al.,* 2015; Weber & Fairhall, 2019). These findings suggest the dependence of neuronal responses on the predictability of the stimulus. Evidence for the differential response to familiar versus novel stimuli has been found at the whole-brain-level, where EEG measurements show different responses to a repeated and a deviant stimulus (termed mismatch negativity; Ross & Hamm, 2020), on the level of population of neurons using calcium imaging (Homann *et al.,* 2022), down to the level of single neurons, where a repeated tone leads to a decrease in neuronal responses (Natan *et al.,* 2015). In my work, we have studied the emergence of adapted and novelty responses in a recurrent network model (Schulz *et al.,* 2021) (see Chapter 3.2).

Synaptic connectivity shapes neuronal activity

Synaptic connections between neurons determine neuronal activity. Studies probing the connection patterns between neurons find evidence for structured, nonrandom connectivity. With different experimental methods, like electrophysiology, optogenetic stimulation, or electron microscopy, the connectivity between different neurons has been probed. Structures have been found on various spatial scales, ranging from large-scale connectivity between brain regions to overrepresented local connectivity motifs (Bassett & Sporns, 2017; Lynn & Bassett, 2019).

A well-studied structure is a neuronal assembly, which I define as a group of strongly recurrently connected neurons, in contrast to a neuronal ensemble, which I define as a group of neurons with correlated activity. Although both concepts are often interpreted as equivalent in the literature, it is important to distinguish them clearly (as we argue in more depth in Miehl *et al.,* 2022). Nevertheless, it is often assumed that the correlated activity of the ensemble is determined by strong recurrent connectivity (i.e. the assembly). Experimental evidence has accumulated over the years, showing that ensembles exist upon stimulus presentation (Berkes *et al.,* 2011; Harris *et al.,* 2003; Miller *et al.,* 2014), during spontaneous activity (Cossart *et al.,* 2003; MacLean *et al.,* 2005; Mao *et al.,* 2001) and has been suggested to underlie behavior (Carrillo-Reid *et al.,* 2019; Choi *et al.,* 2011; Jennings *et al.,* 2019; Josselyn & Frankland, 2018; Liu *et al.,* 2012; Marshel *et al.,* 2019). Experimental work also shows that connectivity is structured, where specifically bidirectional connections seem to be over-represented (Campagnola *et al.,* 2022; Guzman *et al.,* 2016; Jouhanneau *et al.,* 2015; Song *et al.,* 2005; Turner *et al.,* 2022), but see (Lefort *et al.,* 2009). Furthermore, neurons are more likely to be connected if they share common input (Yoshimura *et al.,* 2005), target the same postsynaptic neurons (Brown & Hestrin, 2009), have common neighbors (Perin *et al.,* 2011; Turner *et al.,* 2022), or are similarly tuned (Ko *et al.,* 2011; Lee *et al.,* 2016; Rossi *et al.,* 2020; Wertz *et al.,* 2015). But a clear experimental link between the concept of an assembly and an ensemble is still missing.

It has been hypothesized that assemblies are the basic unit of cortical cognition (Buzsáki, 2010; Eichenbaum, 2018; Huyck & Passmore, 2013; Yuste, 2015). Assemblies have several advantages over the neuron-centric view, where single neu-

rons code for specific features. First, assemblies are more robust, allowing single neurons to malfunction. Second, due to the strong recurrence also weaker stimuli can lead to a strong response. Third, assemblies allow for pattern completion, where an incomplete pattern can activate the whole representation. Fourth, the assembly neurons can stay active after the stimulus is removed. Some computational studies have investigated the power of assemblies as the basis for any computation. This is sometimes referred to as 'assembly calculus' (Papadimitriou & Friederici, 2022; Papadimitriou *et al.,* 2020), where the formation of assemblies and connections between assemblies are mathematically abstracted in basic operations.

If assemblies are the basic unit of computation, they need to be learned. This could be done in two ways: First, assemblies form during early development, and these structures are later used by forming associations between them (Holtmaat & Caroni, 2016). In support of this view, evidence exists that cells stemming from the same progenitor display increased reciprocal connectivity (Tarusawa *et al.,* 2016; Yu *et al.,* 2009). Also, the formation of assemblies based on structured 'spontaneous activity', as evident in the early developmental stages of an organism (Richter & Gjorgjieva, 2017), has been suggested. Second, assemblies are learned or modified during the adult life of an animal, following from the vast experimental literature finding that synapses are plastic (Abbott & Nelson, 2000). Since both possibilities are not mutually exclusive, the brain probably has implemented both strategies, depending on the task and brain region.

Nonrandom connectivity also extends to inhibitory neurons, with experimental work showing, for example, in the auditory cortex that excitatory and inhibitory neurons with similar stimulus tuning are strongly connected (Znamenskiy *et al.,* 2018). In computational studies, this specific inhibitory feedback has been suggested to support network stability (Wu & Zenke, 2021; Znamenskiy *et al.,* 2018), changes in neuronal variability (Rost *et al.,* 2018), and decision making (Najafi *et al.,* 2020).

But do experimentally-verified learning mechanisms support the formation of assemblies? Multiple computational studies have investigated this question. In our review (Miehl *et al.,* 2022) (see Chapter 3.5), we have identified four basic building blocks of assembly formation in the modeling literature: Synaptic plasticity, symmetry breaking, competition, and stability (Figure 1.2). The two closely related concepts of symmetry breaking and competition describe how a specific group of

neurons is selected as an assembly and how this assembly is clearly separated from other assemblies. Stability refers to the question of how learned representations can remain stable over long periods of time. Probably the most important building block for assembly formation is long-term synaptic plasticity at excitatory-to-excitatory synapses.

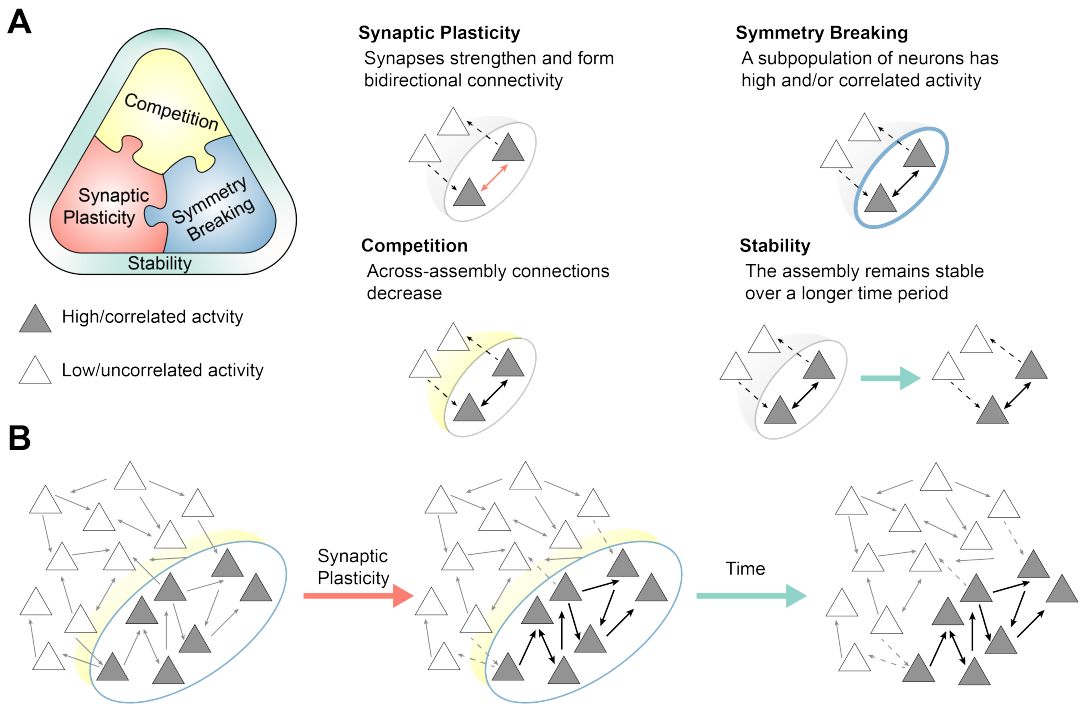


Figure 1.2: The basic building blocks of assembly formation. A. The four key components of assembly formation in computational models are: synaptic plasticity, symmetry breaking, competition, and stability. **B.** Due to symmetry breaking, a subpopulation of neurons fires at a high rate and/or with highly correlated activity compared to the remaining neurons. Synaptic plasticity promotes mutual connections within the assembly, while a competition mechanism decreases the across-assembly weights. The newly formed assembly structure is stable over time. Figure adapted from Miehl *et al.,* 2022.

Excitatory synaptic plasticity

First suggested by Donald Hebb (Hebb, 1949), it is now widely hypothesized that plasticity at excitatory-to-excitatory synapses, often referred to as excitatory plasticity, is the underlying mechanism of structure formation (Abbott & Nelson, 2000; Brea & Gerstner, 2016; Feldman, 2009; Magee & Grienberger, 2020; Suvrathan, 2019). Experimental studies have shown that excitatory long-term

depression (LTD) and long-term potentiation (LTP) can be induced dependent on the firing rates (Kirkwood *et al.,* 1996) or precise spike-timing of the pre- and postsynaptic neuron (Bi & Poo, 1998). In computational studies, phenomenological descriptions of both rate-based and spike-based synaptic plasticity mechanisms have been defined (Fig. 1.3; see Methods and mathematical framework). The finding of rate-dependent plasticity has inspired multiple theoretical learning rules, like Oja's rule (Oja, 1982), the covariance rule (Sejnowski, 1989), and the Bienenstock-Cooper-Munro (BCM) rule (Bienenstock *et al.,* 1982). For example in the BCM rule, rate-dependent synaptic weight change is based on a nonlinear function of the firing rate of the postsynaptic neuron (Fig. 1.3A; see Methods and mathematical framework). Here, if the firing rate is above a certain value (the LTD/LTP threshold) LTP is induced, while if the firing rate is below the LTD/LTP threshold, LTD is induced.

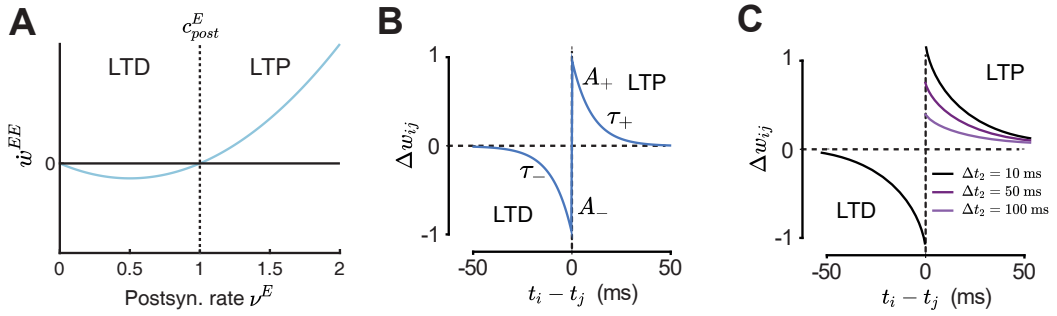


Figure 1.3: Rate-based and spiking-timing-dependent excitatory synaptic plasticity. A. Rate-based plasticity curve of excitatory-to-excitatory weights ($\dot{w}^{EE}$, blue) as a function of the firing rate of the excitatory postsynaptic neuron ν^E. The postsynaptic LTD/LTP threshold c^E_{post} is set to 1. Panel adapted from Miehl & Gjorgjieva, 2022. **B.** Weight change of the synapse from neuron j to neuron i, Δw_{ij}, via the pairwise STDP rule as a function of the time difference between post (t_i) and presynaptic spike (t_j) with long-term potentiation (LTP) parameters A_+ and τ_+ and long-term depression (LTD) parameters A_- and τ_-. Panel adapted from Miehl *et al.,* 2022. **C.** Same as B but for the minimal triplet STDP rule. LTP is induced using triplets of spikes, where the time difference between two postsynaptic spikes Δt_2 shapes the LTP window. Panel modified from Gjorgjieva *et al.,* 2011.

Performing experiments where single pre- and postsynaptic spikes are paired at different time delays uncovered that spike-timing-dependent LTP is induced if a postsynaptic spike follows a presynaptic spike within approximately 10 ms, while spike-timing-dependent LTD is induced if a presynaptic spike follows a postsy-

naptic spike within 10-100 ms (Bi & Poo, 1998; Markram *et al.,* 1997). These observations led modelers to formulate pairwise STDP rules (Gerstner *et al.,* 1996; Kempter *et al.,* 1999). Here, the time difference between a pre- and postsynaptic spike pair indicates the direction and strength of synaptic weight change (Fig. 1.3B; see Methods and mathematical framework). In experiments, the exact shape of the STDP function depends on multiple aspects, like the position of the synapse alongside the dendritic tree (Letzkus *et al.,* 2006), the presence of neuromodulators (like acetylcholine, dopamine, and noradrenaline) (Froemke *et al.,* 2013, 2007), and other factors (Caporale & Dan, 2008). Furthermore, experiments have found that STDP has a firing-rate dependency (Sjöström *et al.,* 2001). To account for this aspect, theoretical studies have incorporated STDP rules based on three spikes, the triplet STDP rule (Gjorgjieva *et al.,* 2011; Pfister & Gerstner, 2006). For the minimal triplet STDP rule, which describes the experimental findings well (Pfister & Gerstner, 2006), LTP is induced by triplets of spikes (one presynaptic and two postsynaptic spikes) where the time difference between the two postsynaptic spikes determines the LTP window (Fig. 1.3C; see Methods and mathematical framework). In my work, we study the formation of assemblies based on the triplet STDP rule (Montangie *et al.,* 2020) (see Chapter 3.6).

Homeostatic mechanisms

A well-known problem of synaptic plasticity at excitatory synapses is that synaptic dynamics are inherently unstable (Miller & MacKay, 1994). An increase in the excitatory synaptic strength leads to an increase in postsynaptic firing rates, which in turn leads to an increase in excitatory synaptic strength, leading to an unstable feedback loop. The instability of excitatory weight dynamics following from Hebbian plasticity is often termed 'Hebbian runaway dynamics' (Turrigiano & Nelson, 2004). Therefore, additional mechanisms are necessary to counteract this uncontrolled growth of excitatory synaptic strength. Both, experimental and theoretical work suggested a multitude of possible homeostatic mechanisms. In computational studies, a straight-forward way to implement stability of weight dynamics is to use a normalization mechanism that preserves the total input (or output) synaptic weight strength (Miller & MacKay, 1994; Oja, 1982), which is

often linked to the mechanism of synaptic scaling or heterosynaptic plasticity.

Synaptic scaling

One solution to counteract Hebbian runaway dynamics is the phenomenon of synaptic scaling, a multiplicative up- or down-scaling of the total synaptic strength as a consequence of prolonged low or high activity levels (Turrigiano, 2008). In experiments, blocking neuronal activity leads to a homeostatic up-scaling of synaptic inputs, while high neuronal activity leads to a homeostatic down-scaling (Turrigiano *et al.,* 1998). Despite the clear advantage of synaptic scaling to stabilize circuit activity, it has been argued that the timescales of synaptic scaling are too slow to counteract Hebbian runaway dynamics (Zenke *et al.,* 2013). While synaptic plasticity can be induced on the timescales of minutes, synaptic scaling usually takes hours to days (Gainey & Feldman, 2017; Keck *et al.,* 2017). This discrepancy is sometimes referred to as the 'temporal paradox' of integrating synaptic scaling and long-term plasticity (Zenke & Gerstner, 2017; Zenke *et al.,* 2017).

Heterosynaptic plasticity

An alternative mechanism to counteract Hebbian runaway dynamics is heterosynaptic plasticity (Bliss & Lømo, 1973; Chistiakova *et al.,* 2015; Lynch *et al.,* 1977). Heterosynaptic plasticity describes synaptic plasticity changes at synapses that were not directly presynaptically stimulated. Computational models have shown that heterosynaptic plasticity can stabilize synaptic weight dynamics (Chen *et al.,* 2013; Kirchner & Gjorgjieva, 2021; Volgushev *et al.,* 1994). Stability in these models follows because heterosynaptic plasticity usually induces weight changes in the opposite direction of homosynaptic plasticity, i.e. homosynaptic LTP (LTD) leads to heterosynaptic LTD (LTP) at nearby synapses, in line with experimental evidence (Royer & Paré, 2003; White *et al.,* 1990). In my work, we study the role of heterosynaptic plasticity at excitatory and inhibitory synapses related the formation of E/I balance set-points (Field *et al.,* 2020) (see Chapter 3.4).

However, the existing experimental and computational literature does not agree on how exactly heterosynaptic plasticity might be implemented at the neuronal level. The different hypotheses can be grouped into five (partially overlapping)

categories: (1) Following experiments measuring the paired-pulse ratio (Volgushev *et al.,* 1997; Volgushev *et al.,* 2000) it is thought that heterosynaptic plasticity depends on the predisposition of the synapse, i.e. strong synapses undergo LTD while weak synapses undergo LTP (Chen *et al.,* 2013; Volgushev *et al.,* 2016). (2) Based on experiments in the Hippocampus (Abraham *et al.,* 2001), an alternative hypothesis argues that presynaptic (spontaneous) activity is critical to induce heterosynaptic plasticity (Benuskova & Abraham, 2007; Jedlicka *et al.,* 2015). (3) Heterosynaptic plasticity follows from internal signaling mechanisms which are triggered by induction of homosynaptic plasticity (Oh *et al.,* 2015). (4) Heterosynaptic plasticity follows from competition for resources (Antunes & Simoes-de-Souza, 2018; Triesch *et al.,* 2018). (5) Alternatively, the synaptic tag-and-capture hypothesis has been suggested to explain how heterosynaptic plasticity might occur (Okuno *et al.,* 2012, 2018).

Intrinsic plasticity

Another alternative homeostatic mechanism is intrinsic plasticity, which describes the adjustment of the intrinsic excitability of single neurons (Debanne *et al.,* 2019; Desai *et al.,* 1999). For example, intrinsic plasticity has been suggested to be important for the recovery of firing rates after sensory deprivation (Wu *et al.,* 2020).

Short-term plasticity

Short-term plasticity is defined as a change in synaptic strength on the timescale of up to 100 milliseconds (Zucker & Regehr, 2002). Input spikes that occur within a short time window can cause short-term facilitation or depression of the postsynaptic potentials (Motanis *et al.,* 2018). These modifications are short-term in the sense that there are no persistent changes to the synaptic signaling machinery. Since short-term plasticity changes with stimulus intensity, i.e. higher input frequencies lead to more short-term depression, is a good candidate to control neuronal firing rates (Reyes, 2011; Tsodyks *et al.,* 1998).

Metaplasticity

Another homeostatic mechanism, first suggested in computational studies, is 'meta-

plasticity' (Bienenstock *et al.,* 1982; Yger & Gilson, 2015). This mechanism induces dynamics of the plasticity rule itself, usually by assuming that the threshold between LTD and LTP is dynamic. This shift in plasticity threshold has later been confirmed in sensory deprivation studies (Cooper & Bear, 2012; Kirkwood *et al.,* 1996; Kuo & Dringenberg, 2009; Philpot *et al.,* 2003), or direct stimulation of input pathways (Abraham, 2008; Huang *et al.,* 1992). A widely-used metaplastic framework is the Bienenstock-Cooper-Munro (BCM) rule (Bienenstock *et al.,* 1982; Cooper & Bear, 2012). In the BCM rule, higher (lower) postsynaptic rates lead to higher (lower) excitatory LTD/LTP thresholds, making it harder to induce LTP (LTD).

Hard- and soft upper bounds

A common strategy in many computational models is to assume upper bounds on synaptic weight strength. A 'hard' upper bound means that synaptic weights can not grow above this maximal synaptic strength value. 'Soft' upper bounds, also termed 'weight-dependent' plasticity, assume that synaptic weight change is proportional to the inverse of the strength of the weights (Gütig *et al.,* 2003; Rubin *et al.,* 2001; van Rossum *et al.,* 2000). Soft bounds have the problem of leading to unimodal weight distributions, hence counteracting learning of connectivity structures (Morrison *et al.,* 2007).

Although these different homeostatic mechanisms are usually discussed separately, evidence exists that they are not mutually exclusive. For example, the concepts of metaplasticity and heterosynaptic plasticity are highly intertwined (Benuskova & Jedlicka, 2012; Jedlicka *et al.,* 2015), and possibly also metaplasticity and synaptic scaling (Keck *et al.,* 2017). Despite recent efforts in computational studies to integrate and study the interaction of Hebbian and various homeostatic mechanisms (Wu *et al.,* 2020; Zenke *et al.,* 2013), many questions remain unresolved, including the problem of timescales, detecting homeostatic plasticity *in vivo*, the overlap and distintiveness of cellular and molecular mechanisms underlying homeostasis, and many more (Fox & Stryker, 2017; Turrigiano, 2017; Yee *et al.,* 2017). One unresolved question is related to the contribution of inhibition and inhibitory plasticity to homeostasis of firing rate and synaptic weight dynamics. In my work, I

have shown that inhibitory plasticity can be a sufficient homeostatic mechanism and that homeostasis via inhibition can be linked to metaplasticity (Miehl & Gjorgjieva, 2022) (see Chapter 3.1).

Inhibitory synaptic plasticity

Long-term synaptic plasticity is also present at inhibitory-to-excitatory synapses. Over the past few years, many experimental studies have investigated inhibitory plasticity in different brain regions and multiple animal models (for reviews, see Capogna *et al.,* 2021; Chen & Nedivi, 2013; Chiu *et al.,* 2019; Froemke, 2015; Gandolfi *et al.,* 2020; Kripkee & Froemke, 2017). To induce inhibitory plasticity, concurrent presynaptic hyperpolarization and postsynaptic depolarization are necessary (Chiu *et al.,* 2018; Mellor, 2018; Song *et al.,* 2022; Udakis *et al.,* 2020; Wang & Maffei, 2014; Woodin *et al.,* 2003). Experiments have shown that inhibitory plasticity can be induced via high-frequency stimulation of input pathways (Caillard *et al.,* 1999; Shew *et al.,* 2000) or pairing of presynaptic and postsynaptic spikes (D'amour & Froemke, 2015; Haas *et al.,* 2006; Holmgren & Zilberter, 2001; Udakis *et al.,* 2020; Woodin *et al.,* 2003).

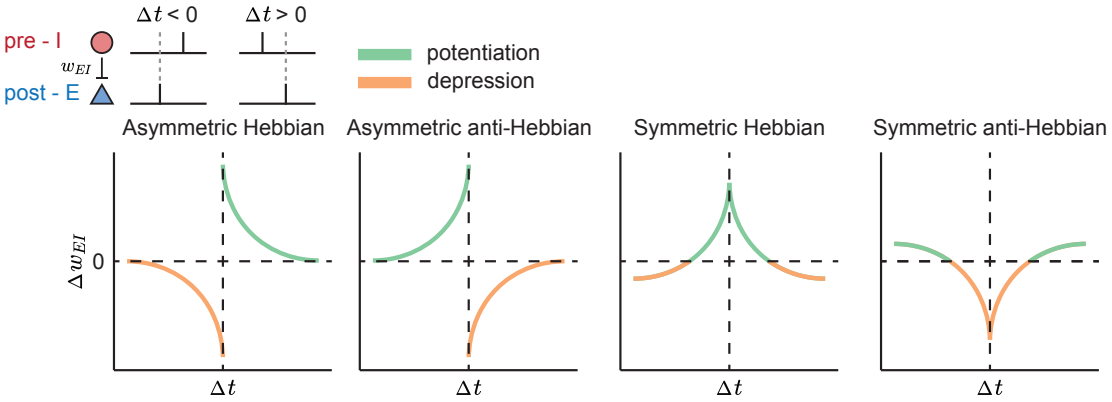


Figure 1.4: Different learning windows of inhibitory spike-timing-dependent plasticity. Inhibitory plasticity can be parameterized into different learning windows as a function of the timing difference between pre- and postsynaptic spikes Δt, leading to either inhibitory long-term potentiation ($\Delta w_{EI} > 0$, green) or inhibitory long-term depression ($\Delta w_{EI} < 0$, orange): asymmetric Hebbian (Kleberg *et al.,* 2014; Luz & Shamir, 2012), asymmetric anti-Hebbian (Kleberg *et al.,* 2014), symmetric Hebbian (Schulz *et al.,* 2021; Vogels *et al.,* 2011), and symmetric anti-Hebbian (Agnes *et al.,* 2020). Figure adapted from Wu *et al.,* 2022.

Computational studies have implemented both, rate-based inhibitory plasticity

(Bourjaily & Miller, 2011; Clopath *et al.,* 2016; Miehl & Gjorgjieva, 2022; Pedrosa & Clopath, 2020; Vogels *et al.,* 2011) or spike-based inhibitory plasticity. Following from the experimental diversity of inhibitory STDP window shapes (Hennequin *et al.,* 2017), computational studies have investigated asymmetric Hebbian (Kleberg *et al.,* 2014; Luz & Shamir, 2012, 2014), asymmetric anti-Hebbian (Kleberg *et al.,* 2014), symmetric Hebbian (Vogels *et al.,* 2011), and symmetric anti-Hebbian (Agnes *et al.,* 2020) shapes (Fig. 1.4).

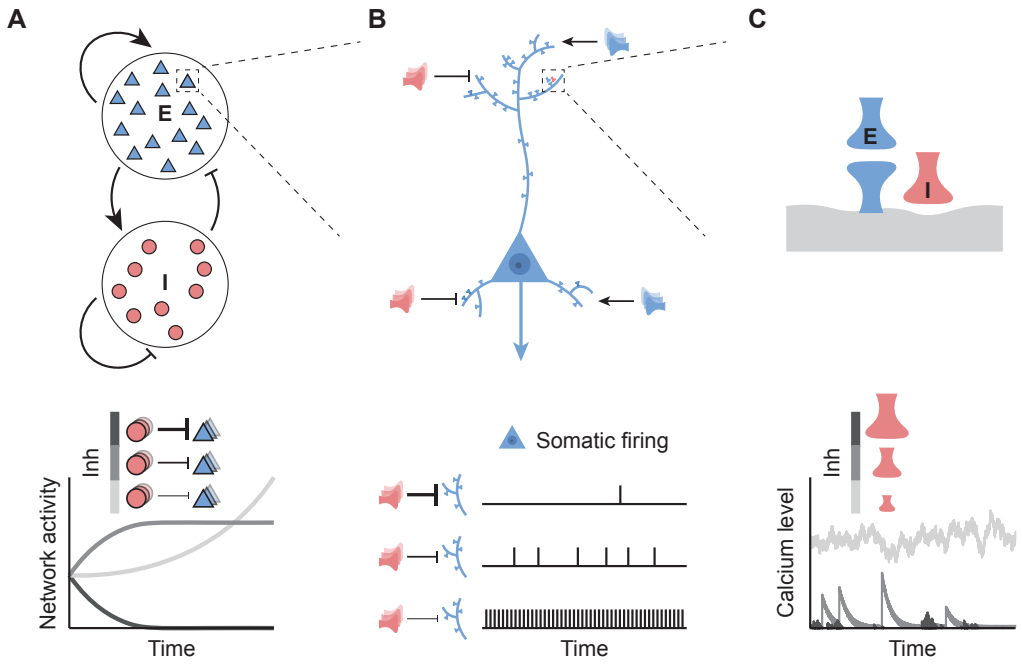


Figure 1.5: Inhibitory control of excitation at different scales. A. At the network level (top), inhibition has a significant effect on network activity (bottom). Excessive inhibition can silence network activity, and insufficient inhibition can lead to the explosion of network activity, while an appropriate amount of inhibition stabilizes network dynamics and maintains network activity at a steady level. **B.** At the single neuron level (top), inhibition has a significant effect on somatic firing (bottom). Excessive inhibition generates very little spiking, insufficient inhibition leads to high levels of spiking, while an appropriate amount of inhibition leads to biologically realistic spiking levels. **C.** At the dendritic level (top), inhibition influences the local calcium level (bottom). Excessive inhibition leads to an extremely low calcium level locally on the dendrite, insufficient inhibition leads to extraordinarily high local calcium level, while an appropriate amount of inhibition leads to an appropriate local calcium level. Figure adapted from Wu *et al.,* 2022.

As outlined in detail in our review (Wu *et al.,* 2022) (see Chapter 3.3), inhibitory plasticity has mostly been suggested to control excitation at different spatiotemporal scales, from the whole network level, single cell level, and the level of dendritic branches (Fig. 1.5). At the network level, inhibition maintains excitatory neurons

in a state of neither hyper- nor hypo-excitability (Turrigiano & Nelson, 2004) (Fig. 1.5A). Hence, inhibitory plasticity operates as a homeostatic mechanism to control network firing rate levels (Gainey & Feldman, 2017). Various computational studies have proposed how inhibitory plasticity regulates firing rates at the single neuron level. They all rely on a negative-feedback mechanism, in which high postsynaptic excitatory firing rates lead to inhibitory LTP, while low postsynaptic excitatory firing rates lead to inhibitory LTD (Akil *et al.,* 2021; Kleberg *et al.,* 2014; Luz & Shamir, 2012; Vogels *et al.,* 2011). Besides regulating firing rates, inhibitory plasticity has a more nuanced role in controlling the firing patterns of single neurons (Fig. 1.5B). Inhibitory plasticity can affect spike generation and spiking statistics, e.g. spike regularity and pairwise correlations (Brunel, 2000; Cardin, 2018; Carvalho & Buonomano, 2009). On the smallest scale, the level of inhibition can control local dendritic calcium levels (Chiu *et al.,* 2013; Hattori *et al.,* 2017; Inglebert *et al.,* 2020) (Fig. 1.5C). Since, as mentioned above, excitatory plasticity depends on factors like pre- and postsynaptic firing rates and spike timing, inhibition has been suggested to affect not only the activity of excitatory neurons but also plasticity at excitatory synapses (Chiu *et al.,* 2019; Hattori *et al.,* 2017; Paulsen & Moser, 1998; Vogels *et al.,* 2013). While these studies suggest a prominent role of inhibition and inhibitory plasticity in controlling excitatory firing rates and synaptic plasticity, a detailed understanding of the interaction of excitatory and inhibitory plasticity have been missing.

2 Methods and mathematical framework

In my work, I used computational models and mathematical approaches of different types. In the following I provide an overview of the methods used in my work. The types of models I used can broadly be split into the type of network, feedforward or recurrent (Fig. 2.1) and the type of neuron model, spiking or rate-based. While spiking neuron models are more biologically-plausible approximations of neuron dynamics, rate-based models allow for a more intuitive mathematical formalism. The choice of the neuron model also determines which implementation of synaptic plasticity I chose, either based on the exact timings of the pre- and postsynaptic spikes or the instantaneous rates of the pre- and postsynaptic neuron.

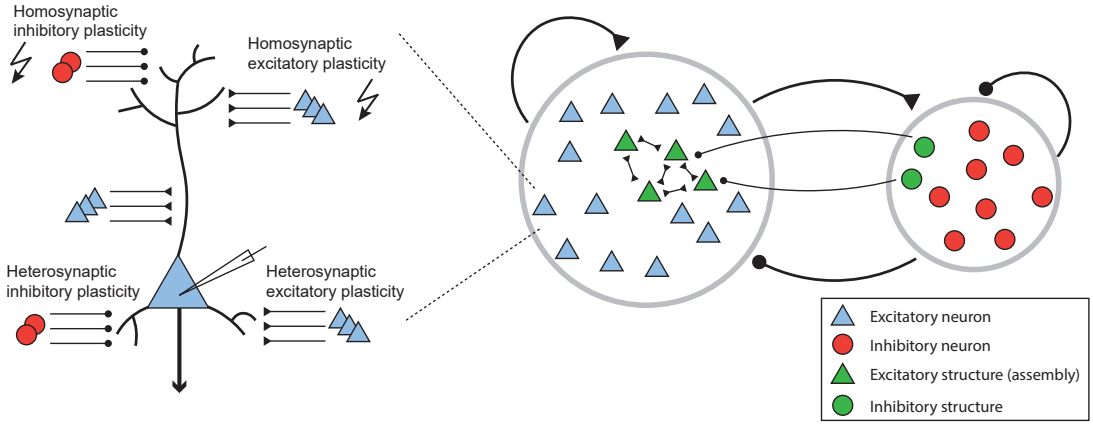


Figure 2.1: Two levels of modeling: single neuron and network level. Left: A single excitatory neuron receives input from excitatory neurons (blue triangles) and inhibitory neurons (red circles). Stimulation of input pathways can induce (homosynaptic) excitatory and inhibitory plasticity, while non-stimulated input pathways can undergo heterosynaptic excitatory and inhibitory plasticity. **Right:** A recurrently connected network of excitatory (blue triangles) and inhibitory (red circles) neurons. In the network, strong recurrent connections among a subset of neurons indicates excitatory structures, so called 'assemblies' (green triangles) and inhibitory structures (green circles).

Firing rate-based models and their plasticity mechanisms

The following description is a summarized version of the Methods used in Miehl & Gjorgjieva, 2022, for more details on the exact implementation and parameters I refer the reader to the article (see Chapter 3.1).

Rate-based neuron models describe how the firing rate of a neuron changes over time. Here, I consider a feedforward neuron model, meaning that a single excitatory neuron receives input from presynaptic excitatory and inhibitory neurons (Fig. 2.1; left). The postsynaptic firing rate ν^E then changes according to:

$$\tau_{FR}^E \dot{\nu}^E = -\nu^E + \left[\sum_{j=1}^{N^E} \rho_j^E w_j^{EE} - \sum_{k=1}^{N^I} \nu_k^I w_k^{EI} \right]_+ , \tag{2.1}$$

where index j refers to all presynaptic excitatory neurons and index k refers to all presynaptic inhibitory neurons. The postsynaptic neuron receives input from N^E presynaptic excitatory neurons with firing rates ρ_j^E through synapses with weight strength w_j^{EE}, and N^I presynaptic inhibitory neurons with firing rates ν_k^I through synapses with weight strength w_k^{EI}. The bracket $[]_+$ denotes a rectification that sets negative values to zero and τ_{FR}^E represents the time constant of excitatory firing rate dynamics.

The firing rate of inhibitory neurons ν_k^I follow similar dynamics:

$$\tau_{FR}^I \dot{\nu}_k^I = -\nu_k^I + \left[\sum_{j=1}^{N^E} \rho_j^E w_j^{IE} + \rho_k^I \right]_+ . \tag{2.2}$$

Here, the inhibitory neurons are driven by the same N^E presynaptic excitatory populations with firing rates ρ_j^E through synapses with weights w_j^{IE} and additional external input with firing rate ρ_k^I. The time constant of inhibitory firing rate dynamics is denoted by τ_{FR}^I.

Following from experimental evidence which suggests that excitatory plasticity depends nonlinearly on the postsynaptic firing rate ν^E (Cooper & Bear, 2012; Kirk-

wood *et al.,* 1996; Philpot *et al.,* 2003), I model (homosynaptic) plasticity at E-to-E synaptic connections w^{EE} as (note that for clarity we drop here the index j):

$$\tau_w^E \dot{w}^{EE} = \rho^E \nu^E \left(\nu^E - c_{post}^E \right), \tag{2.3}$$

where τ_w^E is the timescale of excitatory plasticity (Fig. 1.3A). The variable c_{post}^E is the postsynaptic excitatory LTD/LTP thresholds, which refers to the postsynaptic rate at which excitatory plasticity changes sign.

Similarly, the (homosynaptic) inhibitory plasticity can be described as a function of pre- and postsynaptic firing rates. A classic implementation of plasticity at I-to-E synapses has a linear dependency on the postsynaptic firing rate ν^E (Clopath *et al.,* 2016; Vogels *et al.,* 2011):

$$\tau_w^I \dot{w}^{EI} = \nu^I \left(\nu^E - c_{post}^I \right), \tag{2.4}$$

with timescales of inhibitory plasticity τ_w^I and postsynaptic inhibitory LTD/LTP threshold c_{post}^I. In Miehl & Gjorgjieva, 2022, we suggest a novel form of inhibitory plasticity which, similarly to excitatory plasticity, also depends nonlinearly on the postsynaptic firing rate ν^E:

$$\tau_w^I \dot{w}^{EI} = \nu^I \nu^E \left(\nu^E - c_{post}^I \right). \tag{2.5}$$

The advantage of formulating neuronal dynamics and synaptic plasticity change as a function of firing rates is that it makes the system analytically tractable.

Spiking-based models and their plasticity mechanisms

In contrast to rate-based models, spiking-based models are neuron models which include a spiking mechanism. In my work, I have used three different types of conductance-based spiking models: the leaky integrate-and-fire (LIF) neuron model (Field *et al.,* 2020), the exponential leaky integrate-and fire (EIF) neuron model (Schulz *et al.,* 2021) and the Poisson neuron model (Montangie *et al.,* 2020). In the following, I provide a short overview of these spiking neuron models. For

more details on the implementation and parameters, I refer the reader to the respective articles (see Chapters 3.2, 3.4 and 3.6).

(Exponential) leaky integrate-and-fire neuron model

In LIF and EIF models, the neuron 'spikes' when the membrane potential reaches a certain threshold, and the membrane potential dynamics is modeled based on the following equation (following from Litwin-Kumar & Doiron, 2014; Schulz *et al.,* 2021):

$$\begin{aligned} C\frac{d}{dt}V(t) = &- g_L(V(t) - V_{\text{rest}}) - g^{EE}(t)(V(t) \\ &+ g_L\Delta_T \exp\left(\frac{V(t) - V_T}{\Delta_T}\right) - V^E_{\text{rev}}) - g^{EI}(t)(V(t) - V^I_{\text{rev}}), \end{aligned} \tag{2.6}$$

where $V(t)$ is the membrane potential of the modeled neuron, C the membrane capacitance, g_L the membrane conductance, V_{rest} is the neuron resting potential, $V^{E/I}_{\text{rev}}$ is the excitatory, respectively inhibitory, reversal potential, Δ_T is the slope factor of the exponential rise and V_T the spiking threshold. The terms $g^{EE}(t)$ and $g^{EI}(t)$ are the excitatory or inhibitory conductances:

$$g^{XY}_i(t) = F^Y(t) * \left(J^{XY}_{ext} s^{XY}_{i,ext}(t) + \sum_j J^{XY}_{ij} s^Y_j(t)\right). \tag{2.7}$$

Here, X and Y either represent an excitatory or inhibitory population $(X, Y \in [E, I])$. $s^Y_j(t)$ is the spike train of neuron j in the network and $s^{XY}_{i,ext}$ denotes the spike train of the external input to neuron i, and J^{XY}_{ij} the synaptic strength from recurrent neurons and J^{XY}_{ext} the synaptic strength from the input neurons to the network neurons (instead of the parameter J, I sometimes use the parameter w to indicate synaptic strength). The function $F^Y(t)$ describes the synaptic kernel, which is modeled as a difference of two exponential functions

$$F^Y(t) = \frac{\exp^{\frac{-t}{\tau^Y_{\text{decay}}}} - \exp^{\frac{-t}{\tau^Y_{\text{rise}}}}}{\tau^Y_{\text{decay}} - \tau^Y_{\text{rise}}}. \tag{2.8}$$

with rise and decay times τ^Y_{decay} and τ^Y_{rise}. In Field *et al.,* 2020 I have used a simple exponential decay for the synaptic kernel.

Above equations hold for both, excitatory and inhibitory neurons. However, usually only excitatory neurons are modeled as EIF neurons, and inhibitory neurons as LIF neurons and the parameters between excitatory and inhibitory neurons differ (see publications Field *et al.,* 2020; Schulz *et al.,* 2021 for details). In LIF neurons, the term $g_L \Delta_T \exp\left(\frac{V(t)-V_T}{\Delta_T}\right)$ in Eq. 2.6 is not included. The EIF neuron model is considered to have a more realistic spike generation mechanism compared to the LIF neuron model (Fourcaud-Trocmé *et al.,* 2003).

Spike-timing-dependent plasticity (STDP)

A prominent implementation of synaptic plasticity in spiking neuron models is spike-timing-dependent plasticity (STDP). STDP rules describe how synaptic weights change based on the relative time-difference of the pre- and postsynaptic spike (Fig. 1.3B,C). In my articles, I have used two forms of STDP rules: the pairwise STDP (Field *et al.,* 2020) and the triplet STDP rule (Montangie *et al.,* 2020; Schulz *et al.,* 2021). In contrast to the pairwise STDP rule, the triplet STDP rule can capture the experimentally-verified dependency of plasticity on the firing rates (Gjorgjieva *et al.,* 2011; Pfister & Gerstner, 2006; Sjöström *et al.,* 2001).

In the triplet STDP rule, four spike accumulators, r_1, r_2, o_1, and o_2, increase by one, once a spike at the presynaptic or the postsynaptic neuron occurs at time t^{pre} or t^{post} and otherwise decrease exponentially depending on their respective time constant τ_+, τ_x, τ_-, and τ_y (Pfister & Gerstner, 2006; Schulz *et al.,* 2021):

$$\begin{aligned}
\frac{dr_1(t)}{dt} &= -\frac{r_1(t)}{\tau_+} \text{ if } t = t^{\text{pre}} \text{ then } r_1 \rightarrow r_1 + 1, \\
\frac{dr_2(t)}{dt} &= -\frac{r_2(t)}{\tau_x} \text{ if } t = t^{\text{pre}} \text{ then } r_2 \rightarrow r_2 + 1, \\
\frac{do_1(t)}{dt} &= -\frac{o_1(t)}{\tau_-} \text{ if } t = t^{\text{post}} \text{ then } o_1 \rightarrow o_1 + 1, \\
\frac{do_2(t)}{dt} &= -\frac{o_2(t)}{\tau_y} \text{ if } t = t^{\text{post}} \text{ then } o_2 \rightarrow o_2 + 1.
\end{aligned} \tag{2.9}$$

The weight update is then implemented as

$$\begin{aligned} \Delta J(t) &= -o_1(t)[A_2^- + A_3^- r_2(t-\epsilon)] \text{ if } t = t^{\text{pre}}, \\ \Delta J(t) &= r_1(t)[A_2^+ + A_3^+ o_2(t-\epsilon)] \text{ if } t = t^{\text{post}}, \end{aligned} \tag{2.10}$$

with A^+, A^- corresponds to the LTP or LTD amplitude, and the subscript refers to the triplet (3) or pairwise term (2). The parameter ϵ ensures that the weights are updated before the spike accumulators are updated. Oftentimes, the 'minimal' triplet STDP rule is used, where A_3^- is set to zero (Fig. 1.3C). To implement the pairwise STDP rule, both triplet amplitudes A_3^- and A_3^+ are set to zero (Field *et al.,* 2020; Gerstner *et al.,* 1996) (Fig. 1.3B). The sign of A_2^- and A_2^+ determines the shape of the STDP window, as shown in Fig. 1.4.

At inhibitory synapses, we used a slightly different implementation of the pairwise STDP window (Schulz *et al.,* 2021). Here, the STDP window is symmetric and downshift, which has been first implemented by Vogels *et al.,* 2011 (Fig. 1.4; symmetric Hebbian). Here, the weight updated follows

$$\begin{aligned} \Delta J_{ij}(t) &= o_1(t) - 2r_0\tau_y \text{ if } t = t^{\text{pre}} \\ \Delta J_{ij}(t) &= r_1(t) \text{ if } t = t^{\text{post}}, \end{aligned} \tag{2.11}$$

where r_0 corresponds to the target firing rate of the excitatory neuron.

In Field *et al.,* 2020 we have used a inhibitory STDP window without any downshift, i.e. $r_0 = 0$.

Heterosynaptic plasticity

Heterosynaptic plasticity is defined as synaptic plasticity at synapses which where not directly activated presynaptically by the plasticity protocol (Fig. 2.1; left). Multiple alternative models of heterosynaptic plasticity have been suggested (Chen *et al.,* 2013; Chistiakova *et al.,* 2015; Jedlicka *et al.,* 2015; Triesch *et al.,* 2018; Volgushev *et al.,* 2016). In my model of excitatory and inhibitory heterosynaptic plasticity, we implement the mechanism in close correspondence to experimental findings in L5 pyramidal neurons in the mouse primary auditory cortex, for details see Field *et al.,* 2020 (see Chapter 3.4).

Heterosynaptic potentiation or depression of synapse j is modeled based on an in-

ternal trace $T_j^{E/I}$, which increases for each incoming spike by $T_j^{E/I} \rightarrow T_j^{E/I} + w_j^{E/I}$ and otherwise decreased $\tau_T^{E/I} \frac{dT_j^{E/I}}{dt} = -T_j^{E/I}$ with time constant $\tau_T^{E/I}$. The weights are then updated based on the mean trace per input channel, following: $w_{c,max}^{E/I} \rightarrow w_{c,max}^{E/I} - \eta_{het}^{E/I}[T_c^{E/I}]_{max}$. Here η_{het} is the learning rate of heterosynaptic plasticity, c corresponds to the input channel and max refers to the strongest input channel. To ensure that heterosynaptic plasticity is only induced if also homosynaptic plasticity is induced, we used a learning dependent trace $T_{eLTP} \rightarrow T_{eLTP} + \Delta w_j^E$ which accumulates the excitatory LTP at synapse j, and decreases otherwise $\tau_{TeLTP} \frac{dT_{eLTP}}{dt} = -T_{eLTP}$. When T_{eLTP} reaches the threshold θ_{on}, heterosynaptic plasticity is switched "on", and if T_{eLTP} falls below the threshold θ_{off}, heterosynaptic plasticity is switched "off".

Poisson neuron model and the structural motif framework

Instead of explicitly modeling the membrane potential dynamics as in the LIF or EIF neuron model, an alternative is the Poisson neuron model. Here, spiking activity follows an inhomogenious Poisson process, where the instantaneous firing rate $\lambda(t)$ is the basis of random spike initiation. Considering a network of linear Poisson neurons allows to use the analytical approach of Hawkes processes (Hawkes, 1971). This approach allows to decompose STDP rules into contribution of 'structural motifs', which describe how a spike influences the correlation between two neurons (Jovanović & Rotter, 2016; Montangie *et al.,* 2020; Ravid Tannenbaum & Burak, 2016; Trousdale *et al.,* 2012). We have used this approach to describe the contribution of higher-order correlations on weight changes based on the triplet STDP rule (Montangie *et al.,* 2020). Based on this framework, one can derive the mean synaptic weight change $\langle \dot{w}_{ij} \rangle$ as a sum of structural motif terms of different 'order of contribution' (as we describe in detail in Montangie *et al.,* 2020 and in a more general way in Miehl *et al.,* 2022, see Chapters 3.5 and 3.6):

$$\langle \dot{w}_{ij} \rangle = \langle \dot{w}_{ij} \rangle^{(0)} + \langle \dot{w}_{ij} \rangle^{(1)} + \langle \dot{w}_{ij} \rangle^{(2)} + \dots \tag{2.12}$$

In this framework, the first term describes the contribution of mean pre- and postsynaptic firing rates (r_i, r_j) to the mean synaptic weight change:

$$\langle \dot{w}_{ij} \rangle^{(0)} = r_i r_j M_0, \tag{2.13}$$

where M_0 is the zero-order motif coefficient, which is defined as the area under the curve of the plasticity rule (i.e. integral over the STDP window in Fig. 1.3B). The second term describes the first-order structural motif, how a spike in either the pre- or postsynaptic neuron affects the mean synaptic weight change

$$\langle \dot{w}_{ij} \rangle^{(1)} = r_j w_{ij} M_{1,0} + r_i w_{ji} M_{0,1}, \tag{2.14}$$

where $M_{1,0}$ and $M_{0,1}$ are calculated based on the shape of the EPSC and the STDP parameters (see Montangie *et al.,* 2020 for details).

3 Results

In this dissertation, I investigate the functional roles of synaptic plasticity, with a specific focus on inhibitory plasticity. In total I have contributed to six peer-reviewed journal articles. I am the first or co-first author on four of these articles, where two are original research articles (Miehl & Gjorgjieva, 2022; Schulz *et al.,* 2021) and two articles are reviews (Miehl *et al.,* 2022; Wu *et al.,* 2022), and I am a contributing author on two original research articles (Field *et al.,* 2020; Montangie *et al.,* 2020). In my work, I

1. show how a novel form of inhibitory plasticity, which depends nonlinearly on the postsynaptic firing rate, is a sufficient mechanism to homeostatically stabilize firing rate and excitatory weight dynamics (Miehl & Gjorgjieva, 2022);
2. show based on my work together with my co-author Auguste Schulz how inhibitory plasticity can affect the neuronal responses to familiar versus novel stimuli (Schulz *et al.,* 2021);
3. summarize together with my co-author Yue Kris Wu the literature on how inhibitory plasticity regulates circuit organization and function (Wu *et al.,* 2022);
4. show together with our experimental collaborators how the interaction of homosynaptic and heterosynaptic plasticity at excitatory and inhibitory synapses can determine a set-point for E/I balance (Field *et al.,* 2020);
5. summarize together with my co-author Sebastian Onasch the state-of-the-art literature on the formation of assemblies (Miehl *et al.,* 2022);
6. show as a contributing author together with Dr. Lisandro Montangie how assemblies can spontaneously emerge via the triplet spike-timing-dependent plasticity rule (Montangie *et al.,* 2020).

In the following, I provide a one-page summary for each of these articles, indicate my contribution and include the full article in the Appendix. Finally, I will discuss the relevance, connections and future research directions following from the work contained in this dissertation.

3.1 Stability and learning in excitatory synapses by nonlinear inhibitory plasticity

In Miehl & Gjorgjieva, 2022, we investigate a novel inhibitory plasticity mechanism leading to stability and flexible learning in excitatory synapses. We find that:

1. A (classically studied) linear inhibitory plasticity mechanism fails to robustly stabilize synaptic weight dynamics.
2. Our newly proposed nonlinear inhibitory plasticity mechanism can robustly counteract runaway dynamics of excitatory synaptic weights.
3. We identify two components of robust stabilization of weight dynamics via inhibitory plasticity: dominant inhibition over excitation, and overlapping excitatory and inhibitory LTD/LTP thresholds. Furthermore, we suggest how the assumption of overlapping excitatory and inhibitory LTD/LTP thresholds can be relaxed based on a novel dynamic matching mechanism.
4. We show that inhibitory plasticity regulates the network response to input perturbations, link it to the phenomenon of metaplasticity and to experiments studying sensory deprivation.
5. We further study how the nonlinear inhibitory plasticity mechanism establishes E/I balance and leads to a fixed excitatory to inhibitory weight ratio, similarly as described in experimental results.
6. Finally, we study additional functional consequences of the nonlinear inhibitory plasticity mechanism, showing that feedforward receptive fields and recurrent assembly structures can be gated via a disinhibitory signal.

The work has been completed together with my supervisor Prof. Dr. Julijana Gjorgjieva. I contributed to each section and figure in the article: conceptualization, formal analysis, investigation, methodology, software, visualization, writing - original draft, writing - review and editing.

The full article was published on 2 December 2022 in PLoS Computational Biology and is reproduced in *Appendix I. Stability and learning in excitatory synapses by nonlinear inhibitory plasticity* under the Creative Commons Attribution 4.0 International License.

3.2 The generation of cortical novelty responses through inhibitory plasticity

In Schulz *et al.* (2021), we investigate how inhibitory plasticity can explain differences in neuronal responses to familiar or novel stimuli. We find that:

1. In an a recurrent spiking network model, inhibitory plasticity leads to adaptation of responses to familiar stimuli, while responses to novel stimuli remain high.
2. The novelty response amplitude depends on the exact choice of the stimulation paradigm, in qualitative agreement with experimental findings.
3. Periodicity of the familiar stimuli presentation is not required for the generation of a novelty response, allowing us to formulate experimentally-testable predictions.
4. Inhibitory plasticity allows for adaptation on multiple timescales.
5. The timescale of recovery from adaptation depends on the inter-repetition interval.
6. Stimulus-specific adaptation follows from inhibitory plasticity and tuning of both excitatory and inhibitory neurons.
7. A disinhibitory mechanism can flexibly amplify novelty responses.

The work has been completed together with my co-author Auguste Schulz, my supervisor Prof. Dr. Julijana Gjorgjieva and our experimental collaborator Prof. Dr. Michael J. Berry II. I contributed in equal parts with Auguste Schulz in each section and figure in the article to: conceptualization, resources, software, formal analysis, investigation, visualization, methodology, writing - original draft, writing - review and editing.

The full article was published on 14 October 2021 in eLife and is reproduced in *Appendix II. The generation of cortical novelty responses through inhibitory plasticity* under the Creative Commons Attribution 4.0 International License.

3.3 Regulation of circuit organization and function through inhibitory synaptic plasticity

In Wu *et al.,* 2022, we review the recent literature on the functional role of inhibitory plasticity. We:

1. provide an overview of inhibition throughout development and adulthood.
2. show how inhibitory plasticity controls excitation at different spatiotemporal scales.
3. summarize how inhibition and inhibitory plasticity controls excitatory plasticity.
4. provide evidence that inhibitory plasticity is an important component in the formation of structured networks and their computation.
5. review recent interneuron-specific plasticity mechanisms and their functional implications.
6. provide future perspectives and formulate outstanding questions.

The work has been completed together with my co-author Yue Kris Wu and my supervisor Prof. Dr. Julijana Gjorgjieva. I contributed in equal parts with my co-author Yue Kris Wu in each section and figure in the article to: conceptualization, visualization, writing - original draft, writing - review and editing.

The full article was published on 28 October 2022 in Trends in Neuroscience and is reproduced in *Appendix III. Regulation of circuit organization and function through inhibitory synaptic plasticity* under Creative Commons CC-BY License.

3.4 Heterosynaptic Plasticity Determines the Set Point for Cortical Excitatory-Inhibitory Balance

In Field *et al.,* 2020, we investigate together with our experimental collaborators how interacting homosynaptic and heterosynaptic plasticity at excitatory and inhibitory synapses determines the set-point of excitatory-inhibitory balance. We find that:

1. Heterosynaptic plasticity is induced at excitatory and inhibitory synapses in parallel to homosynaptic plasticity via a paring experiment in layer 5 of mouse auditory cortex.
2. The strongest excitatory and inhibitory inputs onto layer 5 neurons change most drastically via heterosynaptic plasticity.
3. Based on a probabilistic and a biophysical computational model, we find that the experimentally-verified heterosynaptic plasticity mechanism determines the set-point for excitatory-inhibitory balance.
4. In our model, the relative strength of homosynaptic to heterosynaptic plasticity determines the exact set-point of excitatory-inhibitory balance, suggesting a different set-point in adult and young mice.
5. Experiments indicate that calcium release from internal stores is necessary for heterosynaptic plasticity to be induced.

The article is a collaborative work together with experimentalists (Rachel E. Field, Dr. James A. D'amour, Prof. Dr. Robin Tremblay, Prof. Dr. Bernardo Rudy and Prof. Dr. Robert C. Froemke). Together with my supervisor Prof. Dr. Julijana Gjorgjieva, I contributed to the article the computational model (Fig. 3, Suppl. Fig. S7), and to the writing of the manuscript and the reviewing process.

The full article was published on 3 June 2020 in Neuron and is reproduced in *Appendix IV. Heterosynaptic Plasticity Determines the Set Point for Cortical Excitatory-Inhibitory Balance* under the Creative Commons Attribution - Non Commercial - No Derivatives 4.0 International (CC BY-NC-ND 4.0) License.

3.5 Formation and computational implications of assemblies in neural circuits

In Miehl *et al.,* 2022, we review the recent computational literature on the formation and functional implications of assemblies. We:

1. identify the four basic building blocks of assembly formation: synaptic plasticity, symmetry breaking, competition, stability.
2. review experimental evidence of ensembles (group of neurons with correlated activity) and assemblies (group of strong recurrently connected neurons).
3. provide a detailed understanding which types of plasticity rules favor assembly formation.
4. show why symmetry breaking and competition are key components of assembly formation.
5. discuss the importance of stability of representations.
6. provide an overview of recent ideas on the functional role of assemblies.

The work has been completed together with my co-author Sebastian Onasch, with contributing author Dr. Dylan Festa and my supervisor Prof. Dr. Julijana Gjorgjieva. I contributed in equal parts with my co-author Sebastian Onasch in each section and figure in the article to: conceptualization, visualization, writing - original draft, writing - review and editing.

The full article was published on 6 September 2022 in Journal of Physiology and is reproduced in *Appendix V. Formation and computational implications of assemblies in neural circuits* under the Creative Commons Attribution 4.0 International License.

3.6 Autonomous emergence of connectivity assemblies via spike triplet interactions

In Montangie *et al.,* 2020, we investigate the formation of neuronal assemblies without structured external input. We find that:

1. Based on the Hawkes framework, the second and third order cumulant can be reformulated in terms of motif contribution of up to any order, allowing analytical tractability of synaptic weight dynamics based on the triplet spike-timing-dependent plasticity rule.
2. Assembly structures emerge spontaneously when the triplet STDP rule, or the synaptic transmission is modulated.
3. We identify which motifs contribute to the spontaneous formation of assemblies.

The work has been done together with lead author Dr. Lisandro Montangie and my supervisor Prof. Dr. Julijana Gjorgjieva. To this publication I contributed to each section and specifically figures Fig. 5,8,11 and supplementary figures S1-S5 in the article to: formal analysis, investigation, methodology, software, visualization, writing - review and editing.

The full article was published on 8 May 2020 in PLoS Computational Biology and is reproduced in *Appendix VI. Autonomous emergence of connectivity assemblies via spike triplet interactions* under the Creative Commons Attribution 4.0 International License.

4 Discussion

In the articles containing my dissertation, I have used multiple different computational and analytical approaches to answer questions related to the functional role of different plasticity mechanisms, with a specific focus on inhibitory plasticity. In the sections below, I discuss the key points following from my work.

Inhibitory plasticity as a homeostatic mechanism and a solution to the stability-flexibility problem

A problem in computational models is the trade-off between stability and flexibility of neuronal connectivity. On the one side, stimulus representations need to remain stable to ensure reliable interpretation and long-term storage of those stimuli, while on the other hand flexible re-learning, or learning of new representations should remain possible (Fusi, 2017). The question of how connectivity structures can remain stable is tightly linked to the problem of Hebbian runaway dynamics of excitatory plasticity.

In my work, I identify the conditions under which inhibitory plasticity is sufficient to stabilize excitatory weight and rate dynamics (Miehl & Gjorgjieva, 2022). A key component is the 'dominance' of inhibitory over excitatory weight dynamics, meaning that inhibitory synapses change with a higher magnitude than excitatory synapses. This condition is met by our newly proposed nonlinear inhibitory plasticity rule. Interestingly, experiments have found that inhibitory synapses do change more drastically than excitatory synapses when performing plasticity-induction-protocols (D'amour & Froemke, 2015). The second component to ensure stability is that the LTD/LTP thresholds of excitatory and inhibitory plasticity need to be matched, either by having the exact same value or by having a dynamic matching mechanism (as we suggest in Miehl & Gjorgjieva, 2022 or as has been discussed in Keck *et al.,* 2017). These two components ensure stability in the system, i.e. that

for a certain postsynaptic firing rate no synaptic plasticity is induced, or in other words that plasticity is "off" if the neurons fire at the rate of the LTD/LTP threshold. Perturbations of the postsynaptic firing rates, however, can be a flexible way to gate synaptic plasticity "on". In my articles I highlight disinhibition as a possible mechanism allowing to gate synaptic plasticity and learning (Miehl & Gjorgjieva, 2022; Schulz *et al.,* 2021; Wu *et al.,* 2022). This is in line with multiple experimental results, suggesting disinhibition as a gating mechanism for synaptic plasticity and learning (Canto-Bustos *et al.,* 2022; Letzkus *et al.,* 2015). In my models, I stayed agnostic about how disinhibition is triggered, I simply apply an inhibitory input to the inhibitory population. In experiments, disinhibition can be induced either via neuromodulation (Froemke, 2015; Froemke *et al.,* 2007) or via a disinhibitory pathway involving vasoactive intestinal peptide (VIP)-expressing inhibitory neurons and somatostatin (SST)-expressing inhibitory neurons (Adler *et al.,* 2019; Canto-Bustos *et al.,* 2022; Hattori *et al.,* 2017; Krabbe *et al.,* 2019; Williams & Holtmaat, 2019). Therefore, we suggest that gating plasticity "on" or "off" via disinhibition is one stop towards solving the stability-flexibility problem.

Inhibitory plasticity also offers a solution to the 'temporal paradox' of integrating homeostatic mechanisms (like synaptic scaling) and Hebbian plasticity (Zenke & Gerstner, 2017; Zenke *et al.,* 2017). Since inhibitory plasticity acts on the same timescale as excitatory plasticity (D'amour & Froemke, 2015; Field *et al.,* 2020), it does not have the problem of mismatching timescales.

Inhibitory plasticity beyond homeostasis

In my work, I have extended the current understanding of the role of inhibitory plasticity. While previous computational models have mostly focused on inhibitory plasticity as a mechanism to stabilize firing rates and the E/I balance (Baker *et al.,* 2020; Kleberg *et al.,* 2014; Luz & Shamir, 2012; Rubin *et al.,* 2015; Vogels *et al.,* 2011), I could show that a novel nonlinear inhibitory plasticity mechanism can be sufficient to homeostatically regulate also excitatory synaptic weight dynamics (Miehl & Gjorgjieva, 2022). Furthermore, we suggest inhibitory plasticity as the mechanism underlying the difference in cortical responses to familiar and novel stimuli (Schulz *et al.,* 2021). Therefore, I extend recent efforts in computational studies to show additional functional properties of inhibitory plasticity beyond sta-

bilization of excitatory firing rates. Previous models have studied inhibitory plasticity in the context of emergence of receptive fields (Clopath *et al.,* 2016; Kleberg *et al.,* 2014; Luz & Shamir, 2012; Vogels *et al.,* 2011), place and grid fields (Weber & Sprekeler, 2018), ensuring diversity of tuning curves (Larisch *et al.,* 2021), and the formation of recurrent structures as neuronal assemblies (Litwin-Kumar & Doiron, 2014) and chain-like structures (Maes *et al.,* 2020; Zhang *et al.,* 2014).

My analysis indicates that inhibitory plasticity can control excitatory plasticity in a similar manner as the BCM metaplastic rule (Miehl & Gjorgjieva, 2022). This provides additional evidence for recent experimental and theoretical work, which suggests that inhibition can control the sign and amplitude of excitatory plasticity (Hiratani & Fukai, 2017; Paille *et al.,* 2013; Vogels *et al.,* 2013; Wang & Maffei, 2014; Wu *et al.,* 2022). For example, it has been shown in an experiment that application or blockage of inhibition leads to a shift of excitatory LTD/LTP thresholds (Steele & Mauk, 1999), similarly to a metaplastic learning rule. I, therefore, suggest that inhibition and specifically inhibitory plasticity might be one of the underlying mechanisms of the sliding LTD/LTP threshold as used in the BCM rule, potentially acting on fast timescales. The link between inhibitory plasticity and metaplasticity might be specifically interesting in relation to the problem of 'catastrophic forgetting', which is a well-described problem in artificial networks where previously learned tasks are forgotten when new tasks are learned (Parisi *et al.,* 2019). Metaplasticity is a promising candidate solving the problem of catastrophic forgetting (Jedlicka *et al.,* 2022).

Inhibitory plasticity as an alternative mechanism underlying the responses to familiar versus novel stimuli

Most computational studies modeling the different responses to familiar versus novel stimuli have assumed that short-term plasticity at feedforward input synapses is the underlying mechanism (Park & Geffen, 2020; Yarden & Nelken, 2017). Especially in the context of stimulus-specific adaptation, the 'Adaptation of Narrowly Tuned Modules' (ANTM) model has been extensively studied for adaptation to tone frequencies (Hershenhoren *et al.,* 2014; Mill *et al.,* 2011a,b, 2012; Taaseh *et al.,* 2011). In our work, we present an alternative mechanism underlying adaptation to repeated stimuli and generation of novelty responses, which is inhibitory plasticity

(Schulz *et al.,* 2021). The hypothesis that inhibition and inhibitory plasticity is an important mechanism to gain a full understanding of adaptive phenomena is supported by several recent experimental findings in the mammalian cortex (Chen *et al.,* 2015a; Garrett *et al.,* 2020; Hamm & Yuste, 2016; Heintz *et al.,* 2020; Kato *et al.,* 2015; Natan *et al.,* 2015, 2017), in Aplysia (Fischer *et al.,* 1997; Ramaswami, 2014), in *Drosophila melanogaster* (Das *et al.,* 2011; Glanzman, 2011) and by theoretical considerations (Barron *et al.,* 2017; Ramaswami, 2014). We especially suggest inhibitory plasticity being a key component for adaptation on longer timescales of several seconds to minutes and hours. Experimental studies find timescales of adaptation on the order of hundreds of milliseconds to seconds (Lundstrom *et al.,* 2010; Ulanovsky *et al.,* 2004) and in the context of habituation up to multiple days (Haak *et al.,* 2014; Ramaswami, 2014).

The detection of unexpected (or novel) stimuli is related to the predictive coding framework. In this framework, the brain is viewed as a prediction machine, where internal models of the world are constantly compared to incoming sensory inputs (Clark, 2013; Friston, 2018; Rao & Ballard, 1999). Usually, predictive coding is thought of being implemented based on two population of neurons, the prediction error neurons that signal the mismatch between the internal model and the sensory stimulus, and the neurons which reflect the internal model of the world. In our model, we suggest that primary sensory areas might allow for detection of unexpected stimuli without the need for an additional population of error neurons. Our network of excitatory and inhibitory neurons can represent the stimulus features and allow for detection of unexpected events via inhibitory plasticity (Schulz *et al.,* 2021).

Interaction of homo- and heterosynaptic plasticity

We show that the interaction of experimentally-verified homo- and heterosynaptic excitatory and inhibitory plasticity determines the set-point for cortical E/I balance (Field *et al.,* 2020). Based on our model, we suggest that the relative strength of homo- versus heterosynaptic plasticity mechanism can determine such a set-point, where dominant homosynaptic plasticity leads to high E/I ratio set-points and dominant heterosynaptic plasticity leads to low E/I ratio set-points. Interestingly, low and high E/I ratio set-points can be linked to the developmental stage of the

animal. Low (or uncorrelated) excitatory and inhibitory inputs correspond to an early developmental stage of the animal (Dorrn *et al.,* 2010). High correlations between excitatory and inhibitory tuning curves might reflect the maturity of the circuit, allowing for precise spiking of neurons to sensory stimuli. It can be speculated that high E/I correlations, or detailed E/I balance, is an important factor for learned representations, e.g. allowing for precise spiking in response for stimulus presentation.

Excitatory plasticity underlying assembly formation

In the context of excitatory plasticity, we extended the understanding of how assemblies can form under different circumstances. Following from the finding that in the zebrafish larvae optical tectum assemblies form even in the absence of structured external input (i.e. in the case of removed eyes) (Pietri *et al.,* 2017), we show that assemblies can form only based on internal correlations without structured external input (Montangie *et al.,* 2020). We extend existing computational studies which investigate assembly formation based on high external input rates (Clopath *et al.,* 2010; Litwin-Kumar & Doiron, 2014, 2012; Miconi *et al.,* 2016; Schulz *et al.,* 2021; Zenke *et al.,* 2015), correlations in the input spike trains (Gilson *et al.,* 2009b; Ocker & Doiron, 2019; Wu *et al.,* 2020) or without structured external input but with dominant rate-contributions in the plasticity rule (Babadi & Abbott, 2013; Burkitt *et al.,* 2007; Gilson *et al.,* 2009a; Manz & Memmesheimer, 2022; Ocker & Doiron, 2019; Ocker *et al.,* 2015). In a similar approach, Ravid Tannenbaum & Burak, 2016 have shown that assemblies can form without external structured input and a minimal rate-contribution, but the authors required a non-Hebbian STDP rule. We furthermore show that assembly formation based on external correlated input can follow from the interaction of excitatory and inhibitory plasticity, without the need for any fast homeostatic mechanism (Miehl & Gjorgjieva, 2022). This and other findings indicate the importance of inhibition and inhibitory plasticity in the formation of excitatory structures, as discussed in detail in our review Wu *et al.,* 2022.

Open Questions and Outlook

Following my work, many open questions and interesting future research directions emerge. A natural extension of my work is to combine homo- and heterosynaptic excitatory and inhibitory plasticity dynamics also with other mechanisms, like short-term plasticity and intrinsic plasticity. Also, evidence exists for plasticity at excitatory-to-inhibitory and inhibitory-to-inhibitory synapses. Recent computational studies have started implementing synaptic plasticity at all synapses into their models (Eckmann & Gjorgjieva, 2022; Mackwood *et al.,* 2021; Soldado-Magraner *et al.,* 2021), but how all of these mechanisms would interact in the case of e.g. familiar versus novel stimulus presentation is not clear.

Inhibitory neurons are diverse in their electrical properties, connectivity and function (Tremblay *et al.,* 2016), and also recent experimental studies suggest neuron-specific synaptic plasticity rules (Chen *et al.,* 2015b; Chiu *et al.,* 2018; Lagzi *et al.,* 2021; Song *et al.,* 2022; Udakis *et al.,* 2020; Yap *et al.,* 2021). Therefore, extending my work by studying the differential role of PV, SST, and potentially VIP neurons with neuron-specific plasticity rules should be the next step. Specifically adding the VIP-SST or VIP-PV pathway is a biologically-plausible way to induce disinhibition, acting as a gating mechanism for synaptic plasticity.

My models are based on the assumption that the neuronal dynamics can be approximated by point neurons. However, neurons have dendrites and evidence accumulates suggesting that dendrites are important for various computations (see e.g. Larkum, 2022; Poirazi & Papoutsi, 2020). For example, a recent study has highlighted the difference in bottom-up versus top-down signals in apical versus basal dendrites in the context of processing of unexpected (or novel) stimulus features (Gillon *et al.,* 2021). Other computational studies suggest synaptic plasticity is regulated locally on dendritic trees (Ebner *et al.,* 2019; Kirchner & Gjorgjieva, 2021). Especially to fully understand the interaction of excitatory and inhibitory plasticity, local calcium dynamics need to be considered, as shown experimentally (Hayama *et al.,* 2013; Paille *et al.,* 2013) and suggested in computational models (Agnes & Vogels, 2021; Hiratani & Fukai, 2017; Mikulasch *et al.,* 2021). Therefore, adding dendritic compartments to the models used in my work might be an interesting future perspective.

Bibliography

1. Abbott, L. F. & Nelson, S. B. Synaptic plasticity: taming the beast. *Nature Neuroscience* **3,** 1178–1183 (2000).
2. Abraham, W. C. Metaplasticity: Tuning synapses and networks for plasticity. *Nature Reviews Neuroscience* **9,** 387–399 (2008).
3. Abraham, W. C., Mason-Parker, S. E., Bear, M. F., Webb, S. & Tate, W. P. Heterosynaptic metaplasticity in the hippocampus in vivo: A BCM-like modifiable threshold for LTP. *Proceedings of the National Academy of Sciences* **98,** 10924–10929 (2001).
4. Adler, A., Zhao, R., Shin, M. E., Yasuda, R. & Gan, W. B. Somatostatin-Expressing Interneurons Enable and Maintain Learning-Dependent Sequential Activation of Pyramidal Neurons. *Neuron* **102,** 202–216 (2019).
5. Agnes, E. J., Luppi, A. I. & Vogels, T. P. Complementary inhibitory weight profiles emerge from plasticity and allow flexible switching of receptive fields. *Journal of Neuroscience* **40,** 9634–9649 (2020).
6. Agnes, E. J. & Vogels, T. P. Interacting synapses stabilise both learning and neuronal dynamics in biological networks. *bioRxiv* (2021).
7. Akil, A., Rosenbaum, R. & Josic, K. Balanced networks under spike-time dependent plasticity. *PLoS Computational Biology* **17,** e1008958 (2021).
8. Antunes, G. & Simoes-de-Souza, F. M. AMPA receptor trafficking and its role in heterosynaptic plasticity. *Scientific Reports* **8** (2018).
9. Babadi, B. & Abbott, L. F. Pairwise Analysis Can Account for Network Structures Arising from Spike-Timing Dependent Plasticity. *PLoS Computational Biology* **9,** e1002906 (2013).
10. Baker, C., Zhu, V. & Rosenbaum, R. Nonlinear stimulus representations in neural circuits with approximate excitatory-inhibitory balance. *PLOS Computational Biology* **16,** e1008192 (2020).
11. Barron, H. C., Vogels, T. P., Behrens, T. E. & Ramaswami, M. Inhibitory engrams in perception and memory. *Proceedings of the National Academy of Sciences* **114,** 6666–6674 (2017).
12. Bassett, D. S. & Sporns, O. Network neuroscience. *Nature Neuroscience* **20,** 353–364 (2017).

13. Benuskova, L. & Abraham, W. C. STDP rule endowed with the BCM sliding threshold accounts for hippocampal heterosynaptic plasticity. *Journal of Computational Neuroscience* **22,** 129–133 (2007).
14. Benuskova, L. & Jedlicka, P. Computational Modeling of Long-Term Depression of Synaptic Weights: Insights From STDP , Metaplasticity and Spontaneous Activity. *Neural Network World* **22,** 161–180 (2012).
15. Berkes, P., Orbán, G., Lengyel, M. & Fiser, J. Spontaneous Cortical Activity Reveals Hallmarks of an Optimal Internal Model of the Environment. *Science* **331,** 83–88 (2011).
16. Bhatia, A., Moza, S. & Bhalla, U. S. Precise excitation-inhibition balance controls gain and timing in the hippocampus. *eLife* **8,** e43415 (2019).
17. Bi, G.-g. & Poo, M.-M. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience* **18,** 10464–72 (1998).
18. Bienenstock, E. L., Cooper, L. N. & Munro, P. W. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. *The Journal of Neuroscience* **2,** 32–48 (1982).
19. Bliss, T. V. P. & Lømo, T. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *The Journal of Physiology* **232,** 331–356 (1973).
20. Bourjaily, M. A. & Miller, P. Excitatory, Inhibitory, and Structural Plasticity Produce Correlated Connectivity in Random Networks Trained to Solve Paired-Stimulus Tasks. *Frontiers in Computational Neuroscience* **5** (2011).
21. Brea, J. & Gerstner, W. Does computational neuroscience need new synaptic learning paradigms? *Current Opinion in Behavioral Sciences* **11,** 61–66 (2016).
22. Brown, S. P. & Hestrin, S. Intracortical circuits of pyramidal neurons reflect their long-range axonal targets. *Nature* **457,** 1133–1136 (2009).
23. Brunel, N. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of Computational Neuroscience* **8,** 183–208 (2000).
24. Burkitt, A. N., Gilson, M. & Van Hemmen, J. L. Spike-timing-dependent plasticity for neurons with recurrent connections. *Biological Cybernetics* **96,** 533–546 (2007).
25. Buzsáki, G. Neural syntax: cell assemblies, synapsembles, and readers. *Neuron* **68,** 362–385 (2010).
26. Caillard, O., Ben-Ari, Y. & Gaiarsa, J. L. Long-term potentiation of GABAergic synaptic transmission in neonatal rat hippocampus. *Journal of Physiology* **518,** 109–119 (1999).

27. Campagnola, L., Seeman, S. C., Chartrand, T., Kim, L., Hoggarth, A., Gamlin, C., Ito, S., Trinh, J., Davoudian, P., Radaelli, C., Kim, M.-H., Hage, T., Braun, T., Alfiler, L., Andrade, J., Bohn, P., Dalley, R., Henry, A., Kebede, S., Mukora, A., Sandman, D., Williams, G., Larsen, R., Teeter, C., Daigle, T. L., Berry, K., Dotson, N., Enstrom, R., Gorham, M., Hupp, M., Lee, S. D., Ngo, K., Nicovich, R., Potekhina, L., Ransford, S., Gary, A., Goldy, J., McMillen, D., Pham, T., Tieu, M., Siverts, L., Walker, M., Farrell, C., Schroedter, M., Slaughterbeck, C., Cobb, C., Ellenbogen, R., Gwinn, R. P., Keene, C. D., Ko, A. L., Ojemann, J. G., Silbergeld, D. L., Carey, D., Casper, T., Crichton, K., Clark, M., Dee, N., Ellingwood, L., Gloe, J., Kroll, M., Sulc, J., Tung, H., Wadhwani, K., Brouner, K., Egdorf, T., Maxwell, M., McGraw, M., Pom, C. A., Ruiz, A., Bomben, J., Feng, D., Hejazinia, N., Shi, S., Szafer, A., Wakeman, W., Phillips, J., Bernard, A., Esposito, L., D'Orazi, F. D., Sunkin, S., Smith, K., Tasic, B., Arkhipov, A., Sorensen, S., Lein, E., Koch, C., Murphy, G., Zeng, H. & Jarsky, T. Local Connectivity and Synaptic Dynamics in Mouse and Human Neocortex. *Science* **375** (2022).

28. Canto-Bustos, M., Friason, F. K., Bassi, C. & Oswald, A.-M. M. Disinhibitory circuitry gates associative synaptic plasticity in olfactory cortex. *The Journal of Neuroscience* **42,** 2942–2950 (2022).

29. Capogna, M., Castillo, P. E. & Maffei, A. The ins and outs of inhibitory synaptic plasticity: Neuron types, molecular mechanisms and functional roles. *European Journal of Neuroscience* **54,** 6882–6901 (2021).

30. Caporale, N. & Dan, Y. Spike Timing-Dependent Plasticity: A Hebbian Learning Rule. *Annual Review of Neuroscience* **31,** 25–46 (2008).

31. Cardin, J. A. Inhibitory Interneurons Regulate Temporal Precision and Correlations in Cortical Circuits. *Trends in Neurosciences* **41,** 689–700 (2018).

32. Carrillo-Reid, L., Han, S., Yang, W., Akrouh, A. & Yuste, R. Controlling Visually Guided Behavior by Holographic Recalling of Cortical Ensembles. *Cell* **178,** 447–457 (2019).

33. Carvalho, T. P. & Buonomano, D. V. Differential Effects of Excitatory and Inhibitory Plasticity on Synaptically Driven Neuronal Input-Output Functions. *Neuron* **61,** 774–785 (2009).

34. Chen, I.-W., Helmchen, F. & Lütcke, H. Specific early and late oddball-evoked responses in excitatory and inhibitory neurons of mouse auditory cortex. *Journal of Neuroscience* **35,** 12560–12573 (2015).

35. Chen, J.-Y., Lonjers, P., Lee, C., Chistiakova, M., Volgushev, M. & Bazhenov, M. Heterosynaptic Plasticity Prevents Runaway Synaptic Dynamics. *Journal of Neuroscience* **33,** 15915–15929 (2013).

36. Chen, J. L. & Nedivi, E. Highly specific structural plasticity of inhibitory circuits in the adult neocortex. *Neuroscientist* **19,** 384–393 (2013).

37. Chen, S. X., Kim, A. N., Peters, A. J. & Komiyama, T. Subtype-specific plasticity of inhibitory circuits in motor cortex during motor learning. *Nature Neuroscience* **18,** 1109–1115 (2015).

38. Chistiakova, M., Bannon, N. M., Chen, J.-Y., Bazhenov, M. & Volgushev, M. Homeostatic role of heterosynaptic plasticity: models and experiments. *Frontiers in Computational Neuroscience* **9** (2015).

39. Chiu, C. Q., Barberis, A. & Higley, M. J. Preserving the balance: diverse forms of long-term GABAergic synaptic plasticity. *Nature Reviews Neuroscience* **20,** 272–281 (2019).

40. Chiu, C. Q., Lur, G., Morse, T. M., Carnevale, N. T., Ellis-Davies, G. C. R. & Higley, M. J. Compartmentalization of GABAergic inhibition by dendritic spines. *Science* **340,** 759–763 (2013).

41. Chiu, C. Q., Martenson, J. S., Yamazaki, M., Natsume, R., Sakimura, K., Tomita, S., Tavalin, S. J. & Higley, M. J. Input-Specific NMDAR-Dependent Potentiation of Dendritic GABAergic Inhibition. *Neuron* **97,** 368–377 (2018).

42. Choi, G. B., Stettler, D. D., Kallman, B. R., Bhaskar, S. T., Fleischmann, A. & Axel, R. Driving opposing behaviors with ensembles of piriform neurons. *Cell* **146,** 1004–1015 (2011).

43. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* **36,** 181–253 (2013).

44. Clopath, C., Büsing, L., Vasilaki, E. & Gerstner, W. Connectivity reflects coding: a model of voltage-based STDP with homeostasis. *Nature Neuroscience* **13,** 344–352 (2010).

45. Clopath, C., Vogels, T. P., Froemke, R. C. & Sprekeler, H. Receptive field formation by interacting excitatory and inhibitory synaptic plasticity. *bioRxiv* (2016).

46. Cooper, L. N. & Bear, M. F. The BCM theory of synapse modification at 30: interaction of theory with experiment. *Nature Reviews Neuroscience* **13,** 798–810 (2012).

47. Cossart, R., Aronov, D. & Yuste, R. Attractor dynamics of network UP states in the neocortex. *Nature* **423,** 283–288 (2003).

48. D'amour, J. A. & Froemke, R. C. Inhibitory and Excitatory Spike-Timing-Dependent Plasticity in the Auditory Cortex. *Neuron* **86,** 514–528 (2015).

49. Das, S., Sadanandappa, M. K., Dervan, A., Larkin, A., Lee, J. A., Sudhakaran, I. P., Priya, R., Heidari, R., Holohan, E. E., Pimentel, A., Gandhi, A., Ito, K., Sanyal, S., Wang, J. W., Rodrigues, V. & Ramaswami, M. Plasticity of local GABAergic interneurons drives olfactory habituation. *Proceedings of the National Academy of Sciences of the United States of America* **108,** E646–E654 (2011).

50. Debanne, D., Inglebert, Y. & Russier, M. Plasticity of intrinsic neuronal excitability. *Current Opinion in Neurobiology* **54,** 73–82 (2019).

51. Denève, S. & Machens, C. K. Efficient codes and balanced networks. *Nature Neuroscience* **19,** 375–382 (2016).

52. Desai, N. S., Rutherford, L. C. & Turrigiano, G. G. Plasticity in the intrinsic excitability of cortical pyramidal neurons. *Nature Neuroscience* **2,** 515–520 (1999).

53. Dorrn, A. L., Yuan, K., Barker, A. J., Schreiner, C. E. & Froemke, R. C. Developmental sensory experience balances cortical excitation and inhibition. *Nature* **465,** 932–936 (2010).

54. Ebner, C., Clopath, C., Jedlicka, P. & Cuntz, H. Unifying Long-Term Plasticity Rules for Excitatory Synapses by Modeling Dendrites of Cortical Pyramidal Neurons. *Cell Reports* **29,** 4295–4307 (2019).

55. Eckmann, S. & Gjorgjieva, J. Synapse-type-specific competitive Hebbian learning forms functional recurrent networks. *bioRxiv* (2022).

56. Eichenbaum, H. Barlow versus Hebb: When is it time to abandon the notion of feature detectors and adopt the cell assembly as the unit of cognition? *Neuroscience Letters* **680,** 88–93 (2018).

57. Fairhall, A. L. in *The Cognitive Neurosciences* (eds Gazzaniga, M. S. & Mangun, G. R.) 5th ed., 283–294 (MIT Press, 2014).

58. Feldman, D. E. Synaptic mechanisms for plasticity in neocortex. *Annual Review of Neuroscience* **32,** 33–55 (2009).

59. Field, R. E., D'amour, J. A., Tremblay, R., Miehl, C., Rudy, B., Gjorgjieva, J. & Froemke, R. C. Heterosynaptic Plasticity Determines the Set Point for Cortical Excitatory-Inhibitory Balance. *Neuron* **106,** 842–854 (2020).

60. Fischer, T. M., Blazis, D. E., Priver, N. A. & Carew, T. J. Metaplasticity at identified inhibitory synapses in Aplysia. *Nature* **389,** 860–865 (1997).

61. Fourcaud-Trocmé, N., Hansel, D., van Vreeswijk, C. & Brunel, N. How spike generation mechanisms determine the neuronal response to fluctuating inputs. *The Journal of Neuroscience* **23,** 11628–11640 (2003).

62. Fox, K. & Stryker, M. Integrating Hebbian and homeostatic plasticity: Introduction. *Philosophical Transactions of the Royal Society B* **372,** 20160413 (2017).

63. Friston, K. Does predictive coding have a future? *Nature Neuroscience* **21,** 1019–1021 (2018).

64. Froemke, R. C. Plasticity of Cortical Excitatory-Inhibitory Balance. *Annual Review of Neuroscience* **38,** 195–219 (2015).

65. Froemke, R. C., Carcea, I., Barker, A. J., Yuan, K., Seybold, B. A., Martins, A. R. O., Zaika, N., Bernstein, H., Wachs, M., Levis, P. A., Polley, D. B., Merzenich, M. M. & Schreiner, C. E. Long-term modification of cortical synapses improves sensory perception. *Nature Neuroscience* **16,** 79–88 (2013).

66. Froemke, R. C., Merzenich, M. M. & Schreiner, C. E. A synaptic memory trace for cortical receptive field plasticity. *Nature* **450,** 425–429 (2007).

67. Fusi, S. Computational models of long term plasticity and memory. *arXiv* (2017).

68. Fyhn, M., Molden, S., Witter, M. P., Moser, E. I. & Moser, M.-B. Spatial representation in the entorhinal cortex. *Science* **305,** 1258–1264 (2004).

69. Gainey, M. A. & Feldman, D. E. Multiple shared mechanisms for homeostatic plasticity in rodent somatosensory and visual cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372** (2017).

70. Gandolfi, D., Bigiani, A., Porro, C. A. & Mapelli, J. Inhibitory Plasticity: From Molecules to Computation and Beyond. *International Journal of Molecular Sciences* **21** (2020).

71. Garrett, M., Manavi, S., Roll, K., Ollerenshaw, D. R., Groblewski, P. A., Ponvert, N. D., Kiggins, J. T., Casal, L., Mace, K., Williford, A., Leon, A., Jia, X., Ledochowitsch, P., Buice, M. A., Wakeman, W., Mihalas, S. & Olsen, S. R. Experience shapes activity dynamics and stimulus coding of VIP inhibitory cells. *eLife* **9,** e50340 (2020).

72. Gerstner, W., Kempter, R., van Hemmen, J. L. & Wagner, H. A neuronal learning rule for sub-millisecond temporal coding. *Nature* **383,** 76–78 (1996).

73. Gillon, C. J., Pina, J. E., Lecoq, J. A., Ahmed, R., Billeh, Y., Caldejon, S., Groblewski, P., Henley, T. M., Kato, I., Lee, E., Luviano, J., Mace, K., Nayan, C., Nguyen, T., North, K., Perkins, J., Seid, S., Valley, M., Williford, A., Bengio, Y., Lillicrap, T. P., Richards, B. A. & Zylberberg, J. Learning from unexpected events in the neocortical microcircuit. *bioRxiv* (2021).

74. Gilson, M., Burkitt, A. N., Grayden, D. B., Thomas, D. A. & Van Hemmen, J. L. Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. II. Input selectivity-symmetry breaking. *Biological Cybernetics* **101,** 103–114 (2009).

75. Gilson, M., Burkitt, A. N., Grayden, D. B., Thomas, D. A. & van Hemmen, J. L. Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. I. Input selectivity–strengthening correlated input pathways. *Biological Cybernetics* **101,** 81–102 (2009).

76. Gjorgjieva, J., Clopath, C., Audet, J. & Pfister, J.-P. A triplet spike-timing-dependent plasticity model generalizes the Bienenstock-Cooper-Munro rule to higher-order spatiotemporal correlations. *Proceedings of the National Academy of Sciences* **108,** 19383–19388 (2011).

77. Glanzman, D. L. Olfactory habituation: Fresh insights from flies. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 14711–14712 (2011).

78. Gütig, R., Aharonov, R., Rotter, S. & Sompolinsky, H. Learning Input Correlations through Nonlinear Temporally Asymmetric Hebbian Plasticity. *The Journal of Neuroscience* **23,** 3697–3714 (2003).

79. Guzman, S. J., Schlögl, A., Frotscher, M. & Jonas, P. Synaptic mechanisms of pattern completion in the hippocampal CA3 network. *Science* **353,** 1117–1123 (2016).

80. Haak, K. V., Fast, E., Bao, M., Lee, M. & Engel, S. A. Four days of visual contrast deprivation reveals limits of neuronal adaptation. *Current Biology* **24,** 2575–2579 (2014).

81. Haas, J. S., Nowotny, T. & Abarbanel, H. D. I. Spike-Timing-Dependent Plasticity of Inhibitory Synapses in the Entorhinal Cortex. *Journal of Neurophysiology* **96,** 3305–3313 (2006).

82. Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436,** 801–806 (2005).

83. Hamm, J. P. & Yuste, R. Somatostatin Interneurons Control a Key Component of Mismatch Negativity in Mouse Visual Cortex. *Cell Reports* **16,** 597–604 (2016).

84. Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G. & Buzsáki, G. Organization of cell assemblies in the hippocampus. *Nature* **424,** 552–556 (2003).

85. Hattori, R., Kuchibhotla, K. V., Froemke, R. C. & Komiyama, T. Functions and dysfunctions of neocortical inhibitory neuron subtypes. *Nature Neuroscience* **20,** 1199–1208 (2017).

86. Hawkes, A. G. Point spectra of some mutually exciting point processes. *Journal of the Royal Statistical Society: Series B (Methodological)* **33,** 438–443 (1971).

87. Hayama, T., Noguchi, J., Watanabe, S., Takahashi, N., Hayashi-Takagi, A., Ellis-Davies, G. C., Matsuzaki, M. & Kasai, H. GABA promotes the competitive selection of dendritic spines by controlling local Ca 2+ signaling. *Nature Neuroscience* **16,** 1409–1416 (2013).

88. Hebb, D. O. *The organization of behavior; a neuropsychological theory* (Wiley, 1949).

89. Heintz, T. G., Hinojosa, A. J. & Lagnado, L. Opposing forms of adaptation in mouse visual cortex are controlled by distinct inhibitory microcircuits and gated by locomotion. *bioRxiv* **909788** (2020).

90. Hennequin, G., Agnes, E. J. & Vogels, T. P. Inhibitory Plasticity: Balance, Control, and Codependence. *Annual Review of Neuroscience* **40,** 557–579 (2017).

91. Hershenhoren, I., Taaseh, N., Antunes, F. M. & Nelken, I. Intracellular correlates of stimulus-specific adaptation. *Journal of Neuroscience* **34,** 3303–3319 (2014).

92. Hiratani, N. & Fukai, T. Detailed Dendritic Excitatory/Inhibitory Balance through Heterosynaptic Spike-Timing-Dependent Plasticity. *The Journal of Neuroscience* **37,** 12106–12122 (2017).

93. Holmgren, C. D. & Zilberter, Y. Coincident Spiking Activity Induces Long-Term Changes in Inhibition of Neocortical Pyramidal Cells. *The Journal of Neuroscience* **21,** 8270–8277 (2001).

94. Holtmaat, A. & Caroni, P. Functional and structural underpinnings of neuronal assembly formation in learning. *Nature Neuroscience* **19** (2016).

95. Homann, J., Ann, S., Chen, K. S., Tank, D. W. & Berry, M. J. Novel stimuli evoke excess activity in the mouse primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America* **119,** e2108882119 (2022).

96. Huang, Y.-Y., Colino, A., Selig, D. K. & Malenka, R. C. The influence of prior synaptic activity on the induction of long-term potentiation. *Science* **255,** 730–733 (1992).

97. Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology* **160,** 106–154 (1962).

98. Huyck, C. R. & Passmore, P. J. A review of cell assemblies. *Biological Cybernetics* **107,** 263–288 (2013).

99. Inglebert, Y., Aljadeff, J., Brunel, N. & Debanne, D. Altered spike timing-dependent plasticity rules in physiological calcium. *bioRxiv* (2020).

100. Insanally, M. N., Carcea, I., Field, R. E., Rodgers, C. C., DePasquale, B., Rajan, K., DeWeese, M. R., Albanna, B. F. & Froemke, R. C. Spike-timing-dependent ensemble encoding by non-classically responsive cortical neurons. *eLife* **8,** e42409 (2019).

101. Jedlicka, P., Benuskova, L. & Abraham, W. C. A Voltage-Based STDP Rule Combined with Fast BCM-Like Metaplasticity Accounts for LTP and Concurrent "Heterosynaptic" LTD in the Dentate Gyrus In Vivo. *PLOS Computational Biology* **11** (ed Morrison, A.) e1004588 (2015).

102. Jedlicka, P., Tomko, M., Robins, A. & Abraham, W. C. Contributions by metaplasticity to solving the Catastrophic Forgetting Problem. *Trends in Neurosciences* **45,** 656–666 (2022).

103. Jennings, J. H., Kim, C. K., Marshel, J. H., Raffiee, M., Ye, L., Quirin, S., Pak, S., Ramakrishnan, C. & Deisseroth, K. Interacting neural ensembles in orbitofrontal cortex for social and feeding behaviour. *Nature* **565,** 645–649 (2019).

104. Josselyn, S. A. & Frankland, P. W. Memory Allocation: Mechanisms and Function. *Annual Review of Neuroscience* **41,** 389–413 (2018).

105. Jouhanneau, J. S., Kremkow, J., Dorrn, A. L. L. & Poulet, J. F. F. In Vivo Monosynaptic Excitatory Transmission between Layer 2 Cortical Pyramidal Neurons. *Cell Reports* **13,** 2098–2106 (2015).

106. Jovanović, S. & Rotter, S. Interplay between Graph Topology and Correlations of Third Order in Spiking Neuronal Networks. *PLoS Computational Biology* **12,** e1004963 (2016).

107. Kato, H. K., Gillet, S. N. & Isaacson, J. S. Flexible Sensory Representations in Auditory Cortex Driven by Behavioral Relevance. *Neuron* **88,** 1027–1039 (2015).

108. Keck, T., Hübener, M. & Bonhoeffer, T. Interactions between synaptic homeostatic mechanisms: an attempt to reconcile BCM theory, synaptic scaling, and changing excitation/inhibition balance. *Current Opinion in Neurobiology* **43,** 87–93 (2017).

109. Kempter, R., Gerstner, W. & Leo Van Hemmen, J. Hebbian learning and spiking neurons. *Physical Review E* **59,** 4498–4514 (1999).

110. Kirchner, J. H. & Gjorgjieva, J. Emergence of local and global synaptic organization on cortical dendrites. *Nature Communications* **12,** 4005 (2021).

111. Kirkwood, A., Rioult, M. G. & Bear, M. F. Experience-dependent modification of synaptic plasticity in visual cortex. *Nature* **381,** 526–528 (1996).

112. Kleberg, F. I., Fukai, T. & Gilson, M. Excitatory and inhibitory STDP jointly tune feedforward neural circuits to selectively propagate correlated spiking activity. *Frontiers in Computational Neuroscience* **8** (2014).

113. Ko, H., Hofer, S. B., Pichler, B., Buchanan, K. A., Sjöström, P. J. & Mrsic-Flogel, T. D. Functional specificity of local synaptic connections in neocortical networks. *Nature* **473,** 87–91 (2011).

114. Krabbe, S., Paradiso, E., D'Aquin, S., Bitterman, Y., Courtin, J., Xu, C., Yonehara, K., Markovic, M., Müller, C., Eichlisberger, T., Gründemann, J., Ferraguti, F. & Lüthi, A. Adaptive disinhibitory gating by VIP interneurons permits associative learning. *Nature Neuroscience* **22,** 1834–1843 (2019).

115. Kriegeskorte, N. & Diedrichsen, J. Peeling the Onion of Brain Representations. *Annual Review of Neuroscience* **42,** 407–432 (2019).

116. Kripkee, B. & Froemke, R. C. *Organization and Plasticity of Cortical Inhibition* (The Oxford Handbook of Developmental Neural Plasticity, 2017).

117. Kuo, M. C. & Dringenberg, H. C. Short-term (2 to 5 h) dark exposure lowers long-term potentiation (LTP) induction threshold in rat primary visual cortex. *Brain Research* **1276,** 58–66 (2009).

118. Lagzi, F., Canto-Bustos, M., Oswald, A.-M. M. & Doiron, B. Assembly formation is stabilized by Parvalbumin neurons and accelerated by Somatostatin neurons. *bioRxiv* (2021).

119. Larisch, R., Gönner, L., Teichmann, M. & Hamker, F. H. Sensory coding and contrast invariance emerge from the control of plastic inhibition over emergent selectivity. *PLoS Computational Biology* **17,** e1009566 (2021).

120. Larkum, M. Are dendrites conceptually useful? *Neuroscience* (2022).

121. Lee, K. F. H., Soares, C., Thivierge, J.-P. & Béïque, J.-C. Correlated Synaptic Inputs Drive Dendritic Calcium Amplification and Cooperative Plasticity during Clustered Synapse Development. *Neuron* **89,** 784–799 (2016).

122. Lefort, S., Tomm, C., Floyd Sarria, J. C. & Petersen, C. C. The Excitatory Neuronal Network of the C2 Barrel Column in Mouse Primary Somatosensory Cortex. *Neuron* **61,** 301–316 (2009).

123. Letzkus, J. J., Kampa, B. M. & Stuart, G. J. Learning Rules for Spike Timing-Dependent Plasticity Depend on Dendritic Synapse Location. *Journal of Neuroscience* **26,** 10420–10429 (2006).

124. Letzkus, J. J., Wolff, S. B. & Lüthi, A. Disinhibition, a Circuit Mechanism for Associative Learning and Memory. *Neuron* **88,** 264–276 (2015).

125. Litwin-Kumar, A. & Doiron, B. Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nature Communications* **5** (2014).

126. Litwin-Kumar, A. & Doiron, B. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature Neuroscience* **15,** 1498–1505 (2012).

127. Liu, X., Ramirez, S., Pang, P. T., Puryear, C. B., Govindarajan, A., Deisseroth, K. & Tonegawa, S. Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature* **484,** 381–385 (2012).

128. Lopatina, O. L., Malinovskaya, N. A., Komleva, Y. K., Gorina, Y. V., Shuvaev, A. N., Olovyannikova, R. Y., Belozor, O. S., Belova, O. A., Higashida, H. & Salmina, A. B. Excitation/inhibition imbalance and impaired neurogenesis in neurodevelopmental and neurodegenerative disorders. *Reviews in the Neurosciences* **30,** 807–820 (2019).

129. Lundstrom, B. N., Fairhall, A. L. & Maravall, M. Multiple timescale encoding of slowly varying whisker stimulus envelope in cortical and thalamic neurons in vivo. *Journal of Neuroscience* **30,** 5071–5077 (2010).

130. Luz, Y. & Shamir, M. Balancing Feed-Forward Excitation and Inhibition via Hebbian Inhibitory Synaptic Plasticity. *PLoS Computational Biology* **8** (ed Latham, P. E.) e1002334 (2012).

131. Luz, Y. & Shamir, M. The effect of STDP temporal kernel structure on the learning dynamics of single excitatory and inhibitory synapses. *PLoS ONE* **9,** e101109 (2014).

132. Lynch, G. S., Dunwiddie, T. & Gribkoff, V. Heterosynaptic depression: a postsynaptic correlate of long-term potentiation. *Nature* **266,** 737–739 (1977).

133. Lynn, C. W. & Bassett, D. S. The physics of brain network structure, function and control. *Nature Reviews Physics* **1,** 318–332 (2019).

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

135. MacLean, J. N., Watson, B. O., Aaron, G. B. & Yuste, R. Internal dynamics determine the cortical response to thalamic stimulation. *Neuron* **48,** 811–823 (2005).

136. Maes, A., Barahona, M. & Clopath, C. Learning spatiotemporal signals using a recurrent spiking network that discretizes time. *PLoS Computational Biology* **16,** e1007606 (2020).

137. Magee, J. C. & Grienberger, C. Synaptic Plasticity Forms and Functions. *Annual Review of Neuroscience* **43,** 95–117 (2020).

138. Manz, P. & Memmesheimer, R.-M. Purely STDP-based assembly dynamics: stability, learning, overlaps, drift and aging. *bioRxiv* (2022).

139. Mao, B. Q., Hamzei-Sichani, F., Aronov, D., Froemke, R. C. & Yuste, R. Dynamics of spontaneous activity in neocortical slices. *Neuron* **32,** 883–898 (2001).

140. Maravall, M. & Diamond, M. E. Algorithms of whisker-mediated touch perception. *Current Opinion in Neurobiology* **25,** 176–186 (2014).

141. Markram, H., Lübke, J., Frotscher, M. & Sakmann, B. Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs. *Science* **275,** 213–215 (1997).

142. Marshel, J. H., Kim, Y. S., Machado, T. A., Quirin, S., Benson, B., Kadmon, J., Raja, C., Chibukhchyan, A., Ramakrishnan, C., Inoue, M., Shane, J. C., McKnight, D. J., Yoshizawa, S., Kato, H. E., Ganguli, S. & Deisseroth, K. Cortical layer-specific critical dynamics triggering perception. *Science* (2019).

143. Mellor, J. in *Hippocampal Microcircuits* 201–226 (Springer Seriesin Computational Neuroscience, 2018).

144. Miconi, T., McKinstry, J. L. & Edelman, G. M. Spontaneous emergence of fast attractor dynamics in a model of developing primary visual cortex. *Nature Communications* **7** (2016).

145. Miehl, C. & Gjorgjieva, J. Stability and learning in excitatory synapses by nonlinear inhibitory plasticity. *PLoS Computational Biology* **18,** e1010682 (2022).

146. Miehl, C., Onasch, S., Festa, D. & Gjorgjieva, J. Formation and computational implications of assemblies in neural circuits. *Journal of Physiology* (2022).

147. Mikulasch, F. A., Rudelt, L. & Priesemann, V. Local dendritic balance enables learning of efficient representations in networks of spiking neurons. *Proceedings of the National Academy of Sciences* **118,** e2021925118 (2021).

148. Mill, R., Coath, M., Wennekers, T. & Denham, S. L. A neurocomputational model of stimulus-specific adaptation to oddball and markov sequences. *PLoS Computational Biology* **7,** e1002117 (2011).

149. Mill, R., Coath, M., Wennekers, T. & Denham, S. L. Abstract stimulus-specific adaptation models. *Neural Computation* **23,** 435–476 (2011).

150. Mill, R., Coath, M., Wennekers, T. & Denham, S. L. Characterising stimulus-specific adaptation using a multi-layer field model. *Brain Research* **1434,** 178–188 (2012).

151. Miller, J. E. K., Ayzenshtat, I., Carrillo-Reid, L. & Yuste, R. Visual stimuli recruit intrinsically generated cortical ensembles. *Proceedings of the National Academy of Sciences of the United States of America* **111,** E4053–E4061 (2014).

152. Miller, K. D. & MacKay, D. J. C. The Role of Constraints in Hebbian Learning. *Neural Computation* **6,** 100–126 (1994).

153. Montangie, L., Miehl, C. & Gjorgjieva, J. Autonomous emergence of connectivity assemblies via spike triplet interactions. *PLoS Computational Biology* **16,** e1007835 (2020).

154. Morrison, A., Aertsen, A. & Diesmann, M. Spike-timing-dependent plasticity in balanced random networks. *Neural Computation* **19,** 1437–1467 (2007).

155. Motanis, H., Seay, M. J. & Buonomano, D. V. Short-Term Synaptic Plasticity as a Mechanism for Sensory Timing. *Trends in Neurosciences* **41,** 701–711 (2018).

156. Näätänen, R., Simpson, M. & Loveless, N. E. Stimulus deviance and evoked potentials. *Biological Psychology* **14,** 53–98 (1982).

157. Najafi, F., Elsayed, G. F., Cao, R., Pnevmatikakis, E., Latham, P. E., Cunningham, J. & Churchland, A. K. Excitatory and inhibitory subnetworks are equally selective during decision-making and emerge simultaneously during learning. *Neuron* **105,** 165–179 (2020).

158. Natan, R. G., Briguglio, J. J., Mwilambwe-Tshilobo, L., Jones, S. I., Aizenberg, M., Goldberg, E. M. & Geffen, M. N. Complementary control of sensory adaptation by two types of cortical interneurons. *eLife* **4,** e09868 (2015).

159. Natan, R. G., Rao, W. & Geffen, M. N. Cortical Interneurons Differentially Shape Frequency Tuning following Adaptation. *Cell Reports* **21,** 878–890 (2017).

160. O'Keefe, J., Dostrovsky, J. & J. O'Keefe, J. D. The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research* **34,** 171–175 (1971).

161. O'Keefe, J. Place units in the hippocampus of the freely moving rat. *Experimental Neurology* **51,** 78–109 (1976).

162. Ocker, G. K. & Doiron, B. Training and spontaneous reinforcement of neuronal assemblies by spike timing plasticity. *Cerebral Cortex* **29,** 937–951 (2019).

163. Ocker, G. K., Litwin-Kumar, A. & Doiron, B. Self-Organization of Microcircuits in Networks of Spiking Neurons with Plastic Synapses. *PLoS Computational Biology* **11,** e1004458 (2015).

164. Oh, W. C., Parajuli, L. K. & Zito, K. Heterosynaptic Structural Plasticity on Local Dendritic Segments of Hippocampal CA1 Neurons. *Cell Reports* **10,** 162–169 (2015).

165. Oja, E. Simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology* **15,** 267–273 (1982).

166. Okun, M. & Lampl, I. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature Neuroscience* **11,** 535–537 (2008).

167. Okuno, H., Akashi, K., Ishii, Y., Yagishita-Kyo, N., Suzuki, K., Nonaka, M., Kawashima, T., Fujii, H., Takemoto-Kimura, S., Abe, M., Natsume, R., Chowdhury, S., Sakimura, K., Worley, P. F. & Bito, H. Inverse synaptic tagging of inactive synapses via dynamic interaction of Arc/Arg3.1 with CaMKIIβ. *Cell* **149,** 886–898 (2012).

168. Okuno, H., Minatohara, K. & Bito, H. Inverse synaptic tagging: An inactive synapse-specific mechanism to capture activity-induced Arc/arg3.1 and to locally regulate spatial distribution of synaptic weights. *Seminars in Cell and Developmental Biology* **77,** 43–50 (2018).

169. Paille, V., Fino, E., Du, K., Morera-Herreras, T., Perez, S., Kotaleski, J. H. & Venance, L. GABAergic Circuits Control Spike-Timing-Dependent Plasticity. *Journal of Neuroscience* **33,** 9353–9363 (2013).

170. Panzeri, S., Moroni, M., Safaai, H. & Harvey, C. D. The structures and functions of correlations in neural population codes. *Nature Reviews Neuroscience* (2022).

171. Papadimitriou, C. H. & Friederici, A. D. Bridging the Gap Between Neurons and Cognition Through Assemblies of Neurons. *Neural Computation* **34,** 291–306 (2022).

172. Papadimitriou, C. H., Vempala, S. S., Mitropolsky, D., Collins, M. & Maass, W. Brain computation by assemblies of neurons. *Proceedings of the National Academy of Sciences of the United States of America* **117,** 14464–14472 (2020).

173. Parisi, G. I., Kemker, R., Part, J. L., Kanan, C. & Wermter, S. Continual lifelong learning with neural networks: A review. *Neural Networks* **113,** 54–71 (2019).

174. Park, Y. & Geffen, M. N. A circuit model of auditory cortex. *PLoS Computational Biology* **16,** e1008016 (2020).

175. Paulsen, O. & Moser, E. I. A model of hippocampal memory encoding and retrieval: GABAergic control of synaptic plasticity. *Trends in Neurosciences* **21,** 273–278 (1998).

176. Pedrosa, V. & Clopath, C. Interplay between somatic and dendritic inhibition promotes the emergence and stabilization of place fields. *PLoS Computational Biology* **16,** e1007955 (2020).

177. Perin, R., Berger, T. K. & Markram, H. A synaptic organizing principle for cortical neuronal groups. *Proceedings of the National Academy of Sciences of the United States of America* **108,** 5419–5424 (2011).

178. Pfister, J.-P. & Gerstner, W. Triplets of Spikes in a Model of Spike Timing-Dependent Plasticity. *Journal of Neuroscience* **26,** 9673–9682 (2006).

179. Philpot, B. D., Espinosa, J. S. & Bear, M. F. Evidence for altered NMDA receptor function as a basis for metaplasticity in visual cortex. *Journal of Neuroscience* **23,** 5583–5588 (2003).

180. Pietri, T., Romano, S. A., Pérez-Schuster, V., Boulanger-Weill, J., Candat, V. & Sumbre, G. The Emergence of the Spatial Structure of Tectal Spontaneous Activity Is Independent of Visual Inputs. *Cell Reports* **19,** 939–948 (2017).

181. Poirazi, P. & Papoutsi, A. Illuminating dendritic function with computational models. *Nature Reviews Neuroscience* (2020).

182. Ramaswami, M. Network plasticity in adaptive filtering and behavioral habituation. *Neuron* **82,** 1216–1229 (2014).

183. Rao, R. P. N. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience* **2,** 79–87 (1999).

184. Ravid Tannenbaum, N. & Burak, Y. Shaping Neural Circuits by High Order Synaptic Interactions. *PLOS Computational Biology* **12** (ed Latham, P. E.) e1005056 (2016).

185. Reyes, A. D. Synaptic short-term plasticity in auditory cortical circuits. *Hearing Research* **279,** 60–66 (2011).

186. Richter, L. M. A. & Gjorgjieva, J. Understanding neural circuit development through theory and models. *Current Opinion in Neurobiology* **46,** 39–47 (2017).

187. Ross, J. M. & Hamm, J. P. Cortical Microcircuit Mechanisms of Mismatch Negativity and Its Underlying Subcomponents. *Frontiers in Neural Circuits* **14** (2020).

188. Rossi, L. F., Harris, K. D. & Carandini, M. Spatial connectivity matches direction selectivity in visual cortex. *Nature* **588** (2020).

189. Rost, T., Deger, M. & Nawrot, M. P. Winnerless competition in clustered balanced networks: inhibitory assemblies do the trick. *Biological Cybernetics* **112,** 81–98 (2018).

190. Royer, S. & Paré, D. Conservation of total synaptic weight through balanced synaptic depression and potentiation. *Nature* **422,** 518–522 (2003).

191. Rubin, D. B., VanHooser, S. D. & Miller, K. D. The stabilized supralinear network: A unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron* **85,** 402–417 (2015).

192. Rubin, J., Lee, D. D. & Sompolinsky, H. Equilibrium properties of temporally asymmetric Hebbian plasticity. *Physical Review Letters* **86,** 364–367 (2001).

193. Rupprecht, P. & Friedrich, R. W. Precise Synaptic Balance in the Zebrafish Homolog of Olfactory Cortex. *Neuron* **100,** 669–683 (2018).

194. Schulz, A., Miehl, C., Berry II, M. J. & Gjorgjieva, J. The generation of cortical novelty responses through inhibitory plasticity. *eLife* **10,** e65309 (2021).

195. Sejnowski, T. J. The Computer and the Brain Revisited. *{IEEE} Annals of the History of Computing* **11,** 197–201 (1989).

196. Shamma, S. & Fritz, J. Adaptive auditory computations. *Current Opinion in Neurobiology* **25,** 164–168 (2014).

197. Shew, T., Yip, S. & Sastry, B. R. Mechanisms involved in tetanusinduced potentiation of fast IPSCs in rat hippocampal CA1 neurons. *Journal of Neurophysiology* **83,** 3388–3401 (2000).

198. Sjöström, P. J., Turrigiano, G. G. & Nelson, S. B. Rate, Timing, and Cooperativity Jointly Determine Cortical Synaptic Plasticity. *Neuron* **32,** 1149–1164 (2001).

199. Soldado-Magraner, S., Laje, R. & Buonomano, D. V. Orchestrated Excitatory and Inhibitory Learning Rules Lead to the Unsupervised Emergence of Self-sustained and Inhibition-stabilized Dynamics. *bioRxiv* (2021).

200. Song, S., Sjöström, P. J., Reigl, M., Nelson, S. & Chklovskii, D. B. Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biology* **3,** e68 (2005).

201. Song, S. C., Shen, B., Machold, R., Rudy, B., Glimcher, P. W., Louie, K. & Froemke, R. C. Input-Specific Inhibitory Plasticity Improves Decision Accuracy Under Noise. *bioRxiv* (2022).

202. Steele, P. M. & Mauk, M. D. Inhibitory Control of LTP and LTD: Stability of Synapse Strength. *Journal of Neurophysiology* **81,** 1559–1566 (1999).

203. Suvrathan, A. Beyond STDP - towards diverse and functionally relevant plasticity rules. *Current Opinion in Neurobiology* **54,** 12–19 (2019).

204. Taaseh, N., Yaron, A. & Nelken, I. Stimulus-specific adaptation and deviance detection in the rat auditory cortex. *PLoS ONE* **6,** e23369 (2011).

205. Tarusawa, E., Sanbo, M., Okayama, A., Miyashita, T., Kitsukawa, T., Hirayama, T., Hirabayashi, T., Hasegawa, S., Kaneko, R., Toyoda, S., Kobayashi, T., Kato-Itoh, M., Nakauchi, H., Hirabayashi, M., Yagi, T. & Yoshimura, Y. Establishment of high reciprocal connectivity between clonal cortical neurons is regulated by the Dnmt3b DNA methyltransferase and clustered protocadherins. *BMC Biology* **14,** 103 (2016).

206. Tremblay, R., Lee, S. & Rudy, B. GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. *Neuron* **91,** 260–292 (2016).

207. Triesch, J., Vo, A. D. & Hafner, A.-S. Competition for synaptic building blocks shapes synaptic plasticity. *eLife* **7,** e37836 (2018).

208. Trousdale, J., Hu, Y., Shea-Brown, E. & Josić, K. Impact of network structure and cellular response on spike time correlations. *PLoS Computational Biology* **8** (2012).

209. Tsodyks, M., Pawelzik, K. & Markram, H. Neural Networks with Dynamic Synapses. *Neural Computation* **10,** 821–835 (1998).

210. Turner, N. L., Macrina, T., Bae, J. A., Yang, R., Wilson, A. M., Schneider-Mizell, C., Lee, K., Lu, R., Wu, J., Bodor, A. L., Bleckert, A. A., Brittain, D., Froudarakis, E., Dorkenwald, S., Collman, F., Kemnitz, N., Ih, D., Silversmith, W. M., Zung, J., Zlateski, A., Tartavull, I., Yu, S. c., Popovych, S., Mu, S., Wong, W., Jordan, C. S., Castro, M., Buchanan, J. A., Bumbarger, D. J., Takeno, M., Torres, R., Mahalingam, G., Elabbady, L., Li, Y., Cobos, E., Zhou, P., Suckow, S., Becker, L., Paninski, L., Polleux, F., Reimer, J., Tolias, A. S., Reid, R. C., da Costa, N. M. & Seung, H. S. Reconstruction of neocortex: Organelles, compartments, cells, circuits, and activity. *Cell* **185,** 1082–1100 (2022).

211. Turrigiano, G. G. The dialectic of hebb and homeostasis. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372,** 20160258 (2017).

212. Turrigiano, G. G. The Self-Tuning Neuron: Synaptic Scaling of Excitatory Synapses. *Cell* **135,** 422–435 (2008).

213. Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C. & Nelson, S. B. Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature* **391,** 892–896 (1998).

214. Turrigiano, G. G. & Nelson, S. B. Homeostatic plasticity in the developing nervous system. *Nature Reviews Neuroscience* **5,** 97–107 (2004).

215. Udakis, M., Pedrosa, V., Chamberlain, S. E. L., Clopath, C. & Mellor, J. R. Interneuron-specific plasticity at parvalbumin and somatostatin inhibitory synapses onto CA1 pyramidal neurons shapes hippocampal output. *Nature Communications* **11** (2020).

216. Ulanovsky, N., Las, L., Farkas, D. & Nelken, I. Multiple time scales of adaptation in auditory cortex neurons. *Journal of Neuroscience* **24,** 10440–10453 (2004).

217. Van Rossum, M. C. W., Bi, G.-Q. & Turrigiano, G. G. Stable Hebbian Learning from Spike Timing-Dependent Plasticity. *The Journal of Neuroscience* **20,** 8812–8821 (2000).

218. Van Vreeswijk, C. & Sompolinsky, H. Chaos in Neuronal Networks with Balanced Excitatory and Inhibitory Activity. *Science* **274,** 1724–1726 (1996).

219. Vogels, T. P., Froemke, R. C., Doyon, N., Gilson, M., Haas, J. S., Liu, R., Maffei, A., Miller, P., Wierenga, C. J., Woodin, M. A., Zenke, F. & Sprekeler, H. Inhibitory synaptic plasticity: spike timing-dependence and putative network function. *Frontiers in Neural Circuits* **7** (2013).

220. Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C. & Gerstner, W. Inhibitory Plasticity Balances Excitation and Inhibition in Sensory Pathways and Memory Networks. *Science* **334,** 1569–1573 (2011).

221. Volgushev, M., Voronin, L. L., Chistiakova, M. & Singer, W. Relations between long-term synaptic modifications and paired-pulse interactions in the rat neocortex. *European Journal of Neuroscience* **9,** 1656–1665 (1997).

222. Volgushev, M., Balaban, P., Chistiakova, M. & Eysel, U. T. Retrograde signalling with nitric oxide at neocortical synapses. *European Journal of Neuroscience* **12,** 4255–4267 (2000).

223. Volgushev, M., Chen, J.-Y., Ilin, V., Goz, R., Chistiakova, M. & Bazhenov, M. Partial Breakdown of Input Specificity of STDP at Individual Synapses Promotes New Learning. *The Journal of Neuroscience* **36,** 8842–8855 (2016).

224. Volgushev, M., Voronin, L. L., Chistiakova, M. & Singer, W. Induction of LTP and LTD in visual cortex neurons by intracellular tetanization. *NeuroReport* **5,** 2069–2072 (1994).

225. Wang, L. & Maffei, A. Inhibitory Plasticity Dictates the Sign of Plasticity at Excitatory Synapses. *Journal of Neuroscience* **34,** 1083–1093 (2014).

226. Weber, A. I. & Fairhall, A. L. The role of adaptation in neural coding. *Current Opinion in Neurobiology* **58,** 135–140 (2019).

227. Weber, S. N. & Sprekeler, H. Learning place cells, grid cells and invariances with excitatory and inhibitory plasticity. *eLife* **7,** e34560 (2018).

228. Wehr, M. & Zador, A. M. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* **426,** 442–446 (2003).

229. Wertz, A., Trenholm, S., Yonehara, K., Hillier, D., Raics, Z., Leinweber, M., Szalay, G., Ghanem, A., Keller, G., Rózsa, B., Conzelmann, K.-k. & Roska, B. Single-cell–initiated monosynaptic tracing reveals layer-specific cortical network modules. *Science* **349,** 70–74 (2015).

230. White, G., Levy, W. B. & Steward, O. Spatial overlap between populations of synapses determines the extent of their associative interaction during the induction of long-term potentiation and depression. *Journal of Neurophysiology* **64,** 1186–1198 (1990).

231. Whitmire, C. J. & Stanley, G. B. Rapid Sensory Adaptation Redux: A Circuit Perspective. *Neuron* **92,** 298–315 (2016).

232. Williams, L. E. & Holtmaat, A. Higher-Order Thalamocortical Inputs Gate Synaptic Long-Term Potentiation via Disinhibition. *Neuron* **101,** 91–102 (2019).

233. Woodin, M. A., Ganguly, K. & Poo, M.-M. Coincident Pre- and Postsynaptic Activity Modifies GABAergic Synapses by Postsynaptic Changes in Cl- Transporter Activity. *Neuron* **39,** 807–820 (2003).

234. Wu, Y. K., Hengen, K. B., Turrigiano, G. G. & Gjorgjieva, J. Homeostatic mechanisms regulate distinct aspects of cortical circuit dynamics. *Proceedings of the National Academy of Sciences* **117,** 24514–24525 (2020).

235. Wu, Y. K., Miehl, C. & Gjorgjieva, J. Regulation of circuit organization and function through inhibitory synaptic plasticity. *Trends in Neurosciences* **45,** 884–898 (2022).

236. Wu, Y. K. & Zenke, F. Nonlinear transient amplification in recurrent neural networks with short-term plasticity. *eLife* **10,** e71263. (2021).

237. Yap, E. L., Pettit, N. L., Davis, C. P., Nagy, M. A., Harmin, D. A., Golden, E., Dagliyan, O., Lin, C., Rudolph, S., Sharma, N., Griffith, E. C., Harvey, C. D. & Greenberg, M. E. Bidirectional perisomatic inhibitory plasticity of a Fos neuronal network. *Nature* **590,** 115–121 (2021).

238. Yarden, T. S. & Nelken, I. Stimulus-specific adaptation in a recurrent network model of primary auditory cortex. *PLoS Computational Biology* **13,** e1005437 (2017).

239. Yee, A. X., Hsu, Y. T. & Chen, L. A metaplasticity view of the interaction between homeostatic and hebbian plasticity. *Philosophical Transactions of the Royal Society B* **372,** 20160155 (2017).

240. Yger, P. & Gilson, M. Models of Metaplasticity: A Review of Concepts. *Frontiers in Computational Neuroscience* **9** (2015).

241. Yoshimura, Y., Dantzker, J. L. M. & Callaway, E. M. Excitatory cortical neurons form fine-scale functional networks. *Nature* **433,** 868–873 (2005).

242. Yu, Y.-C., Bultje, R. S., Wang, X. & Shi, S.-H. Specific synapses develop preferentially among sister excitatory neurons in the neocortex. *Nature* **458,** 501–504 (2009).

243. Yuste, R. From the neuron doctrine to neural networks. *Nature Reviews Neuroscience* **16,** 487–497 (2015).

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

245. Zenke, F. & Gerstner, W. Hebbian plasticity requires compensatory processes on multiple timescales. *Philosophical Transactions of the Royal Society B: Biological Sciences* **372,** 20160259 (2017).

246. Zenke, F., Gerstner, W. & Ganguli, S. The temporal paradox of Hebbian learning and homeostatic plasticity. *Current Opinion in Neurobiology* **43,** 166–176 (2017).

247. Zenke, F., Hennequin, G. & Gerstner, W. Synaptic Plasticity in Neural Networks Needs Homeostasis with a Fast Rate Detector. *PLoS Computational Biology* **9** (ed Morrison, A.) e1003330 (2013).

248. Zhang, S., Xu, M., Kamigaki, T., Hoang Do, J. P., Chang, W.-C., Jenvay, S., Miyamichi, K., Luo, L. & Dan, Y. Long-range and local circuits for top-down modulation of visual cortex processing. *Science* **345,** 660–665 (2014).

249. Znamenskiy, P., Kim, M.-h., Muir, D. R., Iacaruso, F., Hofer, S. B. & Mrsic-Flogel, T. D. Functional selectivity and specific connectivity of inhibitory neurons in primary visual cortex. *bioRxiv* **294835** (2018).

250. Zucker, R. S. & Regehr, W. G. Short-term synaptic plasticity. *Annual Review of Physiology* **64,** 355–405 (2002).

Acronyms

ANTM	Adaptation of Narrowly Tuned Modules
BCM	Bienenstock-Cooper-Munro
E/I balance	Balance of excitatory and inhibitory inputs
EIF	exponential integrate-and fire
EPSC	excitatory postsynaptic current
LIF	leaky integrate-and fire
LTD	long-term depression
LTP	long-term potentiation
PV	parvalbumin
SST	somatostatin
STDP	spike-timing-dependent plasticity
VIP	vasoactive intestinal peptide

List of scientific communications

Peer-reviewed publications

- Miehl, C. & Gjorgjieva, J. Stability and learning in excitatory synapses by nonlinear inhibitory plasticity. *PLoS Computational Biology* **18,** e1010682 (2022)
- Wu, Y. K., Miehl, C. & Gjorgjieva, J. Regulation of circuit organization and function through inhibitory synaptic plasticity. *Trends in Neurosciences* **45,** 884–898 (2022)
- Miehl, C., Onasch, S., Festa, D. & Gjorgjieva, J. Formation and computational implications of assemblies in neural circuits. *Journal of Physiology* (2022)
- Schulz, A., Miehl, C., Berry II, M. J. & Gjorgjieva, J. The generation of cortical novelty responses through inhibitory plasticity. *eLife* **10,** e65309 (2021)
- Montangie, L., Miehl, C. & Gjorgjieva, J. Autonomous emergence of connectivity assemblies via spike triplet interactions. *PLoS Computational Biology* **16,** e1007835 (2020)
- Field, R. E., D'amour, J. A., Tremblay, R., Miehl, C., Rudy, B., Gjorgjieva, J. & Froemke, R. C. Heterosynaptic Plasticity Determines the Set Point for Cortical Excitatory-Inhibitory Balance. *Neuron* **106,** 842–854 (2020)

Peer-reviewed poster presentations

- Miehl, C., Song, S. C., Froemke, R. C. & Gjorgjieva, J. Mechanisms of plasticity for pup call sounds in the maternal auditory cortex. *Poster presentation at Bernstein Conference, Berlin, Germany* (2022)
- Miehl, C., Song, S. C., Froemke, R. C. & Gjorgjieva, J. Mechanisms of plasticity for pup call sounds in the maternal auditory cortex. *Poster presentation at FENS Conference, Paris, France* (2022)
- Miehl, C., Song, S. C., Froemke, R. C. & Gjorgjieva, J. Mechanisms of plas-

ticity for pup call sounds in the maternal auditory cortex. *Poster presentation at COSYNE Conference, Lisbon, Portugal* (2022)

- Miehl, C., Song, S. C., Froemke, R. C. & Gjorgjieva, J. Short- and long-term plasticity shape inter-stimulus interval-dependent responses in the auditory cortex. *Poster presentation at Bernstein Conference, Online conference* (2021)
- Miehl, C., Schulz, A., Berry II, M. J. & Gjorgjieva, J. The generation of cortical novelty responses through inhibitory plasticit. *Poster presentation at COSYNE Conference, Online Conference* (2021)
- Miehl, C. & Gjorgjieva, J. Inhibitory plasticity as a control mechanism of excitatory weight dynamics. *Poster presentation at Bernstein Conference, Berlin, Germany* (2020)
- Miehl, C. & Gjorgjieva, J. Non-linear inhibitory plasticity can homeostatically regulate excitatory weight dynamics. *Poster presentation at COSYNE Conference, Denver, Colorado, USA* (2020)
- Miehl, C. & Gjorgjieva, J. Functional implications of inhibitory plasticity as a homeostatic mechanism. *Poster presentation at Bernstein Conference, Berlin, Germany* (2019)
- Miehl, C., Field, R. E., D'amour, J. A., Tremblay, R., Kruglikov, I., Rudy, B., Froemke, R. C. & Gjorgjieva, J. Heterosynaptic plasticity determines the set-point for cortical excitatory-inhibitory balance. *Poster presentation at COSYNE Conference, Lisbon, Portugal* (2019)

Oral presentations

- Miehl, C. The role of adaptation in neuronal coding. *Talk at Workshop on "Adaptivity in nonlinear dynamical systems", Potsdam, Germany* (2022)
- Miehl, C. Inhibitory long-term plasticity: homeostasis and beyond. *Talk at CRC autumn event, Frankfurt, Germany* (2022)
- Miehl*, C. & Schulz*, A. The generation of cortical novelty responses through inhibitory plasticity. *Talk at Allen Institute Jorunal Club, online presentation* (2022)
- Miehl, C. Implications of inhibitory plasticity as a homeostatic mechanism. *Talk at CRC autumn event, online presentation* (2021)

Acknowledgments

First and foremost, I would like to thank my supervisor Julijana Gjorgjieva for not only giving me the opportunity to complete my PhD in her group, but also to allow me to develop my own scientific ideas and interests. I am grateful for all the experiences I could gain during my time as a PhD, from visiting multiple conferences all around the world and doing a research stay with our experimental collaborator.

I want to specifically thank all the co-authors from my publications: Lisandro Montangie, Auguste Schulz, Sebastian Onasch, Dylan Festa and Yue Kris Wu. I enjoyed collaborating, working, and thinking together about specific scientific problems with all of you. Furthermore, a big "Thank you!" to all other previous and current members of the Gjorgjieva lab who have influenced me a lot in my personal and scientific development. Special thanks to Shreya Lakhera for providing useful feedback on my thesis.

Furthermore I want to thank Robert Froemke for allowing me to visit his lab at NYU and collaborating with me, and his postdoc Soomin Song for interesting conversations.

I also want to thank the vibrant and active community at the Max Planck Institute for Brain Research.

I would like to thank the members of my thesis advisory committee, Gilles Laurent and Peter Jedlicka for their helpful comments and suggestions. Furthermore, I would like to thank the examiners of my thesis Peter Jedlicka and Matthias Hennig for carefully reading my thesis and providing feedback to its content.

Finally, I would like to thank my family, my newly made friends in Frankfurt, my "old" friends in Austria, and specifically Verena for constant support.

This thesis is dedicated to my father.

Appendix

I. Stability and learning in excitatory synapses by nonlinear inhibitory plasticity

Miehl, C. & Gjorgjieva, J. Stability and learning in excitatory synapses by nonlinear inhibitory plasticity. *PLoS Computational Biology* **18(12)**, e1010682 (2022). https://doi.org/10.1371/journal.pcbi.1010682

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Stability and learning in excitatory synapses by nonlinear inhibitory plasticity

Christoph Miehl[1,2]*, **Julijana Gjorgjieva**[1,2]*

1 Max Planck Institute for Brain Research, Frankfurt am Main, Germany, **2** School of Life Sciences, Technical University of Munich, Freising, Germany

* christoph.miehl@brain.mpg.de (CM); gjorgjieva@tum.de (JG)

Abstract

Synaptic changes are hypothesized to underlie learning and memory formation in the brain. But Hebbian synaptic plasticity of excitatory synapses on its own is unstable, leading to either unlimited growth of synaptic strengths or silencing of neuronal activity without additional homeostatic mechanisms. To control excitatory synaptic strengths, we propose a novel form of synaptic plasticity at inhibitory synapses. Using computational modeling, we suggest two key features of inhibitory plasticity, dominance of inhibition over excitation and a nonlinear dependence on the firing rate of postsynaptic excitatory neurons whereby inhibitory synaptic strengths change with the same sign (potentiate or depress) as excitatory synaptic strengths. We demonstrate that the stable synaptic strengths realized by this novel inhibitory plasticity model affects excitatory/inhibitory weight ratios in agreement with experimental results. Applying a disinhibitory signal can gate plasticity and lead to the generation of receptive fields and strong bidirectional connectivity in a recurrent network. Hence, a novel form of nonlinear inhibitory plasticity can simultaneously stabilize excitatory synaptic strengths and enable learning upon disinhibition.

Author summary

An important task the brain needs to solve is the so-called 'stability-flexibility problem'. On the one hand, any representation in the brain, for example a long-lasting memory, has to be stable for a long time. On the other hand, new representations need to be flexibly learned at any time. Learning and memory formation are implemented through the plasticity of synaptic connections, which describe how the activity in neurons is translated into changes of synaptic strength between these neurons. We propose a novel form of synaptic plasticity at synapses from inhibitory to excitatory neurons as a mechanism to stabilize learned representations, while a gating signal triggers the learning of new representations. We identify the dominance of inhibition over excitation and a nonlinear dependence of inhibitory plasticity on the postsynaptic firing rate as important aspects of our newly proposed plasticity mechanism. Our computational model allows us to uncover the underlying mechanism behind various experimental findings related to synaptic plasticity and sensory perturbations, and we formulate multiple experimentally-testable predictions.

OPEN ACCESS

Citation: Miehl C, Gjorgjieva J (2022) Stability and learning in excitatory synapses by nonlinear inhibitory plasticity. PLoS Comput Biol 18(12): e1010682. https://doi.org/10.1371/journal.pcbi.1010682

Editor: Hugues Berry, Inria, FRANCE

Received: April 5, 2022

Accepted: October 25, 2022

Published: December 2, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pcbi.1010682

Data Availability Statement: All relevant data are within the manuscript and its Supporting information files. All code is available at https://github.com/comp-neural-circuits/Nonlinear-inhibitory-plasticity.

Funding: CM and JG thank the Max Planck Society for funding and a NARSAD Young Investigator Grant from the Brain and Behavior Research Foundation to JG. We also thank the Deutsche Forschungsgemeinschaft (DFG) for funding through the Collaborative Research Centre (CRC) 1080. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Learning and memory formation in the brain are hypothesized to be implemented by synaptic changes undergoing Hebbian plasticity whereby joint pre- and postsynaptic activity increase the strength of synaptic connections [1, 2]. However, Hebbian long-term plasticity of excitatory synapses to other excitatory neurons, referred to as excitatory plasticity, is inherently unstable [3]. Increasing excitatory synaptic strengths leads to an increase in the firing rates of excitatory postsynaptic neurons which in turn further increases synaptic strengths. This positive feedback loop is called 'Hebbian runaway dynamics' [4]. To counteract unstable synaptic growth and control resultant rate dynamics, some form of homeostatic control is needed. Experimental studies have uncovered multiple homeostatic mechanisms. One prominent mechanism is synaptic scaling, where synaptic connections onto a given excitatory neuron potentiate or depress, while preserving relative strengths, to maintain a target level of activity [5, 6]. An alternative mechanism that has gained much recent attention is heterosynaptic plasticity [7, 8], which occurs both at excitatory and inhibitory synapses that have not been directly affected by the induction of plasticity [9]. A third plausible homeostatic mechanism with significant experimental evidence is intrinsic plasticity which affects the intrinsic excitability of single neurons by adjusting the distribution of different ion channel subtypes [10, 11].

Various computational studies have benefited from this plethora of experimental evidence for homeostatic control of firing rates and synaptic strengths, and implemented a range of computational models from purely phenomenological ones to detailed biophysical ones. Some relatively straightforward ways to stabilize firing rates and control synaptic strengths in models include imposing upper bounds on synaptic strengths, applying normalization schemes which adjust synaptic strengths by preserving the total sum of incoming weights into a neuron [3, 12] and assuming that the plasticity mechanism modifying synaptic strengths is itself plastic—called 'metaplasticity' [13–16]. These can often be linked to the above experimentally described homeostatic mechanisms. Computational studies have also begun to uncover the various, often complementary, functional roles of different homeostatic mechanisms, e.g. of synaptic scaling versus intrinsic plasticity [16] or heterosynaptic plasticity [9]. However, how exactly synaptic plasticity and homeostatic mechanisms interact to control synaptic strengths, and yet enable learning, is still partially unresolved [17–19]. Part of the challenge is that the experimentally measured timescales of synaptic scaling are too slow to stabilize the Hebbian runaway dynamics in computational models, where much faster normalization schemes are used instead [16, 20–23]. This is sometimes referred to as the 'temporal paradox' of homeostasis [24–26]. A related problem to the integration of plasticity and homeostasis is the trade-off between stability and flexibility. While stimulus representations need to be stable, for instance to allow long-term memory storage, the system also needs to be flexible to allow re-learning of the same, or learning of new representations [27]. This has been successfully achieved in some circumstances. For example, implementing metaplasticity in the excitatory connections through a sliding threshold between potentiation and depression can generate weight selectivity and firing rate stability [13, 14, 16, 24]. Additionally, heterosynaptic plasticity has been modeled to stabilize synaptic weight dynamics, while still allowing learning [9, 28–30], including behavioral learning [31]. A strong candidate for stabilizing synaptic weights is the induction of homosynaptic LTP (LTD) together with heterosynaptic LTD (LTP) at nearby synapses, referred to as the 'Mexican hat' profile of homo- and heterosynaptic plasticity [32, 33].

Here, we investigate an alternative, under-explored mechanism to control and stabilize excitatory synaptic strengths and their dynamics: long-term plasticity of inhibitory-to-excitatory (I-to-E) synapses, also referred to as inhibitory plasticity. Experimental paradigms have characterized diverse forms of inhibitory plasticity, usually via high-frequency stimulation

[34–36] and via pairing of presynaptic and postsynaptic spikes [37, 38]. Inhibition has been shown to control the plasticity mechanisms regulating connection strengths between excitatory neurons depending on their firing rates [39] as well as precise spike timing [40–42]. Inhibitory plasticity can even dictate the direction of excitatory plasticity, shifting between depression or potentiation [43]. Computational models have shown that different forms of inhibitory plasticity can stabilize excitatory rates [44–46]. Given this potential of inhibitory plasticity to affect so many different aspects of synaptic strength and firing rate dynamics in a network, it remains unclear what properties are important for achieving stability, while still enabling neural circuits to learn.

Using computational modeling, we characterize a novel mechanism of inhibitory plasticity with two key features. First, we propose that inhibitory plasticity should depend nonlinearly on the firing rate of an excitatory postsynaptic neuron to robustly control and stabilize the strengths of excitatory synaptic connections made by that neuron. This means that for low postsynaptic rates, I-to-E synapses should depress, for high postsynaptic rates I-to-E synapses should potentiate and without any postsynaptic activity undergo no plasticity. This nonlinear dependence of inhibitory plasticity on the postsynaptic firing rate is sufficient for stability, without the need for additional homeostatic mechanisms. Second, we require a dominance of inhibition, which can be reflected in the larger number of synaptic connections, faster plasticity dynamics of inhibitory synapses or overall higher firing rates of inhibitory neurons relative to excitatory ones. Dominance of inhibition has already been demonstrated in circuits in the visual cortex which operate as inhibition-stabilized networks (ISNs) [47–49]. A direct consequence from our proposed novel mechanism of nonlinear inhibitory plasticity is the emergence of a fixed ratio of excitatory-to-inhibitory synaptic strengths when input rates are constant, in agreement with experimental data [37]. Besides stability, our proposed inhibitory plasticity mechanism can also support flexible learning of receptive fields and recurrent network structures by gating excitatory plasticity via disinhibition [50, 51]. Therefore, our results provide a plausible solution to the stability-flexibility problem by identifying key aspects of inhibitory plasticity, which provide experimentally testable predictions.

Results

A linear inhibitory plasticity rule fails to robustly stabilize weight dynamics

To investigate the plausibility of inhibitory plasticity as a control mechanism of excitatory synaptic strengths, we initially considered a model based on a feedforward inhibitory motif prominent in many brain circuits (Fig 1A). Here, a population of presynaptic excitatory neurons projects to a population of inhibitory neurons and both populations project to a single postsynaptic excitatory neuron. Such a motif could resemble, for instance, the excitatory input from the thalamus to excitatory and inhibitory neurons in a primary sensory cortical area [52]. We described the activity of neurons by their firing rates. We considered a network consisting of an excitatory postsynaptic neuron with a linear threshold transfer function and firing rate v^E, receiving input from N^E excitatory presynaptic neurons (each with index j) with firing rates ρ_j^E through excitatory weights w_j^{EE}, and from N^I inhibitory presynaptic neurons (each with index k) with firing rates v_k^I through inhibitory weights w_k^{EI}:

$$\tau_{FR}^E \dot{v}^E = -v^E + \left[\sum_{j=1}^{N^E} \rho_j^E w_j^{EE} - \sum_{k=1}^{N^I} v_k^I w_k^{EI} \right]_+ , \quad (1)$$

where $[]_+$ denotes a rectification that sets negative values to zero. The inhibitory neurons follow similar dynamics and are driven by the same N^E presynaptic excitatory neurons with firing

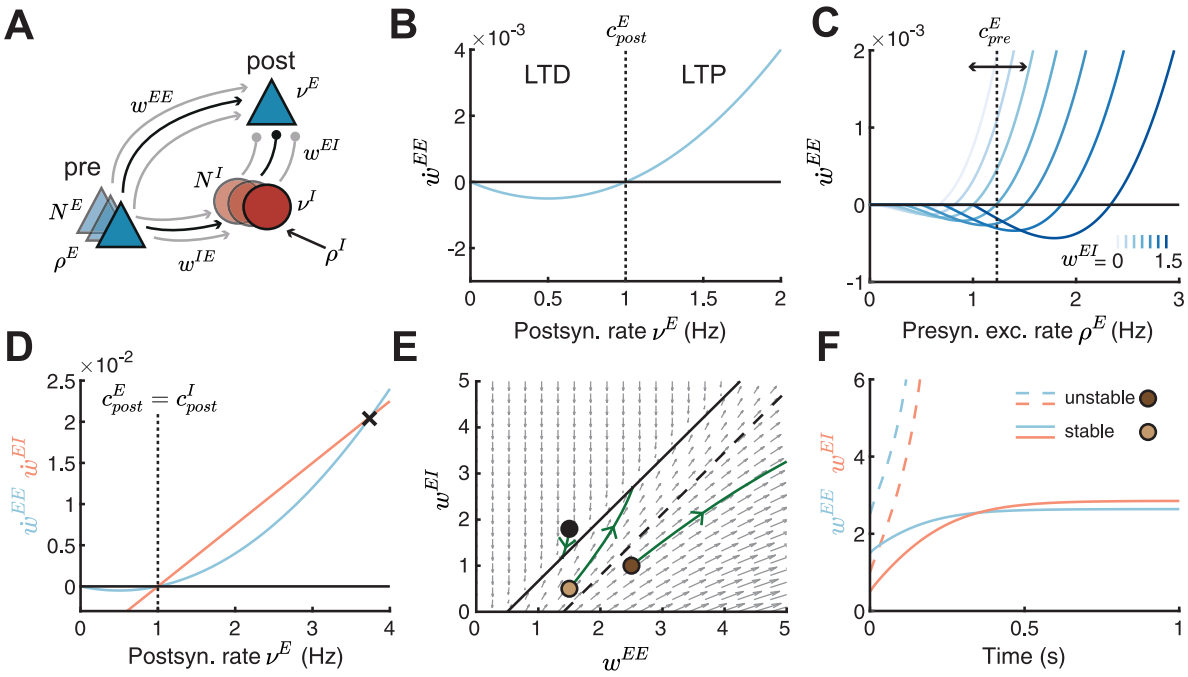


Fig 1. Linear inhibitory plasticity fails to stabilize weights for high postsynaptic firing rates. A. Schematic of a feedforward inhibitory motif. A single postsynaptic excitatory neuron with rate ν^E receives input from N^E excitatory presynaptic neurons, with firing rate ρ^E and weight w^{EE} and N^I inhibitory presynaptic neurons, with firing rate ν^I and weight w^{EI}. The inhibitory neurons receive external excitatory input with rate ρ^I and input from the presynaptic excitatory neurons via w^{IE}. **B**. Plasticity curve of E-to-E weights ($\dot{w}^{EE}$, blue) as a function of the postsynaptic rate ν^E. The postsynaptic LTD/LTP threshold c^E_{post} is set to 1. **C**. E-to-E weight change ($\dot{w}^{EE}$) as a function of the presynaptic excitatory rate ρ^E for different I-to-E weights w^{EI} ranging from 0 to 1.5. The presynaptic LTD/LTP threshold c^E_{pre} is shown for $w^{EI} = 0.75$ (vertical dashed line). **D**. Plasticity curves of E-to-E ($\dot{w}^{EE}$, blue) and I-to-E ($\dot{w}^{EI}$, red) weights as a function of the postsynaptic rate ν^E. The excitatory and inhibitory LTD/LTP thresholds are identical ($c^E_{post} = c^I_{post}$). The black cross marks the postsynaptic rate where the plasticity curves cross beyond which the weight dynamics become unstable. **E**. Phase portrait of the dynamics of E-to-E (w^{EE}) and I-to-E (w^{EI}) weights. Gray arrows indicate the direction of weight evolution over time, points represent three different weight initializations, $[w^{EE}_0, w^{EI}_0] = \{[1.5, 1.8], [1.5, 0.5], [2.5, 1]\}$, and green lines represent the weight evolution for each case. The two colored points represent initial weights in F. Black line indicates the line attractor and the dashed line separates stable from unstable initial conditions (Methods, Eq 20). **F**. E-to-E (w^{EE}, blue) and I-to-E (w^{EI}, red) weights as a function of time for stable (solid lines, $[w^{EE}_0, w^{EI}_0] = [1.5, 0.5]$) and unstable (dashed lines, $[w^{EE}_0, w^{EI}_0] = [2.5, 1]$) initial conditions.

https://doi.org/10.1371/journal.pcbi.1010682.g001

rates ρ^E_j through excitatory weights w^{IE}_j and additional external input with firing rate ρ^I_k,

$$\tau^I_{FR}\dot{\nu}^I_k = -\nu^I_k + \left[\sum_{j=1}^{N^E}\rho^E_j w^{IE}_j + \rho^I_k\right]_+ . \tag{2}$$

Here, τ^E_{FR}, τ^I_{FR} denote the time constants of the firing rate dynamics. For simplicity, we do not use subscripts for neuron identity and interpret all variables as mean values and hence can denote the total excitatory input to the postsynaptic neuron as $N^E\rho^E w^{EE}$ and the total inhibitory input as $N^I\nu^I w^{EI}$. The synaptic weights w^{EE} and w^{EI} are plastic according to different plasticity rules (see below).

Experimental studies have shown that the sign and magnitude of excitatory plasticity depends nonlinearly on the firing rates [53–55]. Inspired by these findings, we implemented plasticity of E-to-E synaptic connections w^{EE} (or weights) as a nonlinear function of the postsynaptic rate ν^E (Fig 1B):

$$\tau^E_w\dot{w}^{EE} = \rho^E\nu^E(\nu^E - c^E_{post}). \tag{3}$$

Here, ρ^E denotes the excitatory presynaptic rate and τ_w^E is the timescale of excitatory plasticity. We refer to the postsynaptic rate at which the plasticity changes sign as the 'postsynaptic LTD/LTP threshold', denoted by c_{post}^E. If the firing rate ν^E is smaller than the threshold c_{post}^E, then the change in synaptic strength is negative leading to long-term depression (LTD), while if ν^E is larger than c_{post}^E, then the change in synaptic strength is positive leading to long-term potentiation (LTP) (Fig 1B and S1 Fig). This means that increasing the excitatory postsynaptic firing rate will lead to potentiation of excitatory weights, and in a positive feedback loop will further increase the neuron's firing rate—known as the classical problem of 'Hebbian runaway dynamics'.

Hence, we wanted to determine a plausible mechanism to counteract excitatory runaway dynamics. We postulated that regulating the inhibitory input into the postsynaptic neuron provides an efficient way to stabilize excitatory weights and firing rates. In our framework, inhibitory neurons can affect excitatory plasticity in three equivalent ways. (1) The number of inhibitory synapses N^I onto the postsynaptic neuron can change, for example, through the growth or removal of synapses via structural plasticity. (2) The strength of I-to-E synapses w^{EI} can change via inhibitory plasticity. (3) Finally, the rate of inhibitory neurons ν^I can also change through the external excitatory input to the inhibitory neurons ρ^I or the excitatory-to-inhibitory weight w^{IE}. Various experimental studies have revealed that the plasticity of I-to-E synapses can be induced via the stimulation of the relevant input pathways [34, 35, 43]. Given this experimental evidence for the plasticity of I-to-E synapses, we examined the influence of changing the strength of I-to-E synapses, w^{EI}, on the strength and magnitude of E-to-E synapses, w^{EE} (Fig 1C).

We found that stronger w^{EI} weights rates require higher presynaptic excitatory rates to induce LTP, while weaker w^{EI} weights require lower presynaptic excitatory rates to induce LTP. This effectively leads to a shift of the threshold between LTD and LTP as a function of the presynaptic excitatory firing rate as w^{EI} changes. We refer to the presynaptic excitatory firing rate at which the plasticity changes sign between potentiation and depression as the 'presynaptic LTD/LTP threshold', denoted by c_{pre}^E (Fig 1C). In contrast to the fixed postsynaptic LTD/LTP threshold, c_{post}^E(Fig 1B), this presynaptic LTD/LTP threshold depends, among others, on the strength of I-to-E synapses (Fig 1C; Methods, Eq 13).

Rather than hand-tuning the strength of I-to-E synapses, here we propose that a particular inhibitory plasticity rule can dynamically adjust their strength as a function of presynaptic inhibitory and postsynaptic excitatory activity. However, the exact form of this plasticity has not yet been mapped experimentally. Therefore, we first investigated an inhibitory plasticity rule widely-used in computational models which depends linearly on the postsynaptic rate ν^E [44, 56] (Fig 1D, $\dot{w}^{EI}$):

$$\tau_w^I \dot{w}^{EI} = \nu^I(\nu^E - c_{post}^I). \tag{4}$$

Here, τ_w^I denotes the timescale of inhibitory plasticity. As for excitatory plasticity, we refer to the postsynaptic rate at which inhibitory plasticity changes from LTD to LTP as the 'inhibitory postsynaptic LTD/LTP threshold', denoted by c_{post}^I. This threshold determines the 'target rate' of the postsynaptic neuron [44]. If the excitatory postsynaptic neuron fires at higher rates than c_{post}^I, inhibitory LTP leads to a decrease of its firing rate, while if the neuron fires at lower rates than c_{post}^I, inhibitory LTD increases its rate. To prevent an unstable scenario where excitatory (Eq 3) and inhibitory plasticity (Eq 4) push the postsynaptic excitatory neuron towards two different firing rates, here we assume that the excitatory and inhibitory thresholds are matched (Fig 1D, $c_{post}^E = c_{post}^I$).

To investigate the effect of this 'linear inhibitory plasticity' mechanism on the temporal evolution of excitatory and inhibitory synaptic weights, w^{EE} and w^{EI}, we plotted the flow field in the phase plane w^{EI} vs. w^{EE} (Fig 1E). We found that the interaction of excitatory and inhibitory plasticity generates a line of stable fixed points (i.e. a line attractor) where both synaptic weights do not change any more (Fig 1E, black solid line; see Methods). The initial weights determine whether the weights ultimately converge to the line attractor and stabilize. When the initial E-to-E weights w^{EE} are much larger than the initial I-to-E weights w^{EI} (Fig 1E, below the dashed line), the weights become unstable (Fig 1E and 1F). Equivalently, the weights become unstable when the postsynaptic rate v^E is beyond the crossover point of the excitatory and inhibitory plasticity curves as a function of the postsynaptic excitatory rate (Fig 1D, black cross). For firing rates beyond this crossover point, the E-to-E weights increase faster than the I-to-E weights, leading to runaway dynamics.

In summary, our results suggest that a well-known form of inhibitory plasticity with a linear dependence on the postsynaptic excitatory firing rate can control excitatory weight changes only for a range of initial conditions. There exists a whole range of initial conditions (specifically where the E-to-E are larger than the I-to-E weights) where the postsynaptic excitatory firing rate is sufficiently large and where the weight dynamics explode. This scenario could be problematic if during normal development in the animal, the E-to-E and I-to-E weights are set up in this range, and implies the need for careful tuning to prevent unlimited weight growth.

A novel nonlinear inhibitory plasticity rule as a robust mechanism to stabilize excitatory weights

To ensure weight stability without fine tuning of the initial E-to-E and I-to-E weights, we proposed a novel inhibitory plasticity rule. The rule depends nonlinearly on the postsynaptic rate v^E, similarly to excitatory plasticity (Eq 3, Fig 2A):

$$\tau_w^I \dot{w}^{EI} = v^I v^E (v^E - c_{post}^I). \tag{5}$$

As before, to prevent a scenario where the two, excitatory and inhibitory, plasticity rules push the postsynaptic excitatory neuron towards two different firing rates, we assume here that the excitatory and inhibitory thresholds are matched $c_{post}^E = c_{post}^I$. However, as we show later, this assumption can be relaxed. Differently from the linear inhibitory plasticity rule (Eq 4), the nonlinear inhibitory plasticity rule ensures that I-to-E synapses do not change in the case where the postsynaptic firing rate is zero (Fig 2B, beyond gray line), as shown in experiments where postsynaptic activity or depolarization is needed to induce inhibitory plasticity [43]. Additionally, the nonlinear rule eliminates the region of initial weight configurations in the phase space where the weights grow out of bound; instead the weights converge to the line attractor (Fig 2B). Indeed, the E-to-E weights, I-to-E weights and the postsynaptic rate reach a stable configuration over time (Fig 2C). We calculated the condition leading to stable weight dynamics (Methods, Eqs 14–17) as a function of the excitatory and inhibitory input rates (v^I, ρ^E), the number of synapses (N^E, N^I) and the timescale of the plasticity mechanisms (τ_w^E, τ_w^I):

$$\frac{N^I (v^I)^2}{\tau_w^I} > \frac{N^E (\rho^E)^2}{\tau_w^E}. \tag{6}$$

This condition ensures stable weight dynamics whenever inhibition is more 'dominant' than excitation, either by having more inhibitory synapses (N^I), higher inhibitory rate (v^I), a faster

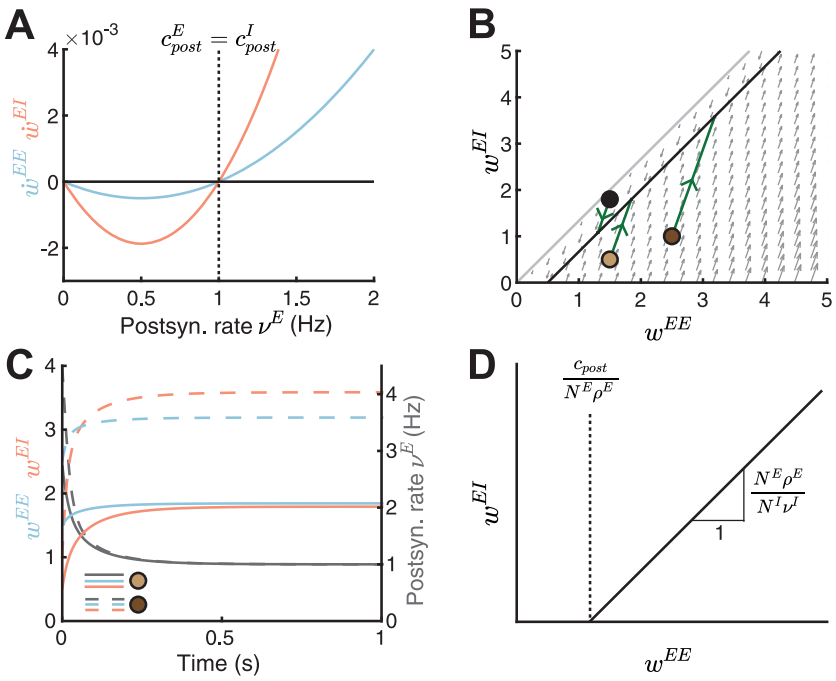


Fig 2. A novel nonlinear inhibitory plasticity rule can counteract runaway dynamics of excitatory-to-excitatory weights. A. Plasticity curves of E-to-E ($\dot{w}^{EE}$, blue) and I-to-E ($\dot{w}^{EI}$, red) weights as a function of the postsynaptic rate ν^E. The excitatory and inhibitory LTD/LTP thresholds are identical ($c^E_{post} = c^I_{post}$). **B**. Phase portrait of the dynamics of E-to-E (w^{EE}) and I-to-E (w^{EI}) weights. Gray arrows indicate the direction of weight evolution over time, points represent three different initial conditions of the weights, $[w_0^{EE}, w_0^{EI}] = \{[1.5, 1.8], [1.5, 0.5], [2.5, 1]\}$, and green lines represent the weight evolution for each initial condition. The two colored points represent initial weights in C. Black line indicates the line attractor and the gray line separates the space at which the postsynaptic firing rate is zero (no dynamics) or larger than zero (Methods, Eq 18). **C**. E-to-E (w^{EE}, blue) and I-to-E (w^{EI}, red) weight dynamics and postsynaptic rate dynamics (ν^E, gray) as a function of time for two initial conditions in B, $[w_0^{EE}, w_0^{EI}] = [1.5, 0.5]$ (solid lines) and $[w_0^{EE}, w_0^{EI}] = [2.5, 1]$ (dashed lines). **D**. The slope and intersection of the line attractor with the abscissa (black line) depend on the number and firing rates of excitatory and inhibitory neurons and the LTD/LTP threshold.

https://doi.org/10.1371/journal.pcbi.1010682.g002

timescale of inhibitory plasticity (τ_w^I) or a combination thereof. From now on, we assume a faster timescale of inhibitory relative to excitatory plasticity (Methods). An alternative way to achieve stability involves a feedback connection from the postsynaptic neuron to the inhibitory population (S2A Fig). In this case, sufficiently strong E-to-I feedforward and feedback weights guarantee stability in the presence of this feedback inhibitory motif (S2B–S2D Fig).

We found that the line attractor depends on several model parameters (see Methods, Eq 14) (Fig 2D)

$$w^{EI} = \frac{N^E\rho^E}{N^I\nu^I} w^{EE} - \frac{c_{post}}{N^I\nu^I}. \tag{7}$$

Under the assumption that the LTD/LTP thresholds of excitatory and inhibitory plasticity are the same, $c_{post} = c^E_{post} = c^I_{post}$, we found that the slope of the line attractor can be written as $N^E\rho^E/(N^I\nu^I)$, while the intersection of the line attractor with the abscissa can be written as $c_{post}/(N^E\rho^E)$. Therefore, by changing any of the network parameters we can predict the stable configuration to which the weights will converge.

Taken together, we have proposed a novel form of nonlinear inhibitory plasticity which can counteract excitatory runaway weight dynamics without the need for fine tuning. The proposed rule eliminates the need for additional homeostatic mechanisms and upper bounds on

the weights to stabilize weight dynamics. Our modeling approach allows us to dissect the exact dependencies of the stability condition on number of synapses, firing rates and plasticity timescales of excitatory and inhibitory neurons.

Dynamic matching of the excitatory and inhibitory postsynaptic thresholds between LTD and LTP

What happens if the postsynaptic thresholds between LTD and LTP for excitatory and inhibitory synapses are not identical, as might be the case in most biological circuits (Fig 3A)? We found that this leads to the disappearance of the line attractor (see Methods Eq 14). When the excitatory postsynaptic threshold is lower than the inhibitory postsynaptic threshold ($c^E_{post} < c^I_{post}$), both E-to-E and I-to-E weighs grow unbounded (Fig 3B). E-to-E weights cannot stabilize as they continue to potentiate ($\dot{w}^{EE} > 0$) even though the postsynaptic neuron is controlled by the fast inhibitory plasticity and approaches the target rate $\nu^E = c^I_{post}$ (Fig 3C).

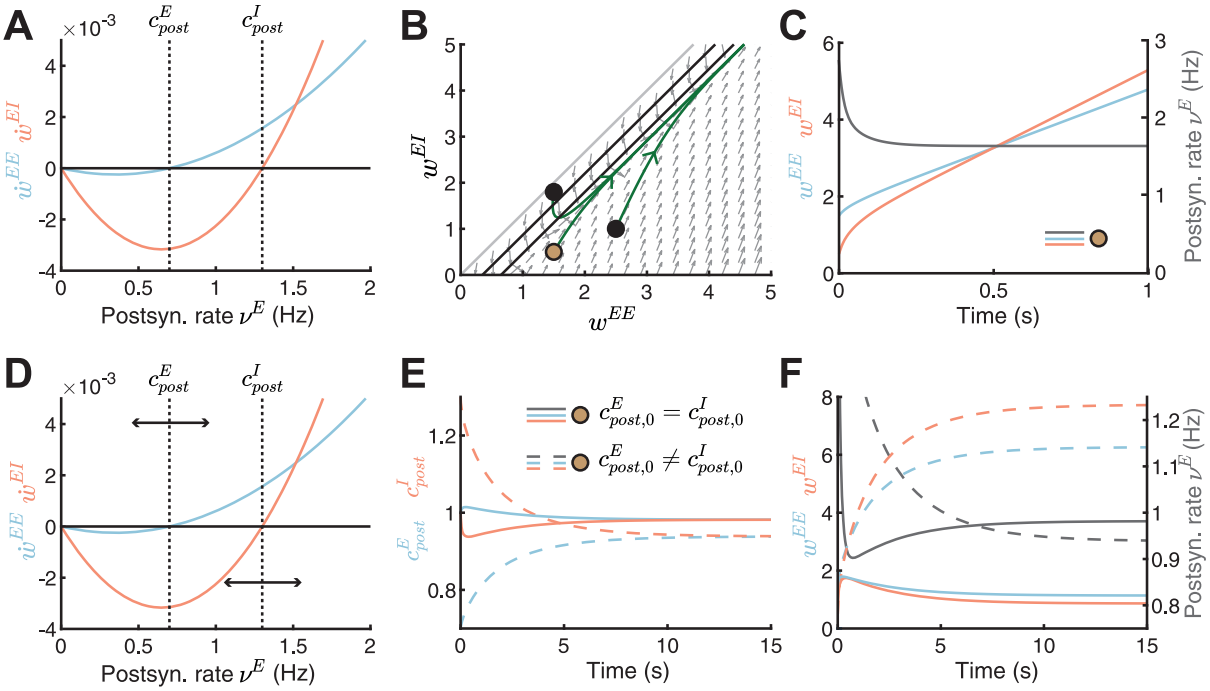


Fig 3. Dynamic matching of the excitatory and inhibitory postsynaptic LTD/LTP thresholds. A. Plasticity curves of E-to-E ($\dot{w}^{EE}$, blue) and I-to-E ($\dot{w}^{EI}$, red) weights as a function of the postsynaptic rate ν^E with static, non-identical LTD/LTP thresholds ($c^E_{post} = 0.7$, $c^I_{post} = 1.3$). **B**. Phase portrait of the dynamics of E-to-E (w^{EE}) and I-to-E (w^{EI}) weights for the scenario with static thresholds in A. Gray arrows indicate the direction of weight evolution over time, points represent three different initial conditions of the weights, $[w^{EE}_0, w^{EI}_0] = \{[1.5, 1.8], [1.5, 0.5], [2.5, 1]\}$, and green lines represent the weight evolution for each initial condition. The colored point represents initial weight in C and E-F. Black lines indicate the nullclines and the gray line separates the space at which the postsynaptic firing rate is zero (no dynamics) or larger than zero (Methods, Eq 18). **C**. Excitatory (w^{EE}, blue) and inhibitory (w^{EI}, red) weight dynamics and postsynaptic rate dynamics (ν^E, gray) for one initial condition in B, $[w^{EE}_0, w^{EI}_0] = [1.5, 0.5]$. The thresholds are static as in A. **D**. Postsynaptic LTD/LTP thresholds c^E_{post} and c^I_{post} shift dynamically depending on recent postsynaptic rate ν^E. For lower postsynaptic rate than the excitatory postsynaptic LTD/LTP threshold ($\nu^E < c^E_{post}$), c^E_{post} decreases, and for $\nu^E > c^E_{post}$, c^E_{post} increases. For higher postsynaptic rate than the inhibitory postsynaptic LTD/LTP threshold ($\nu^E > c^I_{post}$), c^I_{post} decreases, and for $\nu^E < c^I_{post}$, c^I_{post} increases (see Methods). **E**. Evolution of excitatory (c^E_{post}, blue) or inhibitory (c^I_{post}, red) postsynaptic LTD/LTP thresholds for two different initial conditions ($c^E_{post,0} = c^I_{post,0}$, full lines and $c^E_{post,0} = 0.7$, $c^I_{post,0} = 1.3$, dashed lines). Same initial weight condition as in C, $[w^{EE}_0, w^{EI}_0] = [1.5, 0.5]$, but for dynamic thresholds shown in D. **F**. Excitatory (w^{EE}, blue) and inhibitory (w^{EI}, red) weight dynamics and postsynaptic rate dynamics (ν^E, gray) for two different initial conditions ($c^E_{post,0} = c^I_{post,0}$, full lines and $c^E_{post,0} = 0.7$, $c^I_{post,0} = 1.3$, dashed lines). Same initial weight condition as in C, $[w^{EE}_0, w^{EI}_0] = [1.5, 0.5]$, but for dynamic thresholds shown in D. See E for the legend.

https://doi.org/10.1371/journal.pcbi.1010682.g003

Therefore, stability of firing rates does not imply stability of synaptic weights, especially in the case when the postsynaptic thresholds between LTD and LTP are non-equal. In the case of $c^E_{post} > c^I_{post}$, E-to-E and I-to-E weights steadily decrease.

Motivated by experimental findings and theoretical considerations that the excitatory threshold can slide [13, 57], here we proposed that the inhibitory threshold can also be dynamically regulated with both excitatory and inhibitory thresholds shifting into opposite directions (Fig 3D; see Methods). When the postsynaptic rate is lower than the excitatory postsynaptic LTD/LTP threshold ($v^E < c^E_{post}$), the excitatory postsynaptic LTD/LTP threshold should decrease, while when the postsynaptic rate is higher than the threshold ($v^E > c^E_{post}$), the excitatory threshold should increase. Similarly, when the postsynaptic rate is higher than the inhibitory postsynaptic LTD/LTP threshold ($v^E > c^I_{post}$), the inhibitory postsynaptic LTD/LTP threshold should decrease, while when the postsynaptic rate is lower than the threshold ($v^E < c^I_{post}$), the inhibitory threshold should increase. Eventually, these dynamics lead to the matching of excitatory and inhibitory LTD/LTP thresholds (Fig 3E). Therefore, the rates and weights can both be simultaneously stabilized (Fig 3F). The excitatory and inhibitory LTD/LTP thresholds can be matched, and the postsynaptic firing rate and synaptic weights stabilized also for other initializations of the LTD/LTP thresholds (S3A–S3C Fig). Implementing this dynamic threshold adjustment process generates different postsynaptic LTD/LTP threshold configurations (Fig 3E) and postsynaptic rates (Fig 3F, gray lines). Therefore, for different initializations of the LTD/LTP thresholds, a wide variety of stable postsynaptic rates is possible.

The nonlinear inhibitory plasticity rule can regulate the network response to perturbations

Excitatory and inhibitory LTD/LTP thresholds can be dynamically matched under most conditions, even if they are unequal (S3 Fig). Therefore, from now on we assumed that they are equal and static (as shown in Fig 2A). Next, we wanted to investigate how the new nonlinear inhibitory plasticity rule adjusts the network response following a perturbation. Inspired by sensory deprivation experiments [53, 54, 58] or direct stimulation of input pathways [59, 60], we investigated the network response to perturbing the excitatory presynaptic input rate (Fig 4A).

Independent of the direction of the perturbation, we found that the nonlinear inhibitory plasticity rule brings the excitatory postsynaptic rate back to the target rate (Fig 4B). The inhibitory rate v^I also readjusts because the inhibitory population receives input from the perturbed excitatory population. But the new inhibitory rate is different than the rate before the perturbation (Fig 4B). We found that a perturbation which decreases the excitatory input rate, leads to the depression of both type of weights w^{EE} and w^{EI}; in contrast, a perturbation which increases the excitatory input rate leads to their potentiation (Fig 4C). The firing rate response and synaptic weight changes to these perturbations are consistent with previous experimental results [61–66]. Since we used a threshold-linear neuron model (Eqs 1 and 2), our framework can even predict the steady values of the E-to-E and I-to-E synaptic weights, as well as their ratio, by calculating the line attractor in the phase space of w^{EE} and w^{EI} weights as a function of the perturbed parameter (Fig 4D).

Interestingly, we observed that this adjustment occurs by modulation of the presynaptic threshold between LTD and LTP for both excitatory and inhibitory plasticity. Decreasing the excitatory input rate decreases the excitatory presynaptic LTD/LTP threshold, hence limiting the range of presynaptic firing rates that generate depression. The reduction in the LTD/LTP

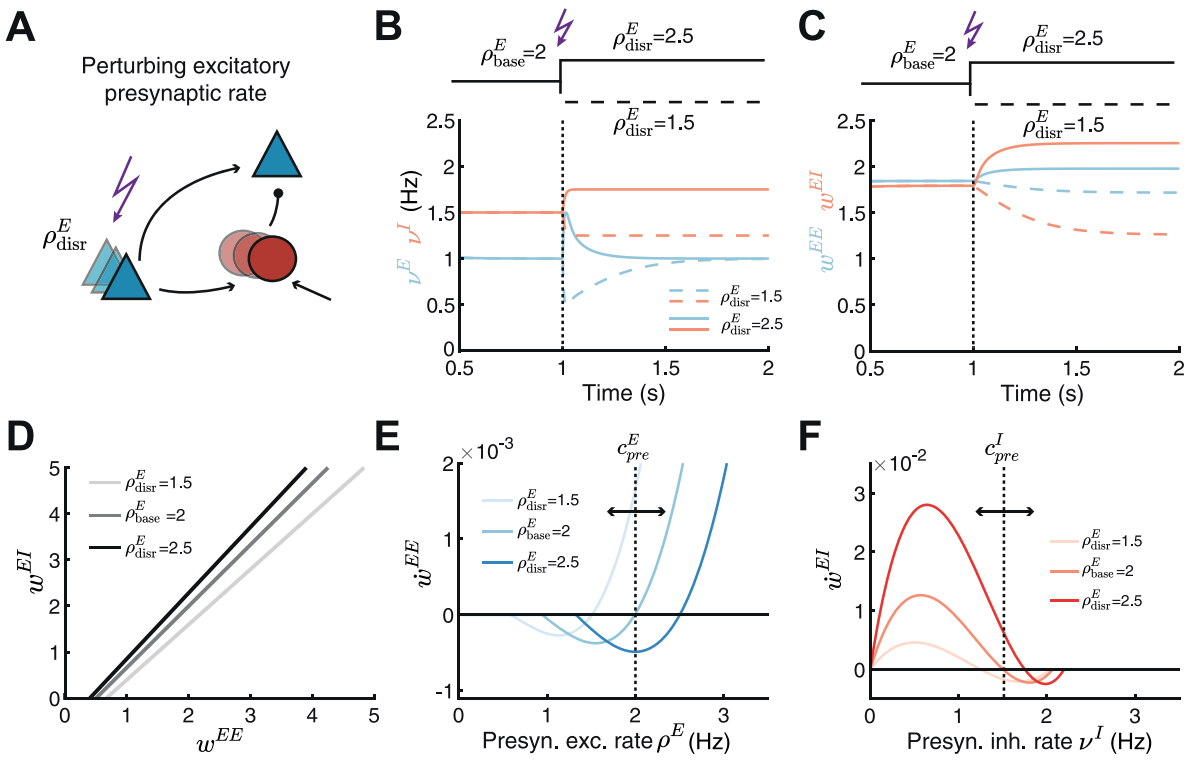


Fig 4. Nonlinear inhibitory plasticity can regulate the network response to perturbations. A. Schematic of perturbing the excitatory presynaptic rate in the inhibitory feedforward motif. We use the nonlinear inhibitory plasticity rule with identical excitatory and inhibitory LTD/LTP thresholds from Fig 2A. **B**. Effect of increasing (solid lines, $\rho^E_{\text{disr}} = 2.5$) or decreasing (dashed lines, $\rho^E_{\text{disr}} = 1.5$) excitatory input rates from a baseline of $\rho^E_{base} = 2$ on excitatory (blue) and inhibitory (red) firing rates. **C**. Same as B but for the w^{EE} and w^{EI} weights. **D**. The line attractor for the baseline input ρ^E_{base} and two input perturbations ρ^E_{disr}. **E**. E-to-E weight change $\dot{w}^{EE}$ as a function of the presynaptic excitatory rate ρ^E for the baseline input ρ^E_{base} and for two input perturbations ρ^E_{disr}. **F**. I-to-E weight change $\dot{w}^{EI}$ as a function of the inhibitory rate ν^I for the baseline input ρ^E_{base} and for two input perturbations ρ^E_{disr}.

https://doi.org/10.1371/journal.pcbi.1010682.g004

threshold follows from the relatively stronger depression of inhibitory compared to excitatory weights allowing the excitatory postsynaptic neuron to fire at the target rate even when the excitatory input is decreased. In contrast, we found that increasing the excitatory input rate increases the LTD/LTP threshold (Fig 4E). Such a shift in the plasticity threshold for excitatory synapses based on presynaptic activity has been measured in sensory deprivation experiments [53, 54, 58], and while restoring vision after sensory deprivation [54, 55] (although deprivation-induced effects occur on much slower timescales than in our plasticity model, see Discussion). Similarly to excitatory plasticity, perturbations in the excitatory input rate also shift the presynaptic threshold between LTD and LTP for inhibitory plasticity (Fig 4F). Since there is no experimental evidence for this effect, we propose it as a prediction for the shift between LTD and LTP for I-to-E weights (w^{EI}) in the presence of these perturbations. Even when implementing the plasticity rules with dynamic thresholds, performing the perturbations still leads to stable weight and rate configurations (S3D–S3F Fig).

In summary, the proposed nonlinear inhibitory plasticity can adjust the network response and synaptic strengths to excitatory input rate perturbations, similar to experimental findings. We predict that this shift occurs by modulating the presynaptic LTD/LTP thresholds for both excitatory and inhibitory plasticity.

The nonlinear inhibitory plasticity rule affects the excitatory-to-inhibitory weight ratio

We next wanted to investigate plausible functional roles of the newly proposed nonlinear inhibitory plasticity besides controlling excitatory and inhibitory firing rates and weights. Given our ability to calculate the steady states of the weights having used a linearly rectified neuron model (Fig 4D), we studied the ratio of E-to-E and I-to-E weights:

$$R^{E/I} = \frac{w^{EE}}{w^{EI}} = \frac{N^I v^I w^{EI} + c_{post}}{N^E \rho^E w^{EI}} = \frac{N^I (N^E \rho^E w^{IE} + \rho^I) w^{EI} + c_{post}}{N^E \rho^E w^{EI}} \tag{8}$$

with $v^I = N^E \rho^E w^{IE} + \rho^I$ (Methods). For strong I-to-E weights w^{EI}, the E/I weight ratio approximates to:

$$R^{E/I}_{\infty} = \frac{N^I v^I}{N^E \rho^E} = \frac{N^I (N^E \rho^E w^{IE} + \rho^I)}{N^E \rho^E} \tag{9}$$

(Fig 5A, inset; see Methods). Therefore, the E/I weight ratio is mainly determined by the ratio of excitatory and inhibitory input rates and the number of synapses, and is independent of the plastic synaptic weights (w^{EE} and w^{EI}). A fixed E/I weight ratio can be reached when the input

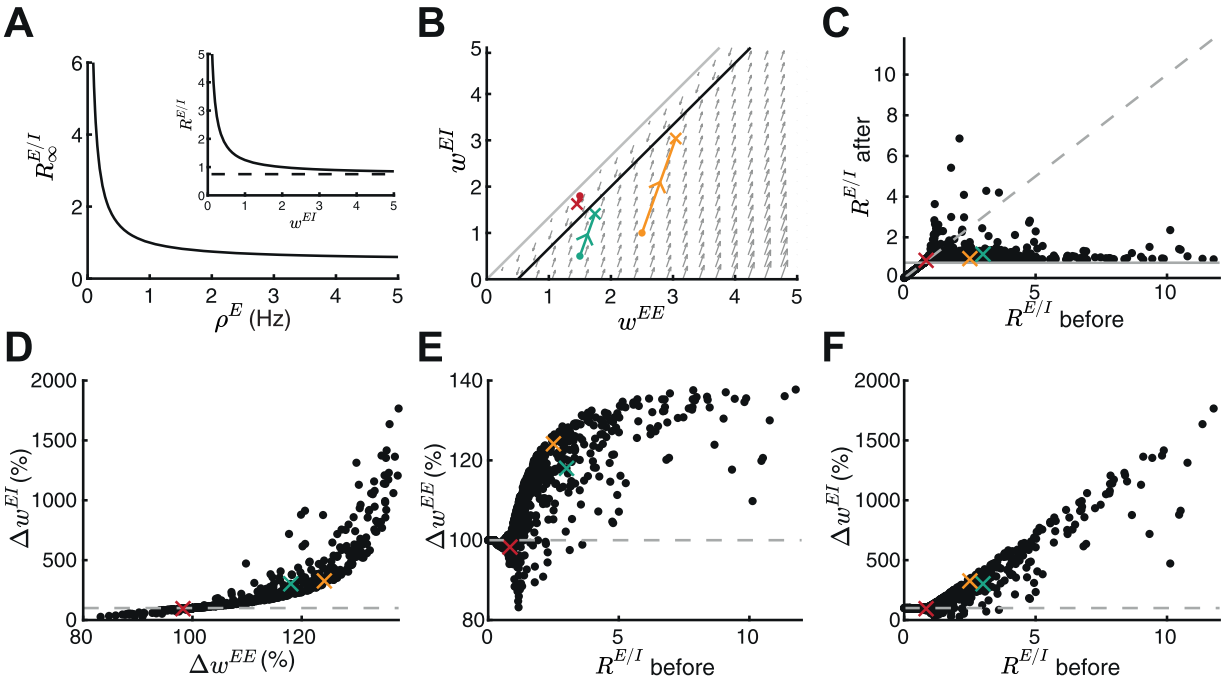


Fig 5. The nonlinear inhibitory plasticity rule maintains an excitatory-to-inhibitory weight ratio. A. The steady state E/I weight ratio R^{EI}_{∞} as a function of the presynaptic excitatory rate ρ^E. Inset: $R^{E/I}$ approaches the steady state $N^I v^I/(N^E \rho^E)$ (dashed line) for large I-to-E weights. **B-F** Based on a random initial weight configuration drawn from a uniform distribution in the range of [0, 3], excitatory and inhibitory plasticity was induced for 100 ms. Extreme initial E/I ratios ($R^{E/I}$ before > 12) were excluded from the analysis. **B**. Phase portrait of the dynamics of E-to-E (w^{EE}) and I-to-E (w^{EI}) weights. Gray arrows indicate the direction of weight evolution over time, colored points represent three different weight initialization, $[w_0^{EE}, w_0^{EI}] = \{[1.5, 1.8], [1.5, 0.5], [2.5, 1]\}$, colored lines represents the weight evolution for each case and the cross marks the weights after plasticity induction. The firing rates dynamics are similar as in Fig 2. **C**. E/I ratio before and after plasticity induction. Crosses indicate examples in B. Gray dashed line indicates the identity line and gray line indicates $R^{E/I}_{\infty}$. **D**. E-to-E weight change Δw^{EE} versus I-to-E weight change Δw^{EI} after plasticity induction in percent of initial synaptic weights. Dashed gray line indicates initial I-to-E weight strength and crosses indicate examples in B. **E**. E-to-E weight change Δw^{EE} as a function of E/I ratio $R^{E/I}$ before plasticity in percent of initial weights. Dashed gray line indicates initial E-to-E weight strength and crosses indicate examples in B. **F**. Same as E but for I-to-E weight change Δw^{EI}.

https://doi.org/10.1371/journal.pcbi.1010682.g005

rates are constant. The E/I ratio decreases as the presynaptic excitatory rate ρ^E increases (Fig 5A; Eq 8). This can be explained by considering that a higher excitatory input rate ρ^E generates more excitatory LTP (Fig 1C), which is counteracted by even more inhibitory LTP to stabilize weight dynamics. Analytically, this corresponds to a line attractor with a steeper slope (Figs 2D and 4D for increasing ρ^E) since the E/I ratio $R_\infty^{E/I}$ corresponds to the slope of the line attractor (Fig 2D; Methods).

Inspired by experiments [37], we evaluated the E/I ratio $R^{E/I}$ before and after inducing excitatory and inhibitory plasticity for multiple initial weight configurations (Fig 5B and 5C; Methods). As predicted analytically (Fig 5A), the E/I ratio after plasticity in these simulations approaches $R_\infty^{E/I}$ (Fig 5C), matching experiments in the mouse auditory cortex where inducing excitatory and inhibitory plasticity generates a fixed E/I ratio [37]. Large E/I ratios before plasticity induction show the most drastic changes, with high postsynaptic firing rates resulting from dominant excitation needing to be overcome by fast and drastic weight changes by nonlinear inhibitory plasticity. Indeed, we observed that the I-to-E weights exhibit more change than E-to-E weights (Fig 5D). This suggests that nonlinear inhibitory plasticity affects the E/I ratio more prominently than excitatory plasticity (Fig 5E and 5F). With the linear inhibitory plasticity rule [44], a fixed E/I ratio for constant input rates is only reached for initial weights which ultimately converge to the line attractor (Fig 1E).

Performance of the nonlinear inhibitory plasticity rule under varying presynaptic input and postsynaptic firing rate

We next investigated the effect of varying the presynaptic input or the postsynaptic firing rate on the stability of weight dynamics. Adding noise or a sinusoidal input to the postsynaptic firing rate ν^E (Methods) maintains synaptic weights within a certain range despite fluctuations (Fig 6A and 6B). We can understand the weight dynamics by studying how a varying input to the postsynaptic neuron affects the line attractors in the phase plane of the w^{EE} and w^{IE} weights. Adding an input to the postsynaptic neuron shifts only the point where the line attractor intersects the abscissa but does not change the slope (Fig 6C; Methods). Therefore, the weights remain constrained within a narrow region, without runaway dynamics. Even when implementing the plasticity rules with dynamic thresholds, adding postsynaptic noise or sinusoidal input leads to stable weight and rate configurations (S4 Fig).

The picture changes when the presynaptic input rate ρ^E varies (Methods). Here, both excitatory and inhibitory weights begin to slowly drift towards higher values while average firing rates remain stable (Fig 6D and 6E). The drift is due to a change in the presynaptic rate which affects the slope of the line attractors (see also Figs 2D and 4D). In the case of presynaptic sinusoidal input rate, the weights slowly increase while oscillating between the line attractors (Fig 6F). Therefore, while on a short timescales the interaction of only excitatory and inhibitory plasticity mechanisms seems to be sufficient to regulate weight and rate stability, we suggest that additional homeostatic mechanisms are necessary to regulate synaptic weight dynamics over longer timescales in the presence of noise or variability in the presynaptic input.

Gating of receptive field formation via a disinhibitory signal

What functional implications does the proposed nonlinear inhibitory plasticity rule have on setting up network circuitry? Other than controlling excitatory and inhibitory rates and weights, here we wanted to examine if the nonlinear inhibitory plasticity rule can also enable flexible learning. Various forms of synaptic plasticity have been observed to support receptive field formation and generate selectivity to stimulus features in the developing cortex [67]. To investigate the function of interacting excitatory and inhibitory plasticity at the network level,

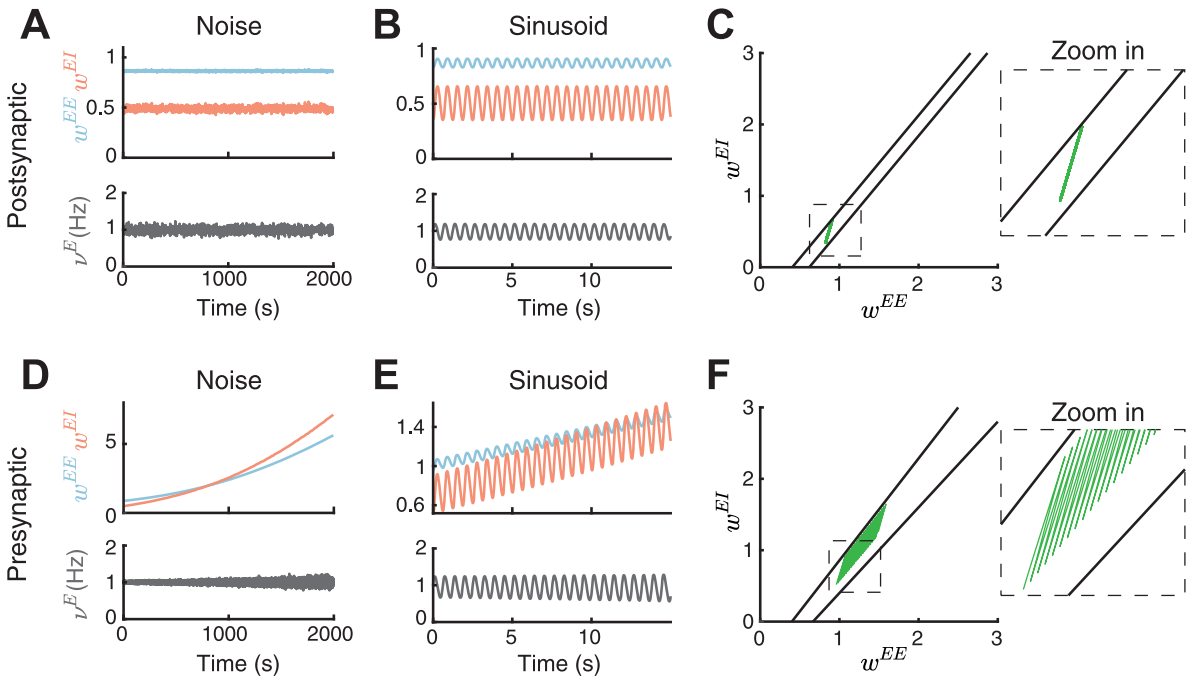


Fig 6. Performance of the nonlinear inhibitory plasticity rule under varying presynaptic input and postsynaptic firing rate. A. Adding noise to the postsynaptic firing rate. Top: E-to-E (w^{EE}, blue) and I-to-E (w^{EI}, red) as a function of time. Bottom: Postsynaptic rate dynamics (ν^E, gray) as a function of time. **B**. Same as A but after adding a sinusiodal input to the postsynaptic firing rate. **C**. Left: The line attractors in the w^{EE} and w^{EI} phase plane at the maximum and minimum of the postsynaptic firing rate after the addition of sinusoidal input (black lines) and the weight dynamics from B (green). Right: Zoom in of the phase plane. **D**. Same as A but after adding the noise to the presynaptic input rate. **E**. Same as B but after using a sinusoid for the presynaptic input rate. **F**. Same as C but after using a sinusoid for the presynaptic input rate with weight dynamics from E (green).

https://doi.org/10.1371/journal.pcbi.1010682.g006

we first extended the feedforward circuit motif to two independent pathways with pathway-specific inhibition (Fig 7A). We found that perturbing the presynaptic excitatory rate of both inputs in opposite directions, decreasing for input 1 and increasing for input 2, differently shifts the input-specific excitatory presynaptic LTD/LTP thresholds and establishes different E/I ratios (Fig 7B). This shift in the model is in agreement with experimental studies in the hippocampus which have shown that the thresholds between the induction of LTD and LTP are synapse-specific [59, 68]. These results suggest that the control of E-to-E weight dynamics via nonlinear inhibitory plasticity is input-specific.

Applying disinhibition by inhibiting the inhibitory population is a widely considered mechanism to 'gate' learning and plasticity [50, 51, 69]. To test the potential of the circuit with nonlinear inhibitory plasticity to learn, we applied a disinhibitory signal by decreasing the external excitatory input onto the inhibitory populations. We found that this decreases the inhibitory input onto the postsynaptic neuron and potentiates E-to-E synapses, w^{EE} (Fig 7C, $\rho^I < 1$). In contrast, increasing the input onto the inhibitory populations depresses E-to-E synapses (Fig 7C, $\rho^I > 1$). Therefore, disinhibition via perturbation of the inhibitory neurons has the capacity to induce plasticity at E-to-E synapses and can gate excitatory plasticity.

How do the current results generalize to larger circuits with multiple independent inputs? In addition to pathway-specific inhibition, in this extended circuit we also introduced an unspecific inhibitory population (Fig 7D). We presented different inputs to each of ten pathways in random order, corresponding to oriented bars in the visual cortex, or different single tone frequencies in the auditory cortex (Methods). We found that disinhibiting via the

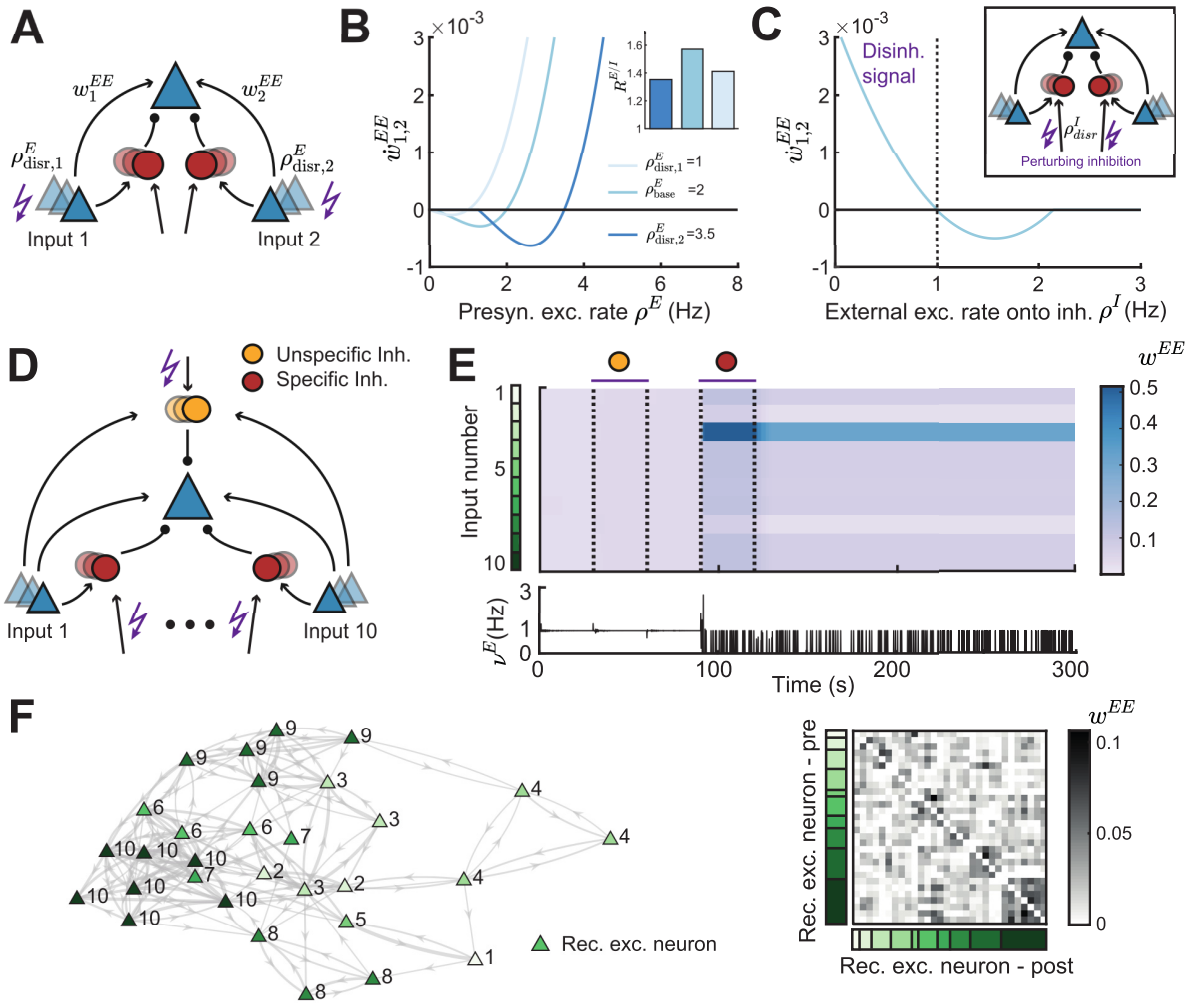


Fig 7. Gating of receptive field formation via a disinhibitory signal. A. Two independent inputs onto the same postsynaptic excitatory neuron. We perturb the presynaptic excitatory rate from input 1 or 2 ($\rho^E_{\mathrm{disr},1,2}$). **B**. Plasticity curve of E-to-E weights for input 1 or 2 ($\dot{w}^{EE}_{1,2}$) as a function of the presynaptic excitatory rate ρ^E for different input-specific perturbations $\rho^E_{\mathrm{disr},1,2}$. Inset: E/I weight ratio $R^{E/I}$ for different input-specific perturbations. **C**. Plasticity curve of E-to-E weights for input 1 and 2 ($\dot{w}^{EE}_{1,2}$) as a function of the external excitatory rate onto the inhibitory neurons ρ^I, corresponding to a perturbation ρ^I_{disr} of the inhibitory populations. Perturbing ρ^I below 1 Hz (dashed line) is interpreted as a disinhibitory signal. Inset: We perturb the external excitatory rate onto the inhibitory neurons ρ^I_{disr}. **D**. Ten independent inputs onto the same postsynaptic excitatory neuron with one inhibitory population unspecific to the input (yellow) and ten inhibitory populations each specific to one input (red). **E**. Top: Evolution of excitatory weights over time. Purple bars indicate the time window where either the unspecific (yellow) or all specific (red) inhibitory populations is disinhibited by applying a negative input onto the inhibitory neurons (Methods). Input number color coded in green. Bottom: Postsynaptic firing rate ν^E over time. **F**. Left: Network connectivity of recurrently connected excitatory neurons (triangles) after disinhibition. The number and the color indicates the input to which each neuron formed a receptive field (10 inputs in total). The thickness of the connection indicates the strength, only weights above 0.03 are shown. Distance and position of neurons is for visualization purposes only. Right: Ordered recurrent E-to-E connectivity matrix. Input number color coded in green as in panel E.

https://doi.org/10.1371/journal.pcbi.1010682.g007

unspecific inhibitory population does not selectively potentiate E-to-E weights, and hence does not generate competition among the different inputs. In this case, the selective potentiation of E-to-E weights corresponding to the inputs stimulated at a given time is counteracted by the potentiation of I-to-E weights specific to the stimulated inputs. This fast cancellation of any input-specific excitatory plasticity by input-specific inhibitory plasticity generates very small changes in the postsynaptic firing rate (Fig 7E, bottom). In contrast, equally disinhibiting

via all ten specific inhibitory populations strongly increases the E-to-E weights corresponding to only a subset of inputs, a process also called receptive field formation (Fig 7E). In this case, the selective potentiation of E-to-E weights corresponding to the inputs stimulated at a given time is counteracted by the potentiation of all unspecific I-to-E weights. Therefore, inhibitory plasticity does not cancel input-specific excitation. The random presentation order of the different inputs generates input-specific differences in excitatory weights and hence leads to competition. The input-specific potentiation is reflected in the fluctuating postsynaptic firing rate which increases only when the winning input is presented (Fig 7E, bottom).

Finally, we implemented a network of 30 recurrently connected excitatory neurons where each neuron in the circuit receives inputs from ten inputs and an unspecific and a specific inhibitory population (as in Fig 7D). In addition to the feedforward excitatory and inhibitory synapses, all recurrent E-to-E weights are also plastic. Similar as with a single postsynaptic neuron, we found that each of the excitatory neurons in the recurrent circuit forms a receptive field by becoming selective to one of the inputs (Fig 7F, left; number next to the neuron). In addition, strong bidirectional connections form among recurrent excitatory neurons with similar receptive fields due to their correlated activity (Fig 7F). This is consistent with strong bidirectional connectivity described in multiple experimental studies [70–72].

In summary, the newly proposed nonlinear inhibitory plasticity rule does not only ensure stable synaptic weights and activity, but also enables the formation of feedforward and recurrent structures upon disinhibition which gates synaptic plasticity.

Discussion

Hebbian excitatory synaptic plasticity is inherently unstable, requiring additional homeostatic mechanisms to control and stabilize excitatory-to-excitatory weight dynamics [4]. Here, we proposed a novel form of inhibitory plasticity (Fig 2), which can control excitatory and inhibitory firing rates and synaptic weights and enable stable and flexible learning of receptive fields in circuit models of the sensory cortex. We identified the dominance of inhibition over excitation (Eq 6) and identical postsynaptic thresholds between LTD and LTP for excitatory and inhibitory plasticity (compare Fig 2A and Fig 3A–3C) as two necessary features for stabilization of weight dynamics in our model. However, the latter requirement can be relaxed with a suitable dynamic mechanism that enables self-adjusting of the plasticity thresholds in opposite directions for excitatory and inhibitory plasticity (Fig 3D–3F). This novel form of nonlinear inhibitory plasticity can also regulate the network response to perturbations of excitatory input rates (Fig 4). Inhibitory plasticity affects the E/I weight ratio and establishes a fixed E/I ratio when input rates are constant (Eq 8), in agreement with experiments in the mouse auditory cortex where inducing excitatory and inhibitory plasticity sets a fixed E/I ratio [37] (Fig 5). We find that varying the presynaptic inputs or the postsynaptic firing rate differently affects stability (Fig 6). Besides stability, the proposed form of inhibitory plasticity enables receptive field formation following disinhibition to input-specific inhibitory populations and in recurrent networks supports the formation of strong bidirectional connectivity among neurons with similar receptive fields (Fig 7), suggesting a possible solution for the stability-flexibility problem.

Inhibitory plasticity as a control mechanism of excitatory-to-excitatory weight dynamics

In the last decades, experimental studies have uncovered multiple possible mechanisms to counteract Hebbian runaway dynamics, including synaptic scaling [5, 73], heterosynaptic plasticity [7, 8], and intrinsic plasticity [10, 11]. At the same time, computational studies have

included multiple homeostatic mechanisms, some of them the same as the experimental ones, to stabilize rates and weight dynamics, including upper bounds on the E-to-E weights, normalization mechanisms [3, 12, 16, 20, 21, 23], metaplastic changes of the plasticity function [13–16, 24], heterosynaptic plasticity [9, 29, 30] and intrinsic plasticity and synaptic scaling [16]. However, the spatial and temporal scales for integrating Hebbian and homeostatic plasticity continue to be subject of investigation [18, 25, 26]. This is especially the case for synaptic scaling which experimentally operates on timescales too slow to counteract the faster Hebbian synaptic plasticity (hours and days, vs. seconds and minutes). Heterosynaptic plasticity has been suggested as a more natural solution to the 'temporal paradox' problem since it operates on a similar timescale as Hebbian plasticity [9, 28, 29].

In our study, we instead proposed a novel inhibitory plasticity rule at inhibitory-to-excitatory synapses which depends nonlinearly on the postsynaptic firing rate as a solution to the temporal paradox problem. While nonlinear excitatory plasticity rules have been identified in experimental studies [53–55], less data is available for inhibitory plasticity. For example, presynaptic stimulation (hyperpolarization) and postsynaptic depolarization, have been shown to be required for inhibitory plasticity induction [74–77]. Additionally, high-frequency stimulation of presynaptic input pathways has been shown to potentiate inhibitory synapses [34–36]. Finally, the amount of inhibitory LTP has been shown to depend on the postsynaptic rate [43]. We designed our nonlinear inhibitory plasticity mechanism to be consistent with these findings: both, pre- and postsynaptic activity is necessary to induce inhibitory plasticity and the amount of LTP depends on the postsynaptic rate. Nonetheless, our rule is inconsistent with some experimental data which found no inhibitory plasticity for very high postsynaptic rates [43].

Several computational models have explored the functional roles of inhibitory spike-timing-dependent plasticity (iSTDP) operating at inhibitory-to-excitatory synapses. A commonly investigated plasticity rule has a symmetric learning window, where pre- and postsynaptic spikes close in time lead to LTP, and spikes further apart lead to LTD [44]. Similar symmetric learning windows have been identified experimentally in the auditory cortex [37], in the orbitofrontal cortex [78], and in the hippocampus [77]. Asymmetric learning windows, in which pre-post spike pairs lead to LTP and post-pre spike pairs lead to LTD have been observed in the entorhinal cortex [79], and also used in computational studies [45, 46]. For an inhibitory plasticity rule to successfully stabilize postsynaptic excitatory firing rates, it needs to implement a negative feedback mechanism whereby for high postsynaptic firing rates the inhibitory synaptic strength increases, while for low rates the inhibitory strength decreases, as is the case for our rule as well as others [44–46]. The nonlinear inhibitory plasticity we propose in our study is probably closest to a recent implementation of inhibitory plasticity via the voltage rule [80], since the voltage rule has a nonlinear dependency on postsynaptic firing rates [81].

Inhibitory plasticity as a metaplastic mechanism

The ability of the proposed nonlinear inhibitory plasticity to control the sign and magnitude of excitatory plasticity resembles metaplasticity, i.e. a plasticity mechanism that is plastic itself [13, 15]. We found that input perturbations modulate the excitatory presynaptic LTD/LTP threshold via a change of the I-to-E weights and inhibitory rates consistent with metaplasticity (Fig 4). Previous computational work has already suggested that a linear inhibitory plasticity rule can implement a metaplastic mechanism [56]. What mechanism underlies the sliding LTD/LTP threshold during the induction of plasticity is still an open question. Some experimental studies have suggested that inhibition can control the sign and magnitude of excitatory plasticity [40, 41, 43, 82]. Most intriguingly, it has been shown that application of

gamma-Aminobutyric acid (GABA) can increase the excitatory LTD/LTP threshold, while blocking GABA can decrease the excitatory LTD/LTP threshold [39], supporting our findings (Fig 1C).

The metaplasticity of excitatory plasticity was first suggested theoretically with the Bienenstock-Cooper-Munro (BCM) rule [13], and was later confirmed in sensory deprivation and restoration experiments [53, 54, 55, 58]. In the BCM rule, the metaplastic mechanism is implemented by a sliding LTD/LTP threshold depending on the excitatory postsynaptic rate [83, 84]. Higher (lower) postsynaptic rates lead to a higher (lower) postsynaptic LTD/LTP threshold making LTP (LTD) induction harder. Various implementations of the BCM rule have demonstrated its ability to achieve weight selectivity and firing rate stability without any inhibitory plasticity [13, 14, 24, 85]. Differently from the BCM model, in our nonlinear inhibitory plasticity rule the metaplastic sliding of the LTD/LTP threshold c^E_{pre} depends on the presynaptic excitatory rate (Fig 1C), whereas the postsynaptic LTD/LTP threshold c^E_{post} is fixed (except in Fig 3D–3F and S3 Fig). This apparent difference can be resolved by assuming that homeostatic mechanisms operate at two different timescales: fast and slow. Slow homeostasis has been linked to synaptic scaling which we (and others, e.g. [57]) hypothesize to be a possible mechanism behind changes in the postsynaptic threshold. It is usually observed on the timescales of many hours to days [6, 86, 87] but can also occur on the timescale of a few hours [88]. Fast homeostasis might be linked to disinhibition and inhibitory plasticity [89], which is induced on the timescale of minutes [9, 37, 90]. We suggest this is the case during sliding of the presynaptic LTD/LTP threshold mediated by our inhibitory plasticity rule. Nonetheless, it is plausible that both, presynaptic and postsynaptic metaplasticity exist in neuronal circuits. An advantage of achieving homeostasis via inhibitory plasticity, rather than a direct influence on the E-to-E weights, might be that there is no interference with stored information in E-to-E connections.

We used the metaplasticity of the nonlinear inhibitory plasticity rule to describe firing rate and weight changes in the model following perturbations of excitatory input (Fig 4) such as during sensory deprivation experiments [53, 54, 58]. For example, the decrease in inhibitory firing rates and weights after decreasing excitatory input in our model is consistent with the decrease in inhibitory activity following sensory deprivation [69, 87, 91]. Specifically, sensory deprivation has been shown to depress inhibitory synaptic strengths, decrease in the number of inhibitory synapses [62–66] (but see [92, 93]) and depress excitatory synaptic strengths [61, 94]. Increasing excitatory input in our model potentiates inhibitory weights, in agreement with experiments where up-regulating activity potentiates I-to-E synapses [95, 96]. We note that the plasticity induced by these sensory deprivation experiments occurs on much longer timescales of hours to days (see e.g. [57, 89]) compared to the shorter plasticity timescales of seconds or minutes in our model, suggesting that other mechanisms than the proposed nonlinear inhibitory plasticity drive the experimentally observed changes. Moreover, in our model we instantaneously and permanently change the input firing rate in contrast to the more complex changes in input patterns occurring during sensory deprivation. Therefore, the applied perturbation in our model could be better related to direct simulation of input pathways when similarly fast LTD/LTP threshold changes have been measured experimentally [59, 60, 97]).

Key features of the nonlinear inhibitory plasticity rule

For the novel inhibitory plasticity rule to stabilize E-to-E weight dynamics, two key features need to be fulfilled. First, I-to-E weight changes need to be more 'dominant' than E-to-E weight changes (Fig 2). More dominant means that I-to-E weights need to change with a higher magnitude at each time step compared to E-to-E weights, for all postsynaptic rates. If

excitatory plasticity exceeds inhibitory plasticity for a certain postsynaptic rate as in the case of linear inhibitory plasticity, weight dynamics will be unstable (Fig 1D–1F). In our model, dominance of nonlinear inhibitory plasticity is guaranteed by the condition in Eq 6, which involves relative number of synapses, presynaptic rates and plasticity timescales of excitation and inhibition to determine stability. Previous experimental work has reported that inhibitory synapses change more drastically than excitatory synapses [37], but inhibitory plasticity may be delayed relative to excitatory plasticity [50].

Second, the excitatory and inhibitory postsynaptic LTD/LTP thresholds need to be matched for stable weight dynamics, whereby excitatory and inhibitory synaptic change occur in the same direction for a given firing rate (Fig 2A–2C versus Fig 3A–3C). However, implementing a mechanism that dynamically shifts these thresholds in the opposite directions for excitatory vs. inhibitory plasticity based on experimental evidence [57], suggests that this match is not needed at all times. An interesting consequence from this dynamic threshold shift is the ability to achieve a range of firing rates. A limitation of the suggested dynamic threshold matching mechanism is that it is non-local whereby the thresholds for all input pathways converge to the same value. While this can still achieve stable weight dynamics and postsynaptic firing rates (S3G–S3I Fig; Methods), it can no longer induce competition among different inputs. Future work needs to investigate whether a different dynamic matching of excitatory and inhibitory LTD/LTP thresholds, perhaps one that is input-specific, can achieve the stable formation of receptive fields.

We found that the newly proposed nonlinear inhibitory plasticity rule achieves a fixed E/I ratio for constant input rates (Fig 5) in agreement with experimental data in the mouse auditory cortex where the induction of excitatory and inhibitory plasticity established a fixed E/I ratio [37]. We observed that inhibitory plasticity is the more dominant mechanism to achieve this. The dominance of inhibitory plasticity suggests a possible solution to the temporal paradox problem of integrating Hebbian excitatory plasticity and homeostasis [25], eliminating the requirement for additional fast stabilizing mechanisms in our model. While the relative timescales of excitatory and inhibitory plasticity mechanisms remain an open question, most computational models agree on the need for faster inhibitory than excitatory plasticity dynamics [25, 98].

Our framework is robust when noise or a varying input is added to the postsynaptic firing rate but not when the presynaptic rate varies (Fig 6). This suggests that additional homeostatic mechanisms are necessary to robustly counteract drift of synaptic weights when the input or the firing rates vary.

Functional implications of the nonlinear inhibitory plasticity rule

The interaction of the nonlinear inhibitory and excitatory plasticity in our model and the overlap of excitatory and inhibitory LTD/LTP thresholds lead to a fixed E/I weight ratio when input rates are constant (Fig 5A and 5C and Eq 8). This is consistent with several experimental studies which have suggested that inhibitory plasticity maintains a stable E/I ratio [9, 37, 43, 50, 96, 99–102]. For example, as our model would predict, some studies have found that the amount of inhibitory plasticity depends on the E/I ratio before plasticity induction (Fig 5F) [37, 103]. In these experiments, a change in E/I ratio is observed on the timescale of induction of plasticity (5–10 min) [37]. When we perturb the excitatory input rate as a model of sensory deprivation the E/I ratio increases (Fig 5A), consistent with sensory deprivation experiments [66, 69, 91, 94]. Despite the ability of the new nonlinear inhibitory plasticity rule to establish and maintain E/I balance, we acknowledge that there are various additional mechanisms that contribute, including heterosynaptic plasticity [9].

The emergence of fixed E/I ratio for constant input rates following from the stabilization of postsynaptic rates driven by the novel inhibitory plasticity rule ensures E/I balance. E/I balance is more broadly defined as the proportionality of total excitatory and inhibitory input onto a neuron [104]. In our model, once the neuron fires with a firing rate equal to the LTD/LTP threshold there is no more synaptic plasticity. To induce further weight changes, an additional gating signal is necessary that perturbs the postsynaptic firing rate. In our model, there are three ways to gate plasticity: (1) directly changing the postsynaptic rate (Fig 1B); (2) perturbing the excitatory input pathway (Fig 4); and (3) perturbing the inhibitory population (Fig 7C). The idea that inhibition gates excitatory plasticity is well-documented in the experimental literature [105–107].

Experimentally, both neuromodulation [50, 108] and disinhibitory circuits [51, 90, 109–111] can directly control the activity of inhibitory neurons and lead to excitatory plasticity. Based on this, we investigated the gating of plasticity via a disinhibitory signal in the context of receptive field formation. While receptive field formation has already been demonstrated in multiple computational studies [13, 45, 56], we propose that it can occur solely from the interaction of excitatory and inhibitory plasticity without any additional mechanism to induce competition among different inputs (Fig 7D and 7E). Recurrently connecting multiple postsynaptic excitatory neurons and allowing the connections between them to be plastic leads to receptive field formation of each excitatory neuron in the recurrent circuit and the formation of strong bidirectional connectivity between neurons with similar receptive fields (Fig 7F). This is in agreement with various experimental data indicating that similarly responsive neurons are more strongly connected [70–72, 112]. The formation of strongly recurrently connected neurons, often referred to as assemblies, via synaptic plasticity has been shown in previous computational studies [20–23, 113]. In contrast to our framework, these studies rely on a fast normalization mechanism in addition to excitatory and inhibitory plasticity to reliably learn assemblies.

We found that gating of receptive field formation via disinhibition depends on the specificity of the targeted inhibitory population to the inputs. While disinhibiting the unspecific population does not form receptive fields, disinhibiting all specific inhibitory populations induces competition between different inputs and forms receptive fields. If inhibitory plasticity counteracts excitatory plasticity in an input-specific way, no competition between input pathways can emerge because small biases in the E-to-E weights in one input are immediately balanced by I-to-E weights in the same input. Therefore, disrupting the specific inhibitory populations allows the strengthening of a subset of inputs. This result is similar to the findings of [56], where receptive field formation was shown to depend on the specificity of the inhibitory neurons.

The inhibitory populations in our model can be linked to the two main inhibitory neuron types in the cortex, somatostatin-expressing (SOM) and parvalbumin-expressing (PV) inhibitory interneurons. Specificity of the inhibitory neuron type to excitatory inputs can be interpreted as tuning of the inhibitory neurons to input features. In the visual [114, 115] and the auditory cortex [116], tuning of SOM interneurons is sharper than PV interneurons, although conflicting evidence exists [117]. Therefore, in our model the specific inhibitory neuron type could represent SOM interneurons while the unspecific inhibitory population could represent PV interneurons. Supporting this interpretation of SOM interneurons being the specific inhibitory population, experimental studies find that a suppression of SOM neurons gates excitatory plasticity [106, 111, 118]. In contrast to this interpretation, the specific inhibitory neurons in our model might be interpreted as PV neurons. This is supported by experimental evidence which shows that PV neurons strongly inhibit pyramidal neurons which have similar selectivity [119].

Predictions

Using rate-based units in our model enabled us to treat it analytically and offered an in-depth mechanistic understanding of the involved processes leading to experimentally testable predictions and making our model assumptions falsifiable. A main feature of our model is that inhibitory plasticity depends nonlinearly on the rate of the postsynaptic excitatory neuron. This can be tested experimentally by inducing inhibitory plasticity while varying the rate of an excitatory neuron and keeping the inhibitory input to this neuron constant. A second feature of our model is that excitatory and inhibitory plasticity have an identical postsynaptic LTD/LTP threshold. This could be tested by inducing plasticity at excitatory and inhibitory pathways onto the same excitatory neuron, and measuring the LTD/LTP thresholds as a function of the rate of that neuron.

Based on the perturbation experiment (Fig 4), we can formulate multiple predictions. First, we hypothesize that the mechanism behind the metaplastic mechanism is a change in the level of inhibition (see Figs 1C and 4E). Therefore, blocking inhibitory plasticity experimentally should also disrupt the metaplastic mechanism. Second, we predict that the shape of inhibitory plasticity as a function of the inhibitory rate is reversed compared to excitatory plasticity, and perturbations of the excitatory input lead to specific metaplastic changes of inhibitory plasticity. Decreasing the excitatory input should lower the inhibitory LTD/LTP threshold as a function of the presynaptic inhibitory rate and decrease the inhibitory LTP magnitude (Fig 4F). Third, since the line of stable fixed point depends on several model parameters (Fig 2C and Eq 7), especially on the excitatory input rate (Fig 4D), we hypothesize that different E/I ratios can be achieved following input perturbations. Decreasing the excitatory input rate should lead to higher E/I ratios, while increasing it to lower E/I ratios.

The proposed rule suggests a new functional role of inhibitory plasticity, namely controlling E-to-E weight dynamics. Therefore, we extend previously studied roles of inhibitory plasticity, which include the stabilization of excitatory rates [44, 98], decorrelation of neuronal responses [120], preventing winner-take-all mechanisms in networks with multiple stable states [20] or generating differences among novel versus familiar stimuli [23]. Recent computational studies also include novel ways of inhibitory influence, like presynaptic inhibition via GABA spillover [121], an input-dependent inhibitory plasticity mechanism [122] and co-dependency of excitatory and inhibitory plasticity rules [123]. Our model includes a single type of inhibitory plasticity. Yet, recent studies have found that cortical circuits have abundance of different inhibitory interneuron types and that inhibitory plasticity depends on the inhibitory neuron type [75–78]. Our result on inhibitory population-dependent effects in gating receptive field formation suggests that subtype-specific plasticity rules might have non-trivial influences on the network, as some recent models have proposed [78, 124]. Furthermore, other homeostatic mechanisms will influence the stability of weight dynamics, E/I ratios and the effect different perturbations have on the network dynamics.

Conclusion

Taken together, our study proposed a novel form of nonlinear inhibitory plasticity which can achieve stable firing rates and synaptic weights without the need for any additional homeostatic mechanisms. Moreover, our proposed plasticity is fast, and hence could provide a solution to the temporal paradox problem because it can counteract fast Hebbian excitatory plasticity. Functionally, our proposed inhibitory plasticity can establish and maintain a fixed E/I ratio for constant input rates at which the postsynaptic firing rate is exactly at the LTD/LTP threshold. For such postsynaptic firing rates, no synaptic plasticity is induced, i.e. plasticity is "off". Perturbing the postsynaptic firing rate, e.g. via disinhibition, can act as a gate,

turning plasticity "on". This enables the competition among input streams leading to receptive field formation in feedforward and recurrent circuits. Therefore, our nonlinear inhibitory plasticity mechanism provides a solution to the stability-flexibility challenge.

Methods

Rate-based model

We studied rate-based neurons to allow us to analytically investigate the dynamics of firing rates and synaptic weights in the model. In the feedforward motif (Fig 1A), we considered a network consisting of one excitatory postsynaptic neuron with a linear threshold transfer function and firing rate v^E, see Eq 1. The inhibitory neurons also follow a similar dynamics, see Eq 2. All parameters are given in Table 1. In the mean-field sense, the number of neurons can be traded-off with the rates or the synaptic weights, hence we assume $N^E = N^I = 1$ (Table 1).

Rate-based plasticity

For the plasticity of E-to-E synaptic weights w^{EE}, we used a learning rule that depends nonlinearly on the postsynaptic rate v^E (Fig 1B) [53–55]:

$$\tau_w^E \dot{w}^{EE} = \rho^E v^E (v^E - c_{post}^E). \tag{10}$$

Here, τ_w^E is the timescale of excitatory plasticity, which can be also thought of as the inverse of the learning rate, with correcting units Hz^2. This timescale is much longer than the timescale of the neuronal dynamics. The plasticity changes sign at the 'postsynaptic LTD/LTP threshold', c_{post}^E. During experimental induction of plasticity, low frequency stimulation (1,3 or 5 Hz) induces LTD, while high frequency stimulation (10–20 Hz) induces LTP [53]. Therefore, a natural value of the LTD/LTP threshold is between 5 and 10 Hz. We chose 1 Hz as the LTD/LTP threshold (Table 1), nonetheless, our findings will still hold with higher LTD/LTP thresholds.

For the plasticity of I-to-E synaptic weights w^{EI}, we used two learning rules. First, we used an inhibitory plasticity rule common in computational models [44, 56], which depends linearly on the postsynaptic rate v^E (Fig 1D, $\dot{w}^{EI}$):

$$\tau_w^I \dot{w}^{EI} = v^I (v^E - c_{post}^I). \tag{11}$$

Table 1. Parameter values for figures, ⋆ denotes that values are provided in the figure captions.

Sym.	Description	Fig 1	Fig 2	Fig 3	Fig 4	Fig 5	Fig 6	Fig 7B and 7C	S2 Fig	S3 Fig
w_0^{EE}	Initial E-to-E weight	⋆			1.5	⋆	1	0.7	1.5	⋆
w_0^{EI}	Initial I-to-E weight	⋆			0.5	⋆	1	0.5		⋆
w^{IE}	E-to-I weight	0.5							⋆	0.5
ρ^E	Presynaptic E rate (Hz)	2			⋆	2		⋆	2	⋆
ρ^I	Ext. E rate onto I neurons (Hz)	0.5						⋆	0.5	
N^E	Number of presyn. E neurons	1								⋆
N^I	Number of I neurons	1								⋆
$\tau_{FR}^{E/I}$	Time const. E/I rate dyn. (s)	0.01								
τ_w^E	Timescale E plasticity (Hz^2)	1							0.5	1
τ_w^I	Timescale I plasticity (Hz or Hz^2)	0.2							1	0.2
c_{post}^E	E postsyn. LTD/LTP thresh. (Hz)	1		⋆	1					⋆
c_{post}^I	I postsyn. LTD/LTP thresh. (Hz)	1		⋆	1					⋆

https://doi.org/10.1371/journal.pcbi.1010682.t001

Here, τ_w^I denotes the timescale of inhibitory plasticity (or the inverse of the learning rate) with correcting units Hz, which again is much longer than the timescale of the neuronal dynamics. As for excitatory plasticity, inhibitory plasticity changes from LTD to LTP at the 'inhibitory postsynaptic LTD/LTP threshold', c_{post}^I, which sets the 'target rate' of the postsynaptic neuron [44]. In our paper, we proposed a novel inhibitory plasticity rule, which also depends nonlinearly on postsynaptic excitatory activity just like excitatory plasticity (Fig 2A):

$$\tau_w^I \dot{w}^{EI} = v^I v^E (v^E - c_{post}^I). \tag{12}$$

For both inhibitory plasticity rules, we assumed that the excitatory and inhibitory thresholds are matched ($c_{post}^E = c_{post}^I$) to prevent excitatory and inhibitory plasticity pushing the postsynaptic excitatory neuron towards two different firing rates. The exception for this was the dynamic mechanism for threshold matching in Fig 3 and S3 Fig.

LTD/LTP plasticity thresholds. As can be see in the equations for excitatory and inhibitory plasticity, the postsynaptic LTD/LTP thresholds, which determine the sign of plasticity as a function of postsynaptic excitatory activity, are fixed. However, in the main text we also introduce the concept of a presynaptic LTD/LTP threshold, defined as the presynaptic excitatory rate at which no synaptic plasticity is induced. We consider v^E at steady state ($v^E = [N^E \rho^E w^{EE} - N^I v^I w^{EI}]_+$) and assume that the dynamics of the rates are in the region where the transfer function is linear. Therefore, we can drop the linear rectifier and solve for ρ^E at which Eq 3 is zero. We derive the presynaptic LTD/LTP threshold as:

$$c_{pre}^E = \frac{c_{post} + N^I v^I w^{EI}}{N^E w^{EE}}. \tag{13}$$

Stability analysis. To investigate the stability of the weights, we first calculated the nullclines, where we assumed that the postsynaptic excitatory rate is at steady state $v^E = [N^E \rho^E w^{EE} - N^I v^I w^{EI}]_+$. By setting Eqs 10 and 12 to zero and dropping the linear rectifier, i.e. $v^E = N^E \rho^E w^{EE} - N^I v^I w^{EI}$, we can write:

$$\begin{aligned} w^{EI} &= \frac{N^E \rho^E w^{EE} - c_{post}^E}{N^I v^I}, \\ w^{EI} &= \frac{N^E \rho^E w^{EE} - c_{post}^I}{N^I v^I}. \end{aligned} \tag{14}$$

We see that the two equations are identical if $c_{post}^E = c_{post}^I$. Therefore, only for identical LTD/LTP thresholds ($c_{post}^E = c_{post}^I$) a line of fixed points emerges. The fixed points are $[w_*^{EE}, w_*^{EI}] = [x, (N^E \rho^E x - c_{post})/(N^I v^I)]$, where we require that $x \geq c_{post}/(N^E \rho^E)$ to avoid negative weights. To calculate the stability of the line of fixed points, we calculate the eigenvalues. We can rewrite Eqs 10 and 12, as:

$$\begin{aligned} \dot{w}^{EE} &= \frac{\rho^E}{\tau_w^E}\left((N^E \rho^E w^{EE})^2 + (N^I v^I w^{EI})^2 - 2N^E N^I \rho^E v^I w^{EE} w^{EI} - N^E \rho^E w^{EE} c_{post} + N^I v^I w^{EI} c_{post}\right) = f \\ \dot{w}^{EI} &= \frac{v^I}{\tau_w^I}\left((N^E \rho^E w^{EE})^2 + (N^I v^I w^{EI})^2 - 2N^E N^I \rho^E v^I w^{EE} w^{EI} - N^E \rho^E w^{EE} c_{post} + N^I v^I w^{EI} c_{post}\right) = g \end{aligned} \tag{15}$$

where we drop the linear rectifier by assuming that the dynamics of the rates are in the region where the transfer function is linear. We find that the entries of the Jacobian matrix

at the fixed points are:

$$J_* = \begin{pmatrix} \frac{\partial f}{\partial w^{EE}} & \frac{\partial f}{\partial w^{EI}} \\ \frac{\partial g}{\partial w^{EE}} & \frac{\partial g}{\partial w^{EI}} \end{pmatrix} = \begin{pmatrix} \frac{N^E(\rho^E)^2 c_{post}}{\tau_w^E} & -\frac{N^I \rho^E v^I c_{post}}{\tau_w^E} \\ \frac{N^E \rho^E v^I c_{post}}{\tau_w^I} & -\frac{N^I (v^I)^2 c_{post}}{\tau_w^I} \end{pmatrix}. \tag{16}$$

The trace of the Jacobian is $Tr(J_*) = \frac{N^E(\rho^E)^2 c_{post}}{\tau_w^E} - \frac{N^I(v^I)^2 c_{post}}{\tau_w^I}$ and the determinant is zero $Det(J_*) = 0$, therefore we find that the eigenvalues are:

$$\lambda_{1,2} = \frac{1}{2}\left(Tr(J_*) \pm \sqrt{Tr(J_*)^2 - 4Det(J_*)}\right) = \begin{cases} Tr(J_*), \\ 0. \end{cases} \tag{17}$$

This means that if $Tr(J_*) < 0$, the system is stable. Reordering this condition leads to the stability condition derived in the main text as Eq 6. By reordering the terms in the nullclines given in Eq 14, we derive the line attractor equation as given in the main text in Eq 7.

The nonlinear excitatory and inhibitory plasticity rules have a fixed point when the postsynaptic excitatory firing rate is $v^E = 0$. Therefore, in the phase plane of w^{EE} and w^{EI} weights there is a region where the total inhibitory input is larger than the total excitatory input, $N^E \rho^E w^{EE} < N^I v^I w^{EI}$, resulting in no postsynaptic firing (Fig 2B, above gray line). The line equation separating the space with and without weight dynamics is:

$$w^{EI} = \frac{N^E \rho^E w^{EE}}{N^I v^I}. \tag{18}$$

In the case of the linear inhibitory plasticity rule, stability depends on the initial weights. The line which separates stable from unstable initial weights can be calculated by taking the ratio of Eqs 10 and 11 and equating that to the slope of the line attractor (Eq 7):

$$\frac{\dot{w}^{EI}}{\dot{w}^{EE}} = \frac{\tau_w^E v^I}{\tau_w^I \rho^E (N^E \rho^E w^{EE} - N^I v^I w^{EI})} = \frac{N^E \rho^E}{N^I v^I} \tag{19}$$

which leads to:

$$w^{EI} = \frac{N^E \rho^E}{N^I v^I} w^{EE} - \frac{v^I \tau_w^E}{N^E (\rho^E)^2 \tau_w^I}, \tag{20}$$

which is the equation of the dashed line in Fig 1E. The slope of the line attractor is the same for linear and nonlinear inhibitory plasticity.

In Eqs 13–20, the inhibitory firing rate can be replaced by its steady state value $v^I = N^E \rho^E w^{IE} + \rho^I$. For the stability condition (Eq 6) this leads to:

$$\frac{N^I (N^E \rho^E w^{IE} + \rho^I)^2}{\tau_w^I} > \frac{N^E (\rho^E)^2}{\tau_w^E}. \tag{21}$$

and for the line attractor (Eq 7) to:

$$w^{EI} = \frac{N^E \rho^E}{N^I (N^E \rho^E w^{IE} + \rho^I)} w^{EE} - \frac{c_{post}}{N^I (N^E \rho^E w^{IE} + \rho^I)}. \tag{22}$$

The perturbations of the presynaptic firing rate ρ^E_{disr} in Fig 4 are defined as instantaneous and permanent increases or decreases from the initial presynaptic firing rate ρ^E_{base}.

Dynamic threshold matching

The equations for the dynamics of the postsynaptic LTD/LTP thresholds in Fig 3D–3F and S3 Fig are:

$$\begin{aligned} \tau_{c^E_{post}} \dot{c}^E_{post} &= \dot{w}^{EE} \\ \tau_{c^I_{post}} \dot{c}^I_{post} &= -\dot{w}^{EI} \end{aligned} \tag{23}$$

and therefore c^E_{post} increases (decreases) if the postsynaptic neuron fires at $v^E > c^E_{post}$ ($v^E < c^E_{post}$) and c^I_{post} decreases (increases) if the postsynaptic neuron fires at $v^E > c^I_{post}$ ($v^E < c^I_{post}$). The amount of increase or decrease of the postsynaptic thresholds is scaled by the amount of plasticity induction, and we used a timescale of $\tau_c^{E/I} = 2\ ms$, which is faster than the timescale of excitatory and inhibitory plasticity (Table 1). We point out that modifications in the LTD/LTP thresholds lead to changes in the induction of plasticity as well as the postsynaptic firing rate.

For two different initializations of the postsynaptic thresholds, $c^E_{post} < c^I_{post}$ and $c^E_{post} > c^I_{post}$, the synaptic weights, postsynaptic firing rate and postsynaptic threshold dynamics can be stabilized (Fig 3D–3F and S3A–S3C Fig). The same also holds when applying input perturbations (S3D–S3F Fig). For multiple input streams (S3G–S3I Fig), the dynamic postsynaptic LTD/LTP thresholds change based on the total excitatory (or inhibitory) weight change, leading to a non-local sliding mechanism which is independent of the input stream. A condition for the stabilization is that the weights do not reach their lower bounds at zero, because zero weights prevent plasticity and promote the continuous increase of LTD/LTP thresholds preventing firing rates from stabilizing.

E/I ratio

We can calculate the E/I weight ratio $R^{E/I}$ in Eq 8 by rewriting Eq 14 and dividing one of the nullclines by w^{EI}. For large weights, or in mathematical terms for $w^{EI} \to \infty$, the E/I ratio becomes $\lim_{w^{EI}\to\infty} R^{E/I} = R^{E/I}_\infty = \frac{N^I v^I}{N^E \rho^E}$. This derivation is only valid for $N^I(N^E\rho^E w^{IE} + \rho^I)w^{EI} \gg c_{post}$. Therefore, the parameters of the input firing rates ρ^E and ρ^I, the synaptic weights w^{EI} and w^{IE}, as well as number of excitatory and inhibitory neurons N^E and N^I need to be chosen appropriately. This inequality is satisfied for the parameters in Fig 5 when the steady state synaptic weights w^{EI} are sufficiently large (Table 1).

The existence of a fixed E/I ratio for constant input rates can be directly related to the line attractor. The line attractor (Eq 7) expresses the I-to-E weight w^{EI} as a multiple of the E-to-E weight w^{EE} minus the offset term $c_{post}/(N^I v^I)$. Therefore, the ratio of excitatory to inhibitory weight strengths, $R^{E/I}$ (Eq 8), can be expressed as the sum of two terms: one constant term equal to the slope of the line attractor, which is independent of the E-to-E and I-to-E weights, w^{EE} and w^{EI}, and a second term, called an offset, which depends on w^{EI}. When this weight is sufficiently large, the offset term can be ignored, leading to an E/I ratio, $R^{E/I}_\infty$, independent from the E-to-E and I-to-E weights.

In the feedforward circuit (Fig 1A), we can write:

$$R^{E/I}_\infty = \frac{N^I v^I}{N^E \rho^E} = \frac{N^I(\rho^I + w^{IE}\rho^E)}{N^E\rho^E} = \frac{N^I}{N^E}\left(\frac{\rho^I}{\rho^E} + w^{IE}\right). \tag{24}$$

Assuming that $N^E = N^I$, for larger excitatory input rate ρ^E the E/I ratio reaches $R^{E/I}_\infty \approx w^{IE}$ (see

Fig 5A, where $w^{IE} = 0.5$). Therefore, the E/I ratio has a lower bound which depends on the strength of the connection from the excitatory to inhibitory population.

In Fig 5, we link our model to the experimental findings on how the interaction of excitatory and inhibitory plasticity can lead to fixed E/I ratios [37]. In [37], the authors induce plasticity with a spike-pairing protocol, in which pre-post spikes elicit excitatory LTP, while post-pre spikes elicit LTD. Inhibitory LTP was induced for short time differences between the pre- and postsynaptic spikes (independent of the order of the spikes) and inhibitory LTD for longer time differences of the spike pairs. Since in the experiments the presynaptic stimulation was done with a stimulation electrode, the excitatory and inhibitory inputs did not have to be functionally related. In the model, we randomly drew initial E-to-E and I-to-E weights and induced plasticity for a limited amount of time (100 ms) based on the rate-based plasticity rules (Eqs 10 and 12). We choose 100 ms so not all synaptic weights have reached the line attractor yet and so we can compare the E/I ratios reached in our model to those measured experimentally [37] which would most likely also not be in steady state.

The E/I balance can also be defined by the total excitatory input divided by the total inhibitory input onto the postsynaptic neuron:

$$R_{tot}^{E/I} = (N^E w^{EE} \rho^E)/(N^I w^{EI} v^I). \tag{25}$$

This leads to:

$$R_{tot}^{E/I} = (N^I (N^E \rho^E w^{IE} + \rho^I) w^{EI} + c_{post})/(N^I (N^E \rho^E w^{IE} + \rho^I) w^{EI}). \tag{26}$$

However, since we calculate the E/I balance at steady state, the total E/I balance is equal to the weight E/I balance multiplied by a constant, i.e.:

$$\tilde{R}^{E/I} = R^{E/I} N^E \rho^E/(N^I v^I). \tag{27}$$

Therefore, the results in Fig 5 also hold with this alternative E/I ratio definition.

Noise and sinusoidal input

In Fig 6, we add a varying input either by modifying the presynaptic input rate ρ^E or adding an additional term to the postsynaptic neuron (adding ρ_{add} to Eq 1):

$$\tau_{FR}^E \dot{v}^E = -v^E + [N^E \rho^E w^{EE} - N^I v^I w^{EI} + \rho_{add}]_+. \tag{28}$$

In the case of postsynaptic noise (Fig 6A), ρ_{add} is a normally distributed random variable with mean zero and standard deviation 0.01. In the case of additional sinusoidal input to the postsynaptic neuron (Fig 6B), $\rho_{add}(t) = 0.25 * \sin(0.01t)$. Recalculating the slope of the line attractor (Eq 7) based on Eq 28 leads to:

$$w^{EI} = \frac{N^E \rho^E}{N^I v^I} w^{EE} - \frac{c_{post} - \rho_{add}}{N^I v^I}, \tag{29}$$

meaning that ρ_{add} only affects the intersection, but not the slope of the line attractor. We note that the line attractor is calculated at steady state firing rates, meaning that the line attractor will actually never be reached by a varying input.

In the case of presynaptic noise (Fig 6D), we add a normally distributed random variable with mean zero and standard deviation 0.3 to the presynaptic firing rate ρ^E. For the sinusoidal input (Fig 6E), we chose $\rho^E(t) = 2 + 0.5 * \sin(0.01t)$.

Gating of receptive field formation and recurrent clustering

Here, we explore a feedforward network with multiple inputs and two inhibitory neuron populations (Fig 7C). To form receptive fields, we provide a random patterned input to the network. An input pattern is defined by a high firing rate of 4 Hz at a subset of four excitatory input neurons for a time of 100 ms. In Eqs 1 and 2, this is reflected by a subset of the N^E inputs having $\rho^E_m = 4$ Hz, where m corresponds to the presynaptic neurons being part of the respective input pattern. After a time of 100 ms, a new subset of four excitatory neurons fire at high firing rates. We then disinhibit the postsynaptic neurons by inhibiting either the total unspecific or specific inhibitory populations for 60 s by inducing an inhibitory input of 2 onto the respective inhibitory neuron population (we set $\rho^I_{spec} = -2$ or $\rho^I_{unsp} = -2$). Disinhibition needs to be applied for a sufficiently long time to ensure that inhibitory plasticity can induce competition and form receptive fields. We model the release of disinhibition for the specific inhibitory population as slow and gradual over a time course of 100 s to avoid complete silencing of the postsynaptic excitatory neurons. We also note that here we used instantaneous integrators, i.e. $\tau^E_{FR} = \tau^I_{FR} = dt$ (Table 2), because we only wanted to focus on the interaction of excitatory and inhibitory plasticity in the model, though see [125].

For the recurrent circuit, we connected recurrently 30 postsynaptic neurons with feedforward circuits with specific and unspecific inhibition as described above (see also Fig 7D and 7E). In addition to feedforward excitatory and inhibitory weights, also recurrent excitatory weights were plastic based on the plasticity mechanism of Eq 10. We allowed the input patterns to each of the recurrent excitatory neuron to be correlated. Initial recurrent excitatory weights were randomly drawn from the interval [0,0.18]. We calculated the mean weight per input pattern and chose the maximum of those to be the input to which the neurons formed a receptive field. The clustering graph in Fig 7F (left) was done with the digraph function in Matlab where the distance between neurons is only used to visualize clusters of neurons with similar tuning.

Table 2. Parameter values for Fig 7E and 7F.

Symbol	Description	Fig 7E	Fig 7F
w^{EE}_0	Initial E-to-E weight	0.03	[0,0.18]
$w^{EI}_{spec,0}$	Initial specific I-to-E weight	0.01	
$w^{EI}_{unsp,0}$	Initial unspecific I-to-E weight	0.01	
w^{IE}_{spec}	Specific E-to-I weight (fixed)	0.2	0.002
w^{IE}_{unsp}	Unspecific E-to-I weight (fixed)	0.02	0.001
ρ^E	Presynaptic E rate (Hz)	1	
ρ^I_{spec}	External E rate onto specific I neurons (Hz)	0	
ρ^I_{unsp}	External E rate onto unspecific I neurons	0	
N^E	Number of presyn. E neurons (Hz)	40	
N^I_{spec}	Number of specific I neurons	20	
N^I_{unsp}	Number of unspecific I neurons	20	
τ^E_{FR}	Timescale for E neuron model (s)	0.0001	
τ^I_{FR}	Timescale for I neuron model (s)	0.0001	
τ^E_w	Timescale for E plasticity (Hz^2)	1	
τ^I_w	Timescale for I plasticity (Hz^2)	0.2	
c^E_{post}	E postsyn. LTD/LTP threshold (Hz)	1	
c^I_{post}	I postsyn. LTD/LTP threshold (Hz)	1	

https://doi.org/10.1371/journal.pcbi.1010682.t002

The simulations were performed using Matlab programming language. Euler integration was implemented using a time step of 0.1. Code implementing our model is available here: https://github.com/comp-neural-circuits/Nonlinear-inhibitory-plasticity.

Supporting information

S1 Fig. Plasticity of excitatory-to-excitatory synapses as a function of presynaptic and postsynaptic firing rates. Excitatory plasticity $\dot{w}^{EE}$ (Eq 3) is normalized to the maximum value of long-term potentiation (1) and the maximum value of long-term depression (−1), respectively.
(EPS)

S2 Fig. Feedback inhibitory motif leads to additional stability. A. Schematic of the feedback inhibitory motif. The inhibitory population receives input from the presynaptic excitatory population with weight strength w_{FF}^{IE} and the excitatory postsynaptic neuron with weight strength w_{FB}^{IE}. **B**. Plasticity of E-to-E ($\dot{w}^{EE}$, blue) and I-to-E ($\dot{w}^{EI}$, red) weights as a function of the postsynaptic rate v^E. The excitatory and inhibitory LTD/LTP thresholds are identical ($c_{post}^E = c_{post}^I$). **C**. E-to-E (w^{EE}, blue) and I-to-E (w^{EI}, red) and rate dynamics of the postsynaptic (gray line) and the inhibitory population (gray dashed line) as a function of time. **D**. Stability of weight dynamics as a function of the excitatory-to-inhibitory weights w_{FB}^{IE} and w_{FF}^{IE}. Star indicates the values shown in panel C.
(EPS)

S3 Fig. Dynamic matching of the excitatory and inhibitory postsynaptic LTD/LTP thresholds and networks response to input perturbations. A. Postsynaptic LTD/LTP thresholds c_{post}^E and c_{post}^I shift dynamically depending on the recent postsynaptic rate v^E. For lower postsynaptic rate than the excitatory postsynaptic LTD/LTP threshold ($v^E < c_{post}^E$), c_{post}^E decreases, and for $v^E > c_{post}^E$, c_{post}^E increases. For higher postsynaptic rate than the inhibitory postsynaptic LTD/LTP threshold ($v^E > c_{post}^I$), c_{post}^I decreases, and for $v^E < c_{post}^I$, c_{post}^I increases (see Methods). **B**. Evolution of excitatory (c_{post}^E, blue) or inhibitory (c_{post}^I, red) postsynaptic LTD/LTP thresholds for initial conditions $c_{post,0}^E = 1.3$, $c_{post,0}^I = 0.7$. **C**. Excitatory (w^{EE}, blue) and inhibitory (w^{EI}, red) weight dynamics and postsynaptic rate dynamics (v^E, gray) for the initial condition $c_{post,0}^E = 1.3$, $c_{post,0}^I = 0.7$. **D**. Effect of increasing (solid lines, $\rho_{disr}^E = 2.5$) or decreasing (dashed lines, $\rho_{disr}^E = 1.5$) excitatory input rates from a baseline of $\rho_{base}^E = 2$ on excitatory (blue) and inhibitory (red) firing rates. **E**. Same as D but for the c_{post}^E and c_{post}^I weights. **F**. Same as D but for the w^{EE} and w^{EI} weights. **G**. Plasticity curve of E-to-E weights for input 1 or 2 ($\dot{w}_{1,2}^{EE}$) as a function of the presynaptic excitatory rate ρ^E for different input-specific perturbations $\rho_{\text{disr},1,2}^E$. **H**. Evolution of excitatory (c_{post}^E, blue) or inhibitory (c_{post}^I, red) postsynaptic LTD/LTP thresholds for the case in G. **I**. Excitatory (w^{EE}, blue) and inhibitory (w^{EI}, red) weight dynamics for the case in G. Compare A-C to Fig 3, D-F to Fig 4 and G-I to Fig 7B.
(EPS)

S4 Fig. Performance of the nonlinear inhibitory plasticity rule under varying postsynaptic firing rate with dynamic excitatory and inhibitory LTD/LTP threshold matching. A. Adding noise to the postsynaptic firing rate. Top: E-to-E (w^{EE}, blue) and I-to-E (w^{EI}, red) as a function of time. Middle: Excitatory (c_{post}^E, blue) and inhibitory (c_{post}^I, red) postsynaptic LTD/LTP threshold as a function of time. Bottom: Postsynaptic rate dynamics (v^E, gray) as a function of time. **B**. Same as A but after adding a sinusiodal input to the postsynaptic firing rate.
(EPS)

Acknowledgments

We thank all members of the 'Computation in Neural Circuits' group, and specifically Yue Kris Wu, for useful discussions and comments on the manuscript.

Author Contributions

Conceptualization: Christoph Miehl, Julijana Gjorgjieva.

Formal analysis: Christoph Miehl.

Funding acquisition: Julijana Gjorgjieva.

Investigation: Christoph Miehl.

Project administration: Julijana Gjorgjieva.

Software: Christoph Miehl.

Supervision: Julijana Gjorgjieva.

Visualization: Christoph Miehl.

Writing – original draft: Christoph Miehl, Julijana Gjorgjieva.

Writing – review & editing: Christoph Miehl, Julijana Gjorgjieva.

References

1. Hebb DO. The organization of behavior; a neuropsychological theory. Wiley; 1949.
2. Abbott LF, Nelson SB. Synaptic plasticity: taming the beast. Nature Neuroscience. 2000; 3:1178–1183. https://doi.org/10.1038/81453 PMID: 11127835
3. Miller KD, MacKay DJC. The Role of Constraints in Hebbian Learning. Neural Computation. 1994; 6:100–126. https://doi.org/10.1162/neco.1994.6.1.100
4. Turrigiano GG, Nelson SB. Homeostatic plasticity in the developing nervous system. Nature Reviews Neuroscience. 2004; 5:97–107. https://doi.org/10.1038/nrn1327 PMID: 14735113
5. Turrigiano GG, Leslie KR, Desai NS, Rutherford LC, Nelson SB. Activity-dependent scaling of quantal amplitude in neocortical neurons. Nature. 1998; 391(6670):892–896. https://doi.org/10.1038/36103 PMID: 9495341
6. Turrigiano GG. The Self-Tuning Neuron: Synaptic Scaling of Excitatory Synapses. Cell. 2008; 135:422–435. https://doi.org/10.1016/j.cell.2008.10.008 PMID: 18984155
7. Lynch GS, Dunwiddie T, Gribkoff V. Heterosynaptic depression: a postsynaptic correlate of long-term potentiation. Nature. 1977; 266(21):737–739. https://doi.org/10.1038/266737a0 PMID: 195211
8. Chistiakova M, Bannon NM, Chen JY, Bazhenov M, Volgushev M. Homeostatic role of heterosynaptic plasticity: models and experiments. Frontiers in Computational Neuroscience. 2015; 9. https://doi.org/10.3389/fncom.2015.00089 PMID: 26217218
9. Field RE, D'amour JA, Tremblay R, Miehl C, Rudy B, Gjorgjieva J, et al. Heterosynaptic Plasticity Determines the Set Point for Cortical Excitatory-Inhibitory Balance. Neuron. 2020; 106(5):842–854. https://doi.org/10.1016/j.neuron.2020.03.002 PMID: 32213321
10. Desai NS, Rutherford LC, Turrigiano GG. Plasticity in the intrinsic excitability of cortical pyramidal neurons. Nature Neuroscience. 1999; 2(6):515–520. https://doi.org/10.1038/9165 PMID: 10448215
11. Debanne D, Inglebert Y, Russier M. Plasticity of intrinsic neuronal excitability. Current Opinion in Neurobiology. 2019; 54:73–82. https://doi.org/10.1016/j.conb.2018.09.001 PMID: 30243042
12. Oja E. Simplified neuron model as a principal component analyzer. Journal of Mathematical Biology. 1982; 15(3):267–273. https://doi.org/10.1007/BF00275687 PMID: 7153672
13. Bienenstock EL, Cooper LN, Munro PW. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. The Journal of Neuroscience. 1982; 2(1):32–48. https://doi.org/10.1523/JNEUROSCI.02-01-00032.1982 PMID: 7054394
14. Gjorgjieva J, Clopath C, Audet J, Pfister JP. A triplet spike-timing-dependent plasticity model generalizes the Bienenstock-Cooper-Munro rule to higher-order spatiotemporal correlations. Proceedings of

the National Academy of Sciences. 2011; 108(48):19383–19388. https://doi.org/10.1073/pnas.1105933108 PMID: 22080608

15. Yger P, Gilson M. Models of Metaplasticity: A Review of Concepts. Frontiers in Computational Neuroscience. 2015; 9(138). https://doi.org/10.3389/fncom.2015.00138 PMID: 26617512
16. Wu YK, Hengen KB, Turrigiano GG, Gjorgjieva J. Homeostatic mechanisms regulate distinct aspects of cortical circuit dynamics. Proceedings of the National Academy of Sciences. 2020; 117(39):24514–24525. https://doi.org/10.1073/pnas.1918368117 PMID: 32917810
17. Fox K, Stryker M. Integrating Hebbian and homeostatic plasticity: Introduction. Philosophical Transactions of the Royal Society B. 2017; 372:20160413. https://doi.org/10.1098/rstb.2016.0413 PMID: 28093560
18. Turrigiano GG. The dialectic of hebb and homeostasis. Philosophical Transactions of the Royal Society B: Biological Sciences. 2017; 372:20160258. https://doi.org/10.1098/rstb.2016.0258 PMID: 28093556
19. Yee AX, Hsu YT, Chen L. A metaplasticity view of the interaction between homeostatic and hebbian plasticity. Philosophical Transactions of the Royal Society B. 2017; 372:20160155. https://doi.org/10.1098/rstb.2016.0155 PMID: 28093549
20. Litwin-Kumar A, Doiron B. Formation and maintenance of neuronal assemblies through synaptic plasticity. Nature Communications. 2014; 5 (5319). PMID: 25395015
21. Zenke F, Agnes EJ, Gerstner W. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. Nature Communications. 2015; 6 (6922). https://doi.org/10.1038/ncomms7922 PMID: 25897632
22. Montangie L, Miehl C, Gjorgjieva J. Autonomous emergence of connectivity assemblies via spike triplet interactions. PLoS Computational Biology. 2020; 16(5):e1007835. https://doi.org/10.1371/journal.pcbi.1007835 PMID: 32384081
23. Schulz A, Miehl C, Berry MJ II, Gjorgjieva J. The generation of cortical novelty responses through inhibitory plasticity. eLife. 2021; 10:e65309. https://doi.org/10.7554/eLife.65309 PMID: 34647889
24. Zenke F, Hennequin G, Gerstner W. Synaptic Plasticity in Neural Networks Needs Homeostasis with a Fast Rate Detector. PLoS Computational Biology. 2013; 9(11):e1003330. https://doi.org/10.1371/journal.pcbi.1003330 PMID: 24244138
25. Zenke F, Gerstner W, Ganguli S. The temporal paradox of Hebbian learning and homeostatic plasticity. Current Opinion in Neurobiology. 2017; 43:166–176. https://doi.org/10.1016/j.conb.2017.03.015 PMID: 28431369
26. Zenke F, Gerstner W. Hebbian plasticity requires compensatory processes on multiple timescales. Philosophical Transactions of the Royal Society B: Biological Sciences. 2017; 372(1715):20160259. https://doi.org/10.1098/rstb.2016.0259 PMID: 28093557
27. Fusi S. Computational models of long term plasticity and memory. arXiv. 2017; https://doi.org/10.48550/arXiv.1706.04946.
28. Chen JY, Lonjers P, Lee C, Chistiakova M, Volgushev M, Bazhenov M. Heterosynaptic Plasticity Prevents Runaway Synaptic Dynamics. Journal of Neuroscience. 2013; 33(40):15915–15929. https://doi.org/10.1523/JNEUROSCI.5088-12.2013 PMID: 24089497
29. Volgushev M, Chen JY, Ilin V, Goz R, Chistiakova M, Bazhenov M. Partial Breakdown of Input Specificity of STDP at Individual Synapses Promotes New Learning. The Journal of Neuroscience. 2016; 36(34):8842–8855. https://doi.org/10.1523/JNEUROSCI.0552-16.2016 PMID: 27559167
30. Kirchner JH, Gjorgjieva J. Emergence of local and global synaptic organization on cortical dendrites. Nature Communications. 2021; 12:4005. https://doi.org/10.1038/s41467-021-23557-3 PMID: 34183661
31. Chasse R, Malyshev A, Fitch RH, Volgushev M. Altered heterosynaptic plasticity impairs visual discrimination learning in adenosine A1 receptor knock-out mice. Journal of Neuroscience. 2021; 41(21):4631–4640. https://doi.org/10.1523/JNEUROSCI.3073-20.2021 PMID: 33849950
32. White G, Levy WB, Steward O. Spatial overlap between populations of synapses determines the extent of their associative interaction during the induction of long-term potentiation and depression. Journal of Neurophysiology. 1990; 64(4):1186–1198. https://doi.org/10.1152/jn.1990.64.4.1186 PMID: 2258741
33. Royer S, Paré D. Conservation of total synaptic weight through balanced synaptic depression and potentiation. Nature. 2003; 422:518–522. https://doi.org/10.1038/nature01530 PMID: 12673250
34. Caillard O, Ben-Ari Y, Gaiarsa JL. Long-term potentiation of GABAergic synaptic transmission in neonatal rat hippocampus. Journal of Physiology. 1999; 518(1):109–119. https://doi.org/10.1111/j.1469-7793.1999.0109r.x PMID: 10373693

35. Shew T, Yip S, Sastry BR. Mechanisms involved in tetanusinduced potentiation of fast IPSCs in rat hippocampal CA1 neurons. Journal of Neurophysiology. 2000; 83(6):3388–3401. https://doi.org/10.1152/jn.2000.83.6.3388 PMID: 10848557

36. Mellor J. Synaptic Plasticity at Hippocampal Synapses: Experimental Background. In: Hippocampal Microcircuits. Springer Seriesin Computational Neuroscience; 2018. p. 201–226.

37. D'amour JA, Froemke RC. Inhibitory and Excitatory Spike-Timing-Dependent Plasticity in the Auditory Cortex. Neuron. 2015; 86:514–528. https://doi.org/10.1016/j.neuron.2015.03.014 PMID: 25843405

38. Hennequin G, Agnes EJ, Vogels TP. Inhibitory Plasticity: Balance, Control, and Codependence. Annual Review of Neuroscience. 2017; 40:557–579. https://doi.org/10.1146/annurev-neuro-072116-031005 PMID: 28598717

39. Steele PM, Mauk MD. Inhibitory Control of LTP and LTD: Stability of Synapse Strength. Journal of Neurophysiology. 1999; 81:1559–1566. https://doi.org/10.1152/jn.1999.81.4.1559 PMID: 10200191

40. Paille V, Fino E, Du K, Morera-Herreras T, Perez S, Kotaleski JH, et al. GABAergic Circuits Control Spike-Timing-Dependent Plasticity. Journal of Neuroscience. 2013; 33(22):9353–9363. https://doi.org/10.1523/JNEUROSCI.5796-12.2013 PMID: 23719804

41. Hiratani N, Fukai T. Detailed Dendritic Excitatory/Inhibitory Balance through Heterosynaptic Spike-Timing-Dependent Plasticity. The Journal of Neuroscience. 2017; 37(50):12106–12122. https://doi.org/10.1523/JNEUROSCI.0027-17.2017 PMID: 29089443

42. Herstel LJ, Wierenga CJ. Network control through coordinated inhibition. Current Opinion in Neurobiology. 2021; 67:1–8. https://doi.org/10.1016/j.conb.2020.08.001 PMID: 32853970

43. Wang L, Maffei A. Inhibitory Plasticity Dictates the Sign of Plasticity at Excitatory Synapses. Journal of Neuroscience. 2014; 34(4):1083–1093. https://doi.org/10.1523/JNEUROSCI.4711-13.2014 PMID: 24453301

44. Vogels TP, Sprekeler H, Zenke F, Clopath C, Gerstner W. Inhibitory Plasticity Balances Excitation and Inhibition in Sensory Pathways and Memory Networks. Science. 2011; 334(6062):1569–1573. https://doi.org/10.1126/science.1211095 PMID: 22075724

45. Luz Y, Shamir M. Balancing Feed-Forward Excitation and Inhibition via Hebbian Inhibitory Synaptic Plasticity. PLoS Computational Biology. 2012; 8(1):e1002334. https://doi.org/10.1371/journal.pcbi.1002334 PMID: 22291583

46. Kleberg FI, Fukai T, Gilson M. Excitatory and inhibitory STDP jointly tune feedforward neural circuits to selectively propagate correlated spiking activity. Frontiers in Computational Neuroscience. 2014; 8(53). https://doi.org/10.3389/fncom.2014.00053 PMID: 24847242

47. Tsodyks M, Skaggs WE, Sejnowski TJ, McNaughton BL. Paradoxical Effects of External Modulation of Inhibitory Interneurons. The Journal of Neuroscience. 1997; 17(11):4382–4388. https://doi.org/10.1523/JNEUROSCI.17-11-04382.1997 PMID: 9151754

48. Sanzeni A, Akitake B, Goldbach HC, Leedy CE, Brunel N, Histed MH. Inhibition stabilization is a widespread property of cortical networks. eLife. 2020; 9:e54875. https://doi.org/10.7554/eLife.54875 PMID: 32598278

49. Ahmadian Y, Miller KD. What is the dynamical regime of cerebral cortex? Neuron. 2021; 109(21):3373–3391. https://doi.org/10.1016/j.neuron.2021.07.031 PMID: 34464597

50. Froemke RC, Merzenich MM, Schreiner CE. A synaptic memory trace for cortical receptive field plasticity. Nature. 2007; 450:425–429. https://doi.org/10.1038/nature06289 PMID: 18004384

51. Letzkus JJ, Wolff SBE, Meyer EMM, Tovote P, Courtin J, Herry C, et al. A disinhibitory microcircuit for associative fear learning in the auditory cortex. Nature. 2011; 480(7377):331–335. https://doi.org/10.1038/nature10674 PMID: 22158104

52. Tremblay R, Lee S, Rudy B. GABAergic Interneurons in the Neocortex: From Cellular Properties to Circuits. Neuron. 2016; 91(2):260–292. https://doi.org/10.1016/j.neuron.2016.06.033 PMID: 27477017

53. Kirkwood A, Rioult MG, Bear MF. Experience-dependent modification of synaptic plasticity in visual cortex. Nature. 1996; 381:526–528. https://doi.org/10.1038/381526a0 PMID: 8632826

54. Philpot BD, Espinosa JS, Bear MF. Evidence for altered NMDA receptor function as a basis for metaplasticity in visual cortex. Journal of Neuroscience. 2003; 23(13):5583–5588. https://doi.org/10.1523/JNEUROSCI.23-13-05583.2003 PMID: 12843259

55. Cooper LN, Bear MF. The BCM theory of synapse modification at 30: interaction of theory with experiment. Nature Reviews Neuroscience. 2012; 13(11):798–810. https://doi.org/10.1038/nrn3353 PMID: 23080416

56. Clopath C, Vogels TP, Froemke RC, Sprekeler H. Receptive field formation by interacting excitatory and inhibitory synaptic plasticity. bioRxiv. 2016; https://doi.org/10.1101/066589.

57. Keck T, Hübener M, Bonhoeffer T. Interactions between synaptic homeostatic mechanisms: an attempt to reconcile BCM theory, synaptic scaling, and changing excitation/inhibition balance. Current Opinion in Neurobiology. 2017; 43:87–93. https://doi.org/10.1016/j.conb.2017.02.003 PMID: 28236778

58. Kuo MC, Dringenberg HC. Short-term (2 to 5 h) dark exposure lowers long-term potentiation (LTP) induction threshold in rat primary visual cortex. Brain Research. 2009; 1276:58–66. https://doi.org/10.1016/j.brainres.2009.04.042 PMID: 19409376

59. Huang YY, Colino A, Selig DK, Malenka RC. The influence of prior synaptic activity on the induction of long-term potentiation. Science. 1992; 255:730–733. https://doi.org/10.1126/science.1346729 PMID: 1346729

60. Abraham WC. Metaplasticity: Tuning synapses and networks for plasticity. Nature Reviews Neuroscience. 2008; 9:387–399. https://doi.org/10.1038/nrn2356 PMID: 18401345

61. Allen CB, Celikel T, Feldman DE. Long-term depression induced by sensory deprivation during cortical map plasticity in vivo. Nature Neuroscience. 2003; 6(3):291–299. https://doi.org/10.1038/nn1012 PMID: 12577061

62. Chen JL, Lin WC, Cha JW, So PT, Kubota Y, Nedivi E. Structural basis for the role of inhibition in facilitating adult brain plasticity. Nature Neuroscience. 2011; 14(5):587–596. https://doi.org/10.1038/nn.2799 PMID: 21478885

63. Keck T, Scheuss V, Jacobsen RI, Wierenga CJ, Eysel UT, Bonhoeffer T, et al. Loss of sensory input causes rapid structural changes of inhibitory neurons in adult mouse visual cortex. Neuron. 2011; 71:869–882. https://doi.org/10.1016/j.neuron.2011.06.034 PMID: 21903080

64. Chen JL, Villa KL, Cha JW, So PTC, Kubota Y, Nedivi E. Clustered Dynamics of Inhibitory Synapses and Dendritic Spines in the Adult Neocortex. Neuron. 2012; 74:361–373. https://doi.org/10.1016/j.neuron.2012.02.030 PMID: 22542188

65. van Versendaal D, Rajendran R, Saiepour HM, Klooster J, Smit-Rigter L, Sommeijer JP, et al. Elimination of Inhibitory Synapses Is a Major Component of Adult Ocular Dominance Plasticity. Neuron. 2012; 74:374–383. https://doi.org/10.1016/j.neuron.2012.03.015 PMID: 22542189

66. Li L, Gainey MA, Goldbeck JE, Feldman DE. Rapid homeostasis by disinhibition during whisker map plasticity. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111(4):1616–1621. https://doi.org/10.1073/pnas.1312455111 PMID: 24474788

67. Thompson A, Gribizis A, Chen C, Crair MC. Activity-dependent development of visual receptive fields. Current Opinion in Neurobiology. 2017; 42:136–143. https://doi.org/10.1016/j.conb.2016.12.007 PMID: 28088066

68. Abraham WC, Bear MF. Metaplasticity: plasticity of synaptic plasticity. Trends Neuroscience. 1996; 19:126–130. https://doi.org/10.1016/S0166-2236(96)80018-X PMID: 8658594

69. Kuhlman SJ, Olivas ND, Tring E, Ikrar T, Xu X, Trachtenberg JT. A disinhibitory microcircuit initiates critical-period plasticity in the visual cortex. Nature. 2013; 501:543–546. https://doi.org/10.1038/nature12485 PMID: 23975100

70. Ko H, Hofer SB, Pichler B, Buchanan KA, Sjöström PJ, Mrsic-Flogel TD. Functional specificity of local synaptic connections in neocortical networks. Nature. 2011; 473:87–91. https://doi.org/10.1038/nature09880 PMID: 21478872

71. Ko H, Cossell L, Baragli C, Antolik J, Clopath C, Hofer SB, et al. The emergence of functional microcircuits in visual cortex. Nature. 2013; 496:96–100. https://doi.org/10.1038/nature12015 PMID: 23552948

72. Miller JEK, Ayzenshtat I, Carrillo-Reid L, Yuste R. Visual stimuli recruit intrinsically generated cortical ensembles. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:E4053–E4061. https://doi.org/10.1073/pnas.1406077111 PMID: 25201983

73. Turrigiano GG. Too Many Cooks? Intrinsic and Synaptic Homeostatic Mechanisms in Cortical Circuit Refinement. Annual Review of Neuroscience. 2011; 34:89–103. https://doi.org/10.1146/annurev-neuro-060909-153238 PMID: 21438687

74. Woodin MA, Ganguly K, Poo MM. Coincident Pre- and Postsynaptic Activity Modifies GABAergic Synapses by Postsynaptic Changes in Cl- Transporter Activity. Neuron. 2003; 39(5):807–820. https://doi.org/10.1016/S0896-6273(03)00507-5 PMID: 12948447

75. Chiu CQ, Martenson JS, Yamazaki M, Natsume R, Sakimura K, Tomita S, et al. Input-Specific NMDAR-Dependent Potentiation of Dendritic GABAergic Inhibition. Neuron. 2018; 97:368–377. https://doi.org/10.1016/j.neuron.2017.12.032 PMID: 29346754

76. Vickers ED, Clark C, Osypenko D, Fratzl A, Kochubey O, Bettler B, et al. Parvalbumin-Interneuron Output Synapses Show Spike-Timing-Dependent Plasticity that Contributes to Auditory Map Remodeling. Neuron. 2018; 99:720–735. https://doi.org/10.1016/j.neuron.2018.07.018 PMID: 30078579

77. Udakis M, Pedrosa V, Chamberlain SEL, Clopath C, Mellor JR. Interneuron-specific plasticity at parvalbumin and somatostatin inhibitory synapses onto CA1 pyramidal neurons shapes hippocampal output. Nature Communications. 2020; 11 (4395). https://doi.org/10.1038/s41467-020-18074-8 PMID: 32879322

78. Lagzi F, Canto-Bustos M, Oswald AMM, Doiron B. Assembly formation is stabilized by Parvalbumin neurons and accelerated by Somatostatin neurons. bioRxiv. 2021; https://doi.org/10.1101/2021.09.06.459211.

79. Haas JS, Nowotny T, Abarbanel HDI. Spike-Timing-Dependent Plasticity of Inhibitory Synapses in the Entorhinal Cortex. Journal of Neurophysiology. 2006; 96:3305–3313. https://doi.org/10.1152/jn.00551.2006 PMID: 16928795

80. Pedrosa V, Clopath C. Voltage-based inhibitory synaptic plasticity: network regulation, diversity, and flexibility. bioRxiv. 2020; https://doi.org/10.1101/2020.12.08.416263.

81. Clopath C, Büsing L, Vasilaki E, Gerstner W. Connectivity reflects coding: a model of voltage-based STDP with homeostasis. Nature Neuroscience. 2010; 13(3):344–352. https://doi.org/10.1038/nn.2479 PMID: 20098420

82. Vogels TP, Froemke RC, Doyon N, Gilson M, Haas JS, Liu R, et al. Inhibitory synaptic plasticity: spike timing-dependence and putative network function. Frontiers in Neural Circuits. 2013; 7. https://doi.org/10.3389/fncir.2013.00119 PMID: 23882186

83. Intrator N, Cooper LN. Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions. Neural Networks. 1992; 5:3–17. https://doi.org/10.1016/S0893-6080(05)80003-6

84. Cooper LN, Intrator N, Blais BS, Shouval HZ. Theory of cortical plasticity. World Scientific Publishing; 2004.

85. Udeigwe LC, Munro PW, Ermentrout GB. Emergent Dynamical Properties of the BCM Learning Rule. The Journal of Mathematical Neuroscience. 2017; 7(1). https://doi.org/10.1186/s13408-017-0044-6 PMID: 28220467

86. Keck T, Keller GB, Jacobsen RI, Eysel UT, Bonhoeffer T, Hübener M. Synaptic scaling and homeostatic plasticity in the mouse visual cortex in vivo. Neuron. 2013; 80:327–334. https://doi.org/10.1016/j.neuron.2013.08.018 PMID: 24139037

87. Hengen KB, Lambo ME, VanHooser SD, Katz DB, Turrigiano GG. Firing rate homeostasis in visual cortex of freely behaving rodents. Neuron. 2013; 80(2):335–342. https://doi.org/10.1016/j.neuron.2013.08.038 PMID: 24139038

88. Ibata K, Sun Q, Turrigiano GG. Report Rapid Synaptic Scaling Induced by Changes in Postsynaptic Firing. Neuron. 2008; 57:819–826. https://doi.org/10.1016/j.neuron.2008.02.031 PMID: 18367083

89. Gainey MA, Feldman DE. Multiple shared mechanisms for homeostatic plasticity in rodent somatosensory and visual cortex. Philosophical Transactions of the Royal Society B: Biological Sciences. 2017; 372 (1715). https://doi.org/10.1098/rstb.2016.0157 PMID: 28093551

90. Canto-Bustos M, Friason FK, Bassi C, Oswald AMM. Disinhibitory circuitry gates associative synaptic plasticity in olfactory cortex. Journal of Neuroscience. 2022; 42(14):2942–2950. https://doi.org/10.1523/JNEUROSCI.1369-21.2021 PMID: 35181596

91. Barnes SJ, Sammons RP, Jacobsen RI, Mackie J, Keller GB, Keck T. Subnetwork-Specific Homeostatic Plasticity in Mouse Visual Cortex In Vivo. Neuron. 2015; 86:1290–1303. https://doi.org/10.1016/j.neuron.2015.05.010 PMID: 26050045

92. Maffei A, Nataraj K, Nelson SB, Turrigiano GG. Potentiation of cortical inhibition by visual deprivation. Nature. 2006; 443(7107):81–84. https://doi.org/10.1038/nature05079 PMID: 16929304

93. Maffei A, Lambo ME, Turrigiano GG. Critical period for inhibitory plasticity in rodent binocular V1. Journal of Neuroscience. 2010; 30(9):3304–3309. https://doi.org/10.1523/JNEUROSCI.5340-09.2010 PMID: 20203190

94. Miska NJ, Richter LMA, Cary BA, Gjorgjieva J, Turrigiano GG. Sensory experience inversely regulates feedforward and feedback excitation-inhibition ratio in rodent visual cortex. eLife. 2018; 7:e38846. https://doi.org/10.7554/eLife.38846 PMID: 30311905

95. Lourenço J, Pacioni S, Rebola N, van Woerden GM, Marinelli S, DiGregorio D, et al. Non-associative Potentiation of Perisomatic Inhibition Alters the Temporal Coding of Neocortical Layer 5 Pyramidal Neurons. PLoS Biology. 2014; 12(7):1–19. https://doi.org/10.1371/journal.pbio.1001903 PMID: 25003184

96. Xue M, Atallah BV, Scanziani M. Equalizing excitation-inhibition ratios across visual cortical neurons. Nature. 2014; 511:596–600. https://doi.org/10.1038/nature13321 PMID: 25043046

97. Abraham WC, Huggett A. Induction and reversal of long-term potentiation by repeated high-frequency stimulation in rat hippocampal slices. Hippocampus. 1997; 7:137–145. https://doi.org/10.1002/(SICI)1098-1063(1997)7:2%3C137::AID-HIPO3%3E3.0.CO;2-K PMID: 9136046

98. Sprekeler H. Functional consequences of inhibitory plasticity: homeostasis, the excitation-inhibition balance and beyond. Current Opinion in Neurobiology. 2017; 43:198–203. https://doi.org/10.1016/j.conb.2017.03.014 PMID: 28500933

99. Maffei A, Turrigiano GG. Multiple modes of network homeostasis in visual cortical layer 2/3. Journal of Neuroscience. 2008; 28(17):4377–4384. https://doi.org/10.1523/JNEUROSCI.5298-07.2008 PMID: 18434516

100. Dorrn AL, Yuan K, Barker AJ, Schreiner CE, Froemke RC. Developmental sensory experience balances cortical excitation and inhibition. Nature. 2010; 465:932–936. https://doi.org/10.1038/nature09119 PMID: 20559387

101. House DRC, Elstrott J, Koh E, Chung J, Feldman DE. Parallel Regulation of Feedforward Inhibition and Excitation during Whisker Map Plasticity. Neuron. 2011; 72:819–831. https://doi.org/10.1016/j.neuron.2011.09.008 PMID: 22153377

102. Adesnik H. Synaptic Mechanisms of Feature Coding in the Visual Cortex of Awake Mice. Neuron. 2017; 95:1147–1159. https://doi.org/10.1016/j.neuron.2017.08.014 PMID: 28858618

103. Aljadeff J, D'amour J, Field RE, Froemke RC, Clopath C. Cortical credit assignment by Hebbian, neuromodulatory and inhibitory plasticity. arXiv. 2019; https://arxiv.org/abs/1911.00307.

104. Isaacson JS, Scanziani M. How Inhibition Shapes Cortical Activity. Neuron. 2011; 72(2):231–243. https://doi.org/10.1016/j.neuron.2011.09.027 PMID: 22017986

105. Dehorter N, Marichal N, Marín O, Berninger B. Tuning neural circuits by turning the interneuron knob. Current Opinion in Neurobiology. 2017; 42:144–151. https://doi.org/10.1016/j.conb.2016.12.009 PMID: 28088067

106. Hattori R, Kuchibhotla KV, Froemke RC, Komiyama T. Functions and dysfunctions of neocortical inhibitory neuron subtypes. Nature Neuroscience. 2017; 20(9):1199–1208. https://doi.org/10.1038/nn.4619 PMID: 28849791

107. Kripkee B, Froemke RC. Organization and Plasticity of Cortical Inhibition. The Oxford Handbook of Developmental Neural Plasticity; 2017.

108. Froemke RC. Plasticity of Cortical Excitatory-Inhibitory Balance. Annual Review of Neuroscience. 2015; 38:195–219. https://doi.org/10.1146/annurev-neuro-071714-034002 PMID: 25897875

109. Letzkus JJ, Wolff SBE, Lüthi A. Disinhibition, a Circuit Mechanism for Associative Learning and Memory. Neuron. 2015; 88(2):264–276. https://doi.org/10.1016/j.neuron.2015.09.024 PMID: 26494276

110. Wang XJ, Yang GR. A disinhibitory circuit motif and flexible information routing in the brain. Current Opinion in Neurobiology. 2018; 49:75–83. https://doi.org/10.1016/j.conb.2018.01.002 PMID: 29414069

111. Williams LE, Holtmaat A. Higher-Order Thalamocortical Inputs Gate Synaptic Long-Term Potentiation via Disinhibition. Neuron. 2019; 101:91–102. https://doi.org/10.1016/j.neuron.2018.10.049 PMID: 30472077

112. Lee WCA, Bonin V, Reed M, Graham BJ, Hood G, Glattfelder K, et al. Anatomy and function of an excitatory network in the visual cortex. Nature. 2016; 532(7599):370–374. https://doi.org/10.1038/nature17192 PMID: 27018655

113. Miehl C, Onasch S, Festa D, Gjorgjieva J. Formation and computational implications of assemblies in neural circuits. The Journal of Physiology. 2022. https://doi.org/10.1113/JP282750 PMID: 36068723

114. Ma WP, Liu BH, Li YT, Huang ZJ, Zhang LI, Tao HW. Visual representations by cortical somatostatin inhibitory neurons—Selective but with weak and delayed responses. Journal of Neuroscience. 2010; 30(43):14371–14379. https://doi.org/10.1523/JNEUROSCI.3248-10.2010 PMID: 20980594

115. Cottam JCH, Smith SL, Häusser M. Target-Specific Effects of Somatostatin-Expressing Interneurons on Neocortical Visual Processing. Journal of Neuroscience. 2013; 33(50):19567–19578. https://doi.org/10.1523/JNEUROSCI.2624-13.2013 PMID: 24336721

116. Li LY, Xiong XR, Ibrahim LA, Yuan W, Tao HW, Zhang LI. Differential Receptive Field Properties of Parvalbumin and Somatostatin Inhibitory Neurons in Mouse Auditory Cortex. Cerebral Cortex. 2015; 25:1782–1791. https://doi.org/10.1093/cercor/bht417 PMID: 24425250

117. Griffen TC, Maffei A. GABAergic synapses: their plasticity and role in sensory cortex. Frontiers in Cellular Neuroscience. 2014; 8(91). https://doi.org/10.3389/fncel.2014.00091 PMID: 24723851

118. Chen SX, Kim AN, Peters AJ, Komiyama T. Subtype-specific plasticity of inhibitory circuits in motor cortex during motor learning. Nature Neuroscience. 2015; 18(8):1109–1115. https://doi.org/10.1038/nn.4049 PMID: 26098758

119. Znamenskiy P, Kim Mh, Muir DR, Iacaruso F, Hofer SB, Mrsic-Flogel TD. Functional selectivity and specific connectivity of inhibitory neurons in primary visual cortex. bioRxiv. 2018;294835.

120. Duarte RCF, Morrison A. Dynamic stability of sequential stimulus representations in adapting neuronal networks. Frontiers in Computational Neuroscience. 2014; 8(124). https://doi.org/10.3389/fncom.2014.00124 PMID: 25374534

121. Naumann LB, Sprekeler H. Presynaptic inhibition rapidly stabilises recurrent excitation in the face of plasticity. PLoS Computational Biology. 2020; 16(8):e1008118. https://doi.org/10.1371/journal.pcbi.1008118 PMID: 32764742

122. Kaleb K, Pedrosa V, Clopath C. Network-centered homeostasis through inhibition maintains hippocampal spatial map and cortical circuit function. Cell Reports. 2021; 36:109577. https://doi.org/10.1016/j.celrep.2021.109577 PMID: 34433026

123. Agnes EJ, Vogels TP. Interacting synapses stabilise both learning and neuronal dynamics in biological networks. bioRxiv. 2021; https://doi.org/10.1101/2021.04.01.437962.

124. Agnes EJ, Luppi AI, Vogels TP. Complementary inhibitory weight profiles emerge from plasticity and allow flexible switching of receptive fields. Journal of Neuroscience. 2020; 40(50):9634–9649. https://doi.org/10.1523/JNEUROSCI.0276-20.2020 PMID: 33168622

125. Gjorgjieva J, Drion G, Marder E. Computational implications of biophysical diversity and multiple timescales in neurons and synapses for circuit performance. Current Opinion in Neurobiology. 2016; 37:44–52. https://doi.org/10.1016/j.conb.2015.12.008 PMID: 26774694

II. The generation of cortical novelty responses through inhibitory plasticity

Schulz*, A., Miehl*, C., Berry II, M.J. & Gjorgjieva, J. The generation of cortical novelty responses through inhibitory plasticity. *eLife* **10**, e65309 (2021). https://doi.org/10.7554/eLife.65309

RESEARCH ARTICLE

The generation of cortical novelty responses through inhibitory plasticity

Auguste Schulz[1,2†], Christoph Miehl[1,3†], Michael J Berry II[4], Julijana Gjorgjieva[1,3*]

[1]Max Planck Institute for Brain Research, Frankfurt, Germany; [2]Technical University of Munich, Department of Electrical and Computer Engineering, Munich, Germany; [3]Technical University of Munich, School of Life Sciences, Freising, Germany; [4]Princeton University, Princeton Neuroscience Institute, Princeton, United States

Abstract Animals depend on fast and reliable detection of novel stimuli in their environment. Neurons in multiple sensory areas respond more strongly to novel in comparison to familiar stimuli. Yet, it remains unclear which circuit, cellular, and synaptic mechanisms underlie those responses. Here, we show that spike-timing-dependent plasticity of inhibitory-to-excitatory synapses generates novelty responses in a recurrent spiking network model. Inhibitory plasticity increases the inhibition onto excitatory neurons tuned to familiar stimuli, while inhibition for novel stimuli remains low, leading to a network novelty response. The generation of novelty responses does not depend on the periodicity but rather on the distribution of presented stimuli. By including tuning of inhibitory neurons, the network further captures stimulus-specific adaptation. Finally, we suggest that disinhibition can control the amplification of novelty responses. Therefore, inhibitory plasticity provides a flexible, biologically plausible mechanism to detect the novelty of bottom-up stimuli, enabling us to make experimentally testable predictions.

***For correspondence:** gjorgjieva@brain.mpg.de

†These authors contributed equally to this work

Competing interests: The authors declare that no competing interests exist.

Funding: See page 23

Received: 30 November 2020
Preprinted: 01 December 2020
Accepted: 22 September 2021
Published: 14 October 2021

Reviewing editor: Maria N Geffen, University of Pennsylvania, United States

Introduction

In an ever-changing environment, animals must rapidly extract behaviorally useful information from sensory stimuli. Appropriate behavioral adjustments to unexpected changes in stimulus statistics are fundamental for the survival of an animal. We still do not fully understand how the brain detects such changes reliably and quickly. Local neural circuits perform computations on incoming sensory stimuli in an efficient manner by maximizing transmitted information or minimizing metabolic cost (***Simoncelli and Olshausen, 2001***; ***Barlow, 2013***). Repeated or predictable stimuli do not provide new meaningful information. As a consequence, one should expect that responses to repeated stimuli are suppressed – a phenomenon postulated by the framework of predictive coding (***Clark, 2013***; ***Spratling, 2017***). Recent experiments have demonstrated that sensory circuits across different modalities can encode a sequence or expectation violation and can detect novelty (***Keller et al., 2012***; ***Natan et al., 2015***; ***Zmarz and Keller, 2016***; ***Hamm and Yuste, 2016***; ***Homann et al., 2017***). The underlying neuronal and circuit mechanisms behind expectation violation and novelty detection, however, remain elusive.

A prominent paradigm used experimentally involves two types of stimuli, the repeated (or frequent) and the novel (or deviant) stimulus (***Näätänen et al., 1982***; ***Fairhall, 2014***; ***Natan et al., 2015***; ***Homann et al., 2017***; ***Weber et al., 2019***). Here, the neuronal responses to repeated stimuli decrease, a phenomenon that is often referred to as adaptation (***Fairhall, 2014***). Adaptation can occur over a wide range of timescales, which range from milliseconds to seconds (***Ulanovsky et al., 2004***; ***Lundstrom et al., 2010***), and to multiple days in the case of behavioral habituation (***Haak et al., 2014***; ***Ramaswami, 2014***). We refer to the elevated neuronal response to a novel stimulus, compared to the response to a repeated stimulus, as a 'novelty response' (***Homann et al., 2017***). Responses to repeated versus novel stimuli, more generally, have also been studied on

different spatial scales spanning the single neuron level, cortical microcircuits and whole brain regions. At the scale of whole brain regions, a widely studied phenomenon is the mismatch negativity (MMN), which is classically detected in electroencephalography (EEG) data and often based on an auditory or visual 'oddball' paradigm (**Näätänen et al., 1982**; **Hamm and Yuste, 2016**). The occasional presentation of the so-called oddball stimulus among frequently repeated stimuli leads to a negative deflection in the EEG signal – the MMN (**Näätänen et al., 2007**).

Experiments at the cellular level typically follow the oddball paradigm with two stimuli that, if presented in isolation, would drive a neuron equally strongly. However, when one stimulus is presented frequently and the other rarely, the deviant produces a stronger response relative to the frequent stimulus (**Ulanovsky et al., 2003**; **Nelken, 2014**; **Natan et al., 2015**). The observed reduction in response to the repeated, but not the deviant, stimulus has been termed stimulus-specific adaptation (SSA) and has been suggested to contribute to the MMN (**Ulanovsky et al., 2003**). SSA has been observed in multiple brain areas, most commonly reported in the primary auditory cortex (**Ulanovsky et al., 2003**; **Yaron et al., 2012**; **Natan et al., 2015**; **Seay et al., 2020**) and the primary visual cortex (**Movshon and Lennie, 1979**; **Hamm and Yuste, 2016**; **Vinken et al., 2017**; **Homann et al., 2017**). Along the visual pathway, SSA has also been found at different earlier stages including the retina (**Schwartz et al., 2007**; **Geffen et al., 2007**; **Schwartz and Berry, 2008**) and the visual thalamic nuclei (**Dhruv and Carandini, 2014**; **King et al., 2016**).

To unravel the link between multiple spatial and temporal scales of adaptation, a variety of mechanisms has been proposed. Most notably, modeling studies have explored the role of adaptive currents, which reduce the excitability of the neuron (**Brette and Gerstner, 2005**), and short-term depression of excitatory feedforward synapses (**Tsodyks et al., 1998**). Most models of SSA in primary sensory areas of the cortex focus on short-term plasticity and the depression of thalamocortical feedforward synapses (**Mill et al., 2011a**; **Mill et al., 2011b**; **Park and Geffen, 2020**). The contribution of other mechanisms has been under-explored in this context. Recent experimental studies suggest that inhibition and the plasticity of inhibitory synapses shape the responses to repeated and novel stimuli (**Chen et al., 2015**; **Kato et al., 2015**; **Natan et al., 2015**; **Hamm and Yuste, 2016**; **Natan et al., 2017**; **Heintz et al., 2020**). Natan and colleagues observed that in the mouse auditory cortex, both parvalbumin-positive (PV) and somatostatin-positive (SOM) interneurons contribute to SSA (**Natan et al., 2015**). Furthermore, neurons that are more strongly adapted receive stronger inhibitory input than less adapted neurons, suggesting potentiation of inhibitory synapses as an underlying mechanism (**Natan et al., 2017**). In the context of habituation, inhibitory plasticity has been previously hypothesized to be the driving mechanism behind the reduction of neural responses to repeated stimuli (**Ramaswami, 2014**; **Barron et al., 2017**). Habituated behavior in *Drosophila*, for example, results from prolonged activation of an odor-specific excitatory subnetwork, which leads to the selective strengthening of inhibitory synapses onto the excitatory subnetwork (**Das et al., 2011**; **Glanzman, 2011**; **Ramaswami, 2014**; **Barron et al., 2017**).

Here, we focus on the role of inhibitory spike-timing-dependent plasticity (iSTDP) in characterizing neuronal responses to repeated and novel stimuli at the circuit level. We base our study on a recurrent spiking neural network model of the mammalian cortex with biologically inspired plasticity mechanisms that can generate assemblies in connectivity and attractors in activity to represent the stimulus-specific activation of specific sub-circuits (**Litwin-Kumar and Doiron, 2014**; **Zenke et al., 2015**; **Wu et al., 2020**). We model excitatory and inhibitory neurons and include stimulus-specific input not only to the excitatory but also to the inhibitory population, as found experimentally (**Ma et al., 2010**; **Griffen and Maffei, 2014**; **Znamenskiy et al., 2018**). This additional assumption readily leads to the formation of specific inhibitory-to-excitatory connections through inhibitory plasticity (**Vogels et al., 2011**), as suggested by recent experiments (**Lee et al., 2014**; **Xue et al., 2014**; **Znamenskiy et al., 2018**; **Najafi et al., 2020**).

We demonstrate that this model network can generate excess population activity when novel stimuli are presented as violations of repeated stimulus sequences. Our framework identifies plasticity of inhibitory synapses as a sufficient mechanism to explain population novelty responses and adaptive phenomena on multiple timescales. In addition, stimulus-specific inhibitory connectivity supports adaptation to specific stimuli (SSA). This finding reveals that the network configuration encompasses computational capabilities beyond those of intrinsic adaptation. Furthermore, we suggest disinhibition to be a powerful regulator of the amplification of novelty responses. Our modeling framework enables us to formulate additional experimentally testable predictions. Most intriguing,

we hypothesize that neurons in primary sensory cortex may not signal the violation of periodicity of a sequence based on bottom-up input, but rather adapt to the distribution of presented stimuli.

Results

A recurrent neural network model with plastic inhibition can generate novelty responses

Recent experimental studies have indicated an essential role of inhibitory circuits and inhibitory plasticity in adaptive phenomena and novelty responses (***Chen et al., 2015***; ***Kato et al., 2015***; ***Natan et al., 2015***; ***Hamm and Yuste, 2016***; ***Natan et al., 2017***; ***Heintz et al., 2020***). To understand if and how plastic inhibitory circuits could explain the emergence of novelty responses, we built a biologically plausible spiking neuronal network model of recurrently connected 4000 excitatory and 1000 inhibitory neurons based on recent experimental findings on tuning, connectivity, and inhibitory and excitatory STDP in the cortex (Materials and methods). Excitatory-to-excitatory (E-to-E) synapses were plastic based on the triplet spike-timing-dependent plasticity (eSTDP) rule (***Sjöström et al., 2001***; ***Pfister and Gerstner, 2006***; ***Gjorgjieva et al., 2011***; ***Figure 1—figure supplement 1***). The triplet STDP rule enabled the formation of strong bidirectional connections among similarly selective neurons (***Gjorgjieva et al., 2011***; ***Montangie et al., 2020***). Plasticity of connections from inhibitory to excitatory neurons was based on an inhibitory STDP (iSTDP) rule measured experimentally (***D'amour and Froemke, 2015***), and shown to stabilize excitatory firing rate dynamics in recurrent networks (***Vogels et al., 2011***; ***Figure 1—figure supplement 1***). In contrast to other frameworks which have found short-term plasticity as key for capturing adaptation phenomena, we only included long-term plasticity and did not explicitly model additional adaptation mechanisms.

We targeted different subsets of excitatory and inhibitory neurons with different external stimuli, to model that these neurons are stimulus-specific ('tuned') to a given stimulus (***Figure 1A***, left, see Materials and methods). One neuron could be driven by multiple stimuli. Starting from an initially randomly connected network, presenting tuned input led to the emergence of excitatory assemblies, which are strongly connected, functionally related subsets of excitatory neurons (***Figure 1—figure supplement 2C***, left). Furthermore, tuned input also led to the stimulus-specific potentiation of inhibitory-to-excitatory connections (***Figure 1—figure supplement 2E***, left). We refer to this part of structure formation as the 'pretraining phase' of our simulations (Materials and methods). This pretraining phase imprinted structure in the network prior to the actual stimulation paradigm as a model of the activity-dependent refinement of structured connectivity during early postnatal development (***Thompson et al., 2017***).

To test the influence of inhibitory plasticity on the emergence of a novelty response, we followed an experimental paradigm used to study novelty responses in layer 2/3 (L2/3) of mouse primary visual cortex (V1) (***Homann et al., 2017***). In ***Homann et al., 2017***, a single stimulus consisted of 100 randomly oriented Garbor patches. Three different stimuli (A, B, and C) were presented in a sequence (ABC) (***Figure 1A***, right). The same sequence (ABC) was then repeated several times in a sequence block. In the second-to-last sequence, the last stimulus was replaced by a novel stimulus (N). In the consecutive sequence block, a new sequence with different stimuli was presented (we refer to this as a unique sequence stimulation paradigm). The novel stimuli were also different for each sequence block. In this paradigm, we observed elevated population activity in the excitatory model population at the beginning of each sequence block ('onset response') and a steady reduction to a baseline activity level for the repeated sequence presentation (***Figure 1B***). Upon presenting a novel stimulus, the excitatory population showed excess activity, clearly discernible from baseline, called the 'novelty response'. This novelty response was comparable in strength to the onset response. Sorting spike rasters according to sequence stimuli revealed that stimulation leads to high firing rates in the neurons that are selective to the presented stimulus (A, B, or C) (***Figure 1C***). When we used a different set of stimuli in the stimulation versus the pretraining phase to better match the randomly oriented Gabor patches presented in ***Homann et al., 2017*** (***Figure 1—figure supplement 3A***, see Materials and methods), we found the same type of responses to repeated and novel stimuli (***Figure 1—figure supplement 3B***). When examining a random subset of neurons, we found general response sparseness and periodicity during sequence repetitions (***Figure 1D***), very

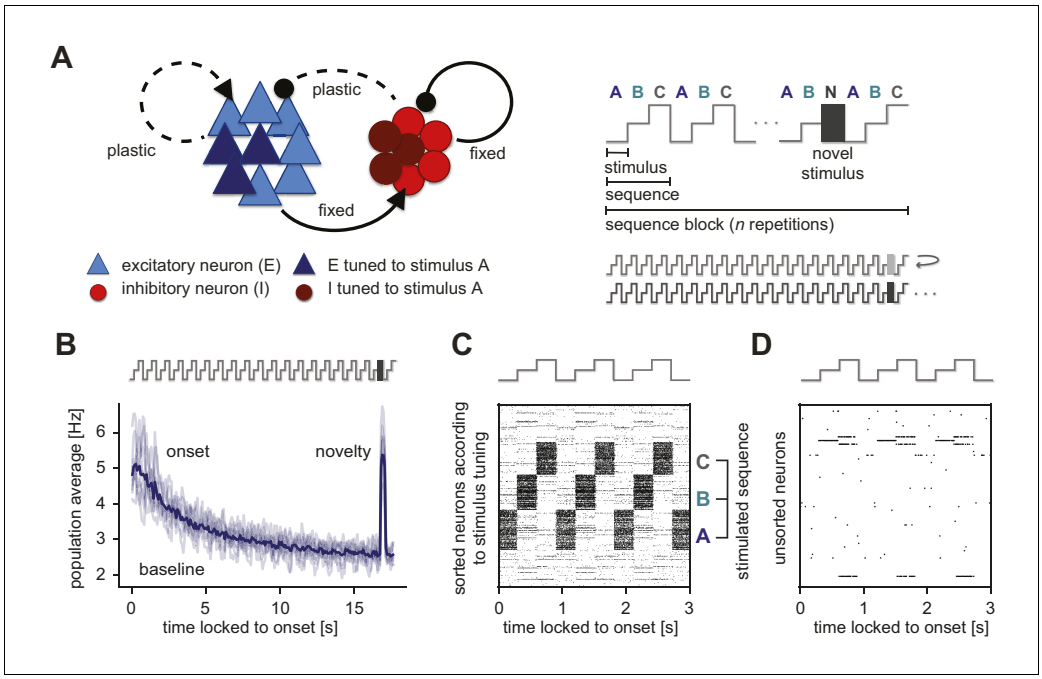


Figure 1. Generation of novelty responses in a recurrent plastic neural network model. (**A**) Left: A recurrently connected network of excitatory (E) neurons (blue triangles) and inhibitory (I) neurons (red circles) receiving tuned input. Excitatory neurons tuned to a sample stimulus A are highlighted in dark blue, the inhibitory counterparts in dark red. E-to-E synapses and I-to-E synapses were plastic, and all other synapses were fixed. Right: Schematic of the stimulation protocol. Multiple stimuli (A, B, and C) were presented in a sequence (ABC). Each sequence was repeated n times in a sequence block. In the second-to-last sequence, the last stimulus was replaced by a novel stimulus (N). Multiple sequence blocks followed each other without interruption, with each block containing sequences of different stimuli. (**B**) Population average firing rate of all excitatory neurons as a function of time after the onset of a sequence block. Activity was averaged (solid line) across multiple non-repeated sequence blocks (transparent lines: individual blocks). A novel stimulus (dark gray) was presented as the last stimulus of the second-to-last sequence. (**C**) Spiking activity in response to a sequence (ABC) in a subset of 1000 excitatory neurons where the neurons were sorted according to the stimulus from which they receive tuned input. A neuron can receive input from multiple stimuli and can appear more than once in this raster plot. (**D**) A random unsorted subset of 50 excitatory neurons from panel C. Time was locked to the sequence block onset.

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Excitatory and inhibitory synaptic plasticity functions for different pairing frequencies.

Figure supplement 2. Strong connections form between excitatory and excitatory, as well as inhibitory and excitatory neuron groups that are tuned to the same stimulus.

Figure supplement 3. Different stimuli in the pretraining and stimulation phases generate similar synaptic weight and firing rate dynamics.

Figure supplement 4. Quantifying response density in the unique sequence stimulation paradigm.

Figure supplement 5. Normalization time step Δt does not affect the occurrence of a novelty response.

similar to experimental findings (***Homann et al., 2017***). More concretely, sparse population activity for repeated stimuli in our model network was the result of each stimulus presentation activating a subset of excitatory neurons in the network, which were balanced by strong inhibitory feedback. Therefore, only neurons that directly received this feedforward drive were highly active, while most other neurons in the network were instead rather silent. Periodicity in the activity of single neurons resulted from the repetition of a sequence.

In the model, the fraction of active excitatory neurons was qualitatively similar for novel, adapted and onset stimuli (***Figure 1—figure supplement 4***). The relatively sparse novelty response in our model was the result of increased inhibition onto all excitatory neurons in the network, with activity remaining mainly in the neurons tuned to the novel stimulus. In contrast, ***Homann et al., 2017*** found that a large fraction of neurons respond to a novel stimulus, suggesting a dense novelty response.

Since the increase in inhibition seems to be responsible for the absence of a dense novelty response in our model, in a later section we suggest disinhibition as a mechanism to achieve the experimentally observed dense novelty responses in our model.

Our results suggest that presenting repeated stimuli (and repeated sequences of stimuli) to a plastic recurrent network with tuned excitatory and inhibitory neurons readily leads to a reduction of the excitatory averaged population response, consistent with the observed adaptation in multiple experimental studies in various animal models and brain regions (***Ulanovsky et al., 2003***; ***Hamm and Yuste, 2016***; ***Homann et al., 2017***). Importantly, the model network generates a novelty response when presenting a novel stimulus by increasing the excitatory population firing rate at the time of stimulus presentation (***Näätänen et al., 2007***).

The dynamics of novelty and onset responses depend on sequence properties

To explore the dynamics of novelty responses, we probed the model network with a modified stimulation paradigm. Rather than fixing the number of sequence repetitions in one sequence block (***Figure 1A***, right), here we presented a random number of sequence repetitions (nine values between 4 and 45 repetitions) for each sequence block. This allowed us to measure the novelty and onset responses as a function of the number of sequence repetitions. Novelty and onset responses were observed after as few as four sequence repetitions (***Figure 2A***). After more than 15 sequence repetitions, the averaged excitatory population activity reached a clear baseline activity level (***Figure 2A***). The novelty response amplitude, measured by the population rate of the novelty peak minus the baseline population rate, increased with the number of sequence repetitions before saturating for a high number of sequence repeats (***Figure 2B***, black dots). The onset response amplitude after the respective sequence block followed the same trend (***Figure 2B***, gray dots). Next, we varied the number of stimuli in a sequence, resulting in different sequence lengths across blocks (3 to 15 stimuli per sequence). By averaging excitatory population responses across sequence blocks with equal length, we found that the decay of the onset response depends on the number of stimuli in a sequence (***Figure 2C***). Upon fitting an exponentially decaying function to the activity of the onset response, we derived a linear relationship between the number of stimuli in a sequence and the decay constant (***Figure 2D***).

In summary, we found that novelty responses arise for different sequence variations. Our model network suggests that certain features of the novelty response depend on the properties of the presented sequences. Changing the number of sequence repetitions modifies the onset and novelty response amplitude (***Figure 2A,B***), while a longer sequence length leads to a longer adaptation time constant (***Figure 2C,D***). Interestingly, both findings are in good qualitative agreement with experimental data that presented similar sequence variations (***Homann et al., 2017***). An exponential fit of the experimental data found a time constant of $\tau = 3.2 \pm 0.7$ repetitions when the number of sequence repetitions was varied (***Homann et al., 2017***). The time constant in our model network was somewhat longer ($\tau = 9 \pm 1$ repetitions), but on a similar order of magnitude (***Figure 2B***). Similarly, our model network produced a linear relationship between the adaptation time constant and sequence length with a slope of $m = 1.6 \pm 0.04$ (***Figure 2D***), very close to the slope extracted from the data ($m = 2.1 \pm 0.3$) (***Homann et al., 2017***). Therefore, grounded on biologically-plausible plasticity mechanisms, and capable of capturing the emergence and dynamics of novelty responses, our model network provides a suitable framework for a mechanistic dissection of the circuit contributions in the generation of a novelty response.

Stimulus periodicity in the sequence is not required for the generation of a novelty response

Experimental studies have often reported novelty or deviant responses by averaging across several trials due to poor signal-to-noise ratios of the measured physiological activity (***Homann et al., 2017***; ***Vinken et al., 2017***). Therefore, we investigated the network response to paradigms with repeated individual sequence blocks (***Figure 3A***), which we refer to as the repeated sequence stimulation paradigm. We randomized the order of the sequence block presentation to avoid additional temporal structure beyond the stimulus composition of the sequences. Repeating sequence blocks dampened the onset response at sequence onset compared to the unique sequence stimulation paradigm

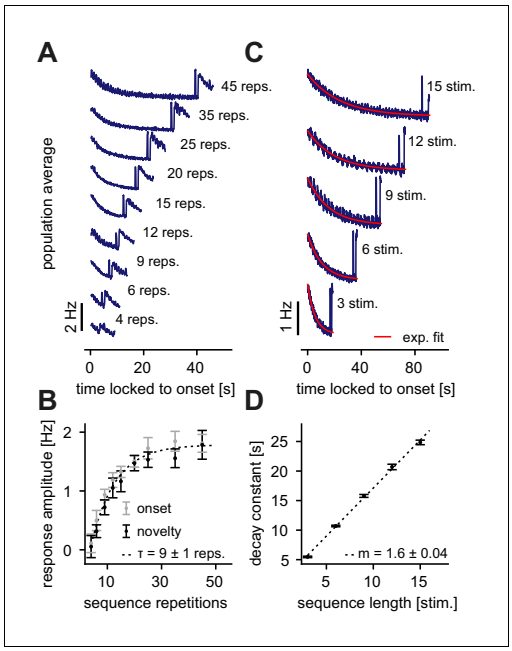


Figure 2. Dependence of the novelty response on the number of sequence repetitions and the sequence length. (**A**) Population average firing rate of all excitatory neurons for a different number of sequence repetitions within a sequence block. Time is locked to the sequence block onset. (**B**) The response amplitude of the onset (gray) and the novelty (black) response as a function of sequence repetitions fit with an exponential with a time constant τ. (**C**) Population average firing rate of all excitatory neurons for varying sequence length fit with an exponential function (red). Time is locked to the sequence block onset. (**D**) The onset decay time constant (fit with an exponential, as shown in panel C) as a function of sequence length. The simulated data was fit with a linear function with slope m. (**B, D**) Error bars correspond to the standard deviation across five simulated instances of the model.

(compare ***Figure 1B*** and ***Figure 2A,B*** with ***Figure 3A***). Next, we wondered whether the excitatory and inhibitory population responses to repeated and novel stimuli are related. We found that both excitatory and inhibitory populations adapt to the repeated stimuli and show a prominent novelty peak that is larger than the respective averaged onset response (***Figure 3B,C***). Based on these findings, we make the following predictions for future experiments: (1) A novelty response is detectable in both the excitatory and inhibitory populations. (2) The sequence onset response is dampened for multiple presentations of the same sequence block compared to the presentation of unique sequence blocks.

Next, we investigated whether the generation of novelty responses observed in the model network depends on the sequence structure. If the novelty responses were to truly signal the violation of the sequence structure or the stimulus predictability in a sequence, we would expect a novelty response to occur if two stimuli in a sequence were swapped, that is, ACB instead of ABC. We found that swapping the last and second-to-last stimulus, instead of presenting a novel stimulus, does not elicit a novelty response (***Figure 3D***). Additionally, we asked whether the periodicity of the stimuli within a sequence influences the novelty response. Shuffling the stimuli within a sequence block still generates a novelty response and adaptation to the repeated stimuli, similar to the strictly periodic case (***Figure 3E***, compare to ***Figure 3A***). Finally, we investigated if the novelty peak depends on the input firing rate of the novel stimulus. We found that a reduction of the input drive decreases the novelty peak, revealing a monotonic dependence of the novelty response on stimulus strength (***Figure 3F***). Based on these results, we make two additional predictions: (3) The periodicity of stimuli in the sequence is not required for the generation of a novelty response. Hence, the novelty response encodes the distribution of presented stimuli, rather than the structure of a sequence. (4) A novelty response depends on the strength of the novel stimulus.

Increased inhibition onto highly active neurons leads to adaptation

To gain an intuitive understanding for the sensitivity of novelty responses to stimulus identity but lack of sensitivity to stimulus periodicity in the sequence, we more closely examined the role of inhibitory plasticity as the leading mechanism behind the novelty responses in our model. We found that novelty responses arise because inhibitory plasticity fails to sufficiently increase inhibitory input and to counteract the excess excitatory input into excitatory neurons upon the presentation of a novel stimulus. In short, novelty responses can be understood as the absence of adaptation in an otherwise adapted response. Adaptation in the network arises through increased inhibition onto highly active neurons through selective strengthening of I-to-E weights (***Figure 4A***).

To determine how inhibitory plasticity drives the generation of novelty responses or, equivalently, adaptation in our model, we studied the evolution of inhibitory weights. The inhibitory weights onto stimulus-specific assemblies tuned to the stimuli in a given sequence increased upon presentation of

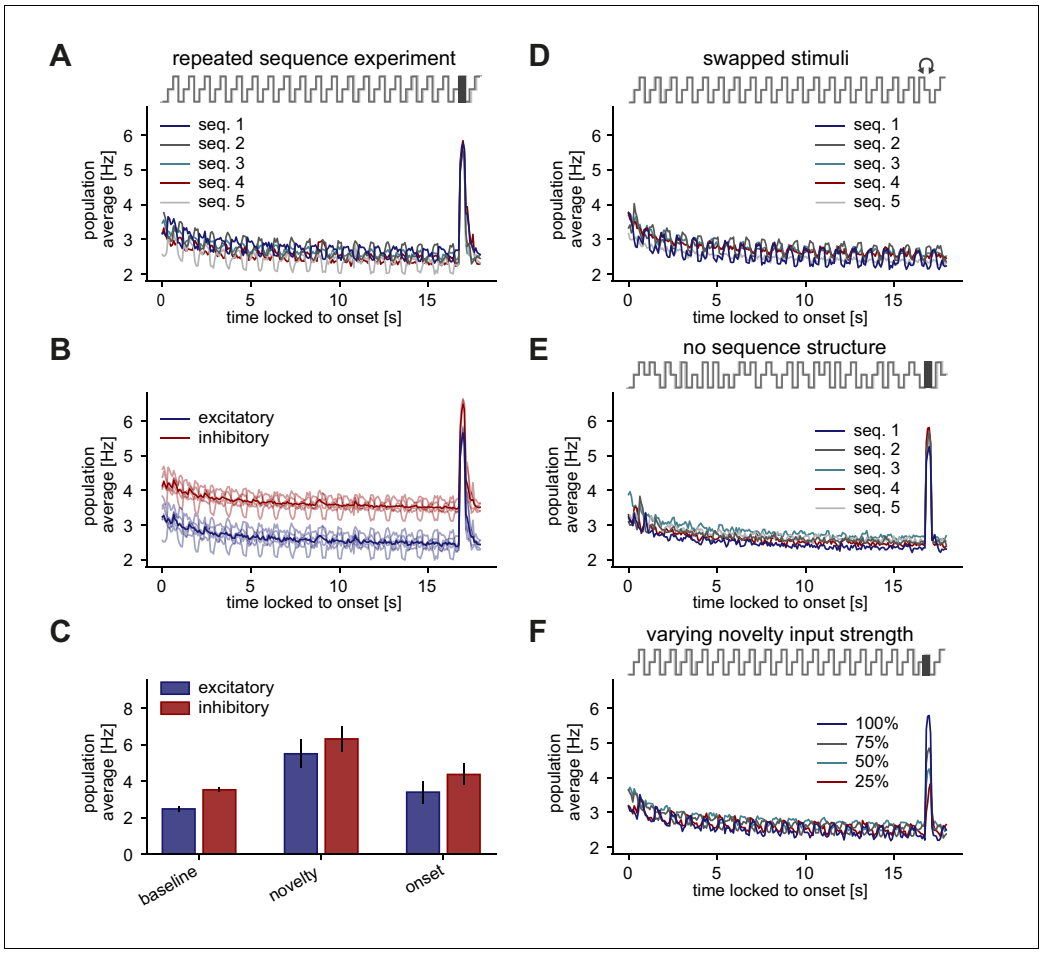


Figure 3. Stimulus periodicity in the sequence is not required for the generation of a novelty response. (**A–F**) Population average firing rate of all excitatory neurons (and all inhibitory neurons in B,C) during the presentation of five different repeated sequence blocks. The population firing rate was averaged across ten repetitions of each sequence block. Time is locked to sequence block onset. (**A**) A novel stimulus was presented as the last stimulus of the second-to-last sequence. (**B**) Same as panel A but for both excitatory and inhibitory populations (transparent lines: individual sequence averages). (**C**) Comparison of baseline, novelty, and onset response for inhibitory and excitatory populations. Error bars correspond to the standard deviation across the five sequence block averages shown in B. (**D**) In the second-to-last sequence, the last and second-to-last stimulus were swapped instead of presenting a novel stimulus. (**E**) Within a sequence, stimuli were shuffled in a pseudo-random manner where a stimulus could not be presented twice in a row. A novel stimulus was presented as the last stimulus of the second-to-last sequence. (**F**) A novel stimulus was presented as the last stimulus of the second-to-last sequence. Each sequence had a different feedforward input drive for the novel stimulus, indicated by the percentage of the typical input drive for the novel stimulus used before.

the corresponding sequence block, and decreased otherwise (***Figure 4B***). The population firing rate during repeated presentation of a sequence decreased (adapted) on the same timescale as the increase of the inhibitory weights related to this sequence (***Figure 4C***). When a stimulus was presented to the network for the first time, the total excitatory input to the corresponding excitatory neurons was initially not balanced by inhibition. Hence, the neurons within the assembly tuned to that stimulus exhibited elevated activity at sequence onset, leading to what we called the 'onset response' (***Figure 1B***). The same was true for the novelty responses as reflected in low inhibitory weights onto novelty assemblies relative to repeated assemblies (***Figure 1—figure supplement 2D, E***). Consequently, the generation of a novelty response did not depend on the specific periodicity of the stimuli within a sequence (***Figure 3***). Swapping two stimuli did not generate a novelty response since the corresponding assemblies of each stimulus were already in an adapted state. Therefore,

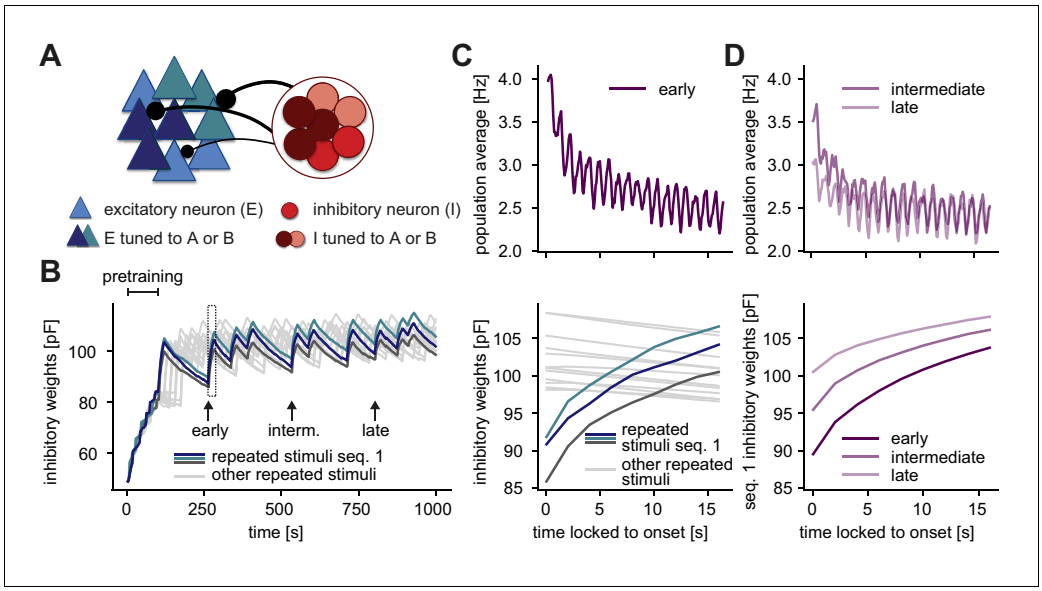


Figure 4. Inhibition onto neurons tuned to repeated stimuli increases during sequence repetitions. (**A**) Schematic of increased inhibitory weights onto two stimulus-specific assemblies upon the repeated presentation of stimuli A and B (indicated in dark blue and turquoise) relative to neurons from other assemblies (light blue). (**B**) Evolution of the average inhibitory weights onto stimulus-specific assemblies. Colored traces mark three stimulus-specific assemblies in sequence 1: A, B, and C. Arrows indicate time points of early, intermediate, and late sequence block presentation shown in C and D. (**C**) Top: Population average firing rate of all excitatory neurons during the repeated presentation of sequence 1 at an early time point (see panel B). Time is locked to sequence onset. Bottom: Close-up of panel B (rectangle). Time is locked to sequence onset. (**D**) Top: Same as panel C (top) but at intermediate and late time points (see panel B). Bottom: Corresponding dynamics of the average inhibitory weights onto all three stimulus-specific assemblies from sequence 1 at early, intermediate and late time points (see panel B). The dark purple trace (early) corresponds to the average of the three colored traces in C (bottom).
The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Pretraining parameters do not qualitatively influence the novelty response.
Figure supplement 2. Fast inhibitory plasticity is key for the generation of a novelty response.

our results suggest that the exact sequence structure of stimulus presentations is not relevant for the novelty response, as long as the overall distribution of stimuli is maintained.

Interestingly, we found that adaptation occurs on multiple timescales in our model. The fastest is the timescale of milliseconds on which inhibitory plasticity operates, the next slowest is the timescale of seconds corresponding to the presentation of a sequence block, and finally the slowest is the timescale of minutes corresponding to the presentation of the same sequence block multiple times (***Figure 4D***, top; also compare ***Figure 1B*** and ***Figure 3A***). The slowest decrease in the population firing rate was the result of long-lasting changes in the average inhibitory weights onto the excitatory neurons tuned to the stimuli within a given sequence. Hence, the average inhibitory weight for a given sequence increased with the number of previous sequence block presentations of that sequence (***Figure 4D***, bottom).

Using a different set of stimuli in the stimulation versus the pretraining phase to match the randomly oriented Gabor patches presented in ***Homann et al., 2017***, led to qualitatively similar firing rate and synaptic weight dynamics (***Figure 1—figure supplement 3C,D***, see also Materials and methods). Differences in the mean inhibitory weights onto different stimulus-specific assemblies in a given sequence were due to random initial differences in assembly size and connection strength (***Figure 4B,C***, see Materials and methods). Differences in early, intermediate, and late inhibitory weight changes, however, were consistent across different experiments and model instantiations (***Figure 4D***, ***Figure 1—figure supplement 3D***, right).

Furthermore, we observed that the dynamics of inhibitory plasticity and the generation of a novelty response did not depend on the exact parameters of the pretraining phase (***Figure 4—figure***

supplement 1). Specifically, increasing the number of repetitions in the pretraining phase increased the height of the novelty peak, but eventually reached a plateau at 10 repetitions (***Figure 4—figure supplement 1A***). Increasing the number of stimuli decreased the height of the novelty peak (***Figure 4—figure supplement 1C***). However, these pretraining parameters only affected some aspects of the novelty response, but preserved the generation of the novelty response. Even without a pretraining phase (zero number of repetitions), a novelty response could be generated.

Based on our result that inhibitory plasticity is the underlying mechanism of adapted and novelty responses in our model, we wondered how fast it needs to be. Hence, we tested the influence of the inhibitory learning rate (η) in the unique sequence stimulation paradigm. We found that inhibitory plasticity needs to be fast for both results, the generation of a novelty response (***Figure 4—figure supplement 2A,B***) and adaptation to repeated stimuli (***Figure 4—figure supplement 2C***). Whether such fast inhibitory plasticity operates in the sensory cortex to underlie the adapted and novelty responses is still unknown.

In summary, we identified the plasticity of connections from inhibitory to excitatory neurons belonging to a stimulus-specific assembly as the key mechanism in our framework for the generation of novelty responses and for the resulting adaptation of the network response to repeated stimuli. This adaptation occurs on multiple timescales, covering the range from the timescale of inhibitory plasticity (milliseconds) to sequence block adaptation (seconds) to the presentation of multiple sequence blocks (minutes).

The adapted response depends on the interval between stimulus presentations

Responses to repeated stimuli do not stay adapted but can recover if the repeated stimulus is no longer presented (***Ulanovsky et al., 2004***; ***Cohen-Kashi Malina et al., 2013***). We investigated the

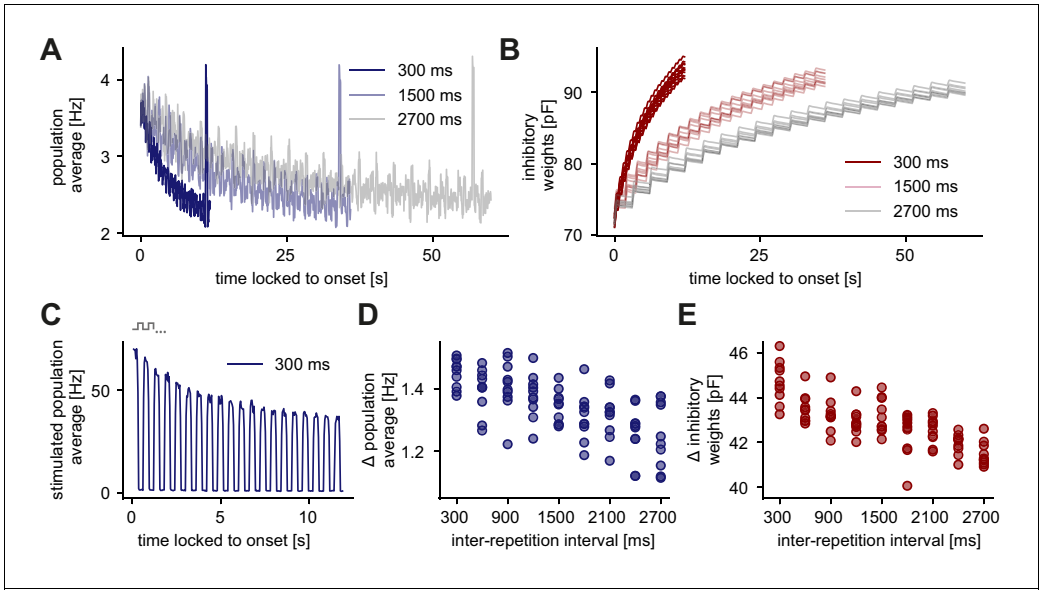


Figure 5. Longer inter-repetition intervals decrease the level of adaptation due to the recovery of inhibitory synaptic weights. (**A**) Population average firing rate of all excitatory neurons in the unique sequence stimulation paradigm for varying inter-repetition intervals (varying sequence length). Time is locked to the sequence block onset. (**B**) Evolution of the average inhibitory weights onto stimulus-specific assembly A (identical in all runs) for varying inter-repetition intervals. Time is locked to the sequence block onset. (**C**) Population average firing rate of stimulated excitatory neurons for a 300 ms inter-repetition interval. Time is locked to the sequence block onset. One step in the schematic corresponds to one stimulus in a presented sequence. (**D**) Difference of the onset population rate (measured at the onset of the stimulation, averaged across runs) and the baseline rate (measured before novelty response) as a function of the inter-repetition interval. (**E**) Absolute change of inhibitory weights onto stimulus-specific assembly A from the start until the end of a sequence block presentation as a function of inter-repetition interval.

recovery of adapted responses in the unique sequence stimulation paradigm (***Figure 5A***). Similar to ***Figure 2C***, we changed the number of stimuli in the sequence, which leads to different inter-repetition intervals of a repeated sequence stimulus (the interval until the same stimulus is presented again). For example, if two repeated stimuli (A, B) are presented, the inter-repetition interval for each stimulus is 300 ms because each stimulus is presented for 300 ms (***Figure 5C***). If four repeated stimuli are presented (A, B, C, D), the inter-repetition interval for each stimulus is 900 ms. We defined the adaptation level as the difference of the onset population rate, measured at the onset of the stimulation, and the baseline rate, measured shortly before the presentation of a novel stimulus. We found that an increase in the inter-repetition interval reduced the adaptation level of the excitatory population (***Figure 5A,D***) due to a decrease of inhibitory synaptic strength onto stimulus-specific assemblies (***Figure 5B,E***). More specifically, the population average of all excitatory neurons tuned to stimulus A was high when stimulus A was presented and low when stimulus B was presented (***Figure 5C***). Hence, inhibitory weights onto stimulus-specific assembly A increased while A was presented and decreased otherwise (***Figure 5B***).

In summary, longer inter-repetition intervals provide more time for the inhibitory weights onto stimulus-specific assemblies to decrease, hence, weakening the adaptation.

Inhibitory plasticity and tuned inhibitory neurons support stimulus-specific adaptation

Next, we investigated whether inhibitory plasticity of tuned inhibitory neurons support additional computational capabilities beyond the generation of novelty responses and adaptation of responses to repeated stimuli on multiple timescales. Therefore, we implemented a different stimulation paradigm to investigate the phenomenon of stimulus-specific adaptation (SSA). At the single-cell level, SSA typically involves a so-called oddball paradigm where two stimuli elicit an equally strong response when presented in isolation, but when one is presented more frequently, the elicited response is weaker than for a rarely presented stimulus (***Natan et al., 2015***).

We implemented a similar paradigm at the network level where the excitatory neurons corresponding to two stimuli A and B were completely overlapping and the inhibitory neurons were partially overlapping (***Figure 6A***). Upon presenting stimulus A several times, the neuronal response gradually adapted to the baseline level of activity, while presenting the oddball stimulus B resulted in an increased population response (***Figure 6B***). Therefore, this network was able to generate SSA. Even though stimuli A and B targeted the same excitatory cells, the network response adapted only to stimulus A, while generating a novelty response for stimulus B. Even after presenting stimulus B, activating stimulus A again preserved the adapted response (***Figure 6B***). This form of SSA exhibited by our model network is in agreement with many experimental findings in the primary auditory cortex, primary visual cortex, and multiple other brain areas and animal models (***Nelken, 2014***). In our model network, SSA could neither be generated with adaptive neurons and static synapses (***Figure 6C***, top; Materials and methods), nor with inhibitory plasticity without inhibitory tuning (***Figure 6C***, bottom). In fact, including an adaptive current in the model neurons (***Brette and Gerstner, 2005***) did not even lead to adaptation of the response to a frequent stimulus since firing rates rapidly adapted during stimulus presentation and completely recovered in the inter-stimulus pause (***Figure 6C***, top).

We investigated the dynamics of inhibitory weights to understand the mechanism behind SSA in our model network. During the presentation of stimulus A, stimulus-specific inhibitory weights corresponding to stimulus A (average weights from inhibitory neurons tuned to stimulus A onto excitatory neurons tuned to stimulus A, see ***Figure 1—figure supplement 2A***, right) increased their strength, while stimulus-specific inhibitory weights corresponding to stimulus B remained low (***Figure 6D***). Hence, upon presenting the oddball stimulus B, the stimulus-specific inhibitory weights corresponding to stimulus B remained sufficiently weak to keep the firing rate of excitatory neurons high, thus resulting in a novelty response.

We next asked about the recovery of the adapted response in this SSA paradigm (***Figure 6—figure supplement 1***). After a 9 s pause, the response remained adapted (***Figure 6—figure supplement 1A***). Only after more than 200 s the response fully recovered (***Figure 6—figure supplement 1B***). In contrast to the results in ***Figure 5***, here, the adaptation level remained high due to the absence of network activity between stimulus presentations. Adaptation slowly recovered as the time between stimulus presentations increased.

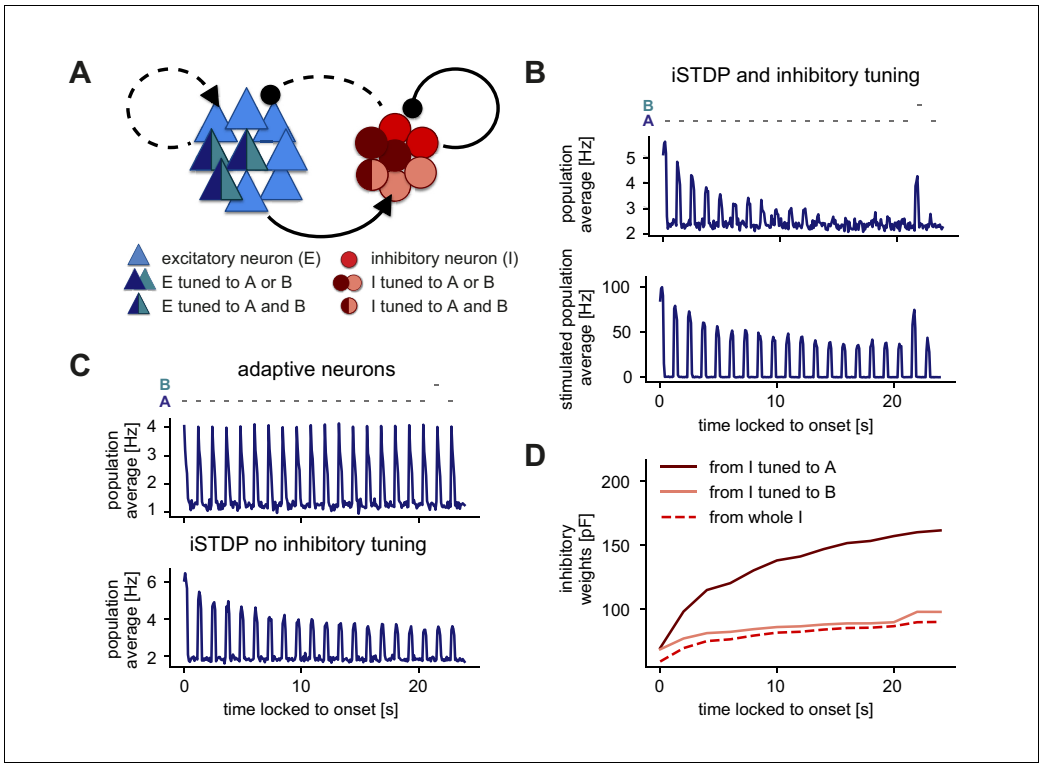


Figure 6. Stimulus-specific adaptation follows from inhibitory plasticity and tuning of both excitatory and inhibitory neurons. (**A**) Stimuli A and B provided input to the same excitatory neurons (dark blue and turquoise). Some neurons in the inhibitory population were driven by both A and B (dark red and rose) and some by only one of the two stimuli (dark red or rose). (**B,C**) Population average firing rate of excitatory neurons over time while stimulus A was presented 20 times. Stimulus B was presented instead of A as the second-to-last stimulus. Time is locked to stimulation onset. (**B**) Top: Population average of all excitatory neurons in the network with inhibitory plasticity (iSTDP) and inhibitory tuning. Bottom: Population average of stimulated excitatory neurons only (stimulus-specific to A and B). (**C**) Top: Same as panel B (top) for neurons with an adaptive current in a non-plastic recurrent network. Bottom: Same as panel B (top) for the network with inhibitory plasticity (iSTDP) and no inhibitory tuning. (**D**) Weight evolution of stimulus-specific inhibitory weights corresponding to stimuli A and B and average inhibitory weights.

The online version of this article includes the following figure supplement(s) for figure 6:

Figure supplement 1. Recovery of adapted responses in the SSA paradigm.

In summary, our results suggest that the combination of inhibitory plasticity and inhibitory tuning can give rise to SSA. Previous work has argued that inhibition or inhibitory plasticity does not allow for SSA (***Nelken, 2014***). However, this is only true if inhibition is interpreted as a 'blanket' without any tuning in the inhibitory population. Including recent experimental evidence for tuned inhibition into the model, (***Lee et al., 2014***; ***Xue et al., 2014***; ***Znamenskiy et al., 2018***), can indeed capture the emergence of SSA.

Disinhibition leads to novelty response amplification and a dense population response

Beyond the bottom-up computations captured by the network response to the different stimuli, we next explored the effect of additional modulations or top-down feedback into our network model. Top-down feedback has been frequently postulated to signal the detection of an error or irregularity in the framework of predictive coding (***Clark, 2013***; ***Spratling, 2017***). Therefore, we specifically tested the effect of disinhibitory signals on sequence violations by inhibiting the population of inhibitory neurons during the presentation of a novel stimulus (***Figure 7A***). Recent evidence has identified a differential disinhibitory effect in sensory cortex in the context of adapted and novelty responses

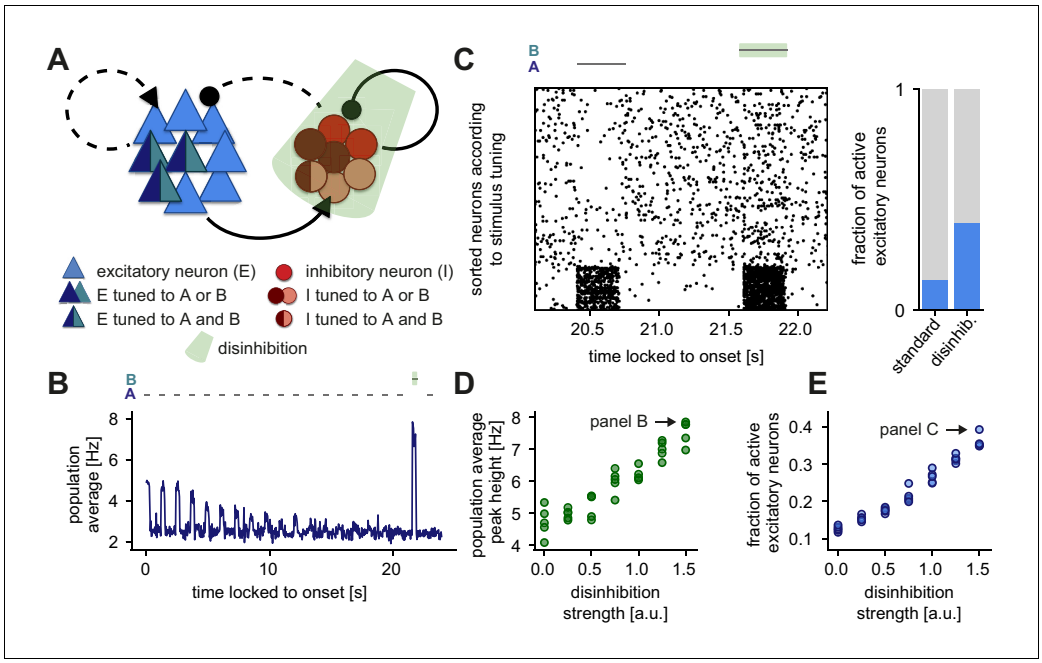


Figure 7. Disinhibition leads to a novelty response amplification and a dense population response. (**A**) Stimuli A and B provided input to the same excitatory neurons (dark blue and turquoise). Some neurons in the inhibitory population were driven by both A and B (dark red and rose) and some by only one of the two stimuli (dark red or rose). Inhibition (light green) of the entire inhibitory population led to disinhibition of the excitatory population. (**B**) Population average firing rate of all excitatory neurons over time while stimulus A is presented 20 times. Stimulus B was presented instead of A as the second-to-last stimulus. During the presentation of B, the inhibitory population was inhibited. Time is locked to stimulation onset. (**C**) Left: Raster plot of 250 excitatory neurons corresponding to the population average shown in panel B. The 50 neurons in the bottom part of the raster plot were tuned to stimuli A and B. Time is locked to stimulation onset. Right: Fraction of active excitatory neurons (at least one spike in a 100 ms window) measured directly after the onset of a stimulus. The raster plot and the fraction of active excitatory neurons are shown for the presentation of stimulus B (with disinhibition) and the preceding presentation of stimulus A (standard). (**D**) Population average peak height during disinhibition and the presentation of stimulus B, as a function of the disinhibition strength. Arrow indicates the population average peak height of the trace shown in panel B. Results are shown for five simulations. (**E**) Fraction of active excitatory neurons during disinhibition as a function of the disinhibition strength. Arrow indicates the data point corresponding to panel C. Results are shown for five simulations.

(***Natan et al., 2015***). However, due to the scarcity of detailed knowledge about higher order feedback signals or within-layer modulations in this context, we did not directly model the source of disinhibition.

When repeating the SSA experiment (***Figure 6***) and applying such a disinhibitory signal (inhibition of the inhibitory population) at the time of the novel stimulus B, our model network amplified the novelty response (***Figure 7B***, shaded green, also compare to ***Figure 6B***, top). Disinhibition also increased the density of the network response which corresponds to the number of active excitatory neurons (***Figure 7C***, left). Indeed, disinhibition increased the fraction of active excitatory neurons, which we defined as the fraction of neurons that spike at least once in a 100 ms window during the presentation of a stimulus (***Figure 7C***, right). Dense novelty responses have been recently reported experimentally, where novel stimuli elicited excess activity in a large fraction of the neuronal population in mouse V1 (***Homann et al., 2017***). Without a disinhibitory signal, the fraction of active neurons for a novel stimulus in our model was qualitatively similar as for repeated stimuli and therefore there was no dense novelty response (***Figure 1—figure supplement 4A***). Given that the inclusion of a disinhibitory signal readily increases the density of the novelty response, we suggest that disinhibition might underlie these experimental findings.

In sum, we found that by controlling the total disinhibitory strength (Materials and methods), disinhibition can flexibly amplify the novelty peak (***Figure 7D***) and increase the density of novelty responses (***Figure 7E***). Therefore, we propose that disinhibition can be a powerful mechanism to modulate novelty responses in a network of excitatory and inhibitory neurons.

Discussion

We developed a recurrent network model with plastic synapses to unravel the mechanistic underpinning of adaptive phenomena and novelty responses. Using the paradigm of repeated stimulus sequences (***Figure 1A***, right), our model network captured the adapted, sparse and periodic responses to repeated stimuli (***Figure 1B–D***) as observed experimentally (***Fairhall, 2014***; ***Homann et al., 2017***). The model network also exhibited a transient elevated population response to novel stimuli (***Figure 1B***), which could be modulated by the number of sequence repetitions and the sequence length in the stimulation paradigm (***Figure 2***), in good qualitative agreement with experimental data (***Homann et al., 2017***). We proposed inhibitory synaptic plasticity as a key mechanism behind the generation of these novelty responses. In our model, repeated stimulus presentation triggered inhibitory plasticity onto excitatory neurons selective to the repeated stimulus, reducing the response of excitatory neurons and resulting in their adaptation (***Figure 4***). In contrast, for a novel stimulus inhibitory input onto excitatory neurons tuned to that stimulus remained low, generating the elevated novelty response. Furthermore, we showed that longer inter-repetition intervals led to the recovery of adapted responses (***Figure 5***).

Based on experimental evidence (***Ohki and Reid, 2007***; ***Griffen and Maffei, 2014***), we included specific input onto both the excitatory and the inhibitory populations (***Figure 1A***, left). Such tuned inhibition (as opposed to untuned, 'blanket' inhibition commonly used in previous models) enabled the model network to generate SSA (***Figure 6***). Additionally, in the presence of tuned inhibition, a top-down disinhibitory signal achieved a flexible control of the amplitude and density of novelty responses (***Figure 7***). Therefore, besides providing a mechanistic explanation for the generation of adapted and novelty responses to repeated and novel sensory stimuli, respectively, our network model enabled us to formulate multiple experimentally testable predictions, as we describe below.

Inhibitory plasticity as an adaptive mechanism

We proposed inhibitory plasticity as the key mechanism that allows for adaptation to repeated stimulus presentation and the generation of novelty responses in our model. Many experimental studies have characterized spike-timing-dependent plasticity (STDP) of synapses from inhibitory onto excitatory neurons (***Holmgren and Zilberter, 2001***; ***Woodin et al., 2003***; ***Haas et al., 2006***; ***Maffei et al., 2006***; ***Wang and Maffei, 2014***; ***D'amour and Froemke, 2015***; ***Field et al., 2020***). In theoretical studies, network models usually include inhibitory plasticity to dynamically stabilize recurrent network dynamics (***Vogels et al., 2011***; ***Litwin-Kumar and Doiron, 2014***; ***Zenke et al., 2015***). In line with recent efforts to uncover additional functional roles of inhibitory plasticity beyond the stabilization of firing rates (***Hennequin et al., 2017***), here, we investigated potential functional consequences of inhibitory plasticity in adaptive phenomena. We were inspired by recent experimental work in the mammalian cortex (***Chen et al., 2015***; ***Kato et al., 2015***; ***Natan et al., 2015***; ***Hamm and Yuste, 2016***; ***Natan et al., 2017***; ***Heintz et al., 2020***), and simpler systems, such as *Aplysia* (***Fischer et al., 1997***; ***Ramaswami, 2014***) and in *Drosophila* (***Das et al., 2011***; ***Glanzman, 2011***) along with theoretical reflections (***Ramaswami, 2014***; ***Barron et al., 2017***), which all point towards a prominent role of inhibition and inhibitory plasticity in the generation of the MMN, SSA, and habituation. For example, Natan and colleagues observed that in the mouse auditory cortex, both PV and SOM interneurons contribute to SSA (***Natan et al., 2015***), possibly due to inhibitory potentiation (***Natan et al., 2017***). In the context of habituation, daily passive sound exposure has been found to lead to an upregulation of the activity of inhibitory neurons (***Kato et al., 2015***). Furthermore, increased activity to a deviant stimulus in the MMN is diminished when inhibitory neurons are suppressed (***Hamm and Yuste, 2016***).

Most experimental studies on inhibition in adaptive phenomena have not directly implicated inhibitory plasticity as the relevant mechanism. Instead, some studies have suggested that the firing rate of the inhibitory neurons changes, resulting in more inhibitory input onto excitatory cells, effectively leading to adaptation (***Kato et al., 2015***). In principle, there can be many other reasons why

the inhibitory input increases: disinhibitory circuits, modulatory signals driving specific inhibition, or increased synaptic strength of excitatory-to-inhibitory connnections, to name a few. However, following experimental evidence (**Natan et al., 2017**) and supported by our results, the plasticity of inhibitory-to-excitatory connections emerges as a top candidate underlying adaptive phenomena. In our model, adaptation to repeated stimuli and the generation of novelty responses via inhibitory plasticity do not depend on the exact shape of the inhibitory STDP learning rule. It is only important that inhibitory plasticity generates a 'negative feedback' whereby high excitatory firing rates lead to net potentiation of inhibitory synapses while low excitatory firing rates lead to net depression of inhibitory synapses. Other inhibitory STDP learning rules can also implement this type of negative feedback (**Luz and Shamir, 2012**; **Kleberg et al., 2014**), and we suspect that they would also generate the adapted and novelty responses as in our model.

One line of evidence to speak against inhibitory plasticity argues that SSA might be independent of NMDA activation (**Farley et al., 2010**). Inhibitory plasticity, on the contrary, seems to be NMDA receptor-dependent (**D'amour and Froemke, 2015**; **Field et al., 2020**). However, there exists some discrepancy in how exactly NMDA receptors are involved in SSA (**Ross and Hamm, 2020**), since blocking NMDA receptors can disrupt the MMN (**Tikhonravov et al., 2008**; **Chen et al., 2015**). These results indicate that a further careful disentanglement of the underlying cellular mechanisms of adaptive phenomena is needed.

In our model, the direction of inhibitory weight change (iLTD or iLTP) depends on the firing rate of the postsynaptic excitatory cells (see **Vogels et al., 2011**). Postsynaptic firing rates above a 'target firing rate' will on average lead to iLTP, while postsynaptic firing rates below the target rate will lead to iLTD. In turn, the average magnitude of inhibitory weight change depends on the firing rate of the presynaptic inhibitory neurons (see **Vogels et al., 2011**). Therefore, if the background activity between stimulus presentations in our model is very low, recovery from adaptation will only happen on a very slow timescale (as in **Figure 6—figure supplement 1**). However, if the activity between stimulus presentations is higher (either because of a higher background firing rate or because of evoked activity from other sources, for example other stimuli), the adapted stimulus can recover faster (as in **Figure 5**). Therefore, we conclude that our model can capture the reduced adaptation for longer inter-stimulus intervals as found in experiments (**Ulanovsky et al., 2004**; **Cohen-Kashi Malina et al., 2013**) when background activity in the inter-stimulus interval is elevated.

Alternative mechanisms can account for adapted and novelty responses

Undoubtedly, mechanisms other than inhibitory plasticity might underlie the difference in network response to repeated and novel stimuli. These mechanisms can be roughly summarized in two groups: mechanisms which are unspecific, and mechanisms which are specific to the stimulus. Two examples of unspecific mechanisms are intrinsic plasticity and an adaptive current. Intrinsic plasticity is a form of activity-dependent plasticity, adjusting the neurons' intrinsic excitability (**Debanne et al., 2019**) and has been suggested to explain certain adaptive phenomena (**Levakova et al., 2019**). Other models at the single neuron level incorporate an additional current variable, the adaptive current, which increases for each postsynaptic spike and decreases otherwise. This adaptive current leads to a reduction of the neuron's membrane potential after a spike (**Brette and Gerstner, 2005**). However, any unspecific mechanism can only account for firing-rate adaptation but not for SSA (**Nelken, 2014**; **Figure 6C**). Examples of stimulus-specific mechanisms are short-term plasticity and long-term plasticity of excitatory synapses. Excitatory short-term depression, usually of thalamocortical synapses, is the most widely hypothesized mechanism to underlie adaptive phenomena in cortex (**Nelken, 2014**).

Short-term plasticity (**Abbott, 1997**; **Tsodyks et al., 1998**) has been implicated in a number of adaptation phenomena in different sensory cortices and contexts. One example is an already established model to explain SSA, namely the 'Adaptation of Narrowly Tuned Modules' (ANTM) model (**Nelken, 2014**; **Khouri and Nelken, 2015**). This model has been extensively studied in the context of adaptation to tone frequencies (**Mill et al., 2011a**; **Taaseh et al., 2011**; **Mill et al., 2012**; **Hershenhoren et al., 2014**). Models based on short-term plasticity have also been extended to recurrent networks (**Yarden and Nelken, 2017**) and multiple inhibitory sub-populations (**Park and Geffen, 2020**). Experimental work has shown that short-term plasticity can be different at the synapses from PV and SOM interneurons onto pyramidal neurons, and can generate diverse temporal responses (facilitated, depressed and stable responses) in pyramidal neurons in the auditory cortex

(*Seay et al., 2020*). Short-term plasticity can also capture the differences in responses to periodic versus random presentation of repeated stimuli in a sequence (*Yaron et al., 2012*; *Chait, 2020*). Finally, short-term plasticity has been suggested to explain a prominent phenomenon in the auditory cortex, named 'forward masking' (*Brosch and Schreiner, 1997*), in which a preceding masker stimulus influences the response to a following stimulus (*Phillips et al., 2017*). This highlights short-term plasticity as a key player in adaptive processes in the different sensory cortices, although it likely works in tandem with long-term plasticity.

Timescales of plasticity mechanisms

The crucial parameter for the generation of adaptation based on short-term plasticity is the timescale of the short-term plasticity mechanism. Experimental studies find adaptation timescales from hundreds of milliseconds to tens of seconds (*Ulanovsky et al., 2004*; *Lundstrom et al., 2010*; *Homann et al., 2017*; *Latimer et al., 2019*), and in the case of habituation even multiple days (*Haak et al., 2014*; *Ramaswami, 2014*). At the same time, the timescales of short-term plasticity can range from milliseconds to minutes (*Zucker and Regehr, 2002*). Hence, explaining the different timescales of adaptive phenomena would likely require a short-term plasticity timescale that can be dynamically adjusted. Our work shows that inhibitory plasticity can readily lead to adaptation on multiple timescales without the need for any additional assumptions (*Figure 4*). However, it is unclear whether inhibitory plasticity can act sufficiently fast to explain adaptation phenomena on the timescale of seconds, as in our model (*Figure 4C,D*). Most computational models of recurrent networks with plastic connections rely on fast inhibitory plasticity to stabilize excitatory rate dynamics (*Sprekeler, 2017*; *Zenke et al., 2017*). Decreasing the learning rate of inhibitory plasticity five-fold eliminates the adaptation to repeated stimuli and the novelty response in our model (*Figure 4—figure supplement 2*). Experimentally, during the induction of inhibitory plasticity, spikes are paired for several minutes and it takes several tens of minutes to reach a new stable baseline of inhibitory synaptic strength (*D'amour and Froemke, 2015*; *Field et al., 2020*). Nonetheless, inhibitory postsynaptic currents increase significantly immediately after the induction of plasticity (see e.g. *D'amour and Froemke, 2015*; *Field et al., 2020*). This suggests that changes of inhibitory synaptic strength already occur while the plasticity induction protocol is still ongoing. Hence, we propose that inhibitory long-term plasticity is a suitable, though not the only, candidate to explain the generation of novelty responses and adaptive phenomena over multiple timescales.

Robustness of the model

We probed our findings against key parameters and assumptions in our model. First, we tested if the specific choice of pretraining parameters and complexity of presented stimuli affects the generation of adapted and novelty responses. Varying the pretraining duration and the number of pretraining stimuli did not qualitatively change the novelty response and its properties (*Figure 4—figure supplement 1*). In addition, presenting different stimuli in the stimulation phase compared to the pretraining phase (Materials and methods) to mimic the scenario of randomly oriented Gabor patches in *Homann et al., 2017*, preserved the adaptation to repeated stimuli and the generation of a novelty response (*Figure 1—figure supplement 3*).

Second, we explored how the timescale of inhibitory plasticity and of the normalization mechanism affects the generation of adapted and novelty responses. In many computational models, normalization mechanisms are often justified by experimentally observed synaptic scaling. In our model, like in most computational work, the timescale of this normalization was much faster than synaptic scaling (*Zenke et al., 2017*). However, slowing normalization down did not affect the generation of adapted and novelty responses (*Figure 1—figure supplement 5*). Since the change in inhibitory synaptic weights through iSTDP is the key mechanism behind the generation of adapted and novelty responses, the speed of normalization was not crucial as it only affected the excitatory and not the inhibitory weights. In contrast, we found that the learning rate of inhibitory plasticity needs to be 'sufficiently fast' . Slow inhibitory plasticity failed to homeostatically stabilize firing rates in the network. Hence, the network no longer showed an adapted response to repeated stimuli and novelty responses became indiscernible from noise (*Figure 4—figure supplement 2*).

Disinhibition as a mechanism for novelty response amplification

Upon including a top-down disinhibitory signal in our model network, we observed: (1) an active amplification of the novelty response (***Figure 7B***); (2) a dense novelty response (***Figure 7C***), similar to experimental findings (***Homann et al., 2017***) (without a disinhibitory signal, the novelty response was not dense, see ***Figure 1—figure supplement 4***); and (3) a flexible manipulation of neuronal responses through a change in the disinhibitory strength (***Figure 7D,E***).

In our model, we were agnostic to the mechanism that generates disinhibition. However, at least two possibilities exist in which the inhibitory population can be regulated by higher-order feedback to allow for disinhibition. First, inhibitory neurons in primary sensory areas can be shaped by diverse neuromodulatory signals, which allow for subtype-specific targeting of inhibitory neurons (***Froemke, 2015***). Second, higher order feedback onto layer 1 inhibitory cells could mediate the behavioral relevance of the adapted stimuli through a disinhibitory pathway (***Letzkus et al., 2011***; ***Wang and Yang, 2018***). Hence, experiments that induce disinhibition either via local mechanisms within the same cortical layer or through higher cortical feedback can provide a test for our postulated role for disinhibition.

In our model, the disinhibitory signal was activated instantaneously. If such additional feedback signals do indeed exist in the brain that signal the detection of higher-order sequence violations, we expect them to arise with a certain delay. Carefully exploring if the dense responses arise with a temporal delay accounting for higher-order processing and projection back to primary sensory areas might shed light on distributed computations upon novel stimuli. These experiments would probably require recording methods on a finer temporal scale than calcium imaging.

Experimental data which points towards a flexible modulation of novelty and adapted responses already exists. The active amplification of novelty responses generated by our model is consistent with some experimental data (***Taaseh et al., 2011***; ***Hershenhoren et al., 2014***; ***Hamm and Yuste, 2016***; ***Harms et al., 2016***), but see also ***Vinken et al., 2017***. Giving a behavioral meaning to a sound through fear conditioning has been shown to modify SSA (***Yaron et al., 2020***). Similarly, contrast adaptation has been shown to reverse when visual stimuli become behaviorally relevant (***Keller et al., 2017***). Other studies have also shown that as soon as a stimulus becomes behaviorally relevant, inhibitory neurons decrease their response and therefore disinhibit adapted excitatory neurons (***Kato et al., 2015***; ***Makino and Komiyama, 2015***; ***Hattori et al., 2017***). Attention might lead to activation of the disinhibitory pathway, allowing for a change in the novelty response compared to the unattended case, as suggested in MMN studies (***Sussman et al., 2014***). Especially in habituation, the idea that a change in context can assign significance to a stimulus and therefore block habituation, leading to 'dehabituation', is widely accepted (***Ramaswami, 2014***; ***Barron et al., 2017***).

Hence, we suggest that disinhibition is a flexible mechanism to control several aspects of novelty responses, including the density of the response, which might be computationally important in signaling change detection to downstream areas (***Homann et al., 2017***). Altogether, our results suggest that disinhibition is capable of accounting for various aspects of novelty responses that cannot be accounted for by bottom-up computations. The functional purpose of a dense response to novel stimuli are yet to be explored.

Functional implications of adapted and novelty responses

In theoretical terms, our model is an attractor network. It differs from classic attractor models where inhibition is considered unspecific (like a 'blanket') (***Amit and Brunel, 1997***). Computational work is starting to uncover the functional role of specific inhibition in static networks (***Rost et al., 2018***; ***Najafi et al., 2020***; ***Rostami et al., 2020***) as well as the plasticity mechanisms that allow for specific connectivity to emerge (***Mackwood et al., 2021***). These studies have argued that inhibitory assemblies can improve the robustness of attractor dynamics (***Rost et al., 2018***) and keep a local balance of excitation and inhibition (***Rostami et al., 2020***). We showed that specific inhibitory connections readily follow from a tuned inhibitory population (***Figure 1A***, ***Figure 1—figure supplement 2***). Our results suggest that adaptation is linked to a stimulus-specific excitatory/inhibitory (E/I) balance. Presenting a novel stimulus leads to a short-term disruption of the E/I balance, triggering inhibitory plasticity, which aims to restore the E/I balance (***Figure 4***; ***Vogels et al., 2011***; ***D'amour and Froemke, 2015***; ***Field et al., 2020***). Disinhibition, which effectively disrupts the E/I balance, allows for flexible control of adapted and novelty responses (***Figure 7***). This links to the notion of

disinhibition as a gating mechanism for learning and plasticity (**Froemke et al., 2007**; **Letzkus et al., 2011**; **Kuhlman et al., 2013**).

A multitude of functional implications have been suggested for the role of adaptation (**Weber et al., 2019**; **Snow et al., 2017**). We showed that one of these roles, the detection of unexpected (or novel) events, follows from the lack of selective adaptation to those events. A second, highly considered functional implication is predictive coding. In the predictive coding framework, the brain is viewed as an inference or a prediction machine. It is thought to generate internal models of the world which are compared to the incoming sensory inputs (**Bastos et al., 2012**; **Clark, 2013**; **Friston, 2018**). According to predictive coding, the overall goal of our brain is to minimize the prediction error, that is the difference between the internal prediction and the sensory input (**Rao and Ballard, 1999**; **Clark, 2013**; **Friston, 2018**). Most predictive coding schemes hypothesize the existence of two populations of neurons. First, prediction error units that signal a mismatch between the internal model prediction and the incoming sensory stimuli. And second, a prediction population unit that reflects what the respective layer 'knows about the world' (**Rao and Ballard, 1999**; **Clark, 2013**; **Spratling, 2017**). Our model suggests that primary sensory areas allow for bottom-up detection of stimulus changes without the need for an explicit population of error neurons or an internal model of the world. However, one could also interpret the state of all inhibitory synaptic weights as an implicit internal model of the recent frequency of various events in the environment.

Predictions and outlook

Our approach to mechanistically understand the generation of adapted and novelty responses leads to several testable predictions. First, the most general implication from our study is that inhibitory plasticity might serve as an essential mechanism underlying many adaptive phenomena. Our work suggests that inhibitory plasticity allows for adaptation on multiple timescales, ranging from the adaptation to sequence blocks on the timescale of seconds to slower adaptation on the timescale of minutes, corresponding to repeating multiple sequence blocks (**Figure 4C,D**). A second prediction follows from the finding that both excitatory and inhibitory neuron populations show adaptive behavior and novelty responses (**Figure 3B,C**). Adaptation of inhibitory neurons on the single-cell level has already been verified experimentally (**Chen et al., 2015**; **Natan et al., 2015**). Third, we further predict that a violation of the sequence order does not lead to a novelty response. Therefore, the novelty response should not be interpreted as signaling a violation of the exact sequence structure (**Figure 3D,E**). However, previous work has found a reduction in the response to repeated stimuli if the stimuli are presented periodically, rather than randomly, in a sequence (**Yaron et al., 2012**) (but see **Mehra et al., 2021**). Fourth, the height of the novelty peak in the population average depends on the input drive, where decreasing the input strength decreases the novelty response (**Figure 3F**). This could be tested, for example, in the visual system, by presenting visual stimuli with different contrasts.

In our modeling approach, we did not distinguish between different subtypes of inhibitory neurons. This assumption is certainly an oversimplification. The main types of inhibitory neurons, parvalbumin-positive (PV), somatostatin-positive (SOM), and vasoactive intestinal peptide (VIP) expressing neurons, differ in their connectivity and their hypothesized functional roles (**Tremblay et al., 2016**). This is certainly also true for adaptation, and computational studies have already started to tackle this problem (**Park and Geffen, 2020**; **Seay et al., 2020**). Studies of the influence of inhibitory neurons on adaptation have shown that different interneuron types have unique contributions to adaptation (**Kato et al., 2015**; **Natan et al., 2015**; **Hamm and Yuste, 2016**; **Natan et al., 2017**; **Garrett et al., 2020**; **Heintz et al., 2020**). It would be interesting to explore the combination of microcircuit connectivity of excitatory neurons, PVs, SOMs, and VIPs with subtype-specific short-term (**Seay et al., 2020**; **Phillips et al., 2017**) and long-term inhibitory plasticity mechanisms (**Agnes et al., 2020**) on the generation and properties of novelty responses.

In sum, we have proposed a mechanistic model for the emergence of adapted and novelty responses based on inhibitory plasticity, and the regulation of this novelty response by top-down signals. Our findings offer insight into the flexible and adaptive responses of animals in constantly changing environments, and could be further relevant for disorders like schizophrenia where adapted responses are perturbed (**Hamm et al., 2017**).

Materials and methods

We built a biologically plausible spiking neuronal network model of the mammalian cortex based on recent experimental findings on tuning, connectivity, and synaptic plasticity. The model consists of 4000 excitatory exponential integrate-and-fire (EIF) neurons and 1000 inhibitory leaky integrate-and-fire (LIF) neurons (**Table 1**). Excitatory (E) and inhibitory (I) neurons were randomly recurrently connected (**Table 2**). Excitatory-to-excitatory and inhibitory-to-excitatory connections were plastic (see below). In addition, excitatory-to-excitatory weight dynamics were stabilized by a homeostatic mechanism (**Fiete et al., 2010**), which preserved the total sum of all incoming synaptic weights into an excitatory neuron. All other synapses in the network were fixed. Both excitatory and inhibitory neurons received an excitatory baseline feedforward input in the form of Poisson spikes. Furthermore, different subsets of excitatory and inhibitory neurons received excess input with elevated Poisson rate to model the presentation of stimuli (see below, **Figure 1A**, left; Table 4).

Dynamics of synaptic conductances and the membrane potential

The membrane dynamics of each excitatory neuron was modeled as an exponential integrate-and-fire (EIF) neuron model (**Fourcaud-Trocmé et al., 2003**):

$$C\frac{d}{dt}V(t) = -g_L(V(t)-V^E_{\rm rest}) + g_L\Delta_T\exp\left(\frac{V(t)-V_T}{\Delta_T}\right) - g^{EE}(t)(V(t)-V^E_{\rm rev}) - g^{EI}(t)(V(t)-V^I_{\rm rev}), \quad (1)$$

where $V(t)$ is the membrane potential of the modeled neuron, C the membrane capacitance, g_L the membrane conductance, and Δ_T is the slope factor of the exponential rise. The membrane potential was reset to $V_{\rm reset}$ once the diverging potential reached the threshold peak voltage $V_{\rm peak}$. Inhibitory neurons were modeled via a leaky-integrate-and-fire neuron model

$$C\frac{d}{dt}V(t) = -g_L(V(t)-V^I_{\rm rest}) - g^{IE}(t)(V(t)-V^E_{\rm rev}) - g^{II}(t)(V(t)-V^I_{\rm rev}). \quad (2)$$

Once the membrane potential reached the threshold voltage $V_{\rm thr}$, the membrane potential was reset to $V_{\rm reset}$. The absolute refractory period was modeled by clamping the membrane voltage of a neuron that just spiked to the reset voltage $V_{\rm reset}$ for the duration $\tau_{\rm abs}$. In this study, we did not model additional forms of adaptation, such as adaptive currents or spiking threshold $V_{\rm T}$ adaptation. To avoid extensive parameter tuning, we used previously published parameter values (**Litwin-Kumar and Doiron, 2014**; **Table 1**).

Table 1. Parameters for the excitatory (EIF) and inhibitory (LIF) membrane dynamics (**Litwin-Kumar and Doiron, 2014**).

Symbol	Description	Value
N^E	Number of E neurons	4000
N^I	Number of I neurons	1000
τ^E, τ^I	E, I neuron resting membrane time constant	20 ms
$V^E_{\rm rest}$	E neuron resting potential	- 70 mV
$V^I_{\rm rest}$	I neuron resting potential	- 62 mV
Δ_T	Slope factor of exponential	2 mV
C	Membrane capacitance	300 pF
g_L	Membrane conductance	C/τ^E
$V^E_{\rm rev}$	E reversal potential	0 mV
$V^I_{\rm rev}$	I reversal potential	- 75 mV
$V_{\rm thr}$	Threshold potential	- 52 mV
$V_{\rm peak}$	Peak threshold potential	20 mV
$V_{\rm reset}$	E, I neuron reset potential	- 60 mV
$\tau_{\rm abs}$	E, I absolute refractory period	1 ms

Table 2. Parameters for feedforward and recurrent connections (***Litwin-Kumar and Doiron, 2014***).

Symbol	Description	Value
p	Connection probability	0.2
τ^{E}_{rise}	Rise time for E synapses	1 ms
τ^{E}_{decay}	Decay time for E synapses	6 ms
τ^{I}_{rise}	Rise time for I synapses	0.5 ms
τ^{I}_{decay}	Decay time for I synapses	2 ms
$\bar{r}^{EE}_{\text{ext}}$	Avg. rate of external input to E neurons	4.5 kHz
$\bar{r}^{IE}_{\text{ext}}$	Avg. rate of external input to I neurons	2.25 kHz
J^{EE}_{min}	Minimum E to E synaptic weight	1.78 pF
J^{EE}_{max}	Maximum E to E synaptic weight	21.4 pF
J^{EE}_{0}	Initial E to E synaptic weight	2.76 pF
J^{EI}_{min}	Minimum I to E synaptic weight	48.7 pF
J^{EI}_{max}	Maximum I to E synaptic weight	243 pF
J^{EI}_{0}	Initial I to E synaptic weight	48.7 pF
J^{IE}	Synaptic weight from E to I	1.27 pF
J^{II}	Synaptic weight from I to I	16.2 pF
J^{EEx}	Synaptic weight from external input population to E	1.78 pF
J^{IEx}	Synaptic weight from external input population to I	1.27 pF

We compared this model to one where we froze plasticity and included adaptive currents w_{adapt} (***Figure 6C***, top). We modeled this by subtracting $w_{adapt}(t)$ on the right hand side of ***Equation 1*** (***Brette and Gerstner, 2005***). Upon a spike, $w_{adapt}(t)$ increased by b_w and the sub-threshold dynamics of the adaptive current were described by $\tau_w \frac{d}{dt} w_{\text{adapt}}(t) = -w_{\text{adapt}}(t) + a_w(V(t) - V^{E}_{\text{rest}})$, where a_w = 4 nS denotes the subthreshold and b_w = 80.5 pA the spike-triggered adaptation. The adaptation time scale was set to τ_w = 150 ms.

The conductance of neuron i which is part of population X and is targeted by another neuron in population Y was denoted with g^{XY}_i. Both X and Y could refer either to the excitatory or inhibitory population, that is $X, Y \in [E, I]$. The shape of the synaptic kernels $F(t)$ was a difference of exponentials and differed for excitatory and inhibitory input depending on the rise and decay times τ^{Y}_{decay} and τ^{Y}_{rise}:

$$F^{Y}(t) = \frac{e^{\frac{-t}{\tau^{Y}_{\text{decay}}}} - e^{\frac{-t}{\tau^{Y}_{\text{rise}}}}}{\tau^{Y}_{\text{decay}} - \tau^{Y}_{\text{rise}}}. \tag{3}$$

This kernel was convolved with the total inputs to neuron i weighted with the respective synaptic strength to yield the total conductance

$$g^{XY}_i(t) = F^{Y}(t) * \left(J^{XY}_{ext} s^{XY}_{i,ext}(t) + \sum_j J^{XY}_{ij} s^{Y}_j(t) \right), \tag{4}$$

where $s^{Y}_j(t)$ is the spike train of neuron j in the network and $s^{XY}_{i,ext}$ denotes the spike train of the external input to neuron i. The external spike trains were generated in an independent homogeneous Poisson process. The synaptic strength from the input neurons to the network neurons, J^{XY}_{ext}, was assumed to be constant.

Excitatory and inhibitory plasticity

We implemented the plasticity from an excitatory to an excitatory neuron J^{EE} based on the triplet spike-time-dependent plasticity rule (triplet STDP), which uses triplets of pre- and postsynaptic

spikes to evoke synaptic change (**Sjöström et al., 2001**; **Pfister and Gerstner, 2006**). The addition of a third spike for the induction of synaptic plasticity modifies the amount of potentiation and depression induced by the classical pair-based STDP, where pairs of pre- and postsynaptic spikes induce plasticity based on their timing and order (**Bi and Poo, 1998**). The triplet eSTDP rule has been shown to capture the dependency of plasticity on firing rates found experimentally, whereby a high frequency of pre- and postsynaptic spike pairs leads to potentiation rather than no synaptic change as predicted by pair-based STDP (**Sjöström et al., 2001**; **Pfister and Gerstner, 2006**; **Gjorgjieva et al., 2011**; **Table 3**). In the triplet rule, four spike accumulators, $r_1, r_2, o_1,$ and o_2, increase by one, once a spike of the corresponding neuron occurs and otherwise decrease exponentially depending on their respective time constant $\tau_+, \tau_x, \tau_-,$ and τ_y:

$$\begin{aligned}
\frac{dr_1(t)}{dt} &= -\frac{r_1(t)}{\tau_+} \text{ if } t = t^{\text{pre}} \text{ then } r_1 \to r_1 + 1,\\
\frac{dr_2(t)}{dt} &= -\frac{r_2(t)}{\tau_x} \text{ if } t = t^{\text{pre}} \text{ then } r_2 \to r_2 + 1,\\
\frac{do_1(t)}{dt} &= -\frac{o_1(t)}{\tau_-} \text{ if } t = t^{\text{post}} \text{ then } o_1 \to o_1 + 1,\\
\frac{do_2(t)}{dt} &= -\frac{o_2(t)}{\tau_y} \text{ if } t = t^{\text{post}} \text{ then } o_2 \to o_2 + 1.
\end{aligned} \tag{5}$$

The E-to-E weights were updated as

$$\begin{aligned}
\Delta J^{EE}(t) &= -o_1(t)[A_2^- + A_3^- r_2(t-\epsilon)] \text{ if } t = t^{\text{pre}},\\
\Delta J^{EE}(t) &= r_1(t)[A_2^+ + A_3^+ o_2(t-\epsilon)] \text{ if } t = t^{\text{post}},
\end{aligned} \tag{6}$$

where the A^+,A^- corresponds to the excitatory LTP or LTD amplitude, and the subscript refers to the triplet (3) or pairwise term (2). The parameter ϵ>0 ensures that the weights are updated prior to increasing the respective spike accumulators by 1. Spike detection was modeled in an all-to-all approach.

The plasticity of inhibitory-to-excitatory connections, J^{EI}, was modeled based on a symmetric inhibitory pairwise STDP (iSTDP) rule, initially suggested on theoretical grounds for its ability to homeostatically stabilize firing rates in recurrent networks (**Vogels et al., 2011**). According to this rule, the timing but not the order of pre- and postsynaptic spikes matters for the induction of synaptic plasticity. Other inhibitory rules have also been measured experimentally, including classical Hebbian and anti-Hebbian (e.g. **Holmgren and Zilberter, 2001**; **Woodin et al., 2003**; **Haas et al., 2006**; for a review see **Hennequin et al., 2017**), and some may even depend on the type of the interneuron (**Udakis et al., 2020**). We chose the iSTDP rule because it can stabilize excitatory firing rate dynamics in recurrent networks (**Vogels et al., 2011**; **Litwin-Kumar and Doiron, 2014**) and was

Table 3. Parameters for the implementation of Hebbian and homeostatic plasticity (**Pfister and Gerstner, 2006**; **Litwin-Kumar and Doiron, 2014**).

Symbol	Description	Value
τ_-	Time constant of pairwise pre-synaptic detector (+)	33.7 ms
τ_+	Time constant of pairwise post-synaptic detector (-)	16.8 ms
τ_x	Time constant of triplet pre-synaptic detector (-)	101 ms
τ_y	Time constant of triplet post-synaptic detector (+)	125 ms
A_2^+	Pairwise potentiation amplitude	7.5×10^{-10} pF
A_3^+	Triplet potentiation amplitude	9.3×10^{-3} pF
A_2^-	Pairwise depression amplitude	7×10^{-3} pF
A_3^-	Triplet depression amplitude	2.3×10^{-4} pF
τ_y^{inhib}	Time constant of low-pass filtered spike train	20 ms
η	Inhibitory plasticity learning rate	1 pF
r_0	Target firing rate	3 Hz

recently verified to operate in the auditory cortex of mice (**D'amour and Froemke, 2015**). The plasticity parameters are shown in **Table 3**. The two spike accumulators $y^{E/I}$, for the inhibitory pre- and the excitatory post-synaptic neuron, have the same time constant τ_y^{inhib}. Their dynamics were described by

$$\begin{aligned}\frac{dy^I(t)}{dt} &= -\frac{y^I(t)}{\tau_y^{\text{inhib}}} \text{ if } t = t^{\text{pre/I}} \text{ then } y^I \to y^I + 1 \text{ and} \\ \frac{dy^E(t)}{dt} &= -\frac{y^E(t)}{\tau_y^{\text{inhib}}} \text{ if } t = t^{\text{post/E}} \text{ then } y^E \to y^E + 1.\end{aligned} \tag{7}$$

The I-to-E weights were updated as

$$\begin{aligned}\Delta J_{ij}^{EI}(t) &= \eta(y_i^E(t) - 2r_0\tau_y^{\text{inhib}}) \text{ if } t = t^{\text{pre/I}} \\ \Delta J_{ij}^{EI}(t) &= \eta y_j^I(t) \text{ if } t = t^{\text{post/E}},\end{aligned} \tag{8}$$

where η is the learning rate, and r_0 corresponds to the target firing rate of the excitatory neuron. In **Figure 4—figure supplement 2** we investigated the inhibitory learning rate η. **Figure 1—figure supplement 1** shows the excitatory and inhibitory STDP rules for different pairing frequencies.

Additional homeostatic mechanisms

Inhibitory plasticity alone is considered insufficient to prevent runaway activity in this network implementation. Hence, additional mechanisms were implemented that also have a homeostatic effect. To avoid unlimited weight increase, the synaptic weights were bound from below and from above, see **Table 2**. Subtractive normalization ensured that the total synaptic input to an excitatory neuron remains constant throughout the simulation. This was implemented by scaling all incoming weights to each neuron every $\Delta t = 20$ ms according to

$$\Delta J_{ij}^{EE}(t) = -\frac{\sum_j J_{ij}^{EE}(t) - \sum_j J_{ij}^{EE}(0)}{N_i^E}, \tag{9}$$

where i is the index of the post-synaptic and j of the pre-synaptic neurons. N_i^E is the number of excitatory connections onto neuron i (**Fiete et al., 2010**). In **Figure 1—figure supplement 5** we investigated the effect of the normalization timestep Δt on the novelty response.

Stimulation protocol

All neurons received external excitatory baseline input. The baseline input to excitatory neurons r_{ext}^E was higher than the input to inhibitory neurons r_{ext}^I (**Table 4**). An external input of $r_{\text{ext}}^E = 4.5$ kHz can be interpreted as 1000 external presynaptic neurons with average firing rates of 4.5 Hz (compare **Litwin-Kumar and Doiron, 2014**).

The stimulation paradigm was inspired by a recent study in the visual system (**Homann et al., 2017**). In **Homann et al., 2017**, the stimulation consisted of images with 100 randomly chosen, superimposed Gabor patches. Rather than explicitly modeling oriented and spatially localized Gabor patches, in our model, stimuli that correspond to Gabor patches of a given orientation were

Table 4. Parameters for the stimulation paradigm and stimulus tuning.

Symbol	Description	Value
r_{ext}^E	External baseline input to E	4.5 kHz
r_{ext}^I	External baseline input to I	2.25 kHz
r_{stim}^E	Additional input to E during stimulus presentation	12 kHz
r_{stim}^I	Additional input to I during stimulus presentation	1.2 kHz
r_{disinh}^I	Additional input to I during disinhibition	−1.5 kHz
p_{member}^E	Probability for an E neuron to be driven by a stimulus	5%
p_{member}^I	Probability for an I neuron to be driven by a stimulus	15%

implemented by simultaneously co-activating subsets of cells by strongly driving them. Hence, the model analog of the presentation of a sensory stimulus, in our experiments, is increased input to a subset of neurons. Every time a particular stimulus is presented again, the same set of neurons receives strong external stimulation, r^E_{stim} and r^I_{stim}. Therefore, while a stimulus in our stimulation paradigm is functionally similar to presenting Gabor patches with similar orientations, it does not represent the Gabor patches themselves.

We first implemented a pretraining phase. In this phase, we sequentially stimulated subsets of neurons that are driven by all stimuli (repeated and novel stimuli) eventually used in the stimulation phase. The stimuli were presented in random order, leading to a change in network connectivity that is only stimulus but not sequence-dependent (***Figure 4B***, first 100 s shown here for five repetitions of each stimulus). Hence, the pretraining phase is a phenomenological model of the development process to generate a structure in the network connections prior to the actual stimulation paradigm. This can be interpreted as imprinting a 'backbone' of orientation selective neurons, where cells which are selective to similar features (e.g. similar orientations) become strongly connected due to synaptic plasticity (as seen in experiments, see for e.g. ***Ko et al., 2011***; ***Ko et al., 2013***).

Next, we implemented a stimulation phase where we presented the same stimuli used during the pretraining phase according to the repeated sequence stimulation paradigm. To match the randomly oriented Gabor patches presented in ***Homann et al., 2017***, we also performed additional simulations where in the stimulation phase we activated different, randomly chosen, subsets of neurons (***Figure 1—figure supplement 3***) (note that there is some overlap with the imprinted orientation selective subsets).

In the standard repeated sequence stimulation paradigm (***Figure 3*** and ***Figure 4***), a total of 65 stimuli were presented (5 x 3 repeated + 5 x 10 novel stimuli) during pretraining. In ***Figure 4—figure supplement 1***, we tested if changes in the pretraining phase, such as a change in the number of repetitions of each stimulus or the total number of stimuli, affect our results.

The timescales of the experimental paradigm in ***Homann et al., 2017*** and the model paradigm were matched, that is the neurons tuned to a stimulus received additional input for 300 ms simulation time. Stimuli were presented without pauses in between, corresponding to continuous stimulus presentation without blank images (visual) or silence (auditory) between sequence blocks. ***Table 4*** lists the stimulus parameters.

In contrast to several previous plastic recurrent networks, we did not only consider the excitatory neurons to have stimulus tuning properties but included inhibitory tuning as well. The probability of an excitatory neuron to be driven by one particular stimulus was 5%, leading to roughly 200 neurons that responded specifically to this stimulus. We modeled inhibitory tuning to be both weaker and broader. The probability of an inhibitory neuron to be driven by one particular stimulus was 15%, leading to roughly 150 neurons that responded specifically to this stimulus. There was overlap in stimulus tuning, that is, one neuron could be driven by multiple stimuli. Given this broader tuning of inhibitory neurons compared to excitatory neurons, a single inhibitory neuron could strongly inhibit multiple excitatory neurons which were selective to different stimuli, effectively implementing lateral inhibition.

Stimulus tuning in both populations led to the formation of stimulus-specific excitatory assemblies due to synaptic plasticity, where the subsets of excitatory neurons receiving the same input developed strong connections among each other as noted above (***Figure 1—figure supplement 2C***) and found experimentally (***Ko et al., 2011***; ***Miller et al., 2014***; ***Lee et al., 2016***). The strong, bidirectional connectivity among similarly selective neurons in our model was a direct consequence of the triplet STDP rule (***Gjorgjieva et al., 2011***; ***Montangie et al., 2020***). Additionally, the connections from similarly tuned inhibitory to excitatory neurons also became stronger, as seen in experiments (***Lee et al., 2014***; ***Xue et al., 2014***; ***Znamenskiy et al., 2018***; ***Najafi et al., 2020***). The number of stimulus-specific assemblies varied depending on the stimulation paradigm and corresponded to the number of unique stimuli presented in a given paradigm. We did not impose topographic organization of these assemblies (for e.g. tonotopy in the auditory cortex) since it would not influence the generation of adapted and novelty responses, but increase model complexity. Such spatial organization could, however, be introduced by allowing the assemblies for neighboring stimuli to overlap.

Disinhibition in the model was implemented via additional inhibiting input to the inhibitory population r^I_{inhib}. This was modeled in a purely phenomenological way, and we are agnostic as to what causes the additional inhibition.

Simulation details

The simulations were performed using the Julia programming language. Further evaluation and plotting was done in Python. Euler integration was implemented using a time step of 0.1 ms. Code implementing our model and generating the stimulation protocols can be found here: https://github.com/comp-neural-circuits/novelty-via-inhibitory-plasticity (***Schulz, 2021***; copy archived at swh:1:rev:d368b14a2368925b290923c2c11411d7b7a40bd1).

Acknowledgements

AS, CM, and JG thank the Max Planck Society for funding and MJB thanks the NEI and the Princeton Accelerator Fund for funding. We thank members of the 'Computation in Neural Circuits' group for useful discussions and comments on the manuscript.

Additional information

Funding

Funder	Grant reference number	Author
Max-Planck-Gesellschaft	Research Group Award to JG	Auguste Schulz Christoph Miehl Julijana Gjorgjieva
NEI and Princeton Accelerator Fund		Michael J Berry

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Auguste Schulz, Christoph Miehl, Conceptualization, Resources, Software, Formal analysis, Investigation, Visualization, Methodology, Writing - original draft, Writing - review and editing; Michael J Berry II, Conceptualization, Methodology, Writing - review and editing; Julijana Gjorgjieva, Conceptualization, Resources, Supervision, Funding acquisition, Methodology, Writing - original draft, Writing - review and editing

Author ORCIDs

Auguste Schulz https://orcid.org/0000-0001-8616-3756
Christoph Miehl http://orcid.org/0000-0001-9094-2760
Michael J Berry II https://orcid.org/0000-0003-4133-7999
Julijana Gjorgjieva https://orcid.org/0000-0001-7118-4079

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.65309.sa1
Author response https://doi.org/10.7554/eLife.65309.sa2

Additional files

Supplementary files

- Transparent reporting form

Data availability

The code to reproduce the figures for this paper has been uploaded on GitHub and be accessed here: https://github.com/comp-neural-circuits/novelty-via-inhibitory-plasticity (copy archived at https://archive.softwareheritage.org/swh:1:rev:d368b14a2368925b290923c2c11411d7b7a40bd1).

References

Abbott LF. 1997. Synaptic depression and cortical gain control. *Science* **275**:221–224. DOI: https://doi.org/10.1126/science.275.5297.221

Agnes EJ, Luppi AI, Vogels TP. 2020. Complementary inhibitory weight profiles emerge from plasticity and allow flexible switching of receptive fields. *The Journal of Neuroscience* **40**:9634–9649. DOI: https://doi.org/10.1523/JNEUROSCI.0276-20.2020, PMID: 33168622

Amit DJ, Brunel N. 1997. Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex* **7**:237–252. DOI: https://doi.org/10.1093/cercor/7.3.237, PMID: 9143444

Barlow HB. 2013. Possible principles underlying the transformations of sensory messages. *Sensory Communication* **1**:216–234. DOI: https://doi.org/10.7551/mitpress/9780262518420.003.0013

Barron HC, Vogels TP, Behrens TE, Ramaswami M. 2017. Inhibitory engrams in perception and memory. *PNAS* **114**:6666–6674. DOI: https://doi.org/10.1073/pnas.1701812114, PMID: 28611219

Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ. 2012. Canonical microcircuits for predictive coding. *Neuron* **76**:695–711. DOI: https://doi.org/10.1016/j.neuron.2012.10.038, PMID: 23177956

Bi GQ, Poo MM. 1998. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of Neuroscience* **18**:10464–10472. PMID: 9852584

Brette R, Gerstner W. 2005. Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of Neurophysiology* **94**:3637–3642. DOI: https://doi.org/10.1152/jn.00686.2005, PMID: 16014787

Brosch M, Schreiner CE. 1997. Time course of forward masking tuning curves in cat primary auditory cortex. *Journal of Neurophysiology* **77**:923–943. DOI: https://doi.org/10.1152/jn.1997.77.2.923, PMID: 9065859

Chait M. 2020. How the brain discovers structure in sound sequences. *Acoustical Science and Technology* **41**:48–53. DOI: https://doi.org/10.1250/ast.41.48

Chen IW, Helmchen F, Lütcke H. 2015. Specific early and late Oddball-Evoked responses in excitatory and inhibitory neurons of mouse auditory cortex. *Journal of Neuroscience* **35**:12560–12573. DOI: https://doi.org/10.1523/JNEUROSCI.2240-15.2015, PMID: 26354921

Clark A. 2013. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* **36**:181–204. DOI: https://doi.org/10.1017/S0140525X12000477, PMID: 23663408

Cohen-Kashi Malina K, Jubran M, Katz Y, Lampl I. 2013. Imbalance between excitation and inhibition in the somatosensory cortex produces postadaptation facilitation. *Journal of Neuroscience* **33**:8463–8471. DOI: https://doi.org/10.1523/JNEUROSCI.4845-12.2013, PMID: 23658183

D'amour JA, Froemke RC. 2015. Inhibitory and excitatory spike-timing-dependent plasticity in the auditory cortex. *Neuron* **86**:514–528. DOI: https://doi.org/10.1016/j.neuron.2015.03.014, PMID: 25843405

Das S, Sadanandappa MK, Dervan A, Larkin A, Lee JA, Sudhakaran IP, Priya R, Heidari R, Holohan EE, Pimentel A, Gandhi A, Ito K, Sanyal S, Wang JW, Rodrigues V, Ramaswami M. 2011. Plasticity of local GABAergic interneurons drives olfactory habituation. *PNAS* **108**:E646–E654. DOI: https://doi.org/10.1073/pnas.1106411108, PMID: 21795607

Debanne D, Inglebert Y, Russier M. 2019. Plasticity of intrinsic neuronal excitability. *Current Opinion in Neurobiology* **54**:73–82. DOI: https://doi.org/10.1016/j.conb.2018.09.001, PMID: 30243042

Dhruv NT, Carandini M. 2014. Cascaded effects of spatial adaptation in the early visual system. *Neuron* **81**:529–535. DOI: https://doi.org/10.1016/j.neuron.2013.11.025, PMID: 24507190

Fairhall AL. 2014. Adaptation and natural stimulus statistics. In: Gazzaniga MS, Mangun GR (Eds). *The Cognitive Neurosciences*. 5th Edn. MIT Press. p. 283–294.

Farley BJ, Quirk MC, Doherty JJ, Christian EP. 2010. Stimulus-specific adaptation in auditory cortex is an NMDA-independent process distinct from the sensory novelty encoded by the mismatch negativity. *Journal of Neuroscience* **30**:16475–16484. DOI: https://doi.org/10.1523/JNEUROSCI.2793-10.2010, PMID: 21147987

Field RE, D'amour JA, Tremblay R, Miehl C, Rudy B, Gjorgjieva J, Froemke RC. 2020. Heterosynaptic plasticity determines the set point for cortical Excitatory-Inhibitory balance. *Neuron* **106**:842–854. DOI: https://doi.org/10.1016/j.neuron.2020.03.002, PMID: 32213321

Fiete IR, Senn W, Wang CZ, Hahnloser RH. 2010. Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. *Neuron* **65**:563–576. DOI: https://doi.org/10.1016/j.neuron.2010.02.003, PMID: 20188660

Fischer TM, Blazis DE, Priver NA, Carew TJ. 1997. Metaplasticity at identified inhibitory synapses in Aplysia. *Nature* **389**:860–865. DOI: https://doi.org/10.1038/39892, PMID: 9349819

Fourcaud-Trocmé N, Hansel D, van Vreeswijk C, Brunel N. 2003. How spike generation mechanisms determine the neuronal response to fluctuating inputs. *The Journal of Neuroscience* **23**:11628–11640. DOI: https://doi.org/10.1523/JNEUROSCI.23-37-11628.2003, PMID: 14684865

Friston K. 2018. Does predictive coding have a future? *Nature Neuroscience* **21**:1019–1021. DOI: https://doi.org/10.1038/s41593-018-0200-7, PMID: 30038278

Froemke RC, Merzenich MM, Schreiner CE. 2007. A synaptic memory trace for cortical receptive field plasticity. *Nature* **450**:425–429. DOI: https://doi.org/10.1038/nature06289, PMID: 18004384

Froemke RC. 2015. Plasticity of cortical excitatory-inhibitory balance. *Annual Review of Neuroscience* **38**:195–219. DOI: https://doi.org/10.1146/annurev-neuro-071714-034002, PMID: 25897875

Garrett M, Manavi S, Roll K, Ollerenshaw DR, Groblewski PA, Ponvert ND, Kiggins JT, Casal L, Mace K, Williford A, Leon A, Jia X, Ledochowitsch P, Buice MA, Wakeman W, Mihalas S, Olsen SR. 2020. Experience shapes

activity dynamics and stimulus coding of VIP inhibitory cells. *eLife* **9**:e50340. DOI: https://doi.org/10.7554/eLife.50340, PMID: 32101169

Geffen MN, de Vries SE, Meister M. 2007. Retinal ganglion cells can rapidly change polarity from off to on. *PLOS Biology* **5**:e65. DOI: https://doi.org/10.1371/journal.pbio.0050065, PMID: 17341132

Gjorgjieva J, Clopath C, Audet J, Pfister JP. 2011. A triplet spike-timing-dependent plasticity model generalizes the Bienenstock-Cooper-Munro rule to higher-order spatiotemporal correlations. *PNAS* **108**:19383–19388. DOI: https://doi.org/10.1073/pnas.1105933108, PMID: 22080608

Glanzman DL. 2011. Olfactory habituation: fresh insights from flies. *PNAS* **108**:14711–14712. DOI: https://doi.org/10.1073/pnas.1111230108, PMID: 21873206

Griffen TC, Maffei A. 2014. GABAergic synapses: their plasticity and role in sensory cortex. *Frontiers in Cellular Neuroscience* **8**:91. DOI: https://doi.org/10.3389/fncel.2014.00091, PMID: 24723851

Haak KV, Fast E, Bao M, Lee M, Engel SA. 2014. Four days of visual contrast deprivation reveals limits of neuronal adaptation. *Current Biology* **24**:2575–2579. DOI: https://doi.org/10.1016/j.cub.2014.09.027, PMID: 25438945

Haas JS, Nowotny T, Abarbanel HD. 2006. Spike-timing-dependent plasticity of inhibitory synapses in the entorhinal cortex. *Journal of Neurophysiology* **96**:3305–3313. DOI: https://doi.org/10.1152/jn.00551.2006, PMID: 16928795

Hamm JP, Peterka DS, Gogos JA, Yuste R. 2017. Altered cortical ensembles in mouse models of schizophrenia. *Neuron* **94**:153–167. DOI: https://doi.org/10.1016/j.neuron.2017.03.019, PMID: 28384469

Hamm JP, Yuste R. 2016. Somatostatin interneurons control a key component of mismatch negativity in mouse visual cortex. *Cell Reports* **16**:597–604. DOI: https://doi.org/10.1016/j.celrep.2016.06.037, PMID: 27396334

Harms L, Michie PT, Näätänen R. 2016. Criteria for determining whether mismatch responses exist in animal models: focus on rodents. *Biological Psychology* **116**:28–35. DOI: https://doi.org/10.1016/j.biopsycho.2015.07.006, PMID: 26196895

Hattori R, Kuchibhotla KV, Froemke RC, Komiyama T. 2017. Functions and dysfunctions of neocortical inhibitory neuron subtypes. *Nature Neuroscience* **20**:1199–1208. DOI: https://doi.org/10.1038/nn.4619, PMID: 28849791

Heintz TG, Hinojosa AJ, Lagnado L. 2020. Opposing forms of adaptation in mouse visual cortex are controlled by distinct inhibitory microcircuits and gated by locomotion. *bioRxiv*. DOI: https://doi.org/10.1101/2020.01.16.909788

Hennequin G, Agnes EJ, Vogels TP. 2017. Inhibitory plasticity: balance, control, and codependence. *Annual Review of Neuroscience* **40**:557–579. DOI: https://doi.org/10.1146/annurev-neuro-072116-031005, PMID: 28598717

Hershenhoren I, Taaseh N, Antunes FM, Nelken I. 2014. Intracellular correlates of stimulus-specific adaptation. *Journal of Neuroscience* **34**:3303–3319. DOI: https://doi.org/10.1523/JNEUROSCI.2166-13.2014, PMID: 24573289

Holmgren CD, Zilberter Y. 2001. Coincident spiking activity induces long-term changes in inhibition of neocortical pyramidal cells. *The Journal of Neuroscience* **21**:8270–8277. DOI: https://doi.org/10.1523/JNEUROSCI.21-20-08270.2001, PMID: 11588198

Homann J, Koay SA, Glidden AM, Tank DW, Berry II MJ. 2017. Predictive coding of novel versus familiar stimuli in the primary visual cortex. *bioRxiv*. DOI: https://doi.org/10.1101/197608

Kato HK, Gillet SN, Isaacson JS. 2015. Flexible sensory representations in auditory cortex driven by behavioral relevance. *Neuron* **88**:1027–1039. DOI: https://doi.org/10.1016/j.neuron.2015.10.024, PMID: 26586181

Keller GB, Bonhoeffer T, Hübener M. 2012. Sensorimotor mismatch signals in primary visual cortex of the behaving mouse. *Neuron* **74**:809–815. DOI: https://doi.org/10.1016/j.neuron.2012.03.040, PMID: 22681686

Keller AJ, Houlton R, Kampa BM, Lesica NA, Mrsic-Flogel TD, Keller GB, Helmchen F. 2017. Stimulus relevance modulates contrast adaptation in visual cortex. *eLife* **6**:e21589. DOI: https://doi.org/10.7554/eLife.21589, PMID: 28130922

Khouri L, Nelken I. 2015. Detecting the unexpected. *Current Opinion in Neurobiology* **35**:142–147. DOI: https://doi.org/10.1016/j.conb.2015.08.003, PMID: 26318534

King JL, Lowe MP, Stover KR, Wong AA, Crowder NA. 2016. Adaptive processes in thalamus and cortex revealed by silencing of primary visual cortex during contrast adaptation. *Current Biology* **26**:1295–1300. DOI: https://doi.org/10.1016/j.cub.2016.03.018, PMID: 27112300

Kleberg FI, Fukai T, Gilson M. 2014. Excitatory and inhibitory STDP jointly tune feedforward neural circuits to selectively propagate correlated spiking activity. *Frontiers in Computational Neuroscience* **8**:53. DOI: https://doi.org/10.3389/fncom.2014.00053, PMID: 24847242

Ko H, Hofer SB, Pichler B, Buchanan KA, Sjöström PJ, Mrsic-Flogel TD. 2011. Functional specificity of local synaptic connections in neocortical networks. *Nature* **473**:87–91. DOI: https://doi.org/10.1038/nature09880, PMID: 21478872

Ko H, Cossell L, Baragli C, Antolik J, Clopath C, Hofer SB, Mrsic-Flogel TD. 2013. The emergence of functional microcircuits in visual cortex. *Nature* **496**:96–100. DOI: https://doi.org/10.1038/nature12015, PMID: 23552948

Kuhlman SJ, Olivas ND, Tring E, Ikrar T, Xu X, Trachtenberg JT. 2013. A disinhibitory microcircuit initiates critical-period plasticity in the visual cortex. *Nature* **501**:543–546. DOI: https://doi.org/10.1038/nature12485, PMID: 23975100

Latimer KW, Barbera D, Sokoletsky M, Awwad B, Katz Y, Nelken I, Lampl I, Fairhall AL, Priebe NJ. 2019. Multiple timescales account for adaptive responses across sensory cortices. *The Journal of Neuroscience* **39**:10019–10033. DOI: https://doi.org/10.1523/JNEUROSCI.1642-19.2019, PMID: 31662427

Lee SH, Marchionni I, Bezaire M, Varga C, Danielson N, Lovett-Barron M, Losonczy A, Soltesz I. 2014. Parvalbumin-positive basket cells differentiate among hippocampal pyramidal cells. *Neuron* **82**:1129–1144. DOI: https://doi.org/10.1016/j.neuron.2014.03.034, PMID: 24836505

Lee WC, Bonin V, Reed M, Graham BJ, Hood G, Glattfelder K, Reid RC. 2016. Anatomy and function of an excitatory network in the visual cortex. *Nature* **532**:370–374. DOI: https://doi.org/10.1038/nature17192, PMID: 27018655

Letzkus JJ, Wolff SB, Meyer EM, Tovote P, Courtin J, Herry C, Lüthi A. 2011. A disinhibitory microcircuit for associative fear learning in the auditory cortex. *Nature* **480**:331–335. DOI: https://doi.org/10.1038/nature10674, PMID: 22158104

Levakova M, Kostal L, Monsempès C, Lucas P, Kobayashi R. 2019. Adaptive integrate-and-fire model reproduces the dynamics of olfactory receptor neuron responses in a moth. *Journal of the Royal Society Interface* **16**: 20190246. DOI: https://doi.org/10.1098/rsif.2019.0246

Litwin-Kumar A, Doiron B. 2014. Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nature Communications* **5**:5319. DOI: https://doi.org/10.1038/ncomms6319, PMID: 25395015

Lundstrom BN, Fairhall AL, Maravall M. 2010. Multiple timescale encoding of slowly varying whisker stimulus envelope in cortical and thalamic neurons in vivo. *Journal of Neuroscience* **30**:5071–5077. DOI: https://doi.org/10.1523/JNEUROSCI.2193-09.2010, PMID: 20371827

Luz Y, Shamir M. 2012. Balancing feed-forward excitation and inhibition via hebbian inhibitory synaptic plasticity. *PLOS Computational Biology* **8**:e1002334. DOI: https://doi.org/10.1371/journal.pcbi.1002334, PMID: 22291583

Ma WP, Liu BH, Li YT, Huang ZJ, Zhang LI, Tao HW. 2010. Visual representations by cortical somatostatin inhibitory neurons - Selective but with weak and delayed responses. *Journal of Neuroscience* **30**:14371–14379. DOI: https://doi.org/10.1523/JNEUROSCI.3248-10.2010, PMID: 20980594

Mackwood O, Naumann LB, Sprekeler H. 2021. Learning excitatory-inhibitory neuronal assemblies in recurrent networks. *eLife* **10**:e59715. DOI: https://doi.org/10.7554/eLife.59715, PMID: 33900199

Maffei A, Nataraj K, Nelson SB, Turrigiano GG. 2006. Potentiation of cortical inhibition by visual deprivation. *Nature* **443**:81–84. DOI: https://doi.org/10.1038/nature05079, PMID: 16929304

Makino H, Komiyama T. 2015. Learning enhances the relative impact of top-down processing in the visual cortex. *Nature Neuroscience* **18**:1116–1122. DOI: https://doi.org/10.1038/nn.4061, PMID: 26167904

Mehra M, Mukesh A, Bandyopadhyay S. 2021. Separate functional subnetworks of excitatory neurons show preference to periodic and random sound structures. *bioRxiv*. DOI: https://doi.org/10.1101/2021.02.13.431077

Mill R, Coath M, Wennekers T, Denham SL. 2011a. A neurocomputational model of stimulus-specific adaptation to oddball and Markov sequences. *PLOS Computational Biology* **7**:e1002117. DOI: https://doi.org/10.1371/journal.pcbi.1002117, PMID: 21876661

Mill R, Coath M, Wennekers T, Denham SL. 2011b. Abstract stimulus-specific adaptation models. *Neural Computation* **23**:435–476. DOI: https://doi.org/10.1162/NECO_a_00077, PMID: 21114400

Mill R, Coath M, Wennekers T, Denham SL. 2012. Characterising stimulus-specific adaptation using a multi-layer field model. *Brain Research* **1434**:178–188. DOI: https://doi.org/10.1016/j.brainres.2011.08.063, PMID: 21955728

Miller JE, Ayzenshtat I, Carrillo-Reid L, Yuste R. 2014. Visual stimuli recruit intrinsically generated cortical ensembles. *PNAS* **111**:E4053–E4061. DOI: https://doi.org/10.1073/pnas.1406077111, PMID: 25201983

Montangie L, Miehl C, Gjorgjieva J. 2020. Autonomous emergence of connectivity assemblies via spike triplet interactions. *PLOS Computational Biology* **16**:e1007835. DOI: https://doi.org/10.1371/journal.pcbi.1007835, PMID: 32384081

Movshon JA, Lennie P. 1979. Pattern-selective adaptation in visual cortical neurones. *Nature* **278**:850–852. DOI: https://doi.org/10.1038/278850a0, PMID: 440411

Näätänen R, Simpson M, Loveless NE. 1982. Stimulus deviance and evoked potentials. *Biological Psychology* **14**: 53–98. DOI: https://doi.org/10.1016/0301-0511(82)90017-5, PMID: 7104425

Näätänen R, Paavilainen P, Rinne T, Alho K. 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology* **118**:2544–2590. DOI: https://doi.org/10.1016/j.clinph.2007.04.026, PMID: 17931964

Najafi F, Elsayed GF, Cao R, Pnevmatikakis E, Latham PE, Cunningham JP, Churchland AK. 2020. Excitatory and inhibitory subnetworks are equally selective during Decision-Making and emerge simultaneously during learning. *Neuron* **105**:165–179. DOI: https://doi.org/10.1016/j.neuron.2019.09.045, PMID: 31753580

Natan RG, Briguglio JJ, Mwilambwe-Tshilobo L, Jones SI, Aizenberg M, Goldberg EM, Geffen MN. 2015. Complementary control of sensory adaptation by two types of cortical interneurons. *eLife* **4**:e09868. DOI: https://doi.org/10.7554/eLife.09868, PMID: 26460542

Natan RG, Rao W, Geffen MN. 2017. Cortical interneurons differentially shape frequency tuning following adaptation. *Cell Reports* **21**:878–890. DOI: https://doi.org/10.1016/j.celrep.2017.10.012, PMID: 29069595

Nelken I. 2014. Stimulus-specific adaptation and deviance detection in the auditory system: experiments and models. *Biological Cybernetics* **108**:655–663. DOI: https://doi.org/10.1007/s00422-014-0585-7, PMID: 24477619

Ohki K, Reid RC. 2007. Specificity and randomness in the visual cortex. *Current Opinion in Neurobiology* **17**: 401–407. DOI: https://doi.org/10.1016/j.conb.2007.07.007, PMID: 17720489

Park Y, Geffen MN. 2020. A circuit model of auditory cortex. *PLOS Computational Biology* **16**:e1008016. DOI: https://doi.org/10.1371/journal.pcbi.1008016, PMID: 32716912

Pfister JP, Gerstner W. 2006. Triplets of spikes in a model of spike timing-dependent plasticity. *Journal of Neuroscience* **26**:9673–9682. DOI: https://doi.org/10.1523/JNEUROSCI.1425-06.2006, PMID: 16988038

Phillips EAK, Schreiner CE, Hasenstaub AR. 2017. Cortical interneurons differentially regulate the effects of acoustic context. *Cell Reports* **20**:771–778. DOI: https://doi.org/10.1016/j.celrep.2017.07.001, PMID: 28746863
Ramaswami M. 2014. Network plasticity in adaptive filtering and behavioral habituation. *Neuron* **82**:1216–1229. DOI: https://doi.org/10.1016/j.neuron.2014.04.035, PMID: 24945768
Rao RP, Ballard DH. 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience* **2**:79–87. DOI: https://doi.org/10.1038/4580, PMID: 10195184
Ross JM, Hamm JP. 2020. Cortical microcircuit mechanisms of mismatch negativity and its underlying subcomponents. *Frontiers in Neural Circuits* **14**:13. DOI: https://doi.org/10.3389/fncir.2020.00013, PMID: 32296311
Rost T, Deger M, Nawrot MP. 2018. Winnerless competition in clustered balanced networks: inhibitory assemblies do the trick. *Biological Cybernetics* **112**:81–98. DOI: https://doi.org/10.1007/s00422-017-0737-7, PMID: 29075845
Rostami V, Rost T, Riehle A, Albada SJv, Nawrot MP. 2020. Spiking neural network model of motor cortex with joint excitatory and inhibitory clusters reflects task uncertainty, reaction times, and variability dynamics. *bioRxiv*. DOI: https://doi.org/10.1101/2020.02.27.968339
Schulz A. 2021. novelty-via-inhibitory-plasticity. *GitHub*. swh:1:rev:d368b14a2368925b290923c2c11411d7b7a40bd1. https://archive.softwareheritage.org/swh:1:dir:25354235d9002a4f0b922bf5226d49d3eec097e4;origin=https://github.com/comp-neural-circuits/novelty-via-inhibitory-plasticity;visit=swh:1:snp:002d92b4645ffdcbcc332fb6e21f7c9a09030095;anchor=swh:1:rev:d368b14a2368925b290923c2c11411d7b7a40bd1
Schwartz G, Harris R, Shrom D, Berry MJ. 2007. Detection and prediction of periodic patterns by the retina. *Nature Neuroscience* **10**:552–554. DOI: https://doi.org/10.1038/nn1887, PMID: 17450138
Schwartz G, Berry MJ. 2008. Sophisticated temporal pattern recognition in retinal ganglion cells. *Journal of Neurophysiology* **99**:1787–1798. DOI: https://doi.org/10.1152/jn.01025.2007, PMID: 18272878
Seay MJ, Natan RG, Geffen MN, Buonomano DV. 2020. Differential Short-Term plasticity of PV and SST neurons accounts for adaptation and facilitation of cortical neurons to auditory tones. *The Journal of Neuroscience* **40**:9224–9235. DOI: https://doi.org/10.1523/JNEUROSCI.0686-20.2020, PMID: 33097639
Simoncelli EP, Olshausen BA. 2001. Natural image statistics and neural representation. *Annual Review of Neuroscience* **24**:1193–1216. DOI: https://doi.org/10.1146/annurev.neuro.24.1.1193, PMID: 11520932
Sjöström PJ, Turrigiano GG, Nelson SB. 2001. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron* **32**:1149–1164. DOI: https://doi.org/10.1016/S0896-6273(01)00542-6, PMID: 11754844
Snow M, Coen-Cagli R, Schwartz O. 2017. Adaptation in the visual cortex: a case for probing neuronal populations with natural stimuli. *F1000Research* **6**:1246. DOI: https://doi.org/10.12688/f1000research.11154.1, PMID: 29034079
Spratling MW. 2017. A review of predictive coding algorithms. *Brain and Cognition* **112**:92–97. DOI: https://doi.org/10.1016/j.bandc.2015.11.003, PMID: 26809759
Sprekeler H. 2017. Functional consequences of inhibitory plasticity: homeostasis, the excitation-inhibition balance and beyond. *Current Opinion in Neurobiology* **43**:198–203. DOI: https://doi.org/10.1016/j.conb.2017.03.014, PMID: 28500933
Sussman ES, Chen S, Sussman-Fort J, Dinces E. 2014. The five myths of MMN: redefining how to use MMN in basic and clinical research. *Brain Topography* **27**:553–564. DOI: https://doi.org/10.1007/s10548-013-0326-6, PMID: 24158725
Taaseh N, Yaron A, Nelken I. 2011. Stimulus-specific adaptation and deviance detection in the rat auditory cortex. *PLOS ONE* **6**:e23369. DOI: https://doi.org/10.1371/journal.pone.0023369, PMID: 21853120
Thompson A, Gribizis A, Chen C, Crair MC. 2017. Activity-dependent development of visual receptive fields. *Current Opinion in Neurobiology* **42**:136–143. DOI: https://doi.org/10.1016/j.conb.2016.12.007, PMID: 28088066
Tikhonravov D, Neuvonen T, Pertovaara A, Savioja K, Ruusuvirta T, Näätänen R, Carlson S. 2008. Effects of an NMDA-receptor antagonist MK-801 on an MMN-like response recorded in anesthetized rats. *Brain Research* **1203**:97–102. DOI: https://doi.org/10.1016/j.brainres.2008.02.006, PMID: 18325485
Tremblay R, Lee S, Rudy B. 2016. GABAergic interneurons in the neocortex: from cellular properties to circuits. *Neuron* **91**:260–292. DOI: https://doi.org/10.1016/j.neuron.2016.06.033, PMID: 27477017
Tsodyks M, Pawelzik K, Markram H. 1998. Neural networks with dynamic synapses. *Neural Computation* **10**:821–835. DOI: https://doi.org/10.1162/089976698300017502, PMID: 9573407
Udakis M, Pedrosa V, Chamberlain SEL, Clopath C, Mellor JR. 2020. Interneuron-specific plasticity at Parvalbumin and somatostatin inhibitory synapses onto CA1 pyramidal neurons shapes hippocampal output. *Nature Communications* **11**:4395. DOI: https://doi.org/10.1038/s41467-020-18074-8, PMID: 32879322
Ulanovsky N, Las L, Nelken I. 2003. Processing of low-probability sounds by cortical neurons. *Nature Neuroscience* **6**:391–398. DOI: https://doi.org/10.1038/nn1032, PMID: 12652303
Ulanovsky N, Las L, Farkas D, Nelken I. 2004. Multiple time scales of adaptation in auditory cortex neurons. *Journal of Neuroscience* **24**:10440–10453. DOI: https://doi.org/10.1523/JNEUROSCI.1905-04.2004, PMID: 15548659
Vinken K, Vogels R, Op de Beeck H. 2017. Recent visual experience shapes visual processing in rats through Stimulus-Specific adaptation and response enhancement. *Current Biology* **27**:914–919. DOI: https://doi.org/10.1016/j.cub.2017.02.024, PMID: 28262485

Vogels TP, Sprekeler H, Zenke F, Clopath C, Gerstner W. 2011. Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science* **334**:1569–1573. DOI: https://doi.org/10.1126/science.1211095, PMID: 22075724

Wang L, Maffei A. 2014. Inhibitory plasticity dictates the sign of plasticity at excitatory synapses. *Journal of Neuroscience* **34**:1083–1093. DOI: https://doi.org/10.1523/JNEUROSCI.4711-13.2014, PMID: 24453301

Wang XJ, Yang GR. 2018. A disinhibitory circuit motif and flexible information routing in the brain. *Current Opinion in Neurobiology* **49**:75–83. DOI: https://doi.org/10.1016/j.conb.2018.01.002, PMID: 29414069

Weber AI, Krishnamurthy K, Fairhall AL. 2019. Coding principles in adaptation. *Annual Review of Vision Science* **5**:427–449. DOI: https://doi.org/10.1146/annurev-vision-091718-014818, PMID: 31283447

Woodin MA, Ganguly K, Poo MM. 2003. Coincident pre- and postsynaptic activity modifies GABAergic synapses by postsynaptic changes in Cl- transporter activity. *Neuron* **39**:807–820. DOI: https://doi.org/10.1016/S0896-6273(03)00507-5, PMID: 12948447

Wu YK, Hengen KB, Turrigiano GG, Gjorgjieva J. 2020. Homeostatic mechanisms regulate distinct aspects of cortical circuit dynamics. *PNAS* **117**:24514–24525. DOI: https://doi.org/10.1073/pnas.1918368117, PMID: 32917810

Xue M, Atallah BV, Scanziani M. 2014. Equalizing excitation-inhibition ratios across visual cortical neurons. *Nature* **511**:596–600. DOI: https://doi.org/10.1038/nature13321, PMID: 25043046

Yarden TS, Nelken I. 2017. Stimulus-specific adaptation in a recurrent network model of primary auditory cortex. *PLOS Computational Biology* **13**:e1005437. DOI: https://doi.org/10.1371/journal.pcbi.1005437, PMID: 28288158

Yaron A, Hershenhoren I, Nelken I. 2012. Sensitivity to complex statistical regularities in rat auditory cortex. *Neuron* **76**:603–615. DOI: https://doi.org/10.1016/j.neuron.2012.08.025, PMID: 23141071

Yaron A, Jankowski MM, Badrieh R, Nelken I. 2020. Stimulus-specific adaptation to behaviorally-relevant sounds in awake rats. *PLOS ONE* **15**:e0221541. DOI: https://doi.org/10.1371/journal.pone.0221541, PMID: 32210448

Zenke F, Agnes EJ, Gerstner W. 2015. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature Communications* **6**:6922. DOI: https://doi.org/10.1038/ncomms7922, PMID: 25897632

Zenke F, Gerstner W, Ganguli S. 2017. The temporal paradox of hebbian learning and homeostatic plasticity. *Current Opinion in Neurobiology* **43**:166–176. DOI: https://doi.org/10.1016/j.conb.2017.03.015, PMID: 28431369

Zmarz P, Keller GB. 2016. Mismatch receptive fields in mouse visual cortex. *Neuron* **92**:766–772. DOI: https://doi.org/10.1016/j.neuron.2016.09.057, PMID: 27974161

Znamenskiy P, Kim M-h, Muir DR, Iacaruso F, Hofer SB, Mrsic-Flogel TD. 2018. Functional selectivity and specific connectivity of inhibitory neurons in primary visual cortex. *bioRxiv*. DOI: https://doi.org/10.1101/294835

Zucker RS, Regehr WG. 2002. Short-Term Synaptic Plasticity. *Annual Review of Physiology* **64**:355–405. DOI: https://doi.org/10.1146/annurev.physiol.64.092501.114547

III. Regulation of circuit organization and function through inhibitory synaptic plasticity

Wu*, Y.K., Miehl*, C., & Gjorgjieva, J. Regulation of circuit organization and function through inhibitory synaptic plasticity. *Trends in Neurosciences* **45(12)**, 884–898 (2022). https://doi.org/10.1016/j.tins.2022.10.006

Trends in Neurosciences

CellPress OPEN ACCESS

Review

Regulation of circuit organization and function through inhibitory synaptic plasticity

Yue Kris Wu [1,2,3], Christoph Miehl [1,2,3], and Julijana Gjorgjieva [1,2,*]

Diverse inhibitory neurons in the mammalian brain shape circuit connectivity and dynamics through mechanisms of synaptic plasticity. Inhibitory plasticity can establish excitation/inhibition (E/I) balance, control neuronal firing, and affect local calcium concentration, hence regulating neuronal activity at the network, single neuron, and dendritic level. Computational models can synthesize multiple experimental results and provide insight into how inhibitory plasticity controls circuit dynamics and sculpts connectivity by identifying phenomenological learning rules amenable to mathematical analysis. We highlight recent studies on the role of inhibitory plasticity in modulating excitatory plasticity, forming structured networks underlying memory formation and recall, and implementing adaptive phenomena and novelty detection. We conclude with experimental and modeling progress on the role of interneuron-specific plasticity in circuit computation and context-dependent learning.

Highlights

Inhibitory synapses are continuously modified by experience through synaptic plasticity. Different learning rules have been proposed to describe the dependence of plasticity on firing rates, spike timing, calcium levels, and membrane potential.

Inhibitory plasticity affects dendritic, cellular, and network dynamics and influences excitatory plasticity at all levels.

Inhibitory plasticity shapes the formation of feedforward receptive fields and structured connectivity in recurrent circuits, supporting the formation and recall of memories and the generation of adaptive and novelty responses. of memories and the generation of adaptive and novelty responses.

Multiple inhibitory neuron subtypes and interneuron-specific plasticity support various computations, including context-dependent processing and pathway-specific selection, and play unique roles in supporting the stability and competition of neural assemblies.

Inhibition throughout development and adulthood

Long-term synaptic plasticity is widely considered to underlie circuit assembly and connectivity refinement during early postnatal development, as well as learning and memory in adulthood [1]. Over the past few decades, extensive studies have characterized the plasticity of synapses between excitatory neurons [2–5]. Consistent with Hebbian principles, coincident pre- and postsynaptic activity potentiates synaptic strength, which enhances the correlation between pre- and postsynaptic activity and further potentiates synaptic strength, potentially leading to runaway synaptic growth and abnormal seizure-like activity [6]. To prevent excessive excitation and maintain stable activity levels, neural circuits employ various mechanisms to dynamically coordinate changes in excitation and inhibition [7,8]. The modulation of inhibitory synapses onto excitatory neurons, called **inhibitory plasticity** (see Glossary), is one such mechanism encountered in different regions of the mammalian brain [9–14] (Box 1). Yet, understanding inhibitory plasticity and its functional implications in shaping network connectivity and dynamics remains challenging because of the different roles inhibitory plasticity might play, depending on the varying demands across an animal's lifetime, as well as the considerable anatomical, electrophysiological, and functional diversity of interneurons, which can undergo different forms of plasticity [15–17].

During early development, it has long been thought that the main inhibitory neurotransmitter in the adult, **gamma-aminobutyric acid (GABA)**, is depolarizing [18,19]. The early excitatory action of GABA has been implicated in the activity-dependent growth and differentiation of neurons and the establishment of neural circuits [20,21]. However, while GABA depolarizes immature cortical neurons *in vivo*, its action at the network level (at least in the neocortex) appears to be inhibitory [22–24]. The maturation of GABAergic synaptic transmission triggers the onset of a critical period in which sensory circuits are highly plastic and sensitive to perturbations [25]. During development and early life, the plasticity of inhibitory GABAergic synapses interacts with **excitatory plasticity** [10]. Multiple computational studies have demonstrated that this interaction shapes

[1]School of Life Sciences, Technical University of Munich, Freising, Germany
[2]Max Planck Institute for Brain Research, Frankfurt, Germany
[3]These authors contributed equally to this work.

*Correspondence:
gjorgjieva@tum.de (J. Gjorgjieva).

Box 1. Inhibitory plasticity in experiments and models

Inhibitory plasticity has been observed in different regions of the mammalian brain [9–12,35]. Experimentally, inhibitory plasticity can be induced by concurrent presynaptic hyperpolarization and postsynaptic depolarization [16,36–39], for instance, via high-frequency stimulation of input pathways [40,41] or pairing of pre- and postsynaptic spikes [16,36,42–44] (see [13] for an extensive summary of experimental studies on inhibitory plasticity).

In computational models, inhibitory plasticity is implemented by phenomenological learning rules, which simplify the underlying complex molecular and biochemical processes [13,14]. In these models, inhibitory synaptic change can depend on firing rates, precise spike times, or membrane potential based on the induction protocol used experimentally [45–49]. A commonly used inhibitory learning rule, which depends on spikes [also called **inhibitory spike-timing-dependent plasticity** (iSTDP)], is the **symmetric Hebbian learning rule** (see Figure I in Box 1). It has a symmetric window as a function of the time difference between pre- and postsynaptic spikes. Spikes near each other in time, independent of their order, lead to inhibitory **long-term potentiation (LTP)** , whereas pre- and postsynaptic spikes far from each other lead to inhibitory **long-term depression (LTD)** [45]. A similar symmetric iSTDP window has been found experimentally in the auditory cortex [44], in the orbitofrontal cortex [50], and in the hippocampus [16]. To account for the diversity of experimentally observed iSTDP windows, computational models have also investigated other learning window shapes, including **asymmetric Hebbian**, where pre-post spike pairs lead to LTP and post-pre spike pairs lead to LTD [51,52], as observed in entorhinal cortex [43]; **asymmetric anti-Hebbian**, where pre-post spike pairs lead to LTD and post-pre spike pairs lead to LTP [52]; and **symmetric anti-Hebbian** window, where spikes near each other in time lead to LTD, while spikes far from each other lead to LTP [53], as observed in hippocampus [36] (see Figure I in Box 1).

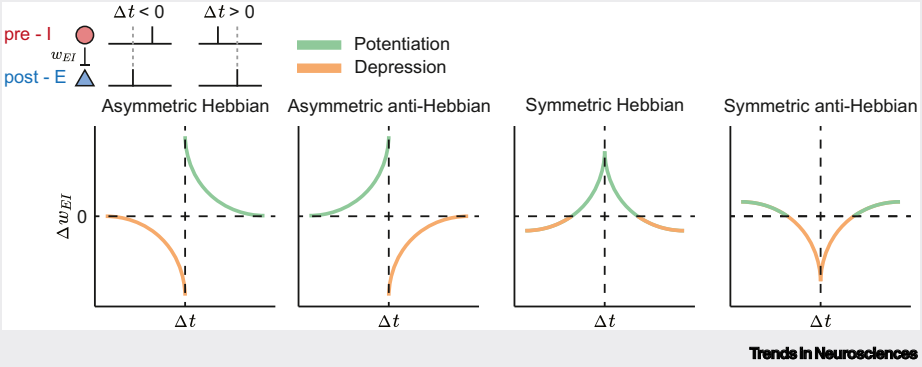


Figure I. Different learning windows of inhibitory spike-timing-dependent plasticity. Inhibitory plasticity can be parameterized into different idealized learning windows as a function of the timing difference between pre- and postsynaptic spikes Δt, leading to either inhibitory long-term potentiation ($\Delta w_{EI} > 0$, green) or inhibitory long-term depression ($\Delta w_{EI} < 0$, orange): asymmetric Hebbian [51,52], asymmetric anti-Hebbian [52], symmetric Hebbian [45], and symmetric anti-Hebbian [53].

network structures and establishes the appropriate network connectivity driven by developmental patterns of spontaneous activity and sensory experience [26–28]. Following sensory deprivation, especially during the critical period, inhibitory plasticity can regulate the balance of excitation and inhibition (E/I balance) and contribute to firing rate homeostasis [29,30]. To adapt to more complex environments, inhibitory plasticity continues to shape learning and network dynamics throughout adulthood. For example, different interneuron subtypes and interneuron-specific plasticity support diverse computations from context-dependent information processing to predictive coding [16,31–34]. Therefore, through plasticity, inhibition can adjust to the needs of the organism at various stages from development to adulthood.

Here, we present recent experimental and theoretical advances on inhibitory plasticity and the control it exerts on circuit connectivity and dynamics. We outline how inhibitory plasticity controls network firing rates and correlations, as well as the plasticity of excitatory connections. We discuss how the interaction of excitatory and inhibitory plasticity can influence the formation of

Glossary

Anti-Hebbian learning rule: a learning rule in which long-term depression is induced by presynaptic followed by postsynaptic spikes, the opposite of Hebb's principle.
Asymmetric Hebbian learning rule: a learning rule that is an asymmetric function of the difference in spike times of pre- and postsynaptic neurons. For asymmetric learning rules, pre-post spike pairs have the opposite impact on the weight change to that of post-pre spike pairs.
Disinhibition: loss or reduction of inhibition. Disinhibition can be induced in multiple ways, for example, via neuromodulators that reduce GABA release from inhibitory neurons onto excitatory neurons, or via increasing inhibition onto inhibitory neurons that target excitatory neurons.
Excitatory plasticity: the plasticity of synapses from an excitatory to another excitatory neuron.
Gamma-aminobutyric acid (GABA): a major inhibitory neurotransmitter in the adult brain.
Hebbian learning rule: a learning rule in which long-term potentiation is induced by presynaptic followed by postsynaptic spikes, in agreement with Hebb's principle.
Inhibition-stabilized network (ISN): a network consisting of excitatory and inhibitory neurons with strong recurrent excitation, which is stabilized by strong feedback inhibition generated in the circuit.
Inhibitory plasticity: the plasticity of synapses from an inhibitory to an excitatory neuron.
Inhibitory spike-timing-dependent plasticity (iSTDP): a process that adjusts the (inhibitory) synaptic strength based on the timing of presynaptic and postsynaptic spikes.
Long-term depression (LTD): a process involving the weakening of synapses between neurons.
Long-term potentiation (LTP): a process involving the strengthening of synapses between neurons.
Symmetric Hebbian learning rule: a learning rule that is a symmetric function of the difference in spike times of pre- and postsynaptic neurons. For symmetric learning rules, pre-post spike pairs have the same impact on the weight change to that of post-pre spike pairs.

different network connectivity structures, including, but not limited to, receptive fields and assemblies, modulate these structures during learning and memory formation, and generate adapted and novelty responses. Based on experimental evidence of different interneuron subtypes and their connectivity profiles, we also present modeling studies that explore differences in the plasticity at these synapses. Throughout, a picture emerges that highlights inhibition and inhibitory plasticity as key factors that control circuit dynamics, ensure appropriate circuit function, and provide a substrate for flexible and complex computations driving behavior throughout the entire life of an organism.

Inhibitory plasticity controls excitation at different spatiotemporal scales

To maintain stable activity levels, inhibitory plasticity can dynamically adjust the amount of inhibition at different spatial and temporal scales during both normal circuit operation and perturbation (Figure 1). At the network level, inhibition is thought to maintain healthy firing rates to prevent runaway dynamics leading to epileptic activity or decreases leading to complete silence (Figure 1A). However, in heavily interconnected neural circuits, the relationship between inhibition and network dynamics is more complicated. In such recurrently dominated networks, strong feedback inhibition generated by the circuit is needed to balance strong recurrent excitation. Both theoretical and experimental studies have put forward such inhibition stabilization as an essential property of cortical networks [54,55]. **Inhibition-stabilized networks (ISNs)** can perform various computations, including input amplification, response normalization, and network multistability [56–58]. A signature of inhibition stabilization is widely considered to be the paradoxical effect, whereby injecting excitatory currents into inhibitory neurons (e.g., via optogenetic stimulation of inhibitory neurons) decreases inhibitory firing [59]. Several circuit aspects, including recurrent excitatory-to-excitatory connection strengths and network activity, can dynamically shape inhibition stabilization [57,60]. For example, in networks where neuronal dynamics are nonlinear, changing the connection from inhibitory to excitatory neurons affects network activity and puts the network in different inhibition-stabilized regimes, as evaluated by the presence of the paradoxical effect (Figure 1B, [57,58,60]). Yet, detecting ISNs via the paradoxical effect is experimentally challenging due to the sensitivity of optogenetic stimulation strength [61] and the complexity introduced by multiple interneuron subtypes [62]. While inhibition stabilization is necessary for various computations, it is still unclear how it can be maintained in the presence of synaptic plasticity, for example, during learning, though recent work addresses this question in the context of balanced excitatory and inhibitory receptive field formation [63].

More broadly, inhibitory plasticity can operate as a homeostatic process and control network activity following perturbation [64,65]. A classical paradigm to explore this process experimentally is elevating or suppressing the activity of cultured neurons, which triggers the potentiation or depression of spontaneous inhibitory synaptic currents into the perturbed neurons [66,67]. In the living animal, a perturbation may involve sensory deprivation, for example, the removal of whiskers in the somatosensory system or the closure of an eye in the visual system [68,69]. Here, inhibitory plasticity could be involved both during the initial circuit response leading to the decrease in network firing rates, as well as later on during their recovery. Initially, the strong potentiation of recurrent inhibition onto excitatory neurons could contribute to the early decrease of network firing rates [30,70,71]. The subsequent gradual upregulation of firing rates could be triggered by the loss of inhibitory synapses onto excitatory neurons [72,73], or the decreased spontaneous inhibitory current frequency [74,75] and amplitude [64,68]. In sum, inhibitory plasticity could act as a common driver behind the homeostatic regulation of network activity immediately after or during a prolonged period following sensory perturbation across sensory cortices.

How could inhibitory plasticity achieve this homeostatic regulation of excitatory firing rates? One answer lies in the concept of E/I balance, which inhibitory plasticity can establish and maintain at

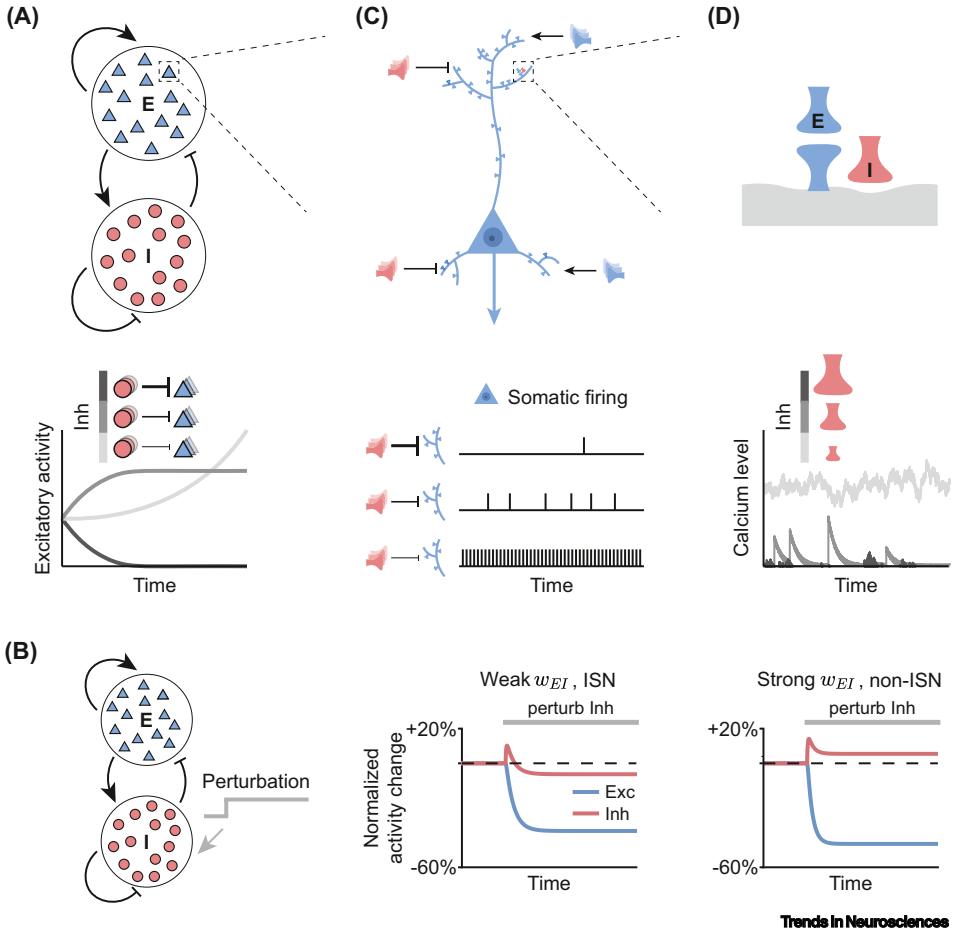


Figure 1. Inhibitory control of excitation at different scales. (A) At the network level (top), inhibition (Inh) affects excitatory population activity (bottom). Excessive inhibition can silence excitatory activity, insufficient inhibition can lead to the explosion of excitatory activity, while the appropriate amount of inhibition stabilizes network dynamics and maintains excitatory activity at a modest level. **(B)** Left: Assessing inhibition stabilization via the paradoxical effect by perturbing the inhibitory population. Middle: For weak inhibitory weights (w_{EI}), network activity is high and the network is in the inhibition-stabilized network (ISN) regime. Injecting additional excitatory currents into inhibitory neurons ('perturb Inh') leads to a paradoxical decrease of the inhibitory population response. Right: For strong w_{EI}, network activity is low and the network is in the non-ISN regime. Injecting additional excitatory currents into inhibitory neurons ('perturb Inh') does not generate a paradoxical response. (C)At the single neuron level (top), inhibition affects somatic firing (bottom). Excessive inhibition generates very little spiking, insufficient inhibition leads to high levels of spiking, while appropriate amount of inhibition leads to appropriate spiking levels. (D) At the dendritic level (top), inhibition influences the local calcium level (bottom). Excessive inhibition leads to extremely low calcium level locally on the dendrite, insufficient inhibition leads to extraordinarily high local calcium level, while the appropriate amount of inhibition leads to an appropriate local calcium level.

the network, cellular, and subcellular level, with different computational implications for circuit processing (Box 2) [29,74–78]. E/I balance is typically quantified by the E/I ratio, defined as the ratio of excitatory to inhibitory input currents. The E/I ratio can in return also affect the amount of inhibitory plasticity, with high initial E/I ratios resulting in stronger inhibitory potentiation, as shown in the mouse auditory cortex [44,79].

Various inhibitory plasticity rules have been proposed to regulate E/I balance in computational models [45,51,52,80–82]. The best-studied model of inhibitory plasticity, which has a symmetric Hebbian learning window (see Figure I in Box 1), can establish a precise E/I balance at the single-

Box 2. Different types of E/I balance

Neural circuits are known to maintain E/I balance [7,10]. E/I balance generally refers to the coregulation of excitation and inhibition and is typically measured by the ratio of excitatory and inhibitory inputs [10]. When excitation and inhibition are balanced at the population level but not necessarily at the single neuron level, the E/I balance is known as global balance [95,102]. Global balance can be achieved via input-dependent inhibitory plasticity rules [84]. If excitatory and inhibitory input currents onto a single neuron are balanced, or co-tuned, across the stimulus space, this is referred to as detailed balance [76–78,103]. Detailed balance can be established via inhibitory plasticity rules, which maintain a target firing rate at the single neuron level [45]. Additionally, when excitatory and inhibitory inputs are balanced also on a millisecond timescale, as observed experimentally [104,105], the E/I balance is known as tight balance, and loose balance otherwise [106]. The coexistence of tight and detailed balance is referred to as precise E/I balance and has been observed in several circuits, such as the zebrafish homolog of olfactory cortex [107] and mammalian hippocampus [108], where it is involved in efficient memory storage, millisecond-range input gating, and subthreshold gain control.

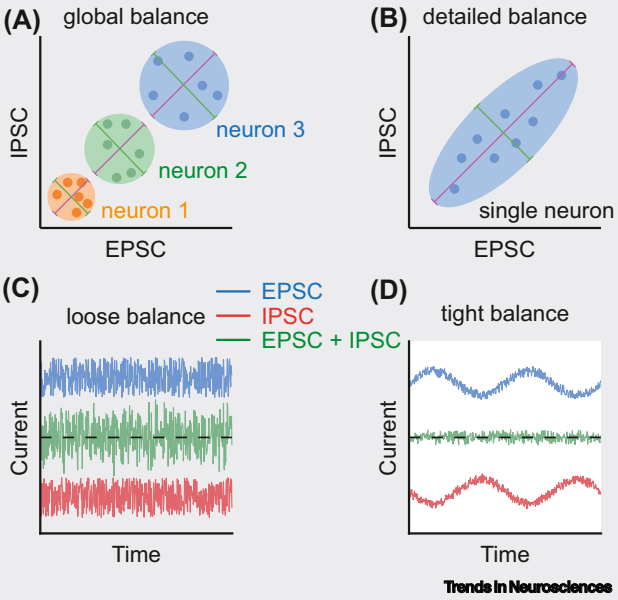


Figure I. Different types of excitation/inhibition (E/I) balance. (A) Global balance is characterized by a high degree of correlation between excitatory postsynaptic currents (EPSCs) and inhibitory postsynaptic currents (IPSCs) at the population level but a low degree of correlation for individual neurons across stimuli. Each dot represents a neuron–stimulus pair. Data for different neurons are marked in different colors. (B) Detailed balance is characterized by a high degree of correlation between EPSCs and IPSCs at the individual neuron level across stimuli. (C) Loose balance is characterized by a low degree of correlation between EPSCs and IPSCs over time. (D) Tight balance is characterized by tightly correlated EPSCs and IPSCs on a millisecond timescale. Panels (A) and (B) are adapted from [107].

neuron level on a millisecond timescale [45,83]. The learning rule achieves the balance by a negative feedback mechanism, which increases inhibitory synaptic strength for high postsynaptic firing rates and decreases inhibitory strength for low firing rates to counteract deviations from a target firing rate (Figure 1C), therefore maintaining a firing rate set-point for each individual neuron. How such a negative feedback mechanism might be implemented biologically remains an open question (see [14] for a discussion of the molecular mechanisms underlying inhibitory plasticity). Due to the resulting robust homeostatic properties, this rule is commonly used in recurrent network models [28,45]. Computational work has proposed several alternatives, including an input-dependent inhibitory plasticity rule [84], or a voltage-dependent plasticity rule [49], both of which can achieve firing rate heterogeneity as observed experimentally [69,85]. One caveat of all these inhibitory plasticity rules is the mismatch between timescales assumed in models and timescales measured in experiments. Most computational models rely on fast inhibitory plasticity

to guarantee homeostasis and establish an E/I balance [48,65]; however, it takes several tens of minutes to reach a stable baseline of inhibitory synaptic strength following plasticity induction in the mouse auditory cortex [44,76,78].

Recent experimental evidence suggests that E/I balance can even extend to local dendritic segments of single neurons [86] (Figure 1D). Inhibitory synapses form and change in strength to complement the dendritic organization of excitatory synaptic inputs, which often form local clusters based on coactivation [87,88], to regulate excitatory synaptic dynamics and plasticity [86,89,90]. For example, in the hippocampus, stimulating clustered excitatory synapses has been shown to trigger the *de novo* formation of inhibitory synapses [91], and a push–pull plasticity mechanism has been found to maintain the balance of local dendritic excitatory and inhibitory strength [92]. Also, inhibitory synapses in the neocortex remain stable if located in the proximity of excitatory synapses during normal visual experience [72]. Thus, while the presence of E/I balance on local stretches of dendrites is supported by experimental data, how it emerges during early postnatal development and how it is maintained during learning and perturbations remains an open question.

Besides regulating E/I balance and firing rates, inhibitory plasticity plays a more nuanced role in controlling the firing patterns of single neurons. By regulating the precise arrival of inhibitory inputs relative to excitatory inputs, experiments in the hippocampus have showed that inhibition can close or open the time window in which a spike is triggered [93]. Inhibitory plasticity can therefore dramatically affect the spike generation properties and spiking statistics of excitatory neurons, including neuronal input–output functions [94], pairwise spike correlations and spiking regularity [95,96], and criticality [97,98]. Both experimental and modeling work have showed that potentiating inhibition can decorrelate network activity [24,99,100] and switch network firing regimes [95] from oscillatory states supporting memory consolidation [101] to asynchronous irregular states supporting high memory capacity, despite the presence of noise [81]. Such switching could occur at different behavioral state transitions (e.g., from sleep to wake). Yet, direct evidence of inhibitory plasticity contributing to a dynamical switching between network firing regimes remains to be examined experimentally.

Inhibitory control of excitatory plasticity

Experimental evidence has revealed that excitatory plasticity is jointly determined by factors like pre- and postsynaptic firing rates [2,4], spike timing [3,4], and dendritic calcium levels [5]. Since inhibition can influence all of these factors, it naturally also affects excitatory plasticity [12,109–111].

In experiments, the frequency of presynaptic stimulation can determine the sign of excitatory synaptic plasticity, with low-frequency stimulation favoring excitatory LTDand high-frequency stimulation inducing excitatory LTP [2]. Decreasing inhibition decreases the excitatory LTD/LTP threshold, making LTP induction easier, while increasing inhibition increases the LTD/LTP threshold and makes LTP induction more difficult [112] (Figure 2A). Based on these results, computational studies have demonstrated that a change of the inhibitory input (e.g., via inhibitory plasticity) can shift the threshold between LTP and LTD [47,48]. By keeping the firing rates exactly at the LTD/LTP threshold, inhibitory plasticity has been suggested as a mechanism to effectively switch excitatory plasticity off [48] (Figure 2A). Any deviation of the firing rates (e.g., via **disinhibition**) can then turn on excitatory plasticity. Such gating of excitatory plasticity has also been modeled at the level of individual inhibitory inputs on dendritic trees by affecting the amplitude of backpropagating action potentials and calcium spikes [113,114] (Figure 2B). Therefore, changes in inhibition can switch excitatory plasticity on or off, regulate how much plasticity is induced, or even dictate the sign of excitatory plasticity [38,115].

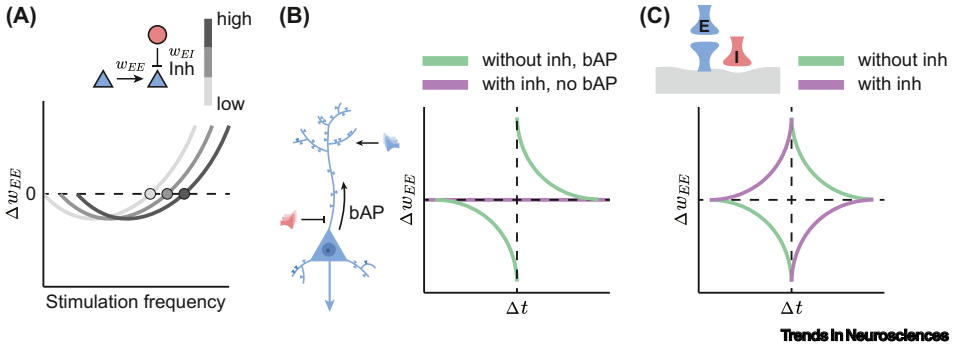


Figure 2. Inhibitory control of excitatory plasticity. (A) The level of inhibition (Inh), modulated by inhibitory weights (w_{EI}) or inhibitory firing rates, controls excitatory plasticity (Δw_{EE}). Higher (lower) level of inhibition leads to higher (lower) long-term depression (LTD)/long-term potentiation (LTP) threshold of excitatory plasticity as a function of the presynaptic stimulation frequency. Different dots represent corresponding LTD/LTP thresholds that separate the depression ($\Delta w_{EE} < 0$) and potentiation ($\Delta w_{EE} > 0$) of excitatory synapses onto excitatory neurons. Different grays represent different levels of inhibition. Panel (A) is adapted from [48,112]. (B) Strong inhibitory input can switch excitatory plasticity on or off via gating of a backpropagating action potential (bAP). In the absence of inhibition, the bAP propagates into the dendrite and spike-timing-dependent plasticity at the excitatory synapse is induced (green). By contrast, in the presence of inhibition, the bAP is suppressed and no synaptic plasticity is induced (purple). Panel (B) is adapted from [113]. **C.** (C) Local inhibitory input can affect calcium concentration in the dendritic spine and flip the excitatory spike-timing-dependent plasticity. Panel (C) is adapted from [115,125].

Multiple experimental studies have suggested disinhibition as a mechanism for the gating of excitatory plasticity [116]. Disinhibition can be induced by neuromodulators, including but not limited to acetylcholine, noradrenalin, and oxytocin [10,76], or by disinhibitory pathways involving multiple interneuron subtypes [117,118] (Box 3). For instance, elevated activity in vasoactive intestinal peptide (VIP)-expressing inhibitory neurons receiving top-down inputs can suppress activity in somatostatin (SST)-expressing inhibitory neurons and, as a result, disinhibit excitatory neurons and control excitatory plasticity [111,117–120].

At the dendritic level, inhibitory input onto the dendrite can affect postsynaptic calcium concentration at nearby excitatory spines [111,121] and, therefore, influence local excitatory plasticity [122,123]. Computational models have proposed that the dynamic local balancing of excitation by inhibition can change the shape of the learning rule for excitatory synapses [124–126]. For example, blocking inhibitory inputs can flip the spike-timing-dependency of excitatory plasticity [125], consistent with previous experimental findings [115] (Figure 2C). Furthermore, local changes in excitatory and inhibitory synapses are coordinated with each other via crosstalk, giving rise to the codependence of excitatory and inhibitory plasticity [7,8]. While these works clearly show that inhibitory synapses can control excitatory plasticity at multiple spatial scales, how this control is used during learning and its impact on behavior remains to be explored.

Inhibitory plasticity in the formation of structured networks and resulting computation

Non-random structure is a hallmark of biological networks. Multiple computational studies have demonstrated that various network structures can form from the coordinated interaction between excitatory and inhibitory plasticity. This includes the emergence of receptive fields [45,47,48], place fields [27], and grid fields [27] through the refinement of feedforward excitatory and inhibitory connectivity, typically in settings with a single postsynaptic neuron based on input statistics [51–53]. In recurrent circuits, inhibitory plasticity also shapes neuronal assemblies [26,48] and chain-like structure [127,128], as well as ensuing tuning diversity and efficient sensory representation [100].

Box 3. Interneuron diversity

Interneurons exhibit high anatomical, electrophysiological, and functional diversity [157,158]. In the mouse neocortex, three major classes of interneurons expressing parvalbumin (PV), somatostatin (SST), and vasoactive intestinal peptide (VIP) constitute more than 80% of GABAergic interneurons [15]. Distinct interneuron subtypes target different domains of pyramidal cells. More specifically, PV neurons preferentially target perisomatic regions of pyramidal neurons, whereas SST neurons target distal dendritic regions of pyramidal neurons that also receive inhibition from neuron-derived neurotrophic factor (NDNF)-expressing interneurons in layer 1 [15,159].

The multiplicity of interneuron subtypes is implicated in diverse computations and cognitive functions, such as locomotion-induced gain modulation [160], selective attention [127], context-dependent modulation [31,33], predictive processing [32,161], and gating of synaptic plasticity [117,120]. For instance, long-range cortico-cortical projections activating upstream VIP neurons in the primary visual cortex exert spatially specific top-down modulation of visual processing, resembling selective attention [127]. In predictive processing framework, mismatches between sensory inputs and internally generated predictive signals evoke the activity of prediction-error neurons [32]. In the layer 2/3 of the primary visual cortex, prediction-error neurons balance inhibitory visual input mediated by SST against excitatory motor-related predictive input targeting VIP [161].

Strongly interconnected groups of excitatory neurons form assemblies, which have been proposed to be the basis of associative memory [129,130]. Inhibition can influence excitatory assemblies in two distinct ways. First, inhibitory neurons may be nonspecific and nonpreferentially target different excitatory assemblies, known as 'blanket of inhibition' [131] (Figure 3A). Second, inhibition may be stimulus-specific if distinct inhibitory neurons receive stimulus-specific feedforward drive, or if excitatory and inhibitory neurons with a similar stimulus tuning connect more strongly and form E/I assemblies, known as stimulus-specific feedback inhibition [132] (Figure 3B).

While many mechanisms are involved in the formation of excitatory assemblies [133], computational models have proposed an important role of inhibitory plasticity in preventing runaway excitation that results from the assemblies' repeated coactivation and preventing winner-take-all dynamics whereby a single assembly is always active [26,28,48]. Specific to forming E/I assemblies, both inhibitory synapses onto excitatory neurons and excitatory synapses onto inhibitory neurons need to be plastic in the recurrent circuit [63,134]. The resulting co-tuned feedback inhibition in networks with E/I assemblies can support network stability [60,132], changes in neuronal variability [135], and decision making in the presence of noise [136].

Irrespective of whether inhibition is unspecific or specific, modeling studies suggest that the plasticity of lateral inhibitory connections across assemblies can ensure that different memories encoded by different assemblies are easily discriminated [50,137]. Concurrently, multiple experimental studies have found evidence for the role of inhibition in memory recall. For instance, inactive memories can be unmasked by suppressing inhibitory neurons [138]. Using E/I assemblies as a model for associative memories, the inactive memories seem to remain in the quiescent state until being recalled by disinhibition [138,139]. Recent work in the human neocortex has further suggested that specific inhibition can avoid inappropriate interference of overlapping memories and permit continual learning [140,141].

The activation of E/I assemblies shaped by inhibitory plasticity has also been hypothesized to underlie the adaptation of behavioral responses to repeated stimulation (i.e., 'habituation') [139,142]. The ability to adapt to repeated stimuli, detect unexpected stimuli in the environment, and identify their relevance to execute appropriate behavioral reactions is important for survival. Inhibitory plasticity has been suggested to be important in shaping adaptation to repeated responses also at the cellular level in the mouse auditory cortex [143]. A recent computational study has provided a mechanistic insight on how inhibitory plasticity can shape the responses to repeated and novel stimuli [144]. While the repeated presentation of a stimulus evokes initially high activity of the excitatory assembly representing the stimulus, the subsequent increase of

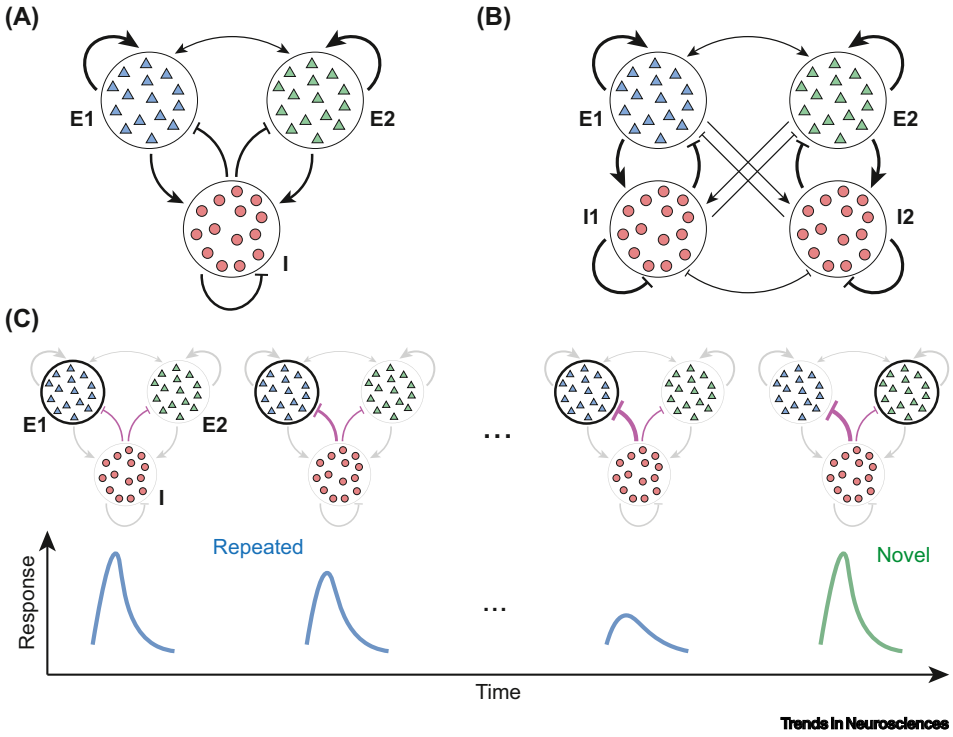


Figure 3. Unspecific versus specific inhibitory connectivity and the generation of adaptive and novelty responses. (A) Network with unspecific inhibition, in which different excitatory assemblies are inhibited by a single inhibitory population. (B) Network with stimulus-specific feedback inhibition, in which distinct excitatory assemblies are inhibited by non-overlapping inhibitory subpopulations. (C)The repeated and novel stimuli activate distinct excitatory assemblies, E1 and E2, respectively (activation marked with bold circles). Repeated presentation of the same stimulus leads to an increase of specific inhibitory synaptic strength onto the E1 assembly and a reduction of the evoked response (blue), while presenting the novel stimulus triggers a high response due to the weak inhibitory synaptic strength onto the E2 assembly (green).

inhibitory synaptic strengths suppresses the ensuing responses upon stimulus repetition. By contrast, a novel stimulus evokes a high response of its corresponding excitatory assembly since the inhibitory synapses onto the assembly do not potentiate (Figure 3C). While both blanket and stimulus-specific inhibition can capture adapted and elevated responses to repeated and novel stimuli, stimulus-specific inhibition is necessary for other adaptive phenomena [144]. This includes stimulus-specific adaptation, whereby excitatory neurons that are equally driven by two stimuli exhibit a higher response to the rarely presented stimulus, but a lower response to the frequently presented stimulus [145].

Interneuron-specific plasticity and its functional implications

Inhibitory neurons can be divided into multiple distinct subtypes based on their electrophysiological, morphological, and transcriptomic properties (Box 3). Accumulating evidence also suggests that synapses from and to different interneuron subtypes undergo distinct forms of synaptic plasticity [16,17,37,146,147]. Computational models have capitalized on these experimental results of interneuron-specific plasticity and explored its role in different settings. In feedforward networks, modeling work has showed that the receptive field of a neuron may not be solely determined by the feedforward excitatory weight profiles, but is heavily modulated by inhibition from different pathways [53]. By exploring several candidate plasticity rules for the different inhibitory pathways, the authors found that the neuron's receptive field strongly depends on the

modulatory state of inhibition as an example of context-dependence [53].

Recent studies in the rodent hippocampus have identified learning rules describing the LTD of parvalbumin (PV) synapses and the LTP of SST synapses onto excitatory pyramidal neurons in CA1 during physiological activity patterns [16]. As PV and SST mainly target perisomatic regions receiving inputs from CA3 and distal dendritic regions receiving inputs from pyramidal neurons in entorhinal cortex, respectively, both experiments and modeling suggest that interneuron-specific plasticity might prioritize inputs from one pathway over another [16] (Figure 4A). As stronger inhibition resulting from the potentiation of SST synapses onto excitatory neurons can limit excitatory plasticity [120], modeling has suggested that interneuron-specific plasticity can promote the stability of place cells [16]. Recent experiments in CA1 suggest that even synapses from different interneurons targeting the same perisomatic regions of excitatory neurons can undergo opposite changes when animals explore novel environments [17] (Figure 4B). Since these two types of interneurons preferentially receiving different inputs fire at different phases of network theta rhythms associated with memory encoding and retrieval [148], the opposite regulation of interneuron-specific plasticity may impact memory formation and maintenance. Future computational models could help uncover how the opposing plasticity mechanisms support long-term memories.

In addition to hippocampus, interneuron-specific plasticity rules based on spike timing have been reported in layer 2/3 of mouse orbitofrontal cortex and implicated in assembly formation in recurrent network models [50]. More specifically, PV synapses onto excitatory neurons follow a symmetric Hebbian learning rule and appear to be important for network stability; by contrast, SST synapses onto excitatory neurons follow an asymmetric Hebbian learning rule and appear to enhance competition between assemblies [50] (Box 1). Although a learning rule has not yet been characterized for neuron-derived neurotrophic factor (NDNF)-expressing interneurons, experimental studies have revealed that inhibition mediated by NDNF interneurons in layer 1 of the auditory cortex changes after associative auditory fear conditioning , and have suggested that NDNF interneurons and their plasticity are involved in the formation of associative memories [149].

While significantly less studied, recent work has begun to explore synapses between inhibitory neurons, including their impact on E/I balance in recent connectomic studies [150], on generating long neuronal timescales that support working memory, and on memory storage in computational models [151,152]. Yet, little is known about the plasticity of these inhibitory-to-inhibitory connections experimentally. Computational models here play an important role in revealing the functional consequences of this type of plasticity. For instance, a two-stage model showed that an initial stage of SST to PV plasticity guides the subsequent plasticity of excitatory-to-excitatory connections in a recurrent network underlying visual stimulus selectivity [153]. Recent modeling work has also begun investigating recurrent network models where multiple synapse types are simultaneously plastic and found that experimentally observed dynamics and computations can emerge from the complex interplay of many plasticity mechanisms. Given the high-dimensional space of learning rule parameters, when such models succeed in finding stable regimes, they can provide predictions for the learning mechanisms in real biological circuits. Deriving learning rules via optimizing a desired function has provided a new promising approach to study plasticity [154,155]. In an elegant example, recent studies derived plasticity rules from the perspective of optimizing a loss function to achieve firing rate set-points; the emergent networks could then generate self-sustained, inhibition-stabilized dynamics [156] and stimulus-specific feedback inhibition [134]. Even without deriving novel learning rules, combining classical Hebbian plasticity with synapse-type-specific competition for synaptic resources can yield novel dynamics such as the development of stimulus selectivity, E/I balance, decorrelated neural activity, assembly structures, and response normalization [63].

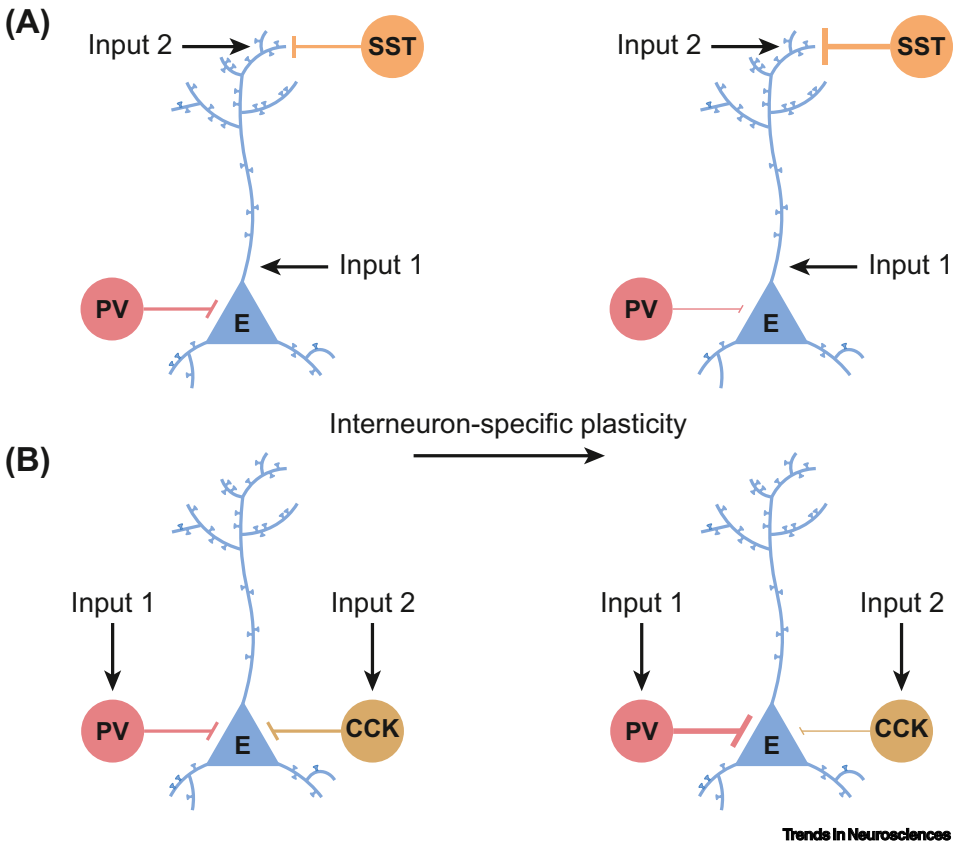


Figure 4. Interneuron-specific plasticity. (A) Inhibitory synapses from parvalbumin (PV)- (red) and somatostatin (SST)- (orange) expressing neurons onto hippocampal CA1 pyramidal neurons (blue) are weakened and enhanced, respectively, during physiological firing patterns [16]. This interneuron-specific plasticity can prioritize proximal input from CA3 over distal input from entorhinal cortex. (B) Perisomatic inhibitory synapses from PV- (red) and cholecystokinin (CCK)-expressing (brown) neurons onto recently activated hippocampal CA1 pyramidal neurons (blue) undergo long-term potentiation and long-term depression, respectively when animals are engaged in novel environments [17].

Concluding remarks and future perspectives

Over the past two decades, our understanding of the inhibitory control of circuit organization and dynamics, as well as the potential to modulate this control via plastic inhibition, has significantly grown. Inhibitory synapses in the brain are highly dynamic and regulated by various plasticity mechanisms, including short-term plasticity operating at the timescale of milliseconds to seconds [162] as well as long-term plasticity acting at the timescale of minutes to hours [44]. Here, we summarized studies on the long-term plasticity of inhibitory-to-excitatory synapses, referred to as inhibitory plasticity. As discussed in this review, abundant evidence suggests that inhibitory plasticity is important for establishing and maintaining E/I balance, achieving firing rate homeostasis, controlling excitatory plasticity, and shaping network connectivity throughout the entire life of an organism. Nonetheless, it remains unclear if the learning rules that characterize inhibitory plasticity in development are the same as those operating in adulthood (see Outstanding questions). Complementary to the growing number of experimental studies on inhibitory plasticity, theoretical and computational approaches have played an important role in synthesizing the available data to reveal how inhibition regulates various aspects of circuit function. This has generated mechanistic insights into the function of inhibitory plasticity at several spatial scales, from the local dendritic regulation of E/I balance, to the cellular control of spiking properties, and the maintenance of stable

Outstanding questions

Neuronal activity during development is typically generated spontaneously in the absence of sensory experience. This activity operates on much slower timescales (hundreds of milliseconds) compared with the sensory-driven activity patterns (few to tens of milliseconds) in adulthood. Do the activity-dependent learning rules that characterize inhibitory plasticity integrate activity at different timescales in development and adulthood?

The phenomenological learning rules that determine how inhibitory plasticity depends on rates and spike timing can be modulated by various external factors. How do different neuromodulators, behavioral states, and environmental perturbations affect inhibitory plasticity rules?

How are phenomenological descriptions of inhibitory plasticity implemented with the biological machinery of molecular interactions?

Distinct forms of E/I balance might be beneficial for different demands in development versus adulthood. How are different types of E/I balance dynamically regulated by inhibitory plasticity over multiple timescales to serve specific goals?

E/I balance also exists at different spatial scales. Are there shared principles underlying the establishment of E/I balance across these different scales? What are the functional implications of breaking E/I balance at some spatial scales but not others?

Interneurons come in diverse subtypes, receive inputs from different pathways, and target excitatory neurons in different locations (e.g., cell body versus dendrite). This diversity is also reflected in the types of plasticity rules experienced at the synapses. How can interneuron-specific plasticity rules be described as a function of firing rates, spike timing, and calcium level?

Inhibitory plasticity rules might also differ across brain regions. How do different brain regions coordinate the potentially different forms of inhibitory plasticity they express to maintain biologically reasonable activity levels and process information?

activity patterns and connectivity structures at the network level. At the same time, we have highlighted that inhibitory control also occurs at multiple temporal scales from the regulation of fast spiking to the slower calcium dynamics and even slower timescales at which measurable changes in synaptic strength can be observed.

Despite this progress, many open challenges remain due to the high diversity of inhibitory neurons and the interneuron-specific plasticity at different synapse types. Experimentally, the development of transgenic and recording techniques opens new possibilities to record activity from multiple interneuron subtypes simultaneously and probe the rules that govern synaptic plasticity. Concurrently, computational models and theories are becoming paramount. First, they are essential to understand the complex interactions of different plasticity mechanisms, especially in highly recurrent circuits with non-intuitive dynamics. Second, models can explore candidate plasticity mechanisms and study their functional implications. Last, theoretical work also enables the exploration of more abstract concepts, like inhibition-stabilization, as general frameworks for circuit processing, which can be established and modulated through inhibitory plasticity.

Different computations, such as selective attention, context-dependent modulation, and predictive processing, typically require diverse interneuron subtypes with specific synaptic connections. How do interneuron-specific plasticity mechanisms establish the network connectivity enabling diverse computations?

Acknowledgments

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. We thank Katharina A. Wilmes, Everton J. Agnes, Elizabeth Herbert, and Dylan Festa for comments on the manuscript.

Declaration of interests

The authors declare no conflicts of interest.

References

1. Magee, J.C. and Grienberger, C. (2020) Synaptic plasticity forms and functions. *Annu. Rev. Neurosci.* 43, 95–117
2. Kirkwood, A. *et al.* (1996) Experience-dependent modification of synaptic plasticity in visual cortex. *Nature* 381, 526–528
3. Bi, G.-g. and Poo, M.-M. (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* 18, 10464–10472
4. Sjöström, P.J. *et al.* (2001) Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron* 32, 1149–1164
5. Inglebert, Y. *et al.* (2020) Synaptic plasticity rules with physiological calcium levels. *Proc. Natl. Acad. Sci. U. S. A.* 117, 33639–33648
6. Turrigiano, G.G. and Nelson, S.B. (2004) Homeostatic plasticity in the developing nervous system. *Nat. Rev. Neurosci.* 5, 97–107
7. Hennequin, G. *et al.* (2017) Inhibitory plasticity: balance, control, and codependence. *Annu. Rev. Neurosci.* 40, 557–579
8. Herstel, L.J. and Wierenga, C.J. (2021) Network control through coordinated inhibition. *Curr. Opin. Neurobiol.* 67, 34–41
9. Chen, J.L. and Nedivi, E. (2013) Highly specific structural plasticity of inhibitory circuits in the adult neocortex. *Neuroscientist* 19, 384–393
10. Froemke, R.C. (2015) Plasticity of cortical excitatory-inhibitory balance. *Annu. Rev. Neurosci.* 38, 195–219
11. Kripkee, B. and Froemke, R.C. (2017) *Organization and Plasticity of Cortical Inhibition*, Oxford University Press
12. Chiu, C.Q. *et al.* (2019) Preserving the balance diverse forms of long-term GABAergic synaptic plasticity. *Nat. Rev. Neurosci.* 20, 272–281
13. Gandolfi, D. *et al.* (2020) Inhibitory plasticity: from molecules to computation and beyond. *Int. J. Mol. Sci.* 21, 1805
14. Capogna, M. *et al.* (2021) The ins and outs of inhibitory synaptic plasticity: neuron types, molecular mechanisms and functional roles. *Eur. J. Neurosci.* 54, 6882–6901
15. Tremblay, R. *et al.* (2016) GABAergic interneurons in the neocortex: from cellular properties to circuits. *Neuron* 91, 260–292
16. Udakis, M. *et al.* (2020) Interneuron-specific plasticity at parvalbumin and somatostatin in-hibitory synapses onto CA1 pyramidal neurons shapes hippocampal output. *Nat. Commun.* 11, 4395
17. Yap, E.L. *et al.* (2021) Bidirectional perisomatic inhibitory plasticity of a Fos neuronal network. *Nature* 590, 115–121
18. Ganguly, K. *et al.* (2001) GABA itself promotes the developmental switch of neuronal GABAergic responses from excitation to inhibition. *Cell* 105, 521–532
19. Ben-Ari, Y. *et al.* (2007) GABA: a pioneer transmitter that excites immature neurons and generates primitive oscillations. *Physiol. Rev.* 87, 1215–1284
20. Cancedda, L. *et al.* (2007) Excitatory GABA action is essential for morphological maturation of cortical neurons in vivo. *J. Neurosci.* 27, 5224–5235
21. Wang, D.D. and Kriegstein, A.R. (2008) GABA regulates excitatory synapse formation in the neocortex via nmda NMDA receptor activation. *J. Neurosci.* 28, 5547–5558
22. Kirmse, K. *et al.* (2015) GABA depolarizes immature neurons and inhibits network activity in the neonatal neocortex in vivo. *Nat. Commun.* 6, 7750
23. Murata, Y. and Colonnese, M.T. (2020) GABAergic interneurons excite neonatal hippocampus in vivo. *Sci. Adv.* 6, 1–11
24. Chini, M. *et al.* (2022) An increase of inhibition drives the developmental decorrelation of neural activity. *eLife* 11, 1–28
25. Hensch, T.K. (2005) Critical period plasticity in local cortical circuits. *Nat. Rev. Neurosci.* 6, 877–888
26. Litwin-Kumar, A. and Doiron, B. (2014) Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nat. Commun.* 5, 5319
27. Weber, S.N. and Sprekeler, H. (2018) Learning place cells, grid cells and invariances with excitatory and inhibitory plasticity. *eLife* 7, e34560
28. Ocker, G.K. and Doiron, B. (2019) Training and spontaneous reinforcement of neuronal assemblies by spike timing plasticity. *Cereb. Cortex* 29, 937–951

29. House, D.R. *et al.* (2011) Parallel regulation of feedforward inhibition and excitation during whisker map plasticity. *Neuron* 72, 819–831
30. Maffei, A. *et al.* (2006) Potentiation of cortical inhibition by visual deprivation. *Nature* 443, 81–84
31. Kuchibhotla, K.V. *et al.* (2017) Parallel processing by cortical inhibition enables context-dependent behavior. *Nat. Neurosci.* 20, 62–71
32. Keller, G.B. and Mrsic-Flogel, T.D. (2018) Predictive processing: a canonical cortical computation. *Neuron* 100, 424–435
33. Keller, A.J. *et al.* (2020) A disinhibitory circuit for contextual modulation in primary visual cortex. *Neuron* 108, 1181–1193
34. Hertäg, L. and Clopath, C. (2022) Prediction-error neurons in circuits with multiple neuron types: formation, refinement and functional implications. *Proc. Natl. Acad. Sci. U. S. A.* 119, e2115699119
35. Vickers, E.D. *et al.* (2018) Parvalbumin-interneuron output synapses show spike-timing-dependent plasticity that contributes to auditory map remodeling. *Neuron* 99, 720–735
36. Woodin, M.A. *et al.* (2003) Coincident pre- and postsynaptic activity modifies GABAergic synapses by postsynaptic changes in Cl- transporter activity. *Neuron* 39, 807–820
37. Chiu, C.Q. *et al.* (2018) Input-specific NMDAR-dependent potentiation of dendritic GABAergic inhibition. *Neuron* 97, 368–377
38. Wang, L. and Maffei, A. (2014) Inhibitory plasticity dictates the sign of plasticity at excitatory synapses. *J. Neurosci.* 34, 1083–1093
39. Mellor, J. (2018) Synaptic plasticity at hippocampal synapses: experimental background. In *Springer Series in Computational Neuroscience* (Cutsuridis, V. *et al.*, eds), pp. 201–226, Springer
40. Caillard, O. *et al.* (1999) Long-term potentiation of GABAergic synaptic transmission in neonatal rat hippocampus. *J. Physiol.* 518, 109–119
41. Shew, T. *et al.* (2000) Mechanisms involved in tetanus induced potentiation of fast IPSCs in rat hippocampal CA1 neurons. *J. Neurophysiol.* 83, 3388–3401
42. Holmgren, C.D. and Zilberter, Y. (2001) Coincident spiking activity induces long-term changes in inhibition of neocortical pyramidal cells. *J. Neurosci.* 21, 8270–8277
43. Haas, J.S. *et al.* (2006) Spike-timing-dependent plasticity of inhibitory synapses in the entorhinal cortex. *J. Neurophysiol.* 96, 3305–3313
44. D'amour, J.A. and Froemke, R.C. (2015) Inhibitory and excitatory spike-timing-dependent plasticity in the auditory cortex. *Neuron* 86, 514–528
45. Vogels, T.P. *et al.* (2011) Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science* 334, 1569–1573
46. Bourjaily, M.A. and Miller, P. (2011) Synaptic plasticity and connectivity requirements to produce stimulus-pair specific responses in recurrent networks of spiking neurons. *PLoS Comput. Biol.* 7, e1001091
47. Clopath, C. *et al.* (2016) Receptive field formation by interacting excitatory and inhibitory synaptic plasticity. *bioRxiv* Published online July 29, 2016. https://doi.org/10.1101/066589
48. Miehl, C. and Gjorgjieva, J. (2022) Stability and learning in excitatory synapses by non-linear inhibitory plasticity. *bioRxiv* Published online March 29, 2022. https://doi.org/10.1101/2022.03.28.486052
49. Pedrosa, V. and Clopath, C. (2020) Voltage-based inhibitory synaptic plasticity: network regulation, diversity, and flexibility. *bioRxiv* Published online December 9, 2020. https://doi.org/10.1101/2020.12.08.416263
50. Lagzi, F. *et al.* (2021) Assembly formation is stabilized by parvalbumin neurons and accelerated by somatostatin neurons. *bioRxiv* Published online September 7, 2021. https://doi.org/10.1101/2021.09.06.459211
51. Luz, Y. and Shamir, M. (2012) Balancing feed-forward excitation and inhibition via Hebbian inhibitory synaptic plasticity. *PLoS Comput. Biol.* 8, e1002334
52. Kleberg, F.I. *et al.* (2014) Excitatory and inhibitory STDP jointly tune feed-forward neural circuits to selectively propagate correlated spiking activity. *Front. Comput. Neurosci.* 8, 53
53. Agnes, E.J. *et al.* (2020) Complementary inhibitory weight profiles emerge from plasticity and allow flexible switching of receptive fields. *J. Neurosci.* 40, 9634–9649
54. Li, N. *et al.* (2019) Spatiotemporal constraints on optogenetic inactivation in cortical circuits. *eLife* 8, e48622
55. Sanzeni, A. *et al.* (2020) Inhibition stabilization is a widespread property of cortical networks. *eLife* 9, e54875
56. Murphy, B.K. and Miller, K.D. (2009) Article balanced amplification a new mechanism of selective amplification of neural activity patterns. *Neuron* 61, 635–648
57. Rubin, D.B. *et al.* (2015) The stabilized supralinear network: a unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron* 85, 402–417
58. Sadeh, S. and Clopath, C. (2021) Inhibitory stabilization and cortical computation. *Nat. Rev. Neurosci.* 22, 21–37
59. Tsodyks, M.V. *et al.* (1997) Paradoxical effects of external modulation of inhibitory interneurons. *J. Neurosci.* 17, 4382–4388
60. Wu, Y.K. and Zenke, F. (2021) Nonlinear transient amplification in recurrent neural networks with short-term plasticity. *eLife* 10, e71263
61. Sadeh, S. *et al.* (2017) Assessing the role of inhibition in stabilizing neocortical networks requires large-scale perturbation of the inhibitory population. *J. Neurosci.* 37, 12050–12067
62. Mahrach, A. *et al.* (2020) Mechanisms underlying the response of mouse cortical networks to optogenetic manipulation. *eLife* 9, 1–37
63. Eckmann, S. and Gjorgjieva, J. (2022) Synapse-type-specific competitive Hebbian learning forms functional recurrent networks. *bioRxiv* Published online March 14, 2022. https://doi.org/10.1101/2022.03.11.483899
64. Gainey, M.A. and Feldman, D.E. (2017) Multiple shared mechanisms for homeostatic plasticity in rodent somatosensory and visual cortex. *Philos. Trans. R. Soc. B. Biol. Sci.* 372, 20160157
65. Sprekeler, H. (2017) Functional consequences of inhibitory plasticity: homeostasis, the excitation-inhibition balance and beyond. *Curr. Opin. Neurobiol.* 43, 198–203
66. Kilman, V. *et al.* (2002) Activity deprivation reduces miniature IPSC amplitude by decreasing the number of postsynaptic GABAA receptors clustered at neocortical synapses. *J. Neurosci.* 22, 1328–1337
67. Hartman, K.N. *et al.* (2006) Activity-dependent regulation of inhibitory synaptic transmission in hippocampal neurons. *Nat. Neurosci.* 9, 642–649
68. Li, L. *et al.* (2014) Rapid homeostasis by disinhibition during whisker map plasticity. *Proc. Natl. Acad. Sci. U. S. A.* 111, 1616–1621
69. Hengen, K.B. *et al.* (2016) Neuronal firing rate homeostasis is inhibited by sleep and promoted by wake. *Cell* 165, 180–191
70. Nahmani, M. and Turrigiano, G.G. (2014) Deprivation-induced strengthening of presynaptic and postsynaptic inhibitory transmission in layer 4 of visual cortex during the critical period. *J. Neurosci.* 34, 2571–2582
71. Miska, N.J. *et al.* (2018) Sensory experience inversely regulates feedforward and feedback excitation-inhibition ratio in rodent visual cortex. *eLife* 7, e38846
72. Chen, J.L. *et al.* (2012) Clustered dynamics of inhibitory synapses and dendritic spines in the adult neocortex. *Neuron* 74, 361–373
73. van Versendaal, D. *et al.* (2012) Elimination of inhibitory synapses is a major component of adult ocular dominance plasticity. *Neuron* 74, 374–383
74. Barnes, S.J. *et al.* (2015) Subnetwork-specific homeostatic plasticity in mouse visual cortex in vivo. *Neuron* 86, 1290–1303
75. Keck, T. *et al.* (2017) Interactions between synaptic homeostatic mechanisms an attempt to reconcile BCM theory, synaptic scaling, and changing excitation/inhibition balance. *Curr. Opin. Neurobiol.* 43, 87–93
76. Froemke, R.C. *et al.* (2007) A synaptic memory trace for cortical receptive field plasticity. *Nature* 450, 425–429
77. Dorn, A.L. *et al.* (2010) Developmental sensory experience balances cortical excitation and inhibition. *Nature* 465, 932–936
78. Field, R.E. *et al.* (2020) Heterosynaptic plasticity determines the set point for cortical excitatory-inhibitory balance. *Neuron* 106, 842–854

79. Aljadeff, J. *et al.* (2019) Cortical credit assignment by Hebbian, neuromodulatory and inhibitory plasticity. *arXiv* Published online November 1, 2019. https://doi.org/10.48550/arxiv.1911.00307
80. Yger, P. *et al.* (2015) Fast learning with weak synaptic plasticity. *J. Neurosci.* 35, 13351–13362
81. Rubin, R. *et al.* (2017) Balanced excitation and inhibition are required for high-capacity, noise-robust neuronal selectivity. *Proc. Natl. Acad. Sci. U. S. A.* 114, E9366–E9375
82. Baker, C. *et al.* (2020) Nonlinear stimulus representations in neural circuits with approximate excitatory-inhibitory balance. *PLoS Comput. Biol.* 16, e1008192
83. Akil, A.E. *et al.* (2021) Balanced networks under spike-time dependent plasticity. *PLoS Comput. Biol.* 17, e1008958
84. Kaleb, K. *et al.* (2021) Network-centered homeostasis through inhibition maintains hippocampal spatial map and cortical circuit function. *Cell Rep.* 36, 109577
85. Buzsáki, G. and Mizuseki, K. (2014) The log-dynamic brain: how skewed distributions affect network operations. *Nat. Rev. Neurosci.* 15, 264–278
86. Iascone, D.M. *et al.* (2020) Whole-neuron synaptic mapping reveals spatially precise excitatory / inhibitory balance limiting dendritic and somatic spiking. *Neuron* 106, 566–578
87. Kleindienst, T. *et al.* (2011) Activity-dependent clustering of functional synaptic inputs on developing hippocampal dendrites. *Neuron* 72, 1012–1024
88. Takahashi, N. *et al.* (2016) Active cortical dendrites modulate perception. *Science* 354, 1159–1165
89. Boivin, J.R. and Nedivi, E. (2018) Functional implications of inhibitory synapse placement on signal processing in pyramidal neuron dendrites. *Curr. Opin. Neurobiol.* 51, 16–22
90. Kirchner, J.H. and Gjorgjieva, J. (2021) Emergence of local and global synaptic organization on cortical dendrites. *Nat. Commun.* 12, 4005
91. Hu, H.Y. *et al.* (2019) Endocannabinoid signaling mediates local dendritic coordination between excitatory and inhibitory synapses. *Cell Rep.* 27, 666–675
92. Liu, G. (2004) Local structural balance and functional interaction of excitatory and inhibitory synapses in hippocampal dendrites. *Nat. Neurosci.* 7, 373–379
93. Pouille, F. and Scanziani, M. (2001) Enforcement of temporal fidelity in pyramidal cells by somatic feed-forward inhibition. *Science* 293, 1159–1163
94. Carvalho, T.P. and Buonomano, D.V. (2009) Differential effects of excitatory and inhibitory plasticity on synaptically driven neuronal input-output functions. *Neuron* 61, 774–785
95. Brunel, N. (2000) Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.* 8, 183–208
96. Cardin, J.A. (2018) Inhibitory interneurons regulate temporal precision and correlations in cortical circuits. *Trends Neurosci.* 41, 689–700
97. Stepp, N. *et al.* (2015) Synaptic plasticity enables adaptive self-tuning critical networks. *PLoS Comput. Biol.* 11, e1004043
98. Ma, Z. *et al.* (2019) Cortical circuit dynamics are homeostatically tuned to criticality in vivo. *Neuron* 104, 655–664
99. Duarte, R.C.F. and Morrison, A. (2014) Dynamic stability of sequential stimulus representations in adapting neuronal networks. *Front. Comput. Neurosci.* 8, 1066828
100. Larisch, R. *et al.* (2021) Sensory coding and contrast invariance emerge from the control of plastic inhibition over emergent selectivity. *PLoS Comput. Biol.* 17, e1009566
101. Buzsáki, G. and Draguhn, A. (2004) Neuronal oscillations in cortical networks. *Science* 304, 1926–1929
102. Van Vreeswijk, C. and Sompolinsky, H. (1996) Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* 274, 1724–1726
103. Wehr, M. and Zador, A.M. (2003) Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* 426, 442–446
104. Wilent, W.B. and Contreras, D. (2005) Dynamics of excitation and inhibition underlying stimulus selectivity in rat somatosensory cortex. *Nat. Neurosci.* 8, 1364–1370
105. Okun, M. and Lampl, I. (2008) Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nat. Neurosci.* 11, 535–537
106. Denève, S. and Machens, C.K. (2016) Efficient codes and balanced networks. *Nat. Neurosci.* 19, 375–382
107. Rupprecht, P. and Friedrich, R.W. (2018) Precise synaptic balance in the zebrafish homolog of olfactory cortex. *Neuron* 100, 669–683
108. Bhatia, A. *et al.* (2019) Precise excitation-inhibition balance controls gain and timing in the hippocampus. *eLife* 8, e43415
109. Paulsen, O. and Moser, E.I. (1998) A model of hippocampal memory encoding and retrieval: GABAergic control of synaptic plasticity. *Trends Neurosci.* 21, 273–278
110. Vogels, T.P. *et al.* (2013) Inhibitory synaptic plasticity: spike timing-dependence and putative network function. *Front. Neural Circ.* 7, 119
111. Hattori, R. *et al.* (2017) Functions and dysfunctions of neocortical inhibitory neuron subtypes. *Nat. Neurosci.* 20, 1199–1208
112. Steele, P.M. and Mauk, M.D. (1999) Inhibitory control of LTP and LTD: stability of synapse strength. *J. Neurophysiol.* 81, 1559–1566
113. Wilmes, K.A. *et al.* (2016) Inhibition as a binary switch for excitatory plasticity in pyramidal neurons. *PLoS Comput. Biol.* 12, e1004768
114. Wilmes, K.A. *et al.* (2017) Spike-timing dependent inhibitory plasticity to learn a selective gating of backpropagating action potentials. *Eur. J. Neurosci.* 45, 1032–1043
115. Paille, V. *et al.* (2013) GABAergic circuits control spike-timing-dependent plasticity. *J. Neurosci.* 33, 9353–9363
116. Letzkus, J.J. *et al.* (2015) Disinhibition, a circuit mechanism for associative learning and nemory. *Neuron* 88, 264–276
117. Krabbe, S. *et al.* (2019) Adaptive disinhibitory gating by VIP interneurons permits associative learning. *Nat. Neurosci.* 22, 1834–1843
118. Canto-Bustos, M. *et al.* (2022) Disinhibitory circuitry gates associative synaptic plasticity in olfactory cortex. *J. Neurosci.* 42, 2942–2950
119. Adler, A. *et al.* (2019) Somatostatin-expressing interneurons enable and maintain learning-dependent sequential activation of pyramidal neurons. *Neuron* 102, 202–216
120. Williams, L.E. and Holtmaat, A. (2019) Higher-order thalamocortical inputs gate synaptic long-term potentiation via disinhibition. *Neuron* 101, 91–102
121. Chiu, C.Q. *et al.* (2013) Compartmentalization of GABAergic inhibition by dendritic spines. *Science* 340, 759–763
122. Hayama, T. *et al.* (2013) GABA promotes the competitive selection of dendritic spines by controlling local Ca2+ signaling. *Nat. Neurosci.* 16, 1409–1416
123. Müllner, F.E. *et al.* (2015) Precision of Inhibition Dendritic inhibition: dendritic inhibition by individual GABAergic synapses on hippocampal pyramidal cells is confined in space and time. *Neuron* 87, 576–589
124. Agnes, E.J. and Vogels, T.P. (2021) Interacting synapses stabilise both learning and neuronal dynamics in biological networks. *bioRxiv* Published online April 4, 2021. https://doi.org/10.1101/2021.04.01.437962
125. Hiratani, N. and Fukai, T. (2017) Detailed dendritic excitatory/inhibitory balance through heterosynaptic spike-timing-dependent plasticity. *J. Neurosci.* 37, 12106–12122
126. Mikulasch, F.A. *et al.* (2021) Local dendritic balance enables learning of efficient representations in networks of spiking neurons. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2021925118
127. Zhang, S. *et al.* (2014) Long-range and local circuits for top-down modulation of visual cortex processing. *Science* 345, 660–665
128. Maes, A. *et al.* (2020) Learning spatiotemporal signals using a recurrent spiking network that discretizes time. *PLoS Comput. Biol.* 16, e1007606
129. Buzsáki, G. (2010) Neural syntax: cell assemblies, synapsembles, and readers. *Neuron* 68, 362–385
130. Carrillo-Reid, L. and Yuste, R. (2020) Playing the piano with the cortex: role of neuronal ensembles and pattern completion in perception and behavior. *Curr. Opin. Neurobiol.* 64, 89–95
131. Fino, E. and Yuste, R. (2011) Dense inhibitory connectivity in neocortex. *Neuron* 69, 1188–1203
132. Znamenskiy, P. *et al.* (2018) Functional selectivity and specific connectivity of inhibitory neurons in primary visual cortex. *bioRxiv* Published online April 4, 2018. https://doi.org/10.1101/294835

133. Miehl, C. *et al.* (2022) Formation and computational implications of assemblies in neural circuits. *J. Physiol.* Published online September 6, 2022. https://doi.org/10.1113/JP282750
134. Mackwood, O. *et al.* (2021) Learning excitatory-inhibitory neuronal assemblies in recurrent networks. *eLife* 10, e59715
135. Rost, T. *et al.* (2018) Winnerless competition in clustered balanced networks: inhibitory assemblies do the trick. *Biol. Cybern.* 112, 81–98
136. Najafi, F. *et al.* (2020) Excitatory and inhibitory subnetworks are equally selective during decision-making and emerge simultaneously during learning. *Neuron* 105, 165–179
137. Herpich, J. and Tetzlaff, C. (2019) Principles underlying the input-dependent formation and organization of memories. *Netw. Neurosci.* 3, 606–634
138. Barron, H.C. *et al.* (2016) Unmasking latent inhibitory connections in human cortex to reveal dormant cortical memories. *Neuron* 90, 191–203
139. Barron, H.C. *et al.* (2017) Inhibitory engrams in perception and memory. *Proc. Natl. Acad. Sci.* 114, 6666–6674
140. Koolschijn, R.S. *et al.* (2019) The hippocampus and neocortical inhibitory engrams protect against memory interference. *Neuron* 101, 528–541
141. Barron, H.C. (2021) Neural inhibition for continual learning and memory. *Curr. Opin. Neurobiol.* 67, 85–94
142. Ramaswami, M. (2014) Network plasticity in adaptive filtering and behavioral habituation. *Neuron* 82, 1216–1229
143. Natan, R.G. *et al.* (2017) Cortical interneurons differentially shape frequency tuning following adaptation. *Cell Rep.* 21, 878–890
144. Schulz, A. *et al.* (2021) The generation of cortical novelty responses through inhibitory plasticity. *eLife* 10, e65309
145. Ulanovsky, N. *et al.* (2003) Processing of low-probability sounds by cortical neurons. *Nat. Neurosci.* 6, 391–398
146. Chen, S.X. *et al.* (2015) Subtype-specific plasticity of inhibitory circuits in motor cortex during motor learning. *Nat. Neurosci.* 18, 1109–1115
147. Song, S. *et al.* (2022) Input-specific inhibitory plasticity improves decision accuracy under noise. *bioRxiv* Published online May 25, 2022. https://doi.org/10.1101/2022.05.24.493332
148. Freund, T.F. and Katona, I. (2007) Perisomatic inhibition. *Neuron* 56, 33–42
149. Abs, E. *et al.* (2018) Learning-related plasticity in dendrite-targeting layer 1 interneurons. *Neuron* 100, 684–699
150. Loomba, S. *et al.* (2022) Connectomic comparison of mouse and human cortex. *Science* 377, eabo0924
151. Kim, R. and Sejnowski, T.J. (2021) Strong inhibitory signaling underlies stable temporal dynamics and working memory in spiking neural networks. *Nat. Neurosci.* 24, 129–139
152. Mongillo, G. *et al.* (2018) Inhibitory connectivity defines the realm of excitatory plasticity. *Nat. Neurosci.* 21, 1463–1470
153. Wilmes, K.A. and Clopath, C. (2019) Inhibitory microcircuits for top-down plasticity of sensory rep-resentations. *Nat. Commun.* 10, 5055
154. Confavreux, B. *et al.* (2020) A meta-learning approach to (re) discover plasticity rules that carve a desired function into a neural network. *bioRxiv* Published online October 25, 2020. https://doi.org/10.1101/2020.10.24.353409
155. Keijser, J. and Sprekeler, H. (2022) Optimizing interneuron circuits for compartment-specific feedback inhibition. *PLoS Comput. Biol.* 18, 1–21
156. Soldado-Magraner, S. *et al.* (2021) Orchestrated excitatory and inhibitory learning rules lead to the unsupervised emergence of self-sustained and inhibition-stabilized dynamics. *bioRxiv* Published online September 13, 2021. https://doi.org/10.1101/2020.12.30.424888
157. Pfeffer, C.K. *et al.* (2013) Inhibition of inhibition in visual cortex the logic of connections between molecularly distinct interneurons. *Nat. Neurosci.* 16, 1068–1076
158. Jiang, X. *et al.* (2015) Principles of connectivity among morphologically defined cell types in adult neocortex. *Science* 350, aac9462
159. Hartung, J. and Letzkus, J.J. (2021) Inhibitory plasticity in layer 1 – dynamic gatekeeper of neocortical associations. *Curr. Opin. Neurobiol.* 67, 26–33
160. Fu, Y. *et al.* (2014) A cortical circuit for gain control by behavioral state. *Cell* 156, 1139–1152
161. Attinger, A. *et al.* (2017) Visuomotor coupling shapes the functional development of mouse visual cortex. *Cell* 169, 1291–1302.e14
162. Campagnola, L. *et al.* (2022) Local connectivity and synaptic dynamics in mouse and human neocortex. *Science* 375, eabj5861

IV. Heterosynaptic Plasticity Determines the Set Point for Cortical Excitatory-Inhibitory Balance

Field*, R.E., D'amour*, J.A., Tremblay, R., Rudy, B., Miehl, C., Gjorgjieva, J. & Froemke, R.C. Heterosynaptic Plasticity Determines the Set Point for Cortical Excitatory-Inhibitory Balance. *Neuron* **106(5)**, 842-854 (2020). https://doi.org/10.1016/j.neuron.2020.03.002

Article

Neuron

Heterosynaptic Plasticity Determines the Set Point for Cortical Excitatory-Inhibitory Balance

Highlights

- Spike pairing induces heterosynaptic excitatory and inhibitory plasticity
- Heterosynaptic plasticity helps adjust excitatory-inhibitory correlations
- Heterosynaptic plasticity determines the set point for excitatory-inhibitory balance
- Input strength can be postsynaptically computed and specifically modified

Authors

Rachel E. Field, James A. D'amour, Robin Tremblay, Christoph Miehl, Bernardo Rudy, Julijana Gjorgjieva, Robert C. Froemke

Correspondence

robert.froemke@med.nyu.edu

In Brief

After induction of synaptic plasticity at specific inputs, Field et al. found that coordinated changes occur across multiple inhibitory and excitatory inputs onto cortical pyramidal neurons. The relative timing and degree of heterosynaptic plasticity determines the overall excitatory-inhibitory correlation, i.e., the set point for excitatory-inhibitory balance.

Field et al., 2020, Neuron *106*, 1–13
June 3, 2020
https://doi.org/10.1016/j.neuron.2020.03.002

CellPress

Please cite this article in press as: Field et al., Heterosynaptic Plasticity Determines the Set Point for Cortical Excitatory-Inhibitory Balance, Neuron (2020), https://doi.org/10.1016/j.neuron.2020.03.002

Neuron
Article

Heterosynaptic Plasticity Determines the Set Point for Cortical Excitatory-Inhibitory Balance

Rachel E. Field,[1,2,3,4,9] James A. D'amour,[1,2,3,4,9] Robin Tremblay,[2,4,5] Christoph Miehl,[6,7] Bernardo Rudy,[2,4,5] Julijana Gjorgjieva,[6,7] and Robert C. Froemke[1,2,3,4,8,10,*]
[1]Skirball Institute for Biomolecular Medicine, New York University School of Medicine, New York, NY 10016, USA
[2]Neuroscience Institute, New York University School of Medicine, New York, NY 10016, USA
[3]Department of Otolaryngology, New York University School of Medicine, New York, NY 10016, USA
[4]Department of Neuroscience and Physiology, New York University School of Medicine, New York, NY 10016, USA
[5]Department of Anesthesiology, New York University School of Medicine, New York, NY 10016, USA
[6]Max Planck Institute for Brain Research, 60438 Frankfurt, Germany
[7]School of Life Sciences, Technical University of Munich, 85354 Freising, Germany
[8]Center for Neural Science, New York University, New York, NY 10003, USA
[9]These authors contributed equally
[10]Lead Contact
*Correspondence: robert.froemke@med.nyu.edu
https://doi.org/10.1016/j.neuron.2020.03.002

SUMMARY

Excitation in neural circuits must be carefully controlled by inhibition to regulate information processing and network excitability. During development, cortical inhibitory and excitatory inputs are initially mismatched but become co-tuned or balanced with experience. However, little is known about how excitatory-inhibitory balance is defined at most synapses or about the mechanisms for establishing or maintaining this balance at specific set points. Here we show how coordinated long-term plasticity calibrates populations of excitatory-inhibitory inputs onto mouse auditory cortical pyramidal neurons. Pairing pre- and postsynaptic activity induced plasticity at paired inputs and different forms of heterosynaptic plasticity at the strongest unpaired synapses, which required minutes of activity and dendritic Ca^{2+} signaling to be computed. Theoretical analyses demonstrated how the relative rate of heterosynaptic plasticity could normalize and stabilize synaptic strengths to achieve any possible excitatory-inhibitory correlation. Thus, excitatory-inhibitory balance is dynamic and cell specific, determined by distinct plasticity rules across multiple excitatory and inhibitory synapses.

INTRODUCTION

In mature cortical networks and elsewhere throughout the adult nervous system, excitation is regulated by a complex set of inhibitory circuits. γ-aminobutyric acid-ergic (GABAergic) inhibition is important for many functions, including spike generation, dendritic integration, synaptic plasticity, sleep, learning, and prevention of pathological activity such as epilepsy (Cossart et al., 2001; Hattori et al., 2017; Isaacson and Scanziani, 2011; Oliveira et al., 2011; Scharfman and Brooks-Kayal, 2014). This requires inhibitory synapses to be calibrated or balanced with the relative strengths of excitatory synapses to ensure that neurons and networks are neither hypo- nor hyper-excitable for prolonged periods. Although the term "excitatory-inhibitory balance" is widely used, it has been difficult to precisely define. In particular, implicit in the concept of balance is a stable set point to which synaptic strengths and/or network activity returns via negative feedback after disruptions of excitability (including positive feedback processes such as excitatory plasticity).

Excitatory-inhibitory balance has been quantified as correlation between excitation and inhibition over a stimulus dimension, such as visual orientation or sound frequencies, or the temporal correlation between patterns of excitation and inhibition. The term "balance" suggests near-perfect matching between excitation and inhibition, and this has been observed experimentally in some systems (Tan and Wehr, 2009), but not in every case. Even in mature circuits (Dorrn et al., 2010; Marlin et al., 2015; Okun and Lampl, 2008; Wehr and Zador, 2003), correlation values are not always perfect (i.e., linear correlation coefficient r: 1.0) but instead are often lower (r: 0.4–0.9). It is unclear whether it is difficult to maintain higher levels of balance in biological neural networks or whether the set point at which excitation and inhibition are in equilibrium is actively maintained at a lower level.

In sensory cortex, inhibitory responses and excitatory-inhibitory balance are established during early postnatal development (Cai et al., 2018; Dorrn et al., 2010; Gandhi et al., 2008; House et al., 2011; Kuhlman et al., 2013; Takesian and Hensch, 2013). Excitatory-inhibitory balance must also be dynamically maintained throughout life, because experience-dependent modification of excitatory synapses requires corresponding changes to inhibition (Dorrn et al., 2010; Froemke, 2015; House et al., 2011; Kuhlman et al., 2013). Computational studies supported by experimental data indicate that disruptions of excitatory-inhibitory balance can rapidly produce epileptiform activity and

seizures (Avoli et al., 2016; Cossart et al., 2001; Dehghani et al., 2016; Ren et al., 2014; Toader et al., 2013), meaning that compensatory mechanisms need to act quickly to re-stabilize neural circuits before pathological activity emerges. At least some homeostatic adjustments take place over hours to days (Lissin et al., 1998; Thiagarajan et al., 2005; Turrigiano et al., 1998; Turrigiano, 2008). It remains unclear whether these processes could correct for changes in excitability on shorter timescales of activity-dependent plasticity (seconds to minutes) in the input-specific manner required to preserve or promote differential computations. This may depend on different set points for excitatory-inhibitory balance, based on the function of the neuron or neural circuit (e.g., single-spike firing versus bursting or narrow versus broad stimulus feature tuning).

An alternative to regulating overall excitability is heterosynaptic plasticity, defined as modifications to inputs not activated during induction of long-term potentiation (LTP) or other forms of long-term plasticity triggered at specific inputs (Chistiakova et al., 2015; Froemke, 2015; Hiratani and Fukai, 2017; Zenke et al., 2017). Heterosynaptic modifications at specific inputs have been observed after excitatory LTP at paired homosynaptic sites (Basu et al., 2016; Christie and Abraham, 1992; Lynch et al., 1977; Muller et al., 1995; Royer and Pare, 2003; Scanziani et al., 1996), including *in vivo*, where these changes affect cortical receptive fields (Dorrn et al., 2010; Froemke et al., 2013) at specific identifiable inputs (El-Boustani et al., 2018). It is unclear whether inhibitory synapses also undergo heterosynaptic modifications or how changes across multiple inputs might be coordinated to alter excitatory-inhibitory balance. Recently, we showed that spike-timing-dependent plasticity (STDP) could be induced at co-activated excitatory and inhibitory synapses (D'amour and Froemke, 2015). Spike pairing induced excitatory and inhibitory LTP, with the degree of inhibitory potentiation depending on the initial amplitude of co-evoked excitatory events. Similar forms of inhibitory plasticity that require activation of excitatory synapses and NMDA receptors have been described in cortex and hippocampus (Chiu et al., 2018; Horn and Nicoll, 2018; Huang et al., 2005). This naturally led to a normalization of the excitation-inhibition ratio at the paired inputs.

Here we ask whether spike pairing also induces heterosynaptic plasticity and whether these changes affect overall organization of excitation and inhibition. If so, inducing synaptic modifications could be used as a bidirectional perturbation to determine the set points for excitatory-inhibitory balance. We aimed to determine the learning rules by which populations of excitatory and inhibitory inputs could be collectively modified, the mechanisms for these changes, and the degree of excitatory-inhibitory co-tuning that could be achieved.

RESULTS

Spike Pairing Induces STDP and Heterosynaptic Excitatory and Inhibitory Plasticity

To examine how homosynaptic and heterosynaptic modifications might synergistically affect cortical excitatory-inhibitory balance, we made 177 whole-cell recordings from layer 5 pyramidal neurons in slices of auditory cortex of young and adult mice. A stimulation electrode array was placed in layer 4 and used to sequentially evoke 4–8 sets of excitatory postsynaptic currents (EPSCs) and inhibitory postsynaptic currents (IPSCs) recorded in voltage clamp (Figure 1A). This recruited separate populations of excitatory and inhibitory presynaptic inputs with a low degree of overlap across channels (Figures 1B and S1), mimicking recruitment of thalamocortical inputs onto cortical neurons *in vivo* by sensory stimulation (Froemke et al., 2007; Hackett et al., 2011; Lee et al., 2004; Miller et al., 2001). The apparent overlap seemed to result mainly from activation of dendritic conductances that led to sublinear summation (Froemke et al., 2010b; Rosenkranz, 2012; Tran-Van-Minh et al., 2015; Urban and Barrionuevo, 1998), instead of shared presynaptic inputs across channels (Figure S1). After measuring baseline events for 5–20 min, recordings were switched to current clamp to pair inputs evoked by one channel with single postsynaptic spikes (Bi and Poo, 1998; D'amour and Froemke, 2015; Feldman, 2000). Other stimulation channels were not activated during pairing. Following pairing, we resumed sequential stimulation of all channels and monitored paired and unpaired EPSCs and IPSCs for >16 min.

Pairing pre- and postsynaptic activity induced long-term synaptic modifications at multiple inputs, including inputs not activated during pairing. Some of these changes could be variable from cell to cell, but we consistently found that the strongest unpaired excitatory and inhibitory inputs (the original best inputs) were specifically modified minutes after pairing. For example, in the recording shown in Figure 1C, repetitively pairing presynaptic activation of channel S4 with postsynaptic spiking (pre→post pairing) induced excitatory and inhibitory LTP at the paired channel (Figure 1C, red), whereas the original best unpaired inputs (excitation at S3 and inhibition at S2) were both depressed (Figure 1C, blue). On average, other unpaired inputs were not substantially affected (Figure 1C, black). Thus, spike pairing induces rapid and specific heterosynaptic modifications, in addition to STDP at paired (homosynaptic) inputs.

These selective modifications to the paired and original best inputs acted together to reorganize the overall profile of excitation and inhibition (i.e., the excitatory-inhibitory balance). As a metric of excitatory-inhibitory balance, we used the linear correlation coefficient r_{ei} of EPSCs and IPSCs evoked across stimulation channels. Linear correlation has previously been used to quantify excitatory-inhibitory balance *in vivo* (Dorrn et al., 2010; Higley and Contreras, 2006; Okun and Lampl, 2008; Tan and Wehr, 2009; Wehr and Zador, 2003) and *in vitro* (Graupner and Reyes, 2013; Xue et al., 2014). For this cell, initial IPSC amplitudes were mostly unrelated to EPSCs across stimulation channels (Figure 1D, left, r_{ei}-before: 0.25). This was unsurprising, because *a priori*, excitatory and inhibitory synapses activated by extracellular stimulation need not be functionally related despite spatial proximity near each electrode. However, correlation increased after pairing as EPSCs and IPSCs evoked by each stimulation site became more similar across channels (Figure 1D, right, r_{ei}-after: 0.48). This was a consequence of coordinated modifications to the paired input (Figure 1D, red arrow) and original best unpaired inputs (Figure 1D, blue arrowheads). Such activity-dependent changes over multiple paired and unpaired synapses, which collectively act to improve excitatory-inhibitory balance, are similar to experience-dependent

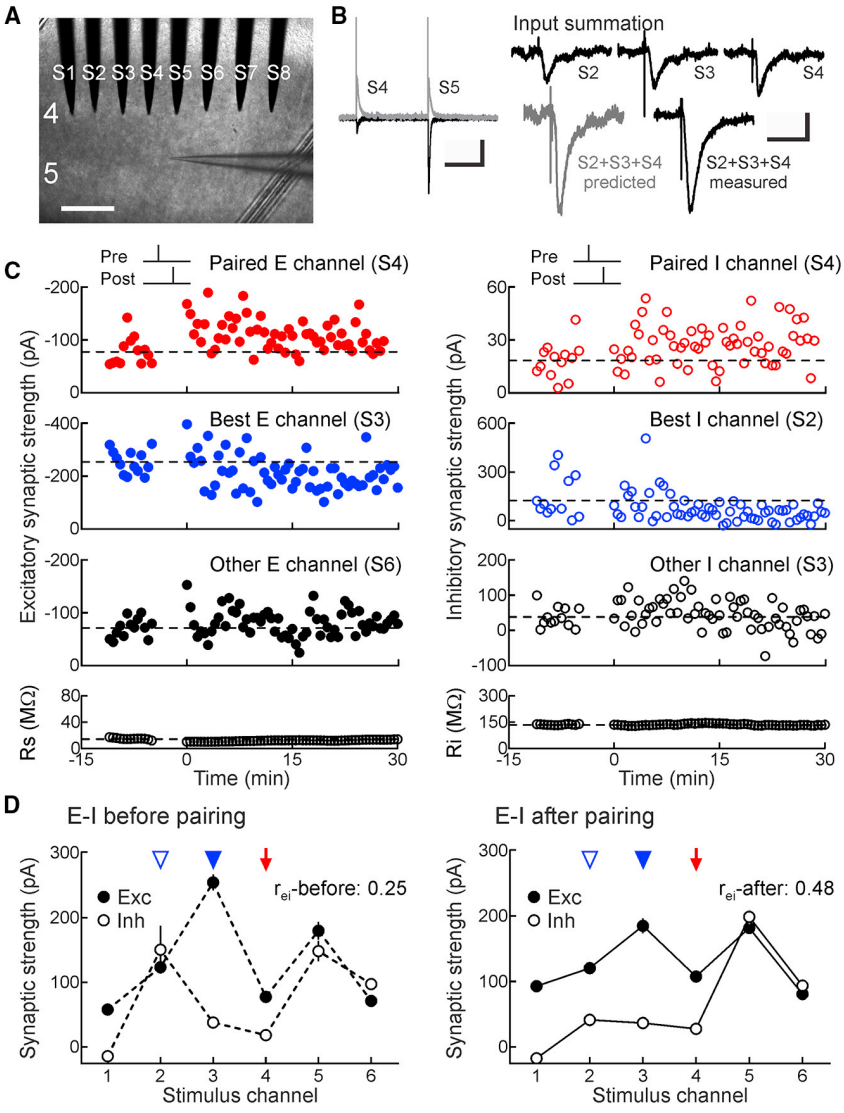


Figure 1. Spike Pairing Modifies Excitation and Inhibition at Paired and Unpaired Inputs

(A) Whole-cell recordings from mouse auditory cortical layer 5 pyramidal cells in slices with 8-electrode stimulation array (channels S1–S8) in layer 4. Scale, 250 μm.

(B) Left: baseline and postpairing EPSCs at −70 mV (black) and IPSCs at −30 mV (gray). Scale: 500 ms, 200 pA. Right: input summation, measuring inputs S3, S4, and S5 separately and together; predicted versus measured summed response. Scale: 50 ms, 100 pA.

(C) Strengths of multiple excitatory (left) and inhibitory inputs (right) onto the same neuron before and after pairing one channel with postsynaptic spiking. Top: excitatory and inhibitory plasticity induced by pre→post pairing at channel S4 (red, Δt: 0.5 ms). Dashed line, prepairing mean. Top middle: heterosynaptic LTD at the strongest unpaired inputs (blue). Bottom middle: other inputs (black). Bottom: series and input resistance.

(D) Increased excitatory-inhibitory balance after pairing; same cell as (C). Excitatory-inhibitory correlation before pairing (r_{ei}-before: 0.25, dashed lines) and after pairing (r_{ei}-after: 0.48, solid lines). Red arrow, paired channel. Blue arrowheads, original best excitation (filled) and inhibition (open). Error bars, SEM.

changes to excitatory and inhibitory synaptic tuning curves in young rodent auditory cortex *in vivo* (Dorrn et al., 2010).

The relative timing of pre- and postsynaptic spiking during pairing determined the sign of heterosynaptic plasticity at the original best inputs. In 25 recordings from developing auditory cortex (post-natal day [P] 12–P26), pre→post pairing induced LTP at paired inputs, together with heterosynaptic long-term depression (LTD) at the original best excitatory and inhibitory inputs (Figures 2A, S2A, and S3A). Although from cell to cell and channel to channel there could be some change to unpaired inputs, no other systematic changes after pairing were detected (Figure 2A, bottom), and such plasticity was unrelated to the degree of apparent channel overlap (Figure S4), indicating that there may be other forms of input-specific heterosynaptic plasticity. In 11 other recordings from young auditory cortex, post→pre pairing induced excitatory LTD and inhibitory LTP at paired inputs, together with heterosynaptic LTP at original best excitatory and inhibitory inputs (Figures 2B, S2B, and S3B). Because pre→post pairing potentiates paired inhibitory inputs, heterosynaptic inhibitory LTD provides a mechanism for bi-directional regulation of inhibitory synaptic strength. Furthermore, heterosynaptic excitatory LTP might compensate for reductions in excitability after homosynaptic LTD at the paired excitatory input.

Heterosynaptic Plasticity Normalizes Excitatory-Inhibitory Correlation

These coordinated synaptic modifications, induced by either pre→post or post→pre pairing, affected overall excitatory-inhibitory correlation r_{ei} in similar ways. When the correlation coefficient was initially low in developing cortex (r_{ei}-before < 0.4), correlation increased after either pre→post or post→pre pairing (Figures 2C and S2). However, when the excitatory-inhibitory correlation was initially high (r_{ei}-before > 0.4), correlation instead decreased after pairing (Figures 2C and S3). In the absence of postsynaptic spiking, no STDP was induced, and excitatory-inhibitory correlation was unchanged regardless as of the initial correlation value (Figure 2C, bottom, no pairing).

Changes in excitatory-inhibitory correlation mainly resulted from heterosynaptic modifications, especially when initial correlation was low. Computing r_{ei}-after assuming only modifications of paired inputs led to smaller correlation changes than only considering modifications to unpaired inputs (Figure 2D). Despite EPSC/IPSC amplitude variability from event to event, correlation values were consistent during the first half versus the second half of the baseline period, as well as the first half

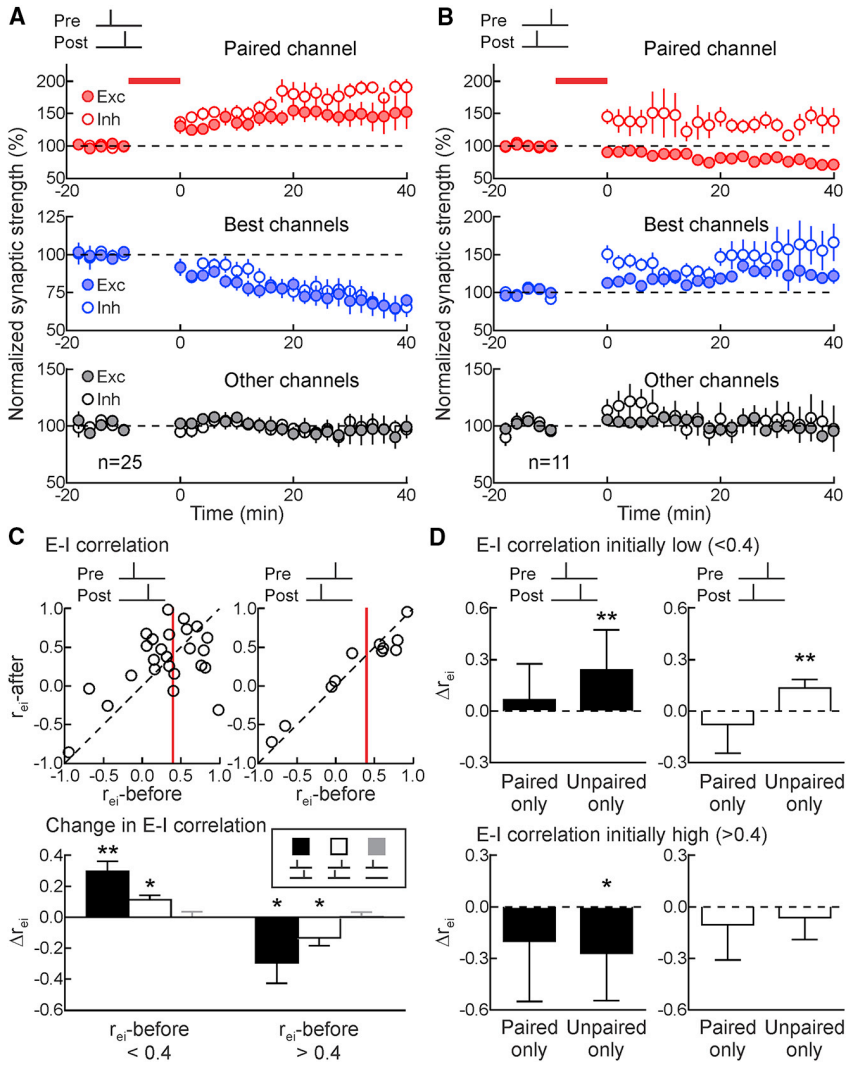


Figure 2. Heterosynaptic Plasticity Normalizes Excitatory-Inhibitory Correlation

(A) Summary of pre→post experiments on paired inputs (top, red; paired EPSCs increased 40.3% ± 10.5% 16–25 min postpairing, n = 25, $p < 0.0009$, Student's paired two-tailed t test, 18/25 cells with significant excitatory LTP; paired IPSCs increased 53.7% ± 13.9%, $p < 0.0008$, 19/25 cells with significant inhibitory LTP), original best inputs (middle, blue; originally largest EPSCs decreased −21.7% ± 4.1%, $p < 10^{-4}$, 21/25 cells with significant heterosynaptic excitatory LTD; originally largest IPSCs decreased −15.4% ± 6.0%, $p < 0.02$, 16/25 cells with significant heterosynaptic inhibitory LTD), and other unpaired inputs (bottom, black; EPSCs increased by 1.4% ± 8.0%, $p > 0.8$; IPSCs increased by 0.7% ± 4.6%, $p > 0.8$). Filled symbols, excitation; open symbols, inhibition. Error bars, SEM.

(B) Summary of post→pre pairing experiments on paired inputs (paired EPSCs decreased −17.0% ± 6.4%, n = 11, $p < 0.03$, 9/11 cells with significant excitatory LTD; paired IPSCs increased 37.9% ± 12.4%, $p < 0.02$, 8/11 cells with significant inhibitory LTP), original best inputs (originally largest EPSCs increased 15.6% ± 4.4%, $p < 0.006$, 7/11 cells with significant heterosynaptic excitatory LTP; originally largest IPSCs increased 25.1% ± 8.6%, $p < 0.02$, 7/11 cells with significant heterosynaptic inhibitory LTP), and other unpaired inputs (bottom, black; EPSCs increased 4.1% ± 5.1%, $p > 0.4$; IPSCs increased 2.7% ± 11.5%, $p > 0.8$). Error bars, SEM.

(C) Normalization of excitatory-inhibitory correlation after pairing. Top, r_{ei}-before versus r_{ei}-after, pre→post pairing (left, n = 25) or post→pre pairing (n = 11). Red line, r_{ei}: 0.4. Bottom, changes in excitatory-inhibitory correlation after pairing (Δr_{ei}; when initially $r < 0.4$ for pre→post pairing: 0.30 ± 0.06, n = 14, $p < 0.0005$; post→pre pairing: 0.11 ± 0.03, n = 5, $p < 0.02$; and no postsynaptic spiking: 0.002 ± 0.03, n = 5, $p > 0.9$; Student's paired two-tailed t test) (when initially $r > 0.4$ for pre→post pairing: −0.29 ± 0.13, n = 11, $p < 0.05$; post→pre pairing: −0.13 ± 0.05, n = 6, $p < 0.05$; and no pairing controls without postsynaptic spiking: 0.006 ± 0.03, n = 10, $p > 0.8$).

(D) Heterosynaptic modifications to unpaired inputs refined excitatory-inhibitory balance. Considered separately, plasticity only at paired inputs was less effective than changes to remaining inputs (paired only; Δr_{ei} when initially $r < 0.4$ for pre→post pairing: 0.07 ± 0.06, n = 14, $p > 0.2$, and post→pre pairing: −0.08 ± 0.08, n = 5, $p > 0.3$, and Δr_{ei} when initially $r > 0.4$ for pre→post pairing: −0.20 ± 0.11, n = 11, $p > 0.1$, and post→pre pairing: −0.10 ± 0.08, n = 6, $p > 0.2$; Student's paired two-tailed t test) (unpaired only; Δr_{ei} when initially $r < 0.4$ for pre→post pairing: 0.24 ± 0.07, $p < 0.005$, and post→pre pairing: 0.13 ± 0.02, $p < 0.005$, and when initially $r > 0.4$ for pre→post pairing: −0.27 ± 0.09, $p < 0.02$, but not post→pre pairing: −0.06 ± 0.05, $p > 0.2$). *$p < 0.05$; **$p < 0.01$. Error bars, SEM.

versus the second half of the postpairing period (Figure S5). This indicates that the change in correlation is not simply regression to the mean but rather a specific consequence of synaptic modifications and directed toward a certain value.

We asked what would happen if pairing was performed at original best inputs (Figure S6). Homosynaptic and heterosynaptic modifications might nullify each other, or perhaps one form of plasticity might win out; either case would inform models relating plasticity rules to excitatory-inhibitory correlation. After pre→post pairing, paired inhibition reliably increased, whereas changes to excitation were more variable (Figures S6A and S6C). By contrast, post→pre pairing led to significant excitatory LTD and inhibitory LTP at original best/paired inputs (Figures S6B and S6C). The second-best but unpaired excitatory and inhibitory inputs were unchanged, indicating that heterosynaptic modifications were not differentially engaged at other inputs instead (Figure S6C). These changes after post→pre pairing did not affect overall correlation r_{ei} (Figure S6D). However, in the absence of other reliable heterosynaptic changes, pre→post pairing at original best inputs greatly increased r_{ei}, beyond the nominal level of 0.4 usually observed at these ages. For 7/9 recordings, r_{ei}-before began <0.7; in each case, after changes predominantly to homosynaptic inputs, r_{ei} increased by 0.36 ± 0.14 ($p < 0.04$).

Thus, spike pairing rapidly induces heterosynaptic plasticity to effectively normalize excitatory-inhibitory balance in developing auditory cortex, adjusting the relation of inhibition to excitation to

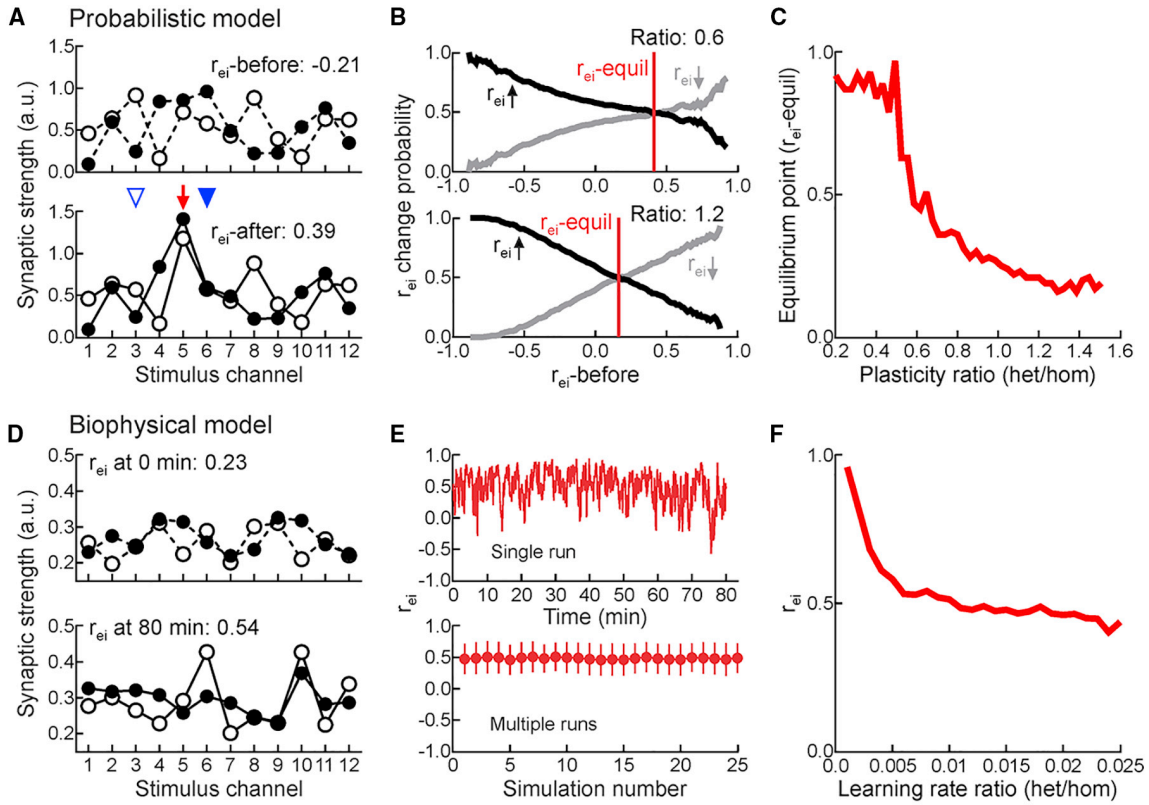


Figure 3. Heterosynaptic Plasticity Determines the Set Point for Excitatory-Inhibitory Balance

(A) Example tuning curves for a probabilistic model before and after synaptic weight adjustment.

(B) Results of all simulations; probability of r_{ei} increasing (black) or decreasing (gray) after plasticity as function of initial correlation. Where lines cross at probability 0.5 is the equilibrium point (r_{ei}-equil) at which homosynaptic and heterosynaptic plasticity are balanced and r_{ei} values stabilize. Top, ratio of heterosynaptic to homosynaptic plasticity: 0.6. Bottom, plasticity ratio: 1.2.

(C) Equilibrium point (r_{ei}-equil) as a function of the heterosynaptic-to-homosynaptic plasticity ratio.

(D) Example tuning curves for a biophysical model of plasticity at time 0 and after 80 min.

(E) r_{ei} over time during a single simulation (top) and mean r_{ei} for 25 different tuning curve initializations (bottom). Ratio of heterosynaptic to homosynaptic learning rates: $(\eta^E_{het}/\eta^E_w) = 1.3 * 10^{-2}$ ($\eta^E_{het} = 1.3 * 10^{-5} ms^{-1}$ and $\eta^I_{het} = 1.3 * 10^{-4}\ ms^{-1}$). Error bars, SD.

(F) r_{ei} depends on the excitatory heterosynaptic-to-homosynaptic learning rate ratio (η^E_{het}/η^E_w).

promote correlation of ~0.4. This value is close to that observed in rat auditory cortex *in vivo* during the critical period for frequency tuning (Dorrn et al., 2010), suggesting this value is a set point actively maintained by an orchestrated array of plasticity mechanisms during this stage of cortical development. Intuitively, when the excitatory-inhibitory correlation was initially low, this occurred at least partly because the original best excitatory and inhibitory inputs were activated by different channels (in 12/14 pre→post and 5/5 post→pre pairing recordings). Heterosynaptic plasticity at the best excitatory and inhibitory inputs would naturally make those inputs more similar, because they were both depressed after pre→post pairing and potentiated after post→pre pairing. Moreover, when excitatory-inhibitory correlation was initially too high, changes to the paired channel normalized the correlation. These results show that single neurons have mechanisms for sensing and selectively modifying input strengths to achieve a range of excitatory-inhibitory co-tuning. It may be computationally advantageous to not perfectly match excitation and inhibition, especially during developmental critical periods when cortical plasticity is important for initializing sensory processing circuits.

Heterosynaptic Plasticity Determines the Set Point for Excitatory-Inhibitory Balance

To quantitatively assess this capacity in a theoretical framework, we simulated homosynaptic and heterosynaptic plasticity onto a model postsynaptic neuron driven by 12 excitatory and inhibitory inputs. We first considered the effects of pre→post pairing in a probabilistic model, in which 50,000 excitatory and inhibitory tuning curves were generated randomly by sampling from a uniform distribution across channels (Figure 3A, r_{ei}-before). This resulted in initial correlation r_{ei}-before values ranging from −0.9 to 0.9. One channel was chosen as the paired channel (excitation and inhibition were increased), and the original best excitatory and inhibitory channels were decreased by a fixed amount (Figure 3A, r_{ei}-after). The degree of homosynaptic plasticity was similar to the experimentally measured increase (~65%; Figure 2A), whereas the magnitude of simulated heterosynaptic

plasticity varied across different runs of the model (decreasing between −14% and −98%). Following weight modification, we recomputed excitatory-inhibitory correlation r_{ei} across channels. As expected, the probability of r_{ei} increasing or decreasing strongly depended on the initial correlation r_{ei}-before. When homosynaptic plasticity was stronger than the heterosynaptic changes, the probability of r_{ei} increasing was higher than the probability of decreasing. However, with sufficiently strong heterosynaptic plasticity, a crossover occurred between the probability of r_{ei} increasing and the probability of r_{ei} decreasing. This value of the ratio between heterosynaptic and homosynaptic plasticity is an equilibrium point at which excitatory-inhibitory correlation would eventually settle as increases and decreases of r_{ei} were balanced (Figure 3B). As in the experiments (Figure 2), correlation values initially higher than this set point (r_{ei}-equil) were likely to decrease, whereas correlation values initially lower than r_{ei}-equil were more likely to increase. The main influence on r_{ei}-equil was determined by the strength of heterosynaptic relative to homosynaptic plasticity (Figure 3C). This equilibrium point decreased as heterosynaptic plasticity strength was increased relative to homosynaptic plasticity strength. Thus, by titrating the relative strengths of heterosynaptic and homosynaptic plasticity, the system can in principle achieve nearly any correlation value, i.e., an arbitrary set point for stable excitatory-inhibitory balance.

To ask whether this relationship between excitatory-inhibitory correlation and relative strengths of heterosynaptic versus homosynaptic plasticity holds under more realistic conditions and over multiple consecutive pairings, we simulated a single postsynaptic integrate-and-fire neuron driven by 12 excitatory and inhibitory input channels. Each channel consisted of 10 excitatory and 10 inhibitory presynaptic conductance-based inputs, with weights modified by homosynaptic versus heterosynaptic activity-dependent plasticity (Figures 3D and S7A). During the simulation, we made paired and unpaired channels fire at different rates to elicit postsynaptic spiking only during paired channel activation. Homosynaptic and heterosynaptic plasticity were implemented with biophysical traces that tracked pre- and postsynaptic activation online, and we presented an alternating sequence of consecutive paired and unpaired stimulation phases. Despite high correlation variability during the simulation, r_{ei} fluctuated around a constant mean (Figure 3E, top), consistent across different initial conditions (Figure 3E, bottom). This finding indicates that heterosynaptic plasticity can normalize excitatory-inhibitory correlation over the course of multiple pairings. As indicated in the probabilistic model (Figures 3A–3C), excitatory-inhibitory correlation converged to a value that depended on the relative learning rates of heterosynaptic versus homosynaptic plasticity (Figure 3F).

In particular, when homosynaptic plasticity was dominant (i.e., the homosynaptic learning rate was faster than the heterosynaptic rate), r_{ei} was high and the excitatory and inhibitory weights gradually increased over the simulation. In contrast, when heterosynaptic plasticity was dominant, r_{ei} was low and the excitatory and inhibitory weights during training gradually decreased. Reducing the rate of homosynaptic LTD also led to dominance of homosynaptic LTP and higher r_{ei} set points (Figure S7B). When the effective strengths (i.e., rates) of homosynaptic and heterosynaptic plasticity were approximately balanced, excitatory and inhibitory weights were relatively stable during an extended period of training (Figure S7C) and r_{ei} converged to 0.45–0.5, close to the values observed experimentally. Note that balanced rates here means that the heterosynaptic modifications are necessarily slower than homosynaptic changes. These simulations demonstrate that heterosynaptic plasticity can powerfully control the positive feedback of homosynaptic plasticity and achieve a range of possible correlation r_{ei} values by simply adjusting the degree of heterosynaptic modifications relative to homosynaptic plasticity.

Plasticity Rates Determine the Excitatory-Inhibitory Set Point in Young and Adult Cortex

This model predicts that homosynaptic and heterosynaptic plasticity learning rates are dissociable and affect overall change in r_{ei}, especially for pre→post pairing. Specifically, when heterosynaptic plasticity is rapid and strong (relative to a nominal amount of homosynaptic plasticity), the set point for r_{ei} should be lower; conversely, when heterosynaptic plasticity is slower and weaker, then the homosynaptic changes dominate and r_{ei} should be higher.

Therefore, we experimentally determined learning rates for expression of synaptic modification at the paired and best inputs (Figure 4). Given both the predictions of the model and the results of pairing at the best inputs (Figure S6), we focused on effects of pre→post pairing, but not post→pre pairing. Rates of modification were quantified in two ways, both by determining the earliest time point of continued (3+ min) statistically different strengths after pairing compared with baseline and by fitting single exponentials to excitatory and inhibitory strengths over time. Although each method might be noisy, there was general agreement between these approaches.

After pre→post pairing in developing auditory cortex, homosynaptic changes to excitation and inhibition were faster than heterosynaptic changes. For the cell from Figure 1, significant excitatory potentiation was detected by the fourth minute after pairing and maintained thereafter (Figure 4A). The single exponential fitted to this process had a time constant of ∼1.0 min. Similarly, paired inhibition was significantly increased starting at the ninth minute after pairing, and the exponential time constant was ∼0.6 min. Heterosynaptic modifications were considerably slower; changes to the original best channel were significant only after 20 min for excitation and 15 min for inhibition, with longer time constants of 4.8 and 9.1 min for exponential fits to the synaptic weights (Figure 4A). Over the 25 pre→post pairing experiments, rates of heterosynaptic modifications were slower than rates of homosynaptic changes (Figure 4B). Furthermore, across recordings, relative rates of heterosynaptic versus homosynaptic modifications were related to the excitatory-inhibitory correlation after pairing, both for excitatory plasticity (Figure 4C, top) and for inhibitory plasticity (Figure 4C, bottom). This closely matches the results of simulations in Figure 3.

Correlations between excitatory and inhibitory responses *in vivo* are generally higher in adult than in developing auditory cortex (Dorrn et al., 2010). We asked whether plasticity might lead to higher correlation values after spike pairing *in vitro* in

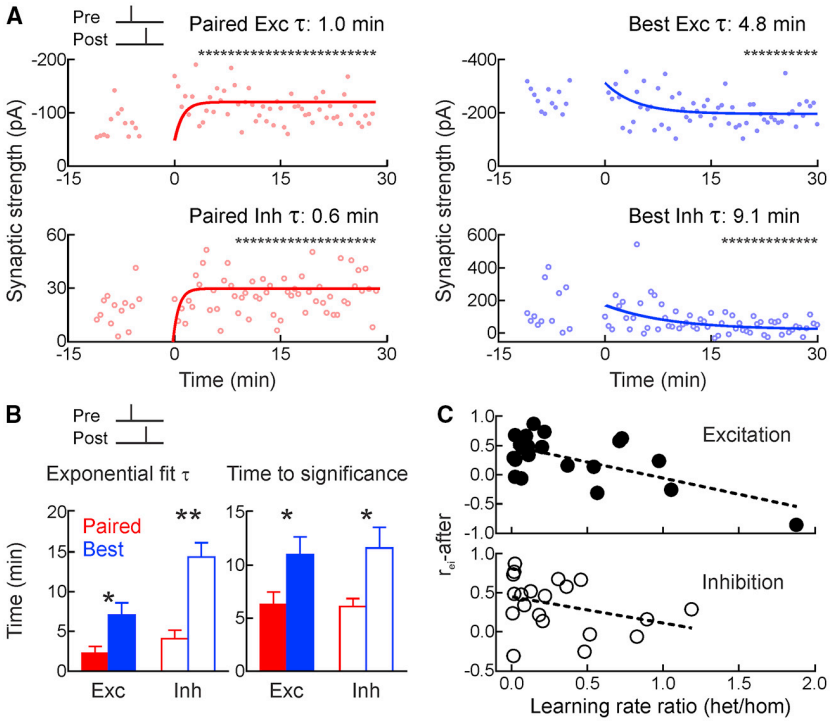


Figure 4. Heterosynaptic Modifications Lag Changes to Paired Inputs

(A) Time course of changes to paired EPSCs (top left) and IPSCs (bottom left) for the cell from Figure 1, measured with single exponentials and running t tests (asterisks, $p < 0.05$ versus baseline) to determine when significant modifications were first expressed.

(B) Summary for paired homosynaptic (red) and original best heterosynaptic modifications (blue); exponential fits (left; paired excitation τ: 2.3 ± 0.9 min, original best excitation τ: 7.0 ± 1.6 min, $p < 0.02$; paired inhibition τ: 4.1 ± 1.1 min, original best inhibition τ: 14.4 ± 1.8 min, $p < 0.001$) and running t tests (right; paired excitation: 6.3 ± 1.2 min, original best excitation: 11.0 ± 1.7 min, $p < 0.03$; paired inhibition: 6.1 ± 0.8 min, original best inhibition: 11.6 ± 1.9 min, $p < 0.03$). Error bars, SEM.

(C) r_{ei}-after inversely correlated with the ratio of heterosynaptic versus homosynaptic τs; higher r_{ei} values were associated with weaker/slower heterosynaptic plasticity, and lower r_{ei} values were associated with faster heterosynaptic modifications for excitation (r: −0.63) and inhibition (r: −0.34).

adult mouse auditory cortex (animals aged 2–3 months). We found that pre→post pairing induced LTP of paired excitatory and inhibitory inputs in adult cortex. Heterosynaptic modifications, while present, were minimal in adult cortex, and changes to the original best excitatory and inhibitory inputs were not statistically significant (Figures 5A–5C). Regardless, excitatory-inhibitory correlation values were greatly increased after pairing to higher levels than in younger auditory cortex (Figure 5D). For the 8/13 adult cells for which r_{ei}-before < 0.7, changes to paired inputs alone contributed about twice as much to r_{ei}-after as changes to unpaired inputs (Figure 5E). This was qualitatively different from in young cortex, where excitatory-inhibitory correlation change mainly resulted from heterosynaptic modifications. Thus, homosynaptic plasticity may be more reliable and heterosynaptic plasticity less pervasive in mature cortical circuits, leading to different set points for overall excitatory-inhibitory balance.

Heterosynaptic Plasticity Requires Dendritic Ca^{2+} Signaling and Internal Stores

We next examined biological mechanisms that enable selective heterosynaptic plasticity at original best unpaired inputs. We used two-photon Ca^{2+} imaging to measure dendritic Ca^{2+} events in layer 5 pyramidal cells during spike pairing (Figure S8A). Both pre→post and post→pre pairing led to broader backpropagating action potential-evoked Ca^{2+} transients (Figures S8B and S8C; normal solution). This enhanced Ca^{2+} signaling triggered by spike pairing might be related to Ca^{2+}-induced Ca^{2+} release from internal stores (Larkum et al., 2003; Lee et al., 2016), which would provide a rapid signal for intracellular communication across disparate synapses and is implicated in heterosynaptic modifications in amygdala (Royer and Pare, 2003) and hippocampus (Nishiyama et al., 2000). We found that depleting internal Ca^{2+} stores via intracellular perfusion with thapsigargin (10 μM) prevented broadening of the Ca^{2+} event such that transients evoked during pre→post and post→pre pairing were no different from transients due to postsynaptic spikes alone (Figures S8B and S8C; thapsigargin).

Ca^{2+}-induced Ca^{2+} release was also the major mechanism for heterosynaptic plasticity (Figure 6). Either intracellular thapsigargin (10 μM, Figures 6A and 6E) or ruthenium red (which blocks Ca^{2+} release from internal stores, 20 μM; Figures 6D and S9) prevented heterosynaptic modifications but spared changes to paired excitatory and inhibitory inputs after pre→post or post→pre pairing (Figure 6B). Long-term synaptic modifications required NMDA receptors, because bath application of APV (50 μM) prevented all changes to paired and unpaired inputs (Figure 6C). Therefore, intracellular Ca^{2+} signaling initiated by activation of NMDA receptors at paired excitatory synapses triggered other modifications to paired inhibitory synapses and original best unpaired excitatory and inhibitory synapses, perhaps via Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII) activation and broader patterns of Ca^{2+} release from internal stores that interact with large synaptic events in a winner-takes-all manner for heterosynaptic depression (Figure 7).

Heterosynaptic Plasticity Is Induced at Relative Best Inputs Minutes after Pairing

These results show that heterosynaptic plasticity can be selectively induced at a specific subset of excitatory and inhibitory inputs onto individual postsynaptic neurons. The original best inputs are not necessarily globally maximal, because only a fraction of the total inputs received by these neurons were activated by the stimulation electrodes. Because heterosynaptic changes were expressed ~20 min after pairing, we hypothesized that

Please cite this article in press as: Field et al., Heterosynaptic Plasticity Determines the Set Point for Cortical Excitatory-Inhibitory Balance, Neuron (2020), https://doi.org/10.1016/j.neuron.2020.03.002

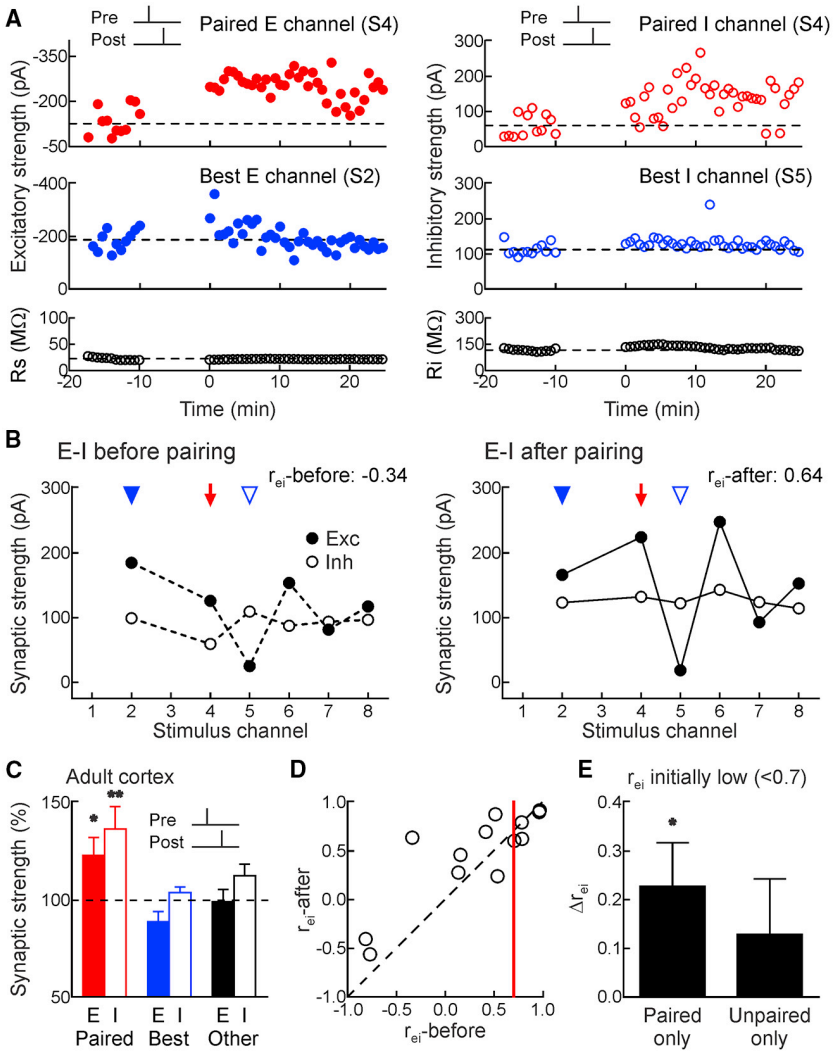


Figure 5. Pairing Increased Excitatory-Inhibitory Correlation in Adult Cortex via Homosynaptic Changes

(A) Top: example of excitatory LTP (left) and inhibitory LTP (right) induced in adult cortex by pre→post pairing at channel S4 (red, $\Delta t = 4$ ms). Middle: original best inputs were minimally affected (blue). Bottom: series and input resistance.

(B) Increased r_{ei}; same cell as (A) (r_{ei}-before: −0.34; r_{ei}-after: 0.64). Red arrow, paired channel. Blue arrowheads, original best excitation (filled) and inhibition (open).

(C) Adult excitatory and inhibitory STDP after pre→post pairing (paired EPSCs increased 23.1% ± 9.2%, n = 13, $p < 0.03$; paired IPSCs increased 36.7% ± 11.5%, $p < 0.008$; originally largest unpaired EPSCs decreased −11.2% ± 5.4%, $p > 0.05$; originally largest unpaired IPSCs increased 4.0% ± 2.8%, $p > 0.1$; other unpaired EPSCs decreased −0.8% ± 6.4%, $p > 0.9$; other unpaired IPSCs increased 12.6% ± 6.0%, $p > 0.05$). Error bars, SEM.

(D) Pre→post pairing and Δr_{ei} in adult cortex (n = 13). Red line, r_{ei}: 0.7 (when r_{ei}-before < 0.7, change of r_{ei}: 0.30 ± 0.12, n = 8, $p < 0.05$; Student's paired two-tailed t test).

(E) In adult neurons, mainly homosynaptic modifications increased r_{ei} (paired only: Δr_{ei} when initially $r < 0.7$: 0.23 ± 0.09, n = 8, $p < 0.04$, Student's paired two-tailed t test; unpaired only: Δr_{ei} when initially $r < 0.7$ for pre→post pairing: 0.13 ± 0.11, $p > 0.2$).

these locally maximal inputs were computed by postsynaptic cells within this brief postpairing period. To test this prediction, we performed a final set of experiments in which for 10 min immediately following pairing, the original best excitatory and inhibitory inputs (selected to be on the same input channel) were not stimulated.

We found that during this 10-min period, the second-largest inputs (relative best inputs), rather than the original best inputs, were selectively affected by heterosynaptic modifications for both pre→post pairing (Figure 8) and post→pre pairing (Figure S10). In the recording shown in Figure 8A, channel S8 evoked the originally largest EPSCs and IPSCs, channel S6 evoked the second-largest EPSCs and IPSCs, and channel S4 was the paired channel. After pre→post pairing, channel S8 was turned off for 10 min. During that period, the paired EPSCs and IPSCs increased, while heterosynaptic LTD was induced at the relative best EPSCs and IPSCs evoked by channel S6. When channel S8 was reactivated, the EPSCs and IPSCs at that channel remained at their initial amplitudes and were stable until the end of the recording. Over all of these recordings, the relative best inputs, rather than the original best inputs, were selectively affected by heterosynaptic modifications (Figure 8B). Similarly, when the original best input was not presented after post→pre pairing, the relative best input instead experienced heterosynaptic plasticity—in this case, heterosynaptic LTP of excitation and inhibition (Figure S10).

This experiment demonstrates that heterosynaptic plasticity can be specifically directed to occur at whichever inputs were most strongly activated in a restricted postpairing period. Furthermore, these results show that cortical neurons have a Ca^{2+}-dependent mechanism for determining and adjusting overall excitation and excitatory-inhibitory balance in a rapid and stimulus-specific manner.

DISCUSSION

Excitatory-inhibitory balance is a fundamental feature of neural networks (Froemke, 2015; Takesian and Hensch, 2013; Wehr and Zador, 2003; Xue et al., 2014). However, it has remained unclear how this organization is set up and calibrated online in response to changes of excitatory synapses important for learning and memory. Here we described how forms of long-term homosynaptic and heterosynaptic plasticity selectively adjust populations of inputs onto cortical pyramidal neurons to achieve a particular set point for excitatory-inhibitory balance. Instead of a slower global optimization process, which might be difficult to implement biologically, our results demonstrate

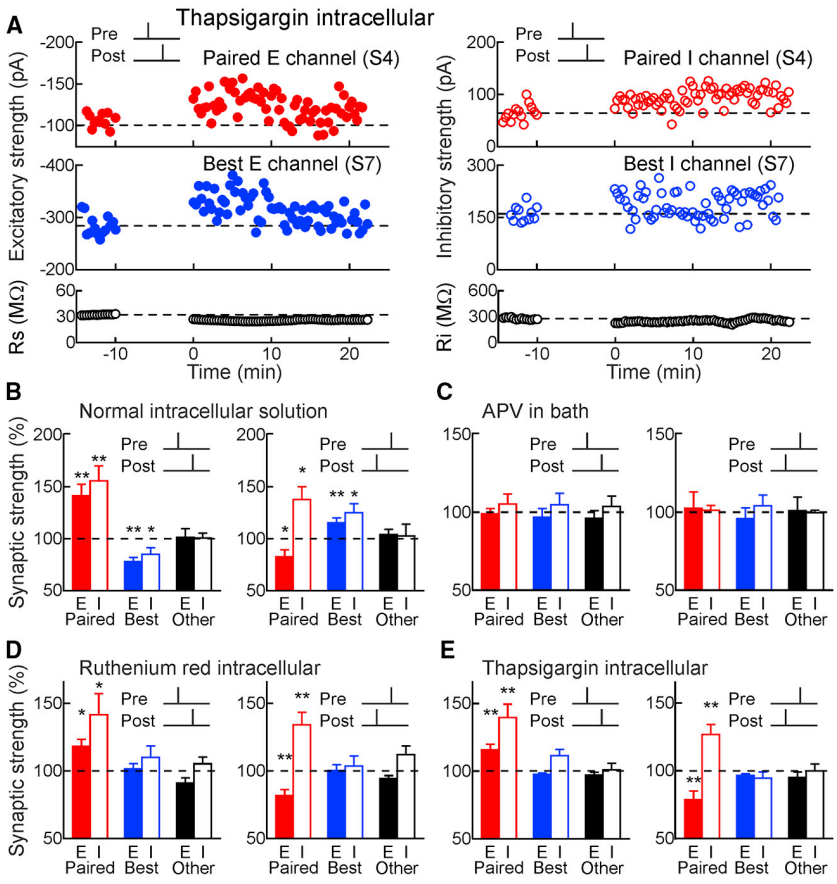


Figure 6. Mechanisms of Input-Specific Heterosynaptic Plasticity

(A) Thapsigargin in a whole-cell pipette (10 μM) prevented heterosynaptic excitatory and inhibitory LTD after pre→post pairing. Top: excitatory inhibitory LTP induced by pre→post pairing at channel S4 (red, Δt = 4 ms). Middle, thapsigargin prevented heterosynaptic LTD. Bottom: R_s and R_i.

(B) Spike pairing with normal solutions and artificial cerebrospinal fluid (ACSF) on paired inputs (red), original best unpaired inputs (blue), and other unpaired inputs (black); same recordings as Figures 2A and 2B. Filled bars, excitation; open bars, inhibition. Error bars, SEM.

(C) Blocking NMDA receptors (50 μm APV in bath) prevented plasticity (pre→post, n = 6: paired EPSCs, $p > 0.7$, Student's paired two-tailed t test; paired IPSCs, $p > 0.4$; originally largest unpaired EPSCs, $p > 0.6$; originally largest unpaired IPSCs, $p > 0.5$; and post→pre, n = 4: paired EPSCs, $p > 0.8$; paired IPSCs, $p > 0.7$; originally largest unpaired EPSCs, $p > 0.6$; originally largest unpaired IPSCs, $p > 0.5$). Error bars, SEM.

(D) Intracellular ruthenium red (20 μm) spared homosynaptic plasticity but prevented heterosynaptic plasticity at original best inputs (pre→post, n = 9: paired EPSCs, $p < 0.02$; paired IPSCs, $p < 0.05$; originally largest unpaired EPSCs, $p > 0.6$; originally largest unpaired IPSCs, $p > 0.3$; and post→pre, n = 9: paired EPSCs, $p < 0.004$; paired IPSCs, $p < 0.006$; originally largest unpaired EPSCs, $p > 0.9$; originally largest unpaired IPSCs, $p > 0.6$). Error bars, SEM.

(E) Intracellular thapsigargin (10 μm) spared homosynaptic plasticity but prevented heterosynaptic plasticity (pre→post, n = 12: paired EPSCs, $p < 0.006$; paired IPSCs, $p < 0.003$; originally largest unpaired EPSCs, $p > 0.05$; originally largest unpaired IPSCs, $p > 0.1$; and post→pre, n = 10: paired EPSCs, $p < 0.01$; paired IPSCs, $p < 0.006$; originally largest unpaired EPSCs, $p > 0.05$; originally largest unpaired IPSCs, $p > 0.2$). Error bars, SEM.

that a restricted set of activity-dependent changes is sufficient to normalize excitatory-inhibitory balance within minutes, enhancing the relation between inhibition and excitation when mismatched or reducing this value if inhibition is too restrictive. Our theoretical analysis indicates that the definition of excitatory-inhibitory balance can be dynamic, and the set point is determined by the relative degree to which heterosynaptic modifications are engaged. Consequentially, heterosynaptic plasticity and inhibitory plasticity work together to reorganize cortical inputs after induction of long-term excitatory modifications to update information storage and enable flexible computation without disrupting overall network function.

Cortical excitation and inhibition are not perfectly matched in all cases, especially before extensive exposure or experience with particular stimuli. For frequency tuning curves measured in the young adult and adult rodent auditory cortex *in vivo*, magnitudes of tone-evoked excitatory and inhibitory responses can be highly correlated, with average values of 0.7 to >0.9 (Froemke et al., 2007; Tan and Wehr, 2009; Wehr and Zador, 2003), although the range across the population can be quite variable (Dorrn et al., 2010). In younger animals, however, frequency tuning tends to be initially broad or erratic; excitatory inputs mature within the first 1–2 weeks of postnatal life in rodents, but inhibitory tuning requires experience over weeks 2–4 to balance excitation (Chang et al., 2005; de Villers-Sidani et al., 2007; Dorrn et al., 2010). In developing rat auditory cortex *in vivo*, repetitive sensory stimulation generally increases excitatory-inhibitory correlation levels to higher levels regardless of the initial baseline correlation (Dorrn et al., 2010). Here we identified a complementary mechanism in young mouse auditory cortex in which paired pre- and postsynaptic spiking can increase correlations when initially quite low but otherwise seems to maintain the excitatory-inhibitory correlations at intermediate levels before adulthood. Although there could be species differences in the learning rules or excitatory-inhibitory set points, a more likely hypothesis is that repetitive patterned stimulation with pure tones *in vivo* more aggressively engages homosynaptic plasticity, which predominates over heterosynaptic modifications. This is consistent with the findings of Dorrn et al. (2010) in terms of heterosynaptic potentiation and increases of excitatory-inhibitory correlations and consistent with the model presented here: when homosynaptic plasticity is faster and/or stronger than heterosynaptic plasticity, the set point for excitatory-inhibitory correlation is higher.

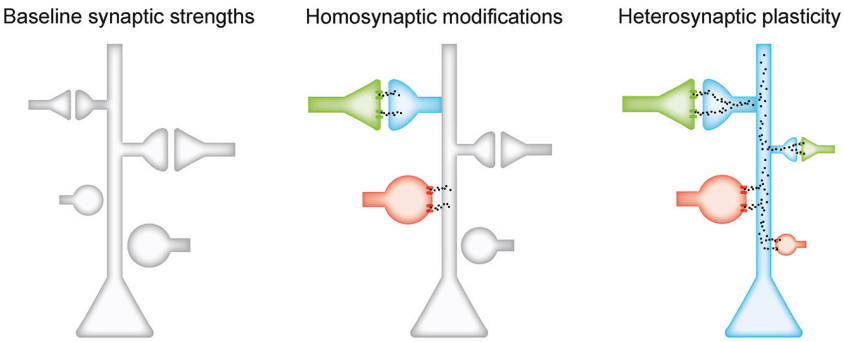


Figure 7. Hypothesized Plasticity Mechanisms

Green, excitatory input; red, inhibitory input. Homosynaptic modifications depend on NMDA receptors, L-type Ca^{2+} channels, and kinase activation. Integrated over minutes, Ca^{2+} release from internal stores is sensitive to the largest inputs in a winner-takes-all manner, inducing input-specific heterosynaptic depression.

Even in adult animals, correlated excitatory and inhibitory responses to complex sounds such as vocalizations can require experience. Spiking responses to infant mouse distress calls are weak in adult virgin female auditory cortex because of imbalanced (uncorrelated) excitation and inhibition; after maternal experience with pups, excitation and inhibition become more closely matched to enable reliable action potential generation (Marlin et al., 2015). Even inputs that are patently artificial can lead to meaningful neural and behavioral responses, perhaps partly because of mechanisms of cortical plasticity. Rodents can learn to use intracortical electrical microstimulation as a behaviorally meaningful input (Long and Carmena, 2013), and analogously, humans can learn to use cochlear implants despite what might be initially random patterns of electrically evoked activity (Wilson, 2015; Glennon et al., 2020).

Our experiments might emulate how novel inputs recruit initially unrelated populations of excitatory and inhibitory synapses, becoming functionally coupled via experience-dependent plasticity. One caveat is that these recordings were made at the soma, perhaps electrotonically far from the sites of activated inputs. Thus, somatic values of r_{ei} might not be the most relevant for regulating NMDA receptors or generating dendritic spikes, although presumably these values are more accurate in terms of excitatory-inhibitory control of spike generation at the axon hillock. Although inputs evoked by each stimulation channel may not initially be functionally related, these inputs become bound together via repetitive co-activation, together with postsynaptic spiking. Initially high response variability might also facilitate this plasticity. In particular, relatively imbalanced inhibition might make it easier for incoming input to activate NMDA receptors, leading to long-term modifications, which in turn balance inhibition with excitation (D'amour and Froemke, 2015). Regulation of cortical inhibition in this way is believed to be important for the opening and closing of developmental critical periods (Dorrn et al., 2010; Hensch and Fagiolini, 2005; Kuhlman et al., 2013; Takesian et al., 2018). Our results indicate that these phenomena are not independently induced (which might pose challenges for dynamic control of excitatory-inhibitory balance) but are effectively coordinated over a timescale of minutes by activity-dependent mechanisms.

Part of this process involves computing local maxima of incoming inputs for selective modifications of specific synapses. Combined with slower forms of homeostatic plasticity (Turrigiano, 2008), individual cortical neurons have the capability to integrate or accumulate recent activity over minutes to hours, enabling flexible representations of external stimuli and control over excitability on multiple short and long timescales. Different patterns of coordinated pre- and postsynaptic spiking might engage distinct mechanisms or forms of synaptic plasticity, such as those seen here for pairing at non-best versus best inputs. Long-term plasticity depends on many different variables, including baseline amplitude of synaptic strengths, number and frequency of pre- and postsynaptic spiking, postsynaptic membrane potential, and the dendritic location of synaptic inputs (Sjöström et al., 2001; Froemke et al., 2005, 2010a; Wang et al., 2005). At high spiking rates or levels of postsynaptic depolarization, LTP is reliably induced irrespective of precise spike timing; other more global homeostatic mechanisms for normalizing overall synaptic strengths might then be engaged. Similarly, synaptic plasticity might be regulated by other factors such as neuromodulation or critical periods (Froemke, 2015), and we observed that heterosynaptic modifications were less prevalent in older than in developing auditory cortex. Regardless, the results of our models might be generally applicable, suggesting that as long as there are analogous forms of plasticity, there can be stable set points for excitatory-inhibitory inputs to be balanced in potentially any system. This is reminiscent of findings that many forms of inhibitory STDP can lead to balanced networks and equilibrium states in simulations (Vogels et al., 2011; Luz and Shamir, 2012).

In terms of mechanism, CaMKII activation because of Ca^{2+} influx through NMDA receptors enhances excitatory transmission through AMPA receptors (Malenka and Nicoll, 1999; Froemke, 2015), and a growing literature implicates CaMKII in potentiation of inhibitory transmission (Huang et al., 2005; Chiu et al., 2018). These local phenomena affecting paired synapses must then set in motion a more wide-ranging process involving release from thapsigargin-sensitive internal stores to selectively downregulate the largest unpaired incoming events. Consequentially, the total synaptic strength is roughly conserved (Royer and Pare, 2003; Froemke et al., 2013), while fine-scale patterns of co-activated excitatory and inhibitory inputs become relatively larger or smaller together. Beyond the paired and original best inputs, certain other inputs seem to be modified, but these might vary from cell to cell. The detailed mechanisms by which these modifications occur remain to be determined, including how specific inhibitory events are adjusted after excitatory synaptic activation and how heterosynaptic plasticity is regulated over development or by experience to allow the set point for excitatory-inhibitory balance to be dynamic.

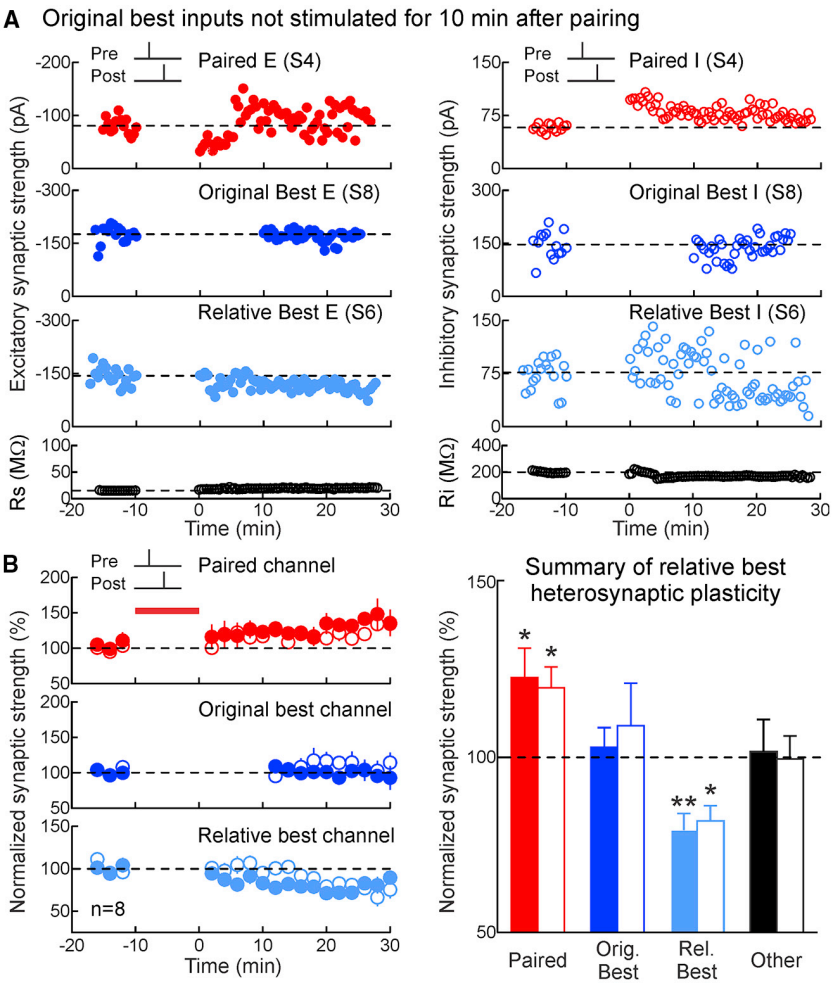


Figure 8. Postsynaptic Neurons Compute Maximally Strong Inputs

(A) Deactivating the original best input channel led to heterosynaptic excitatory and inhibitory LTD at the second-best (relative best) channel after pre→post pairing. Top: excitatory and inhibitory LTP induced by pre→post pairing at channel S4 (red, $\Delta t = 4$ ms). Top middle: original best inputs at S8 were unaltered when inactivated for 10 min after pairing. Bottom middle: LTD was induced at the relative best inputs at S6. Bottom: R_s and R_i.

(B) Pre→post experiments with the original best input channel deactivated for 10 min after pairing (red: paired EPSCs: 22.7% ± 8.4%, n = 8, $p < 0.04$, Student's paired two-tailed t test, and paired IPSCs: 19.8% ± 6.0%, $p < 0.02$; dark blue: originally largest EPSCs: 2.8% ± 5.6%, $p > 0.6$, and originally largest IPSCs: 9.0% ± 12.1%, $p > 0.4$; light blue: relative best EPSCs: −20.6% ± 4.8%, $p < 0.004$, and relative best IPSCs: −18.2% ± 4.4%, $p < 0.02$; black: other EPSCs: 1.6% ± 9.2%, $p > 0.8$, and other IPSCs: −0.5% ± 6.6%, $p > 0.9$). Filled, excitation; open, inhibition. Error bars, SEM.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
  - Materials Availability Statement
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Slice preparation- mouse auditory cortex
  - Electrophysiology
  - Two-photon Ca^{2+} imaging
  - Simulations: probabilistic model
  - Simulations: biophysical model
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.neuron.2020.03.002.

ACKNOWLEDGMENTS

We thank K. Kuchibhotla, M. Jin, E. Morina, D. Talos, R.W. Tsien, and N. Zaika for comments, discussions, and technical assistance. This work was funded by grants from the Max Planck Society and a NARSAD Young Investigator Grant from the Brain and Behavior Research Foundation to J.G.; the NINDS (NS074972) to B.R. and R.C.F.; the NIDCD (DC009635 and DC012557), NICHD (HD088411), and NIH BRAIN Initiative (NS107616); and a Sloan Research Fellowship, a Klingenstein Fellowship, and a Howard Hughes Medical Institute Faculty Scholarship to R.C.F. Art in Figure 7 was produced by Samantha Holmes (https://www.samantha-holmes.com/).

AUTHOR CONTRIBUTIONS

All authors designed the studies and wrote the paper. R.E.F., J.A.D., and R.T. performed the experiments and analyzed the data. C.M. and J.G. performed the modeling.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 10, 2018
Revised: December 27, 2019
Accepted: March 3, 2020
Published: March 25, 2020

REFERENCES

Avoli, M., de Curtis, M., Gnatkovsky, V., Gotman, J., Köhling, R., Lévesque, M., Manseau, F., Shiri, Z., and Williams, S. (2016). Specific imbalance of excitatory/inhibitory signaling establishes seizure onset pattern in temporal lobe epilepsy. J. Neurophysiol. *115*, 3229–3237.

Basu, J., Zaremba, J.D., Cheung, S.K., Hitti, F.L., Zemelman, B.V., Losonczy, A., and Siegelbaum, S.A. (2016). Gating of hippocampal activity, plasticity, and memory by entorhinal cortex long-range inhibition. Science *351*, aaa5694.

Bi, G.Q., and Poo, M.M. (1998). Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J. Neurosci. *18*, 10464–10472.

Cai, D., Han, R., Liu, M., Xie, F., You, L., Zheng, Y., Zhao, L., Yao, J., Wang, Y., Yue, Y., et al. (2018). A critical role of inhibition in temporal processing maturation in the primary auditory cortex. Cereb. Cortex *28*, 1610–1624.

Chang, E.F., Bao, S., Imaizumi, K., Schreiner, C.E., and Merzenich, M.M. (2005). Development of spectral and temporal response selectivity in the auditory cortex. Proc. Natl. Acad. Sci. USA *102*, 16460–16465.

Chistiakova, M., Bannon, N.M., Chen, J.Y., Bazhenov, M., and Volgushev, M. (2015). Homeostatic role of heterosynaptic plasticity: models and experiments. Front. Comput. Neurosci. *9*, 89.

Chiu, C.Q., Martenson, J.S., Yamazaki, M., Natsume, R., Sakimura, K., Tomita, S., Tavalin, S.J., and Higley, M.J. (2018). Input-specific NMDAR-dependent potentiation of dendritic GABAergic inhibition. Neuron *97*, 368–377.e3.

Christie, B.R., and Abraham, W.C. (1992). NMDA-dependent heterosynaptic long-term depression in the dentate gyrus of anaesthetized rats. Synapse *10*, 1–6.

Cossart, R., Dinocourt, C., Hirsch, J.C., Merchan-Perez, A., De Felipe, J., Ben-Ari, Y., Esclapez, M., and Bernard, C. (2001). Dendritic but not somatic GABAergic inhibition is decreased in experimental epilepsy. Nat. Neurosci. *4*, 52–62.

D'amour, J.A., and Froemke, R.C. (2015). Inhibitory and excitatory spike-timing-dependent plasticity in the auditory cortex. Neuron *86*, 514–528.

de Villers-Sidani, E., Chang, E.F., Bao, S., and Merzenich, M.M. (2007). Critical period window for spectral tuning defined in the primary auditory cortex (A1) in the rat. J. Neurosci. *27*, 180–189.

Dehghani, N., Peyrache, A., Telenczuk, B., Le Van Quyen, M., Halgren, E., Cash, S.S., Hatsopoulos, N.G., and Destexhe, A. (2016). Dynamic balance of excitation and inhibition in human and monkey neocortex. Sci. Rep. *6*, 23176.

Dorrn, A.L., Yuan, K., Barker, A.J., Schreiner, C.E., and Froemke, R.C. (2010). Developmental sensory experience balances cortical excitation and inhibition. Nature *465*, 932–936.

El-Boustani, S., Ip, J.P.K., Breton-Provencher, V., Knott, G.W., Okuno, H., Bito, H., and Sur, M. (2018). Locally coordinated synaptic plasticity of visual cortex neurons *in vivo*. Science *360*, 1349–1354.

Feldman, D.E. (2000). Timing-based LTP and LTD at vertical inputs to layer II/III pyramidal cells in rat barrel cortex. Neuron *27*, 45–56.

Froemke, R.C. (2015). Plasticity of cortical excitatory-inhibitory balance. Annu. Rev. Neurosci. *38*, 195–219.

Froemke, R.C., Poo, M.M., and Dan, Y. (2005). Spike-timing-dependent synaptic plasticity depends on dendritic location. Nature *434*, 221–225.

Froemke, R.C., Merzenich, M.M., and Schreiner, C.E. (2007). A synaptic memory trace for cortical receptive field plasticity. Nature *450*, 425–429.

Froemke, R.C., Debanne, D., and Bi, G.Q. (2010a). Temporal modulation of spike-timing-dependent plasticity. Front. Synaptic Neurosci. *2*, 19.

Froemke, R.C., Letzkus, J.J., Kampa, B.M., Hang, G.B., and Stuart, G.J. (2010b). Dendritic synapse location and neocortical spike-timing-dependent plasticity. Front. Synaptic Neurosci. *2*, 29.

Froemke, R.C., Carcea, I., Barker, A.J., Yuan, K., Seybold, B.A., Martins, A.R., Zaika, N., Bernstein, H., Wachs, M., Levis, P.A., et al. (2013). Long-term modification of cortical synapses improves sensory perception. Nat. Neurosci. *16*, 79–88.

Gandhi, S.P., Yanagawa, Y., and Stryker, M.P. (2008). Delayed plasticity of inhibitory neurons in developing visual cortex. Proc. Natl. Acad. Sci. USA *105*, 16797–16802.

Glennon, E., Svirsky, M.A., and Froemke, R.C. (2020). Auditory cortical plasticity in cochlear implant users. Curr. Opin. Neurobiol. *60*, 108–114.

Graupner, M., and Reyes, A.D. (2013). Synaptic input correlations leading to membrane potential decorrelation of spontaneous activity in cortex. J. Neurosci. *33*, 15075–15085.

Hackett, T.A., Barkat, T.R., O'Brien, B.M., Hensch, T.K., and Polley, D.B. (2011). Linking topography to tonotopy in the mouse auditory thalamocortical circuit. J. Neurosci. *31*, 2983–2995.

Hattori, R., Kuchibhotla, K., Froemke, R.C., and Komiyama, T. (2017). Functions and dysfunctions of neocortical inhibitory neuron subtypes. Nat. Neurosci. *20*, 1199–1208.

Hensch, T.K., and Fagiolini, M. (2005). Excitatory-inhibitory balance and critical period plasticity in developing visual cortex. Prog. Brain Res. *147*, 115–124.

Higley, M.J., and Contreras, D. (2006). Balanced excitation and inhibition determine spike timing during frequency adaptation. J. Neurosci. *26*, 448–457.

Hiratani, N., and Fukai, T. (2017). Detailed dendritic excitatory/inhibitory balance through heterosynaptic spike-timing-dependent plasticity. J. Neurosci. *37*, 12106–12122.

Horn, M.E., and Nicoll, R.A. (2018). Somatostatin and parvalbumin inhibitory synapses onto hippocampal pyramidal neurons are regulated by distinct mechanisms. Proc. Natl. Acad. Sci. USA *115*, 589–594.

House, D.R.C., Elstrott, J., Koh, E., Chung, J., and Feldman, D.E. (2011). Parallel regulation of feedforward inhibition and excitation during whisker map plasticity. Neuron *72*, 819–831.

Huang, C.S., Shi, S.H., Ule, J., Ruggiu, M., Barker, L.A., Darnell, R.B., Jan, Y.N., and Jan, L.Y. (2005). Common molecular pathways mediate long-term potentiation of synaptic excitation and slow synaptic inhibition. Cell *123*, 105–118.

Isaacson, J.S., and Scanziani, M. (2011). How inhibition shapes cortical activity. Neuron *72*, 231–243.

Kuhlman, S.J., Olivas, N.D., Tring, E., Ikrar, T., Xu, X., and Trachtenberg, J.T. (2013). A disinhibitory microcircuit initiates critical-period plasticity in the visual cortex. Nature *501*, 543–546.

Larkum, M.E., Watanabe, S., Nakamura, T., Lasser-Ross, N., and Ross, W.N. (2003). Synaptically activated Ca2+ waves in layer 2/3 and layer 5 rat neocortical pyramidal neurons. J. Physiol. *549*, 471–488.

Lee, C.C., Imaizumi, K., Schreiner, C.E., and Winer, J.A. (2004). Concurrent tonotopic processing streams in auditory cortex. Cereb. Cortex *14*, 441–451.

Lee, K.F.H., Soares, C., Thivierge, J.P., and Béïque, J.-C. (2016). Correlated synaptic inputs drive dendritic calcium amplification and cooperative plasticity during clustered synapse development. Neuron *89*, 784–799.

Lissin, D.V., Gomperts, S.N., Carroll, R.C., Christine, C.W., Kalman, D., Kitamura, M., Hardy, S., Nicoll, R.A., Malenka, R.C., and von Zastrow, M. (1998). Activity differentially regulates the surface expression of synaptic AMPA and NMDA glutamate receptors. Proc. Natl. Acad. Sci. USA *95*, 7097–7102.

Long, J.D., 2nd, and Carmena, J.M. (2013). Dynamic changes of rodent somatosensory barrel cortex are correlated with learning a novel conditioned stimulus. J. Neurophysiol. *109*, 2585–2595.

Luz, Y., and Shamir, M. (2012). Balancing feed-forward excitation and inhibition via Hebbian inhibitory synaptic plasticity. PLoS Comput. Biol. *8*, e1002334.

Lynch, G.S., Dunwiddie, T., and Gribkoff, V. (1977). Heterosynaptic depression: a postsynaptic correlate of long-term potentiation. Nature *266*, 737–739.

Malenka, R.C., and Nicoll, R.A. (1999). Long-term potentiation—a decade of progress? Science *285*, 1870–1874.

Marlin, B.J., Mitre, M., D'amour, J.A., Chao, M.V., and Froemke, R.C. (2015). Oxytocin enables maternal behaviour by balancing cortical inhibition. Nature *520*, 499–504.

Miller, L.M., Escabí, M.A., Read, H.L., and Schreiner, C.E. (2001). Functional convergence of response properties in the auditory thalamocortical system. Neuron *32*, 151–160.

Muller, D., Hefft, S., and Figurov, A. (1995). Heterosynaptic interactions between LTP and LTD in CA1 hippocampal slices. Neuron *14*, 599–605.

Nishiyama, M., Hong, K., Mikoshiba, K., Poo, M.-M., and Kato, K. (2000). Calcium stores regulate the polarity and input specificity of synaptic modification. Nature *408*, 584–588.

Okun, M., and Lampl, I. (2008). Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. Nat. Neurosci. *11*, 535–537.

Oliveira, M.S., Pacheco, L.F., Mello, C.F., Cavalheiro, E.A., and Garrido-Sanabria, E.R. (2011). Epileptiform activity in the limbic system. Front. Biosci. (Schol. Ed.) *3*, 565–593.

Ren, H., Shi, Y.J., Lu, Q.C., Liang, P.J., and Zhang, P.M. (2014). The role of the entorhinal cortex in epileptiform activities of the hippocampus. Theor. Biol. Med. Model. *11*, 14.

Rosenkranz, J.A. (2012). *In vivo* voltage-dependent influences on summation of synaptic potentials in neurons of the lateral nucleus of the amygdala. Neuroscience *226*, 101–118.

Royer, S., and Pare, D. (2003). Conservation of total synaptic weight through balanced synaptic depression and potentiation. Nature *422*, 518–522.

Scanziani, M., Malenka, R.C., and Nicoll, R.A. (1996). Role of intercellular interactions in heterosynaptic long-term depression. Nature *380*, 446–450.

Scharfman, H.E., and Brooks-Kayal, A.R. (2014). Is plasticity of GABAergic mechanisms relevant to epileptogenesis? Adv. Exp. Med. Biol. *813*, 133–150.

Sjöström, P.J., Turrigiano, G.G., and Nelson, S.B. (2001). Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. Neuron *32*, 1149–1164.

Takesian, A.E., and Hensch, T.K. (2013). Balancing plasticity/stability across brain development. Prog. Brain Res. *207*, 3–34.

Takesian, A.E., Bogart, L.J., Lichtman, J.W., and Hensch, T.K. (2018). Inhibitory circuit gating of auditory critical-period plasticity. Nat. Neurosci. *21*, 218–227.

Tan, A.Y., and Wehr, M. (2009). Balanced tone-evoked synaptic excitation and inhibition in mouse auditory cortex. Neuroscience *163*, 1302–1315.

Thiagarajan, T.C., Lindskog, M., and Tsien, R.W. (2005). Adaptation to synaptic inactivity in hippocampal neurons. Neuron *47*, 725–737.

Toader, O., Forte, N., Orlando, M., Ferrea, E., Raimondi, A., Baldelli, P., Benfenati, F., and Medrihan, L. (2013). Dentate gyrus network dysfunctions precede the symptomatic phase in a genetic mouse model of seizures. Front. Cell. Neurosci. *7*, 138.

Tran-Van-Minh, A., Cazé, R.D., Abrahamsson, T., Cathala, L., Gutkin, B.S., and DiGregorio, D.A. (2015). Contribution of sublinear and supralinear dendritic integration to neuronal computations. Front. Cell. Neurosci. *9*, 67.

Turrigiano, G.G. (2008). The self-tuning neuron: synaptic scaling of excitatory synapses. Cell *135*, 422–435.

Turrigiano, G.G., Leslie, K.R., Desai, N.S., Rutherford, L.C., and Nelson, S.B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. Nature *391*, 892–896.

Urban, N.N., and Barrionuevo, G. (1998). Active summation of excitatory postsynaptic potentials in hippocampal CA3 pyramidal neurons. Proc. Natl. Acad. Sci. USA *95*, 11450–11455.

Vogels, T.P., Sprekeler, H., Zenke, F., Clopath, C., and Gerstner, W. (2011). Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. Science *334*, 1569–1573.

Wang, H.X., Gerkin, R.C., Nauen, D.W., and Bi, G.Q. (2005). Coactivation and timing-dependent integration of synaptic potentiation and depression. Nat. Neurosci. *8*, 187–193.

Wehr, M., and Zador, A.M. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. Nature *426*, 442–446.

Wilson, B.S. (2015). Getting a decent (but sparse) signal to the brain for users of cochlear implants. Hear. Res. *322*, 24–38.

Xue, M., Atallah, B.V., and Scanziani, M. (2014). Equalizing excitation-inhibition ratios across visual cortical neurons. Nature *511*, 596–600.

Zenke, F., Gerstner, W., and Ganguli, S. (2017). The temporal paradox of Hebbian learning and homeostatic plasticity. Curr. Opin. Neurobiol. *43*, 166–176.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Thapsigargin	Tocris	Cat#: 1138
Ruthenium red	Millipore Sigma	Cat#: R2751
APV	Millipore Sigma	Cat#: A8054
Alexa Fluor	Tocris	Cat#: 6625
Fluo-4	Tocris	Cat#: 6255
Experimental Models: Organisms/Strains		
C57BL/6 mice	Jackson Labs	Stock No. 000664
Software and Algorithms		
Model code	This manuscript	https://github.com/cmiehl/heterosynplast2018

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Robert C. Froemke (robert.froemke@med.nyu.edu).

Materials Availability Statement

This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All procedures were approved under NYU School of Medicine IACUC protocols, in accordance with NIH guidelines. Animals were housed in fully-equipped facilities in either the NYU School of Medicine Skirball Institute or Science Building (New York City). The facilities were operated by the NYU Division of Comparative Medicine. Wild-type C57BL/6 mice (Jackson Labs; Stock No. 000664) of both sexes were used in all experiments; animals were between P9-P90.

METHOD DETAILS

Slice preparation- mouse auditory cortex

Acute slices of auditory cortex were prepared from juvenile (P9-35) and adult (P60-90) C57BL/6 mice, an age range spanning the critical period for excitatory-inhibitory balancing in rodent auditory cortex (de Villers-Sidani et al., 2007; Dorrn et al., 2010). Animals were deeply anesthetized with a 1:1 ketamine/xylazine cocktail and decapitated. The brain was rapidly placed in ice-cold dissection buffer containing (in mM): 87 NaCl, 75 sucrose, 2 KCl, 1.25 NaH_2PO_4, 0.5 $CaCl_2$, 7 $MgCl_2$, 25 $NaHCO_3$, 1.3 ascorbic acid, and 10 dextrose, bubbled with 95%/5% O_2/CO_2 (pH 7.4). Slices (300–400 μm thick) were prepared with a vibratome (Leica), placed in warm (33-35°C) dissection buffer for ~10 min, then transferred to a holding chamber containing warm artificial cerebrospinal fluid (ACSF, in mM: 124 NaCl, 2.5 KCl, 1.5 $MgSO_4$, 1.25 NaH_2PO_4, 2.5 $CaCl_2$, and 26 $NaHCO_3$,). Slices were kept at room temperature (22-24°C) for at least 30 minutes before use. For experiments, slices were transferred to the recording chamber and perfused (2–2.5 mL min^{-1}) with oxygenated ACSF at 33°C.

Electrophysiology

Somatic whole-cell recordings were made from layer 5 pyramidal cells in current-clamp and voltage-clamp mode with a Multiclamp 700B amplifier (Molecular Devices) using IR-DIC video microscopy (Olympus). Patch pipettes (3-8 MΩ) were filled with intracellular solution (in mM: 135 K-gluconate, 5 NaCl, 10 HEPES, 5 MgATP, 10 phosphocreatine, and 0.3 GTP). For pharmacological studies, either thapsigargin (10 μM) or ruthenium red (20 μM) was included in the internal solution. In one experiment, 1 μM thapsigargin was added directly to the bath solution. Mean resting potential was −68.1 ± 6.4 mV (standard deviation, SD), series resistance (Rs) was 26.9 ± 12.0 MΩ, and input resistance (Ri) was 295.91 ± 129.6 MΩ, determined by monitoring cells with hyperpolarizing pulses (50 pA or 5-10 mV for 100 msec). Recordings were excluded from analysis if R_i changed > 30% compared to the baseline

period. Data were filtered at 2 kHz, digitized at 10 kHz, and analyzed with Clampfit 10 (Molecular Devices). Focal extracellular stimulation (0.033-0.2 Hz) was applied with a monopolar metal electrode 8-channel array (AMPI Master-9, stimulation strengths of 0-10 V for 6-300 μsec) located < 150 μm from the recording electrode. Cells were held in voltage clamp at two membrane potentials alternating between –40 to –30 mV for measuring IPSCs and –80 to –70 mV for measuring EPSCs. Mean peak EPSC (2 msec window) was used to measure excitatory strength. For IPSCs, a larger window (5-20 msec) was used. The 'best' inputs were not pre-selected, but determined by analysis after each recording.

To determine the synaptic overlap between different stimulation channels, recordings were performed in voltage-clamp mode; in some experiments we used a different internal solution (in mM: 130 Cs-methanesulfonate, 1 QX-314, 4 TEA-Cl, 0.5 BAPTA, 4 MgATP, 0.3 Na-GTP, 10 phosphocreatine, 10 HEPES, pH 7.2). We interleaved stimulation of all channels individually with summation of the paired channel plus one other channel, and compared the measured summed E/IPSC to the predicted sum based on the amplitudes of each event individually (Figure S1). On average, the degree of synaptic overlap was minimal (~10%–20%), and lower in the experiments containing the Cs+/QX-314-based internal solution (~5%–10%), indicating that these channels activated separate inputs (Froemke et al., 2005; Tran-Van-Minh et al., 2015; Urban and Barrionuevo, 1998).

For monitoring long-term changes in synaptic strength, stable baselines were first established with 5-20 min of stimulation. Synaptic strength after induction was measured 16-25 min after the end of the induction protocol. During induction, postsynaptic spiking was evoked with brief depolarizing current pulses. Presynaptic spike timing was defined as EPSP onset, and postsynaptic spike timing was measured at the peak of the action potential. All statistics and error bars are reported as means ± SEM and statistical significance assessed with paired two-tailed Student's t test, unless otherwise noted.

Two-photon Ca^{2+} imaging

Whole-cell recordings were performed with current-clamp intracellular solution containing Alexa Fluor (100 μM) to visualize the dendritic arbor and Fluo-4 (100-200 μM) to monitor Ca^{2+} signals. In some experiments, thapsigargin (10 μM) was also added to the internal solution. Ca^{2+} imaging began at least 30 min after breakin to allow for dye diffusion, equilibration, and assessing stability of the recording. Two-photon laser scanning microscopy of Ca^{2+} signals was performed using an upright microscope (BX61WI, Olympus), equipped with a slice recording chamber, 40X, 0.8 NA water immersion objective, and a Ti:Sapphire (MaiTai DeepSee, Spectra-Physics, Mountain View, CA) laser tuned to 810 nm to excite both Alexa Fluor 594 and Fluo-4. Imaging of dendritic segments was acquired with Fluoview software (Olympus) at 4X digital zoom, every 50 ms. Images were analyzed in ImageJ (NIH, Bethesda, MD, USA).

Simulations: probabilistic model

We modeled the interaction between homosynaptic and heterosynaptic plasticity in a probabilistic model with 12 excitatory and inhibitory inputs onto a single postsynaptic neuron. Excitatory and inhibitory tuning curves were initialized by generating the individual weights from a uniform distribution, where each value represented the total synaptic excitatory (or inhibitory) strength of one channel. For each tuning curve, one channel was chosen as the 'paired' channel where excitation and inhibition were increased, and the best excitatory and inhibitory channels were decreased by a fixed amount. The amount of increase due to homosynaptic plasticity for both excitatory and inhibitory channels was fixed at 65%, and the amount of decrease due to heterosynaptic plasticity was varied on each trial over the range −14 to −98% depression. We compared the Pearson correlation coefficient between excitatory and inhibitory weights before induction of any plasticity ('r_{ei}-before') and after synaptic weight adjustments ('r_{ei}-after'). This procedure was repeated for 50,000 pseudo-random tuning curve initializations. From all initializations, we computed the probability that the excitatory-inhibitory correlation r_{ei}-after was greater than or less than r_{ei}-before. All code for simulations can be found at: https://github.com/cmiehl/heterosynplast2018

Simulations: biophysical model

Similar to the probabilistic model, we modeled 12 input channels, each consisting of 10 excitatory and 10 inhibitory neurons, onto a single postsynaptic neuron. These channels represented the extracellular stimulation of a population of excitatory and inhibitory neurons converging onto the postsynaptic neuron. The postsynaptic neuron was modeled as a conductance-based leaky integrate-and-fire model:

$$\tau_m \frac{dV}{dt} = E_{leak} - V + \sum_j g_j^E \left(E_{rev}^E - V\right) + \sum_j g_j^I \left(E_{rev}^I - V\right).$$

When the membrane potential reached a certain threshold V_{thresh}, a spike was fired and the membrane potential was reset to V_{reset}. Each synaptic conductance increased with an input spike by: $g_j^{E/I} \rightarrow g_j^{E/I} + w_j^{E/I}$ and otherwise decreased: $\tau_g^{E/I}(dg_j^{E/I}/dt) = -g_j^{E/I}$.

Changes to excitatory and inhibitory synaptic strength were based on a pair-based STDP plasticity rule. For the excitatory learning window we used a classical asymmetric learning window where pre→post spike pairing $(\Delta t = t_{post} - t_{pre} \geq 0)$ led to excitatory LTP and post→pre spike pairing led to excitatory LTD $(\Delta t < 0)$:

$$W^E(\Delta t) = \begin{cases} A^E e^{-\Delta t/\tau^E}, for\ \Delta t \geq 0 \\ -A^E e^{\Delta t/\tau^E}, for\ \Delta t < 0 \end{cases}$$

For the inhibitory window we used a symmetric window where both pre→post and post→pre spike pairings led to inhibitory LTP:

$$W^I(\Delta t) = \begin{cases} A^I e^{-\Delta t/\tau^I}, for\ \Delta t \geq 0 \\ A^I e^{\Delta t/\tau^I}, for\ \Delta t < 0. \end{cases}$$

The synaptic weights evolved as: $w_j^{E/I} \rightarrow w_j^{E/I} + \eta_w^{E/I} W^{E/I}(\Delta t)$ with learning rates η_w^E for excitatory and η_w^I for inhibitory synaptic weights. The heterosynaptic decrease of synaptic weights was modeled based on an internal trace. The trace of each synapse increased with an incoming spike: $T_j^{E/I} \rightarrow T_j^{E/I} + w_j^{E/I}$ and otherwise decreased: $\tau_T^{E/I}(dT_j^{E/I}/dt) = -T_j^{E/I}$. Based on the mean trace per input channel $T_c^{E/I}$ (where the channel index c ranges from 1 to 12), the synaptic weights corresponding to the maximum trace per channel were decreased by: $w_{c,max}^{E/I} \rightarrow w_{c,max}^{E/I} - \eta_{het}^{E/I}[T_c^{E/I}]_{max}$.

Occasionally, when the synaptic weights $w_j^{E/I}$ for several channels were similar, this mechanism induced heterosynaptic plasticity at the channel which was not the best-tuned channel – this was the result of the internal trace $T_j^{E/I}$ not being a perfect measure of the synaptic weight strength. Due to the imbalance between potentiation and depression achieved by the STDP rules (namely, excitatory STDP can give rise to both potentiation and depression, while inhibitory STDP can only give rise to potentiation), the inhibitory heterosynaptic plasticity was faster, $\eta_{het}^E < \eta_{het}^I$. To enable the induction of heterosynaptic plasticity only after homosynaptic plasticity, we introduced a learning dependent trace T_{eLTP}, which could switch the heterosynaptic plasticity "on" or "off" based on accumulated excitatory LTP. Following the induction of LTP, $T_{eLTP} \rightarrow T_{eLTP} + \Delta w_j^E$ and otherwise decayed exponentially: $\tau_{TeLTP}(dT_{eLTP}/dt) = -T_{eLTP}$. Whenever T_{eLTP} reached the threshold θ_{on}, heterosynaptic plasticity was switched "on" and implemented as described above. Following the drop of the learning-dependent trace T_{eLTP} below the threshold θ_{off}, heterosynaptic plasticity was switched "off" again.

The inputs were modeled as Poisson spike trains. In the paired phase, the firing rate of the activated channel was 75 Hz for each input (no activation of the other channels). In the unpaired phase, all channels other than the channel which was activated during paring, had a firing rate of 0.5 Hz. These values led to postsynaptic activation only during the pairing phase, with very few postsynaptic spikes induced during the unpaired phase. The paired phase lasted for 1.5 s, the unpaired phase for 6 s and we presented multiple alternating sequences of paired and unpaired stimulation phases to the postsynaptic neuron. The initial values of the synaptic weights per channel for both excitatory and inhibitory synapses were drawn randomly from the interval [0.2-0.35]. All code for simulations can be found at: https://github.com/cmiehl/heterosynplast2018

Biophysical Model Parameters

Parameter	Description	Value
A^E	Excitatory STDP learning amplitude	1
A^I	Inhibitory STDP learning amplitude	1
τ^E	Excitatory learning time constant	20 ms
τ^I	Inhibitory learning time constant	20 ms
τ_T^E	Excitatory trace time constant	1 s
τ_T^I	Inhibitory trace time constant	1 s
τ_{TeLTP}	Learning-dependent trace time constant	5 s
θ_{on}	Threshold above which heterosynaptic plasticity is "on"	0.7
θ_{off}	Threshold below which heterosynaptic plasticity is "off"	0.1
η_w^E	Excitatory learning rate	1 s^{-1}
η_w^I	Inhibitory learning rate	1 s^{-1}
η_{het}^E	Excitatory heterosynaptic learning rate	varied
η_{het}^I	Inhibitory heterosynaptic learning rate	10 η_{het}^E
τ_m	Membrane time constant	20 ms
E_{rev}^E	Excitatory reversal potential	0 mV
E_{rev}^I	Inhibitory reversal potential	−80 mV
E_{leak}	Leak reversal potential	−70 mV
V_{thresh}	Spiking threshold	−50 mV
V_{reset}	Reset membrane potential	−70 mV
τ_g^E	Excitatory conductance decaying constant	5 ms
τ_g^I	Inhibitory conductance decaying constant	5 ms

QUANTIFICATION AND STATISTICAL ANALYSIS

Student's t test was used for comparisons between two groups, with paired or unpaired tests used when appropriate. One- or two-way analysis of variance (ANOVA) was used for analysis between three or more groups. Statistical analyses were performed using Prism 6.0 GraphPad and MATLAB (MathWorks). Statistical tests used, p values, and the number of cells are reported in the main text describing each figure. All quantifications are the result of data from at least 3 different animals, unless otherwise indicated. Data reported in the text are generally shown as mean ± standard error of the mean (s.e.m), unless otherwise indicated.

DATA AND CODE AVAILABILITY

Upon request to the Lead Contact, data are immediately available.

Neuron, Volume *106*

Supplemental Information

Heterosynaptic Plasticity Determines the Set Point for Cortical Excitatory-Inhibitory Balance

Rachel E. Field, James A. D'amour, Robin Tremblay, Christoph Miehl, Bernardo Rudy, Julijana Gjorgjieva, and Robert C. Froemke

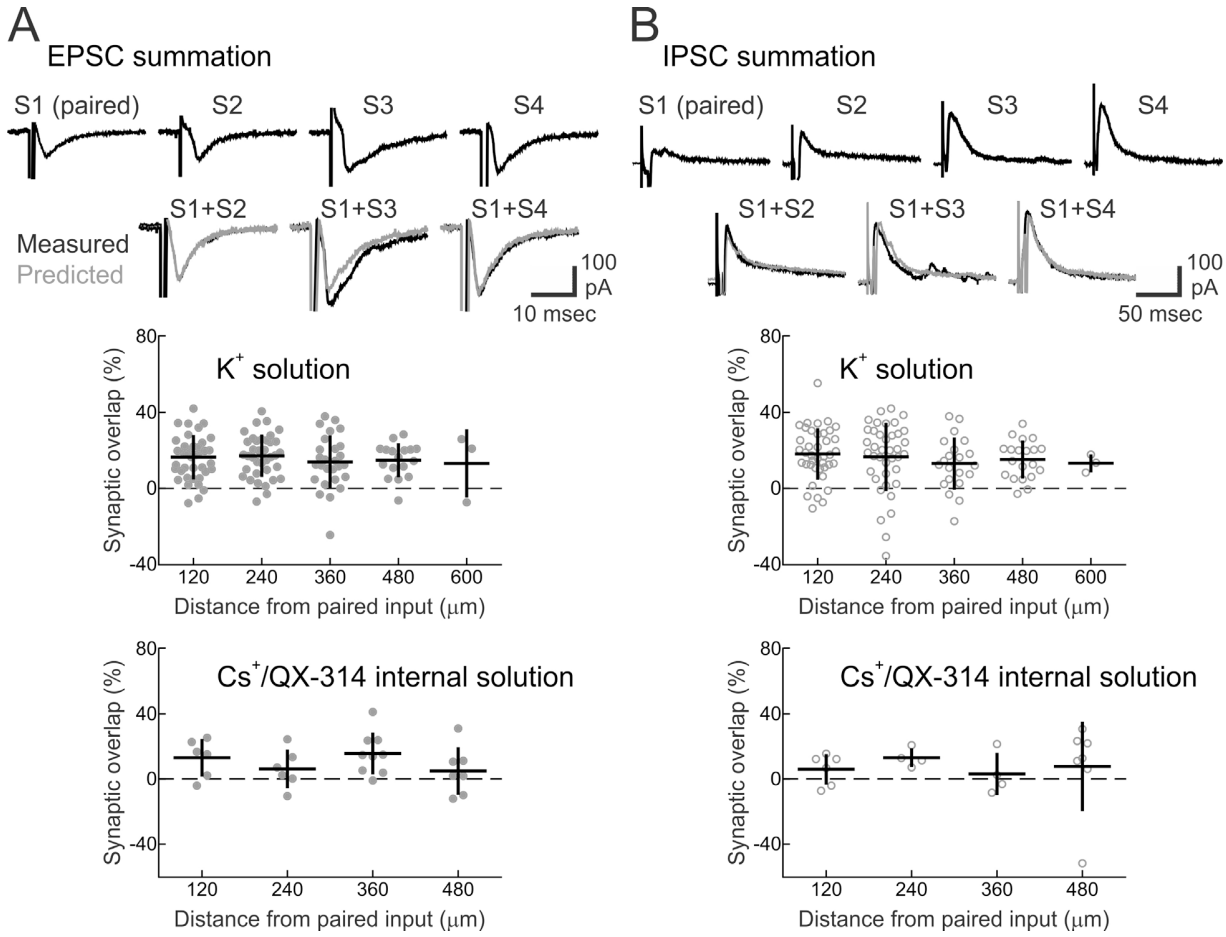


Figure S1. Assessment of synaptic overlap between different stimulation channels in auditory cortical slices, Related to **Figure 1**.

(**A**) EPSC summation across channels. Top, example traces for one recording; there were four active stimulation electrodes (S1-S4), and S1 was the paired channel. Activating S1 together with either S2, S3, or S4 showed approximately linear summation of EPSC pairs, indicating minimal overlap between inputs activated by S1 and S2, S3, or S4. Scale: 10 msec, 100 pA. Middle, summary of channel overlap percentage across all cells with 'current-clamp' K^+-based internal pipette solution used for STDP experiments (120 μm inter-channel distance); 0% overlap indicates independent summation (120 μm from paired channel: 16.5±11.7% overlap, mean±SD, n=40; 240 μm from paired channel: 17.2±11.1% overlap, n=37; 360 μm from paired channel: 13.9±14.0% overlap, n=27; 480 μm from paired channel: 14.9±8.9% overlap, n=17; 600 μm from paired channel: 13.2±18.0% overlap, n=3). Bottom, summary of channel overlap percentage in different cells recorded with 'voltage-clamp' Cs^+-and QX-314-based internal pipette solution (120 μm from

paired channel: 13.1±4.7% overlap, n=6; 240 µm from paired channel: 6.2±4.9% overlap, n=6; 360 µm from paired channel: 15.6±4.3% overlap, n=9; 480 µm from paired channel: 5.0±5.5% overlap, n=7).

(**B**) IPSC summation across channels. Top, example traces from same cell as in **A**, showing IPSCs evoked by stimulation of the four active channels. Scale: 50 msec, 100 pA. Middle, summary of IPSC overlap across channels (120 µm from paired channel: 18.2±13.6% overlap, n=40; 240 µm from paired channel: 16.7±18.0% overlap, n=39; 360 µm from paired channel: 13.1±13.7% overlap, n=21; 480 µm from paired channel: 15.3±10.1% overlap, n=20; 600 µm from paired channel: 13.2±4.8% overlap, n=3). Bottom, channel overlap percentage with 'voltage-clamp' solution (120 µm from paired channel: 5.9±3.9% overlap, n=6; 240 µm from paired channel: 13.1±2.9% overlap, n=4; 360 µm from paired channel: 3.1±6.5% overlap, n=4; 480 µm from paired channel: 7.8±10.4% overlap, n=7).

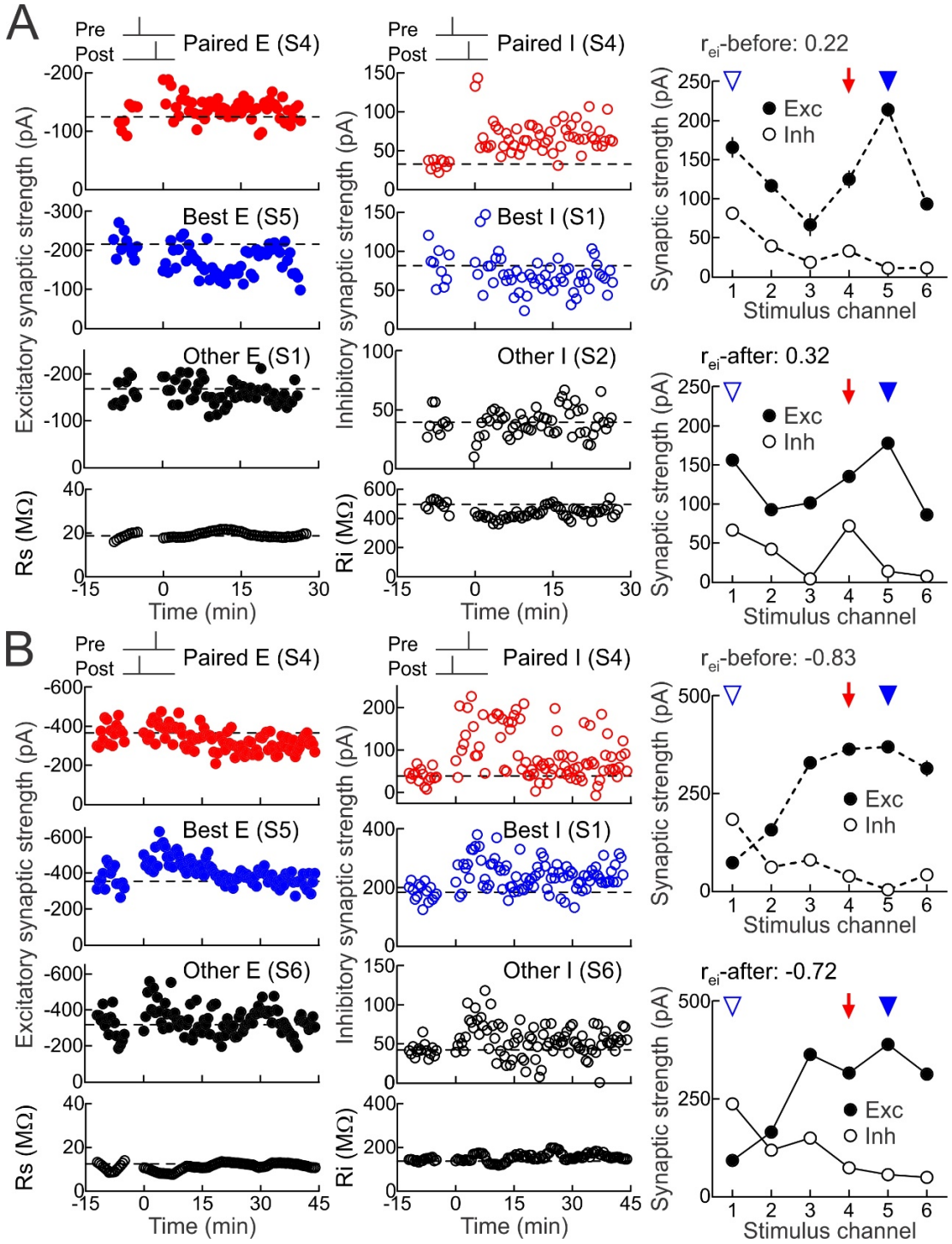


Figure S2. Examples of paired and heterosynaptic STDP increasing excitatory-inhibitory correlation when initially low, Related to **Figure 2**.

(**A**) Example of excitatory and inhibitory plasticity induced by pre→post pairing at channel S4 (red, Δt: 4 msec; EPSCs before pairing: −124.8±7.5 pA; EPSCs after pairing: −135.2±4.4 pA,

increase of 8.3%; IPSCs before pairing: 32.7±2.0 pA; IPSCs after pairing: 71.7±2.4 pA, increase of 119.2%). Dashed line, pre-pairing mean. Upper middle, heterosynaptic LTD at the strongest unpaired inputs onto this cell (blue, EPSCs at best channel S5 before: −214.2±9.8 pA, EPSCs after: −178.0±7.0 pA, decrease of −16.9%; IPSCs at best channel S1 before: 81.3±7.7 pA, IPSCs after: 66.4±4.0 pA, decrease of −18.4%). Lower middle, example other inputs (black, EPSCs at channel S1 before: −165.9±8.4 pA, EPSCs after: −156.2±4.3 pA, decrease of −5.8%; IPSCs at channel S2 before: 39.3±3.6 pA, IPSCs after: 42.1±3.2 pA, increase of 7.1%). Bottom, series and input resistances were stable (Rs before: 19.0±0.5 MΩ, Rs after: 20.7±0.2 MΩ, increase of 9.0%; Ri before: 496.8±11.2 MΩ, Ri after: 448.9±7.4 MΩ, decrease of −9.6%). Right, increase in excitatory-inhibitory correlation across all channels (r_{ei}-before: 0.22; r_{ei}-after: 0.32). Red arrow, paired channel. Blue arrowheads, original best excitation (filled) and inhibition (open).

(**B**) Example of plasticity induced by post→pre pairing at channel S4 (red, Δt: −4 msec; EPSCs before pairing: −363.8±14.2 pA; EPSCs after pairing: −316.7±11.9 pA, decrease of −12.9%; IPSCs before pairing: 38.4±4.3 pA; IPSCs after pairing: 72.9±13.2 pA, increase of 89.7%). Upper middle, heterosynaptic LTP at the strongest unpaired inputs onto this cell (blue, EPSCs at best channel S5 before: −370.1±15.6 pA, EPSCs after: −390.2±11.4 pA, increase of 5.4%; IPSCs at best channel S1 before: 183.9±7.4 pA, IPSCs after: 237.1±10.4 pA, increase of 28.9%). Lower middle, example other inputs (black, EPSCs at channel S6 before: −314.7±20.3 pA, EPSCs after: −313.7±15.6 pA, decrease of −0.3%; IPSCs at channel S6 before: 41.8±2.2 pA, IPSCs after: 48.7±5.0 pA, increase of 16.5%). Bottom, series and input resistances were stable (Rs before: 10.9±0.5 MΩ, Rs after: 12.5±0.2 MΩ, increase of 15.0%; Ri before: 143.8±1.6 MΩ, Ri after: 161.4±2.8 MΩ, increase of 12.3%). Right, increase in excitatory-inhibitory correlation across all channels (r_{ei}-before: −0.83; r_{ei}-after: −0.72).

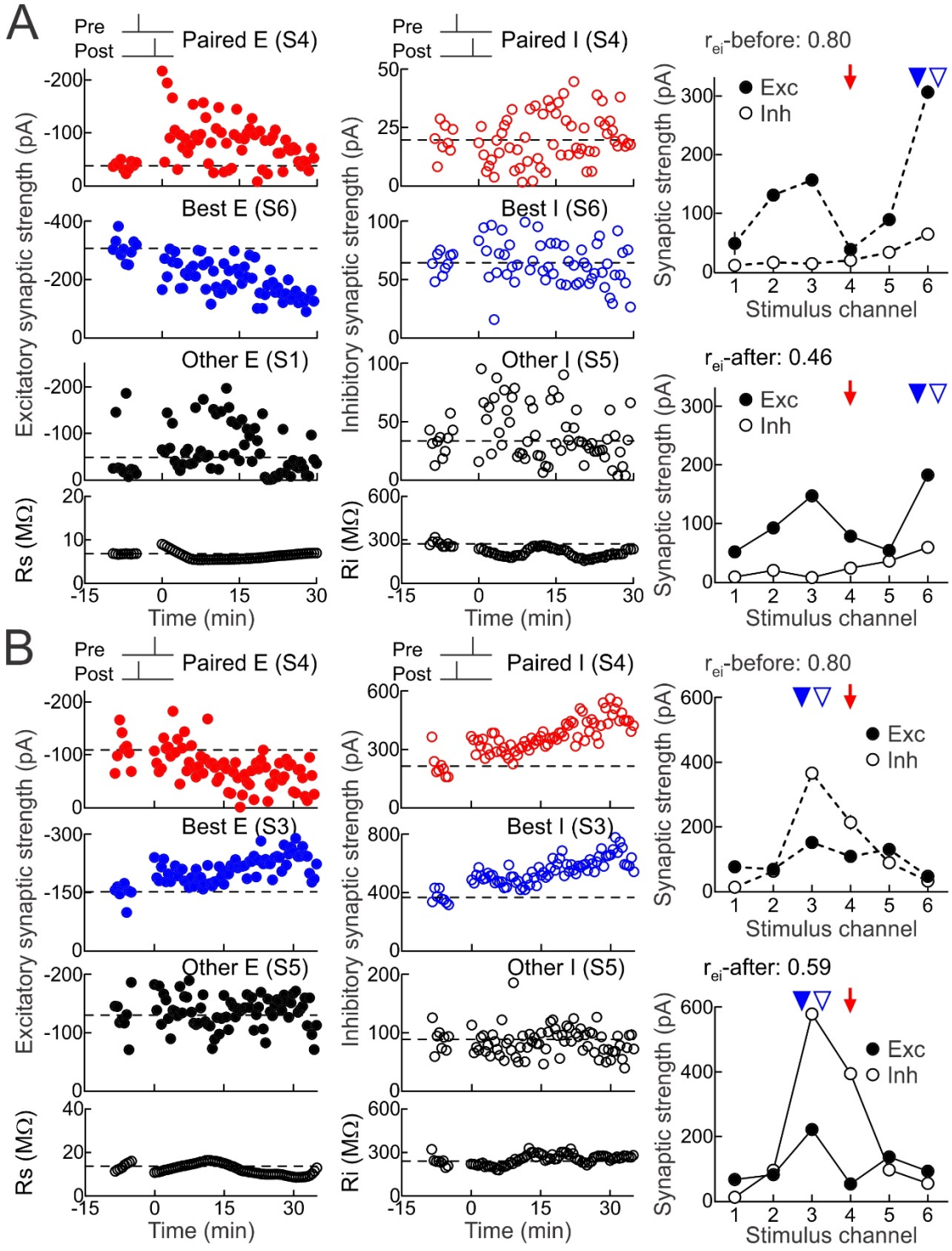


Figure S3. Examples of paired and heterosynaptic STDP decreasing excitatory-inhibitory correlation when initially high, Related to **Figure 2**.

(**A**) Example of excitatory and inhibitory plasticity induced by pre→post pairing at channel S4 (red, Δt: 4 msec; EPSCs before pairing: –38.0±2.6 pA; EPSCs after pairing: –78.3±7.1 pA,

increase of 105.9%; IPSCs before pairing: 19.7±2.4 pA; IPSCs after pairing: 23.8±2.4 pA, increase of 20.8%). Upper middle, heterosynaptic LTD at the strongest unpaired inputs onto this cell (blue, EPSCs at best channel S6 before: −306.5±12.4 pA, EPSCs after: −182.6±11.0 pA, decrease of −40.4%; IPSCs at best channel S6 before: 64.4±3.1 pA, IPSCs after: 59.3±3.9 pA, decrease of −7.9%). Lower middle, example other inputs (black, EPSCs at channel S1 before: −48.6±19.9 pA, EPSCs after: −51.5±10.6 pA, increase of 6.1%; IPSCs at channel S5 before: 33.5±4.1 pA, IPSCs after: 35.8±4.0 pA, increase of 6.8%). Bottom, series and input resistances were stable (Rs before: 6.8±0.02 MΩ, Rs after: 6.0±0.1 MΩ, decrease of −11.2%; Ri before: 272.4±6.5 MΩ, Ri after: 193.4±6.3 MΩ, decrease of −28.9%). Right, decrease in excitatory-inhibitory correlation across all channels (r_{ei}-before: 0.80; r_{ei}-after: 0.46). Red arrow, paired channel. Blue arrowheads, original best excitation (filled) and inhibition (open).

(**B**) Example of plasticity induced by post→pre pairing at channel S4 (red, Δt: −5 msec; EPSCs before pairing: −108.9±12.1 pA; EPSCs after pairing: −52.9±6.6 pA, decrease of −51.4%; IPSCs before pairing: 214.3±23.7 pA; IPSCs after pairing: 394.8±12.4 pA, increase of 84.2%). Upper middle, heterosynaptic LTP at the strongest unpaired inputs onto this cell (blue, EPSCs at best channel S3 before: −151.8±8.2 pA, EPSCs after: −222.0±5.0 pA, increase of 46.2%; IPSCs at best channel S3 before: 366.6±15.6 pA, IPSCs after: 580.2±9.9 pA, increase of 58.3%). Lower middle, example other inputs (black, EPSCs at channel S5 before: −130.0±11.7 pA, EPSCs after: −136.6±5.0 pA, increase of 5.1%; IPSCs at channel S5 before: 88.8±7.4 pA, IPSCs after: 96.7±3.9 pA, increase of 8.9%). Bottom, series and input resistances were stable (Rs before: 13.7±0.6 MΩ, Rs after: 11.3±0.3 MΩ, decrease of −17.8%; Ri before: 240.1±13.2 MΩ, Ri after: 263.5±6.5 MΩ, increase of 9.7%). Right, decrease in excitatory-inhibitory correlation across all channels (r_{ei}-before: 0.80; r_{ei}-after: 0.59).

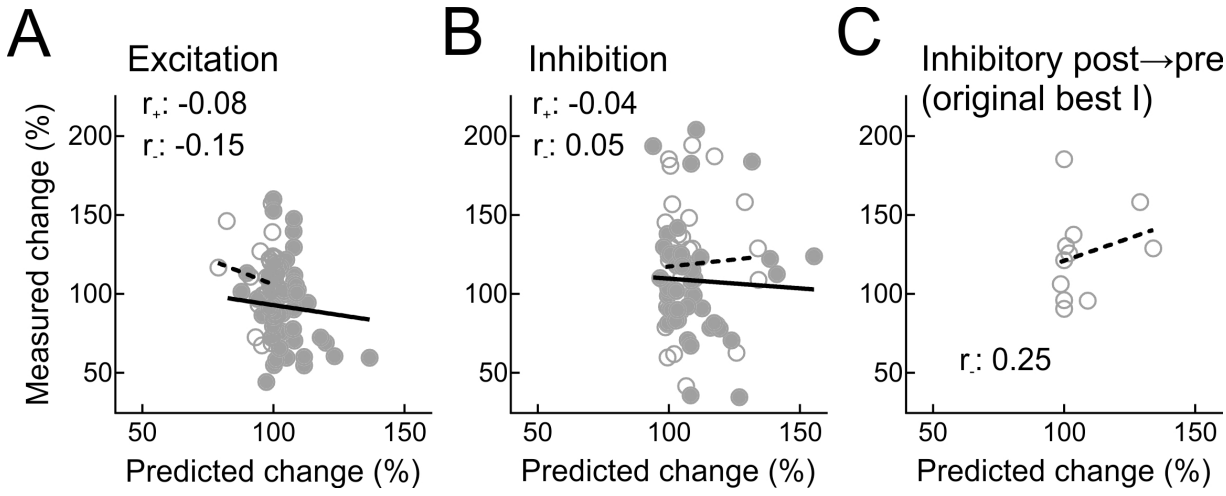


Figure S4. No significant correlations between synaptic overlap and heterosynaptic plasticity across channels, Related to **Figure 2**.

(**A**) Predicted amount of synaptic modification for EPSCs computed from overlap with paired channel for unpaired channels (i.e., sublinear summation as in **Figure S1** across inputs and cells) vs. experimentally-measured modification of each EPSC after pairing. Filled circles, pre→post pairing experiments ('r_+': –0.08, p=0.58, solid line); open circles, post→pre pairing experiments ('r_-': –0.15, p=0.43, dashed line).

(**B**) As in **A** but for unpaired IPSCs ('r_+': –0.04, p=0.79, solid line; 'r_-': 0.05, p=0.81, dashed line).

(**C**) For post→pre pairing, paired IPSCs and the original best IPSCs were both potentiated on average (**Figure 2B**); there was no significant correlation between the predicted inhibitory LTP (from the measured amount of synaptic overlap between paired and best inhibitory inputs) vs. the experimentally-measured heterosynaptic LTP ('r_-': 0.25, p=0.46).

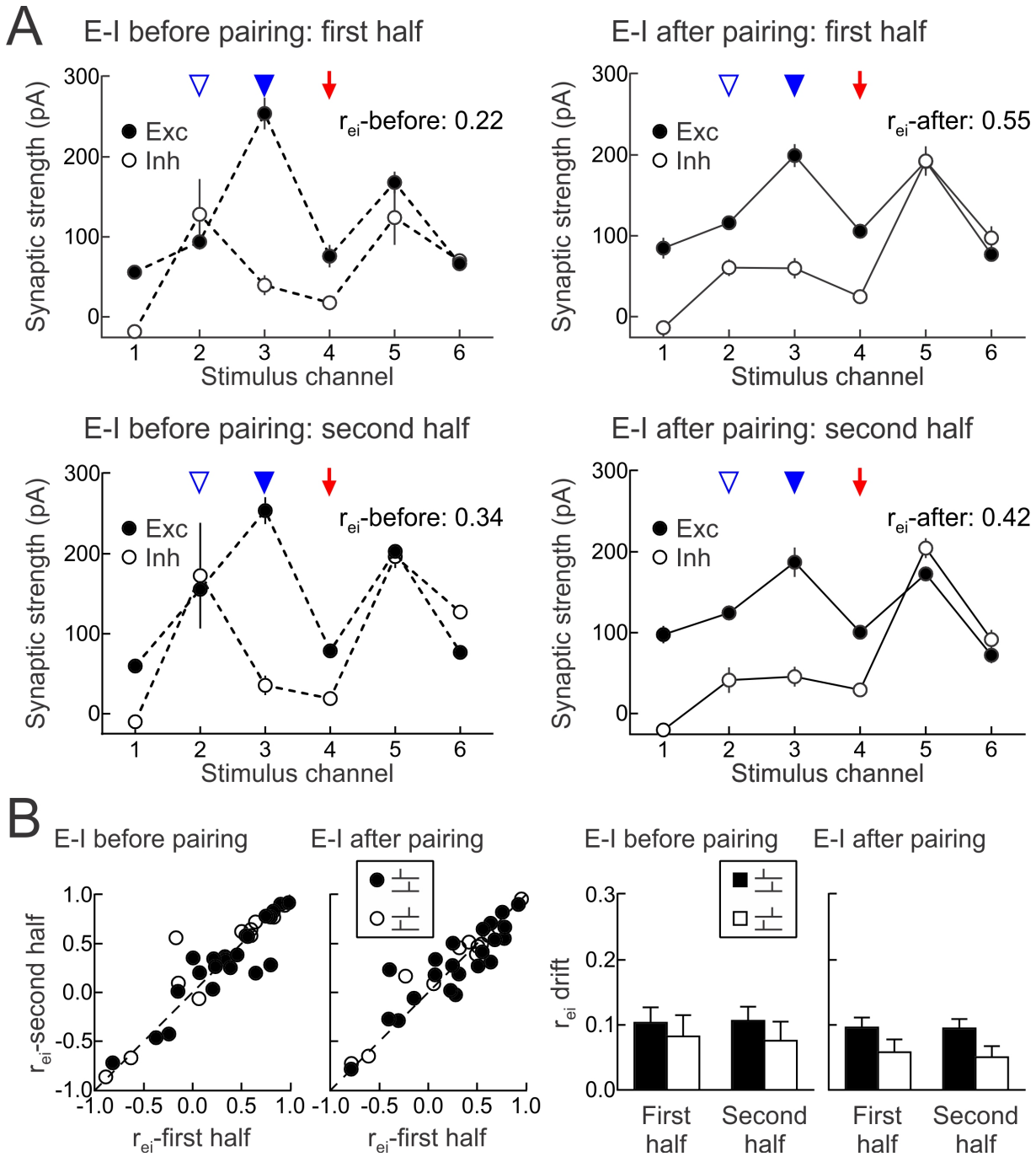


Figure S5. Minimal drift in excitatory-inhibitory correlation during baseline and 16-25 minutes after pairing, Related to **Figure 2**.

(**A**) Excitatory and inhibitory strengths from cell in **Figure 1**. Left, during first half of baseline period (top, r_{ei}-first half: 0.22) and second half of baseline period (top, r_{ei}-second half: 0.34) before pairing. Full baseline r_{ei}-before: 0.25. Right, r_{ei} during 16-20 minutes after pairing (top, r_{ei}-first

half: 0.55) and 21-25 minutes after pairing (top, r_{ei}-second half: 0.42). Full 16-25 minutes r_{ei}-after: 0.48.

(**B**) Differences between r_{ei} for first half and second half of baseline period and 16-25 minutes after pairing compared to value from entire period ('r_{ei} drift'). Left, individual recordings from **Figure 2**. Right, average difference in r_{ei} between first or second halves of baseline period and full baseline (r_{ei} drift pre→post pairing, first half: 0.10±0.02, second half: 0.11±0.02; r_{ei} drift post→pre pairing, first half: 0.08±0.03, second half: 0.08±0.03), and for first and second halves of post-pairing period compared to full 16-25 minutes (r_{ei} drift pre→post pairing, first half: 0.10±0.02, second half: 0.09±0.01; r_{ei} drift post→pre pairing, first half: 0.06±0.02, second half: 0.05±0.02). Filled symbols and bars, pre→post pairing; open symbols and bars, post→pre pairing.

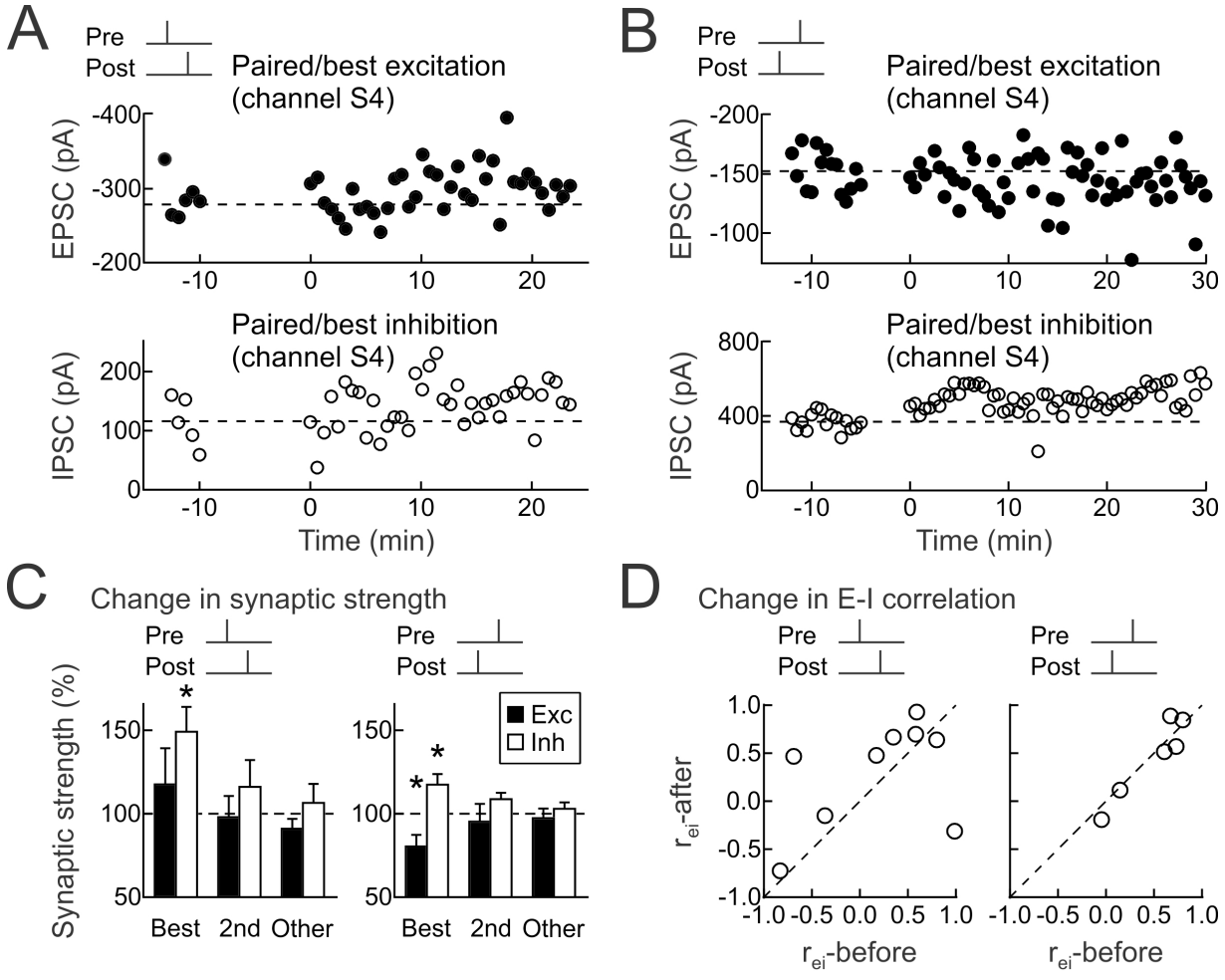


Figure S6. Spike pairing at the best inputs, Related to **Figure 2**.

(**A**) Example of pre→post pairing at the original best channel S4 for both excitation and inhibition (*Δt*: 4 msec; EPSCs before pairing: –277.9±6.4 pA; EPSCs after pairing: –308.5±9.5 pA, increase of 11.0%; IPSCs before pairing: 115.9±18.8 pA; IPSCs after pairing: 153.8±7.7 pA, increase of 32.7%).

(**B**) Example of post→pre pairing at the original best channel S4 for both excitation and inhibition (*Δt*: –4 msec; EPSCs before pairing: –151.1±4.9 pA; EPSCs after pairing: –144.9±5.1 pA, decrease of –4.1%; IPSCs before pairing: 370.2±12.9 pA; IPSCs after pairing: 498.2±10.0 pA, increase of 34.6%).

(**C**) Summary of experiments with pairing at original best input channels for pre→post pairing (left, paired/best EPSCs increased by 17.7±21.7%, n=9, p=0.44, Student's paired two-tailed t-test;

2^{nd} best but unpaired EPSCs decreased by −2.1±12.8%, p=0.88; paired/best IPSCs increased by 49.2±14.9%, p<0.02; 2^{nd} best but unpaired IPSCs increased by 16.1±16.1%, p=0.35; unpaired EPSCs decreased by −8.9±5.9%, p=0.17; unpaired IPSCs increased by 6.5±11.4%, p=0.59), and post→pre pairing (right, paired/best EPSCs decreased by −19.5±6.9%, n=6, p<0.04; 2^{nd} best but unpaired EPSCs decreased by −4.7±10.6%, p=0.68; paired/best IPSCs increased by 17.4±6.3%, p<0.05; 2^{nd} best but unpaired IPSCs increased by 8.7±3.8%, p=0.07; unpaired EPSCs decreased by −2.6±5.7%, p=0.67; unpaired IPSCs increased by 3.0±3.8%, p=0.65).

(**D**) Excitatory-inhibitory correlation before (r_{ei}-before) and after (r_{ei}-after) pre→post pairing (left, n=9) or post→pre pairing (n=6).

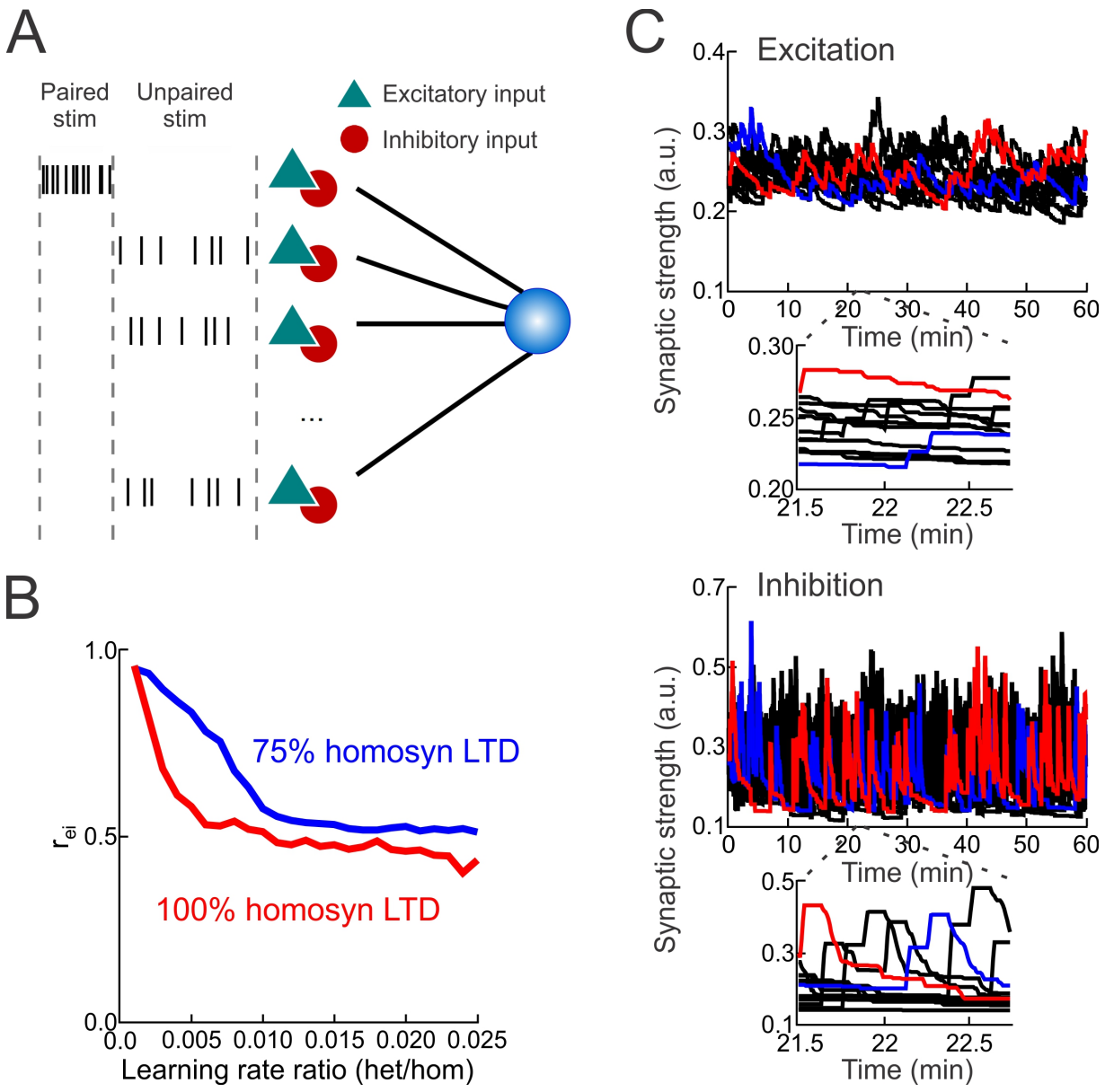


Figure S7. Biophysical model of homosynaptic and heterosynaptic plasticity, Related to **Figure 3**.

(**A**) Schematic of the biophysical model with a single postsynaptic neuron receiving inputs from different channels consisting of excitatory (green) and inhibitory (red) populations, with alternating sequences of paired and unpaired stimulation phases.

(**B**) Reducing the amount of homosynaptic LTD to 75% of original leads to higher r_{ei} set-points.

(**C**) Weight dynamics of a simulation with an excitatory heterosynaptic to homosynaptic learning rate ratio of: $\frac{\eta_{het}^{E}}{\eta_{w}^{E}} = 1.3 * 10^{-2}$, with $\eta_{het}^{E} = 1.3 * 10^{-5}\, ms^{-1}$, $\eta_{het}^{I} = 1.3 * 10^{-4}\, ms^{-1}$. Left, excitatory inputs; right, inhibitory inputs. Insets, weight dynamics between timepoints 21.5-22.75 minutes. Inhibitory weight dynamics usually change the order of the strongest to weakest channel faster than the excitatory weights, leading to changes in r_{ei}. Color is used just to highlight dynamics of two different channels (red, channel #4; blue, channel #7).

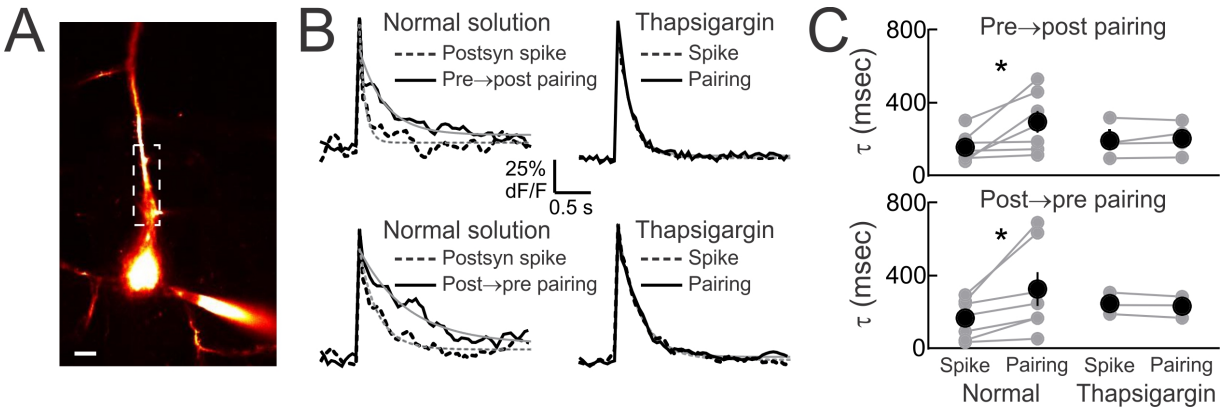


Figure S8. Spike pairing leads to release of dendritic Ca^{2+} from internal stores, Related to **Figure 6**.

(**A**) Example of two-photon imaging of pairing-induced Ca^{2+} signals in apical dendrites of layer 5 pyramidal neurons; dashed box, analysis region. Scale: 15 µm.

(**B**) Examples of Ca^{2+} transients evoked by single postsynaptic spikes (black dashed line) and postsynaptic spikes paired with presynaptic stimulation (black solid line). Shown are single exponential fits to each transient (thin gray lines). Thapsigargin (10 µM) included in the whole-cell pipette prevented the broadening of dendritic Ca^{2+} after both pre→post pairing and post→pre pairing.

(**C**) Summary of pairing-induced Ca^{2+} release quantified by time constant τ of single exponential fits. With normal intracellular solution, pairing broadened the evoked Ca^{2+} signals for both pre→post pairing (top, postsynaptic spike alone, τ: 155.2±29.9 msec; pairing, τ: 295.3±60.6 msec, n=7, $p<0.04$) and post→pre pairing (bottom, postsynaptic spike alone, τ: 164.1±40.4 msec; pairing, τ: 324.9±92.9 msec, n=7, $p<0.04$). This broadening of the Ca^{2+} event was prevented by thapsigargin (10 µM) in the internal solution (pre→post pairing, postsynaptic spike alone, τ: 191.6±46.5 msec; pairing, τ: 203.6±43.1 msec, n=4, $p>0.4$; post→pre pairing, postsynaptic spike alone, τ: 244.7±34.6 msec; pairing, τ: 229.7±34.1 msec, n=3, $p>0.1$).

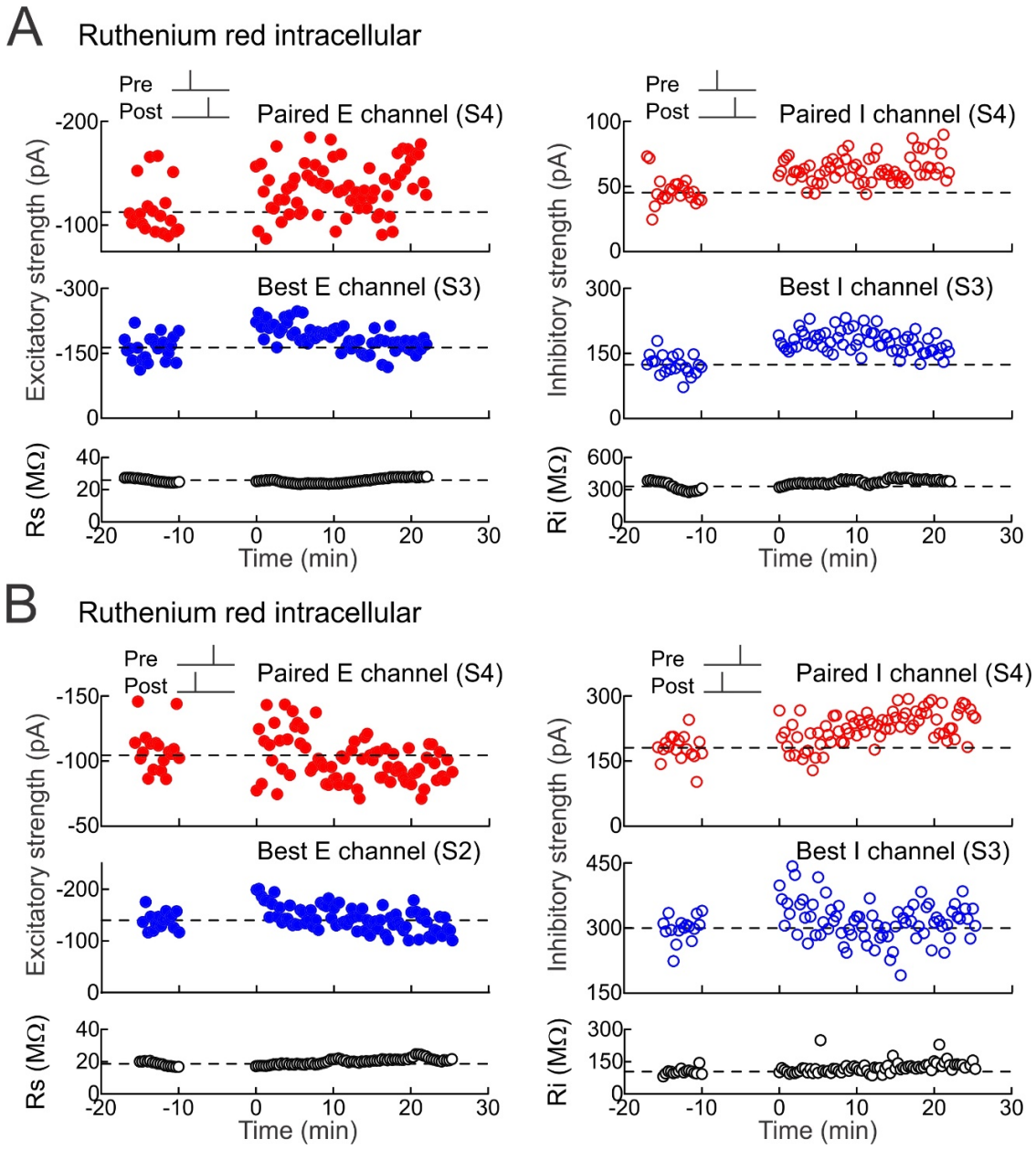


Figure S9. Intracellular blockade of Ca^{2+} store signaling with ruthenium red prevents heterosynaptic changes at the best inputs but spares STDP at paired inputs, Related to **Figure 6**. (**A**) Ruthenium red in the whole-cell pipette (20 μM) prevents heterosynaptic excitatory and inhibitory LTD after pre→post pairing. Top, example of excitatory LTP (left) and inhibitory LTP (right) induced by pre→post pairing at channel S4 (red, Δt=1 msec; EPSCs before pairing: −112.6±5.8 pA, EPSCs after pairing: −143.3±6.3 pA, increase of 27.3%; IPSCs before pairing: 45.2±2.0 pA, IPSCs after pairing: 67.8±2.7 pA, increase of 49.9%). Dashed line, pre-pairing mean.

Middle, ruthenium red prevented heterosynaptic LTD at the strongest unpaired inputs onto this cell (blue, EPSCs at channel S3 before: −163.1±6.1 pA, EPSCs after: −167.2±5.0 pA, increase of 2.5%; IPSCs at channel S3 before: 123.9±4.9 pA, IPSCs after: 164.3±4.9 pA, increase of 32.6%). Bottom, series and input resistance were stable (R_s before: 25.9±0.6 MΩ, R_s after: 27.6±0.3 MΩ, increase of 7.0%; R_i before: 330.4±20.7 MΩ, R_i after: 405.3±22.2 MΩ, increase of 13.8%).

(**B**) Ruthenium red prevents heterosynaptic excitatory and inhibitory LTP after post→pre pairing. Top, example of excitatory LTD (left) and inhibitory LTP (right) induced by post→pre pairing at channel S4 (red, Δt=−2 msec; EPSCs before pairing: −108.6±3.9 pA, EPSCs after pairing: −93.8±2.1 pA, decrease of −13.7%; IPSCs before pairing: 180.3±7.2 pA, IPSCs after pairing: 249.1±6.3 pA, increase of 38.1%). Middle, ruthenium red prevented heterosynaptic LTP at the strongest unpaired inputs onto this cell (blue, EPSCs at channel S2 before: −139.7±4.3 pA, EPSCs after: −132.4±4.1 pA, decrease of −5.2%; IPSCs at channel S3 before: 299.2±7.3 pA, IPSCs after: 318.9±6.9 pA, increase of 6.6%). Bottom, series and input resistance were stable (R_s before: 18.7±0.4 MΩ, R_s after: 21.8±0.2 MΩ, increase of 16.9%; R_i before: 103.6±3.4 MΩ, R_i after: 147.6±13.6 MΩ, increase of 13.6%).

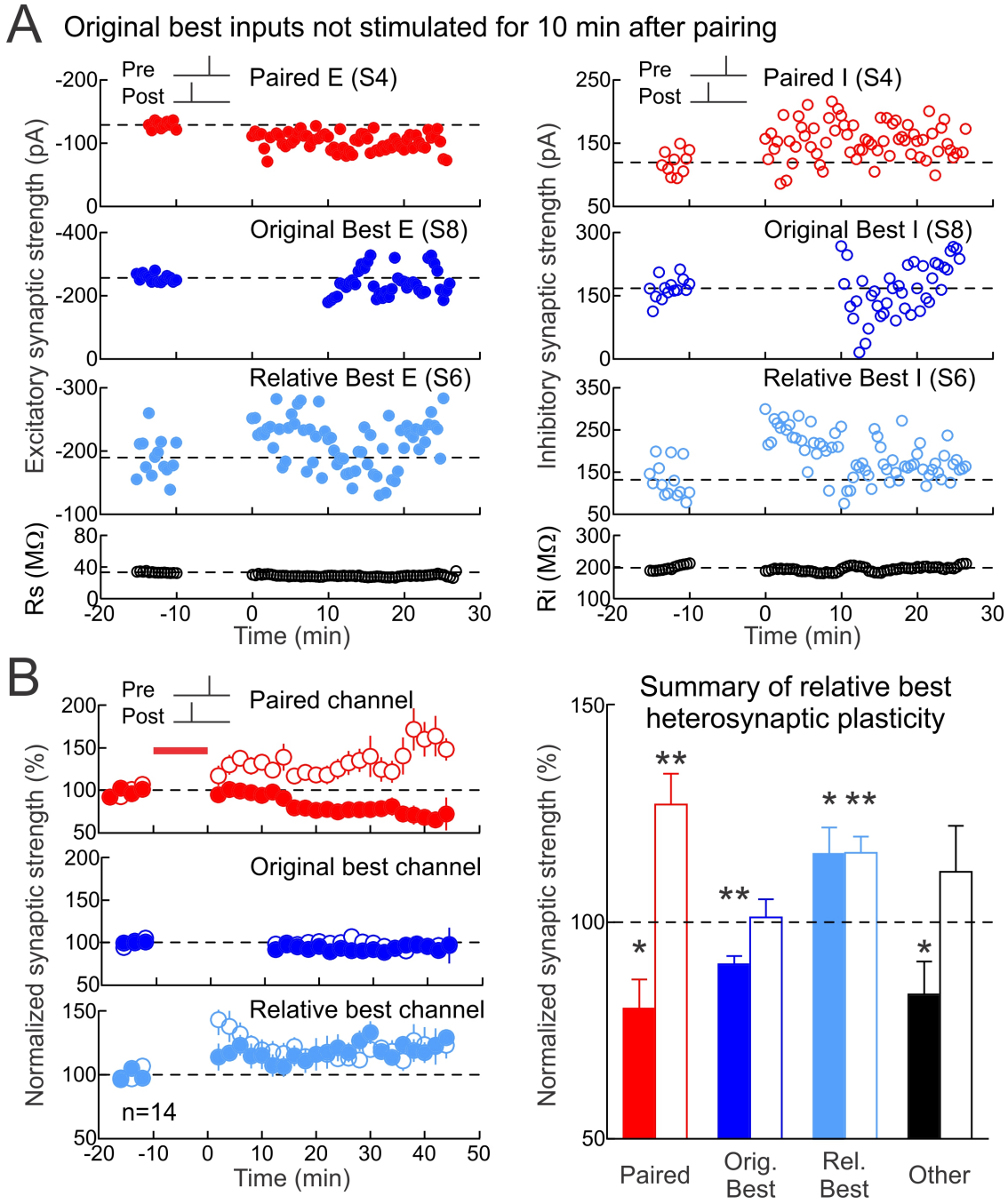


Figure S10. Heterosynaptic modifications to relative best inputs when original best input not presented in ten minutes following post→pre pairing, Related to **Figure 8**.

(**A**) Deactivating original best input channel led to heterosynaptic excitatory and inhibitory LTP at the second best ('relative best') channel after post→pre pairing. Top, example of excitatory LTD (left) and inhibitory LTP (right) induced by post→pre pairing at channel S4 (red, Δt=4 msec;

EPSCs before pairing: −129.0±1.9 pA, EPSCs after pairing: −99.4±2.6 pA, decrease of −22.9%; IPSCs before pairing: 119.7±5.9 pA, IPSCs after pairing: 149.1±7.2 pA, increase of 24.6%). Dashed line, pre-pairing mean. Upper middle, original best inputs evoked by S8 were unaltered when this channel was turned off for ten minutes immediately after pairing (dark blue, original best EPSCs before: −256.6±3.2 pA, original best EPSCs after: −242.6±9.0 pA, decrease of −5.5%; original best IPSCs before: 167.8±6.8 pA, original best IPSCs after: 183.2±10.5 pA, increase of 9.2%). Lower middle, heterosynaptic depression was induced at the relative best inputs evoked by S6 (light blue, relative best EPSCs before: −189.7±8.4 pA, relative best EPSCs after: −208.5±8.3 pA, increase of 9.9%; relative best IPSCs before: 131.8±11.0 pA, relative best IPSCs after: 167.2±7.2 pA, increase of 26.8%). Bottom, series and input resistance (R_s before: 33.3±0.2 MΩ, R_s after: 28.7±0.3 MΩ, decrease of −13.8%; R_i before: 197.6±2.3 MΩ, R_i after: 196.6±0.7 MΩ, decrease of −0.5%).

(**B**) Summary of post→pre experiments with original best input channel deactivated for the ten-minute after-pairing period. Shown are changes to paired inputs (red; paired EPSCs decreased by −19.9±6.7% at 16-25 minutes post-pairing, n=14, $p<0.02$, Student's paired two-tailed t-test; paired IPSCs increased by 27.2±7.1%, $p<0.005$), original best inputs (dark blue; originally-largest EPSCs decreased by −9.7±1.9% at 16-25 minutes post-pairing, $p<0.01$; originally-largest IPSCs increased by 1.2±4.2%, $p>0.7$), relative best inputs (light blue; EPSCs increased by 15.8±6.1% at 16-25 minutes post-pairing, $p<0.04$; IPSCs increased by 16.1±3.7%, $p<0.002$), and averaged other inputs (black; EPSCs decreased by −16.6±7.6% at 16-25 minutes post-pairing, $p<0.03$; IPSCs increased by 11.6±10.6%, $p>0.2$). Filled symbols, excitation; open symbols, inhibition. Left, time course (compare with **Fig. 2B**); right, summary of changes at 16-25 minutes after pairing.

V. Formation and computational implications of assemblies in neural circuits

Miehl*, C., Onasch*, S., Festa, D. & Gjorgjieva, J. Formation and computational implications of assemblies in neural circuits. *Journal of Physiology* (2022). https://doi.org/10.1113/JP282750

J Physiol 0.0 (2022) pp 1–20

The Journal of Physiology

TOPICAL REVIEW

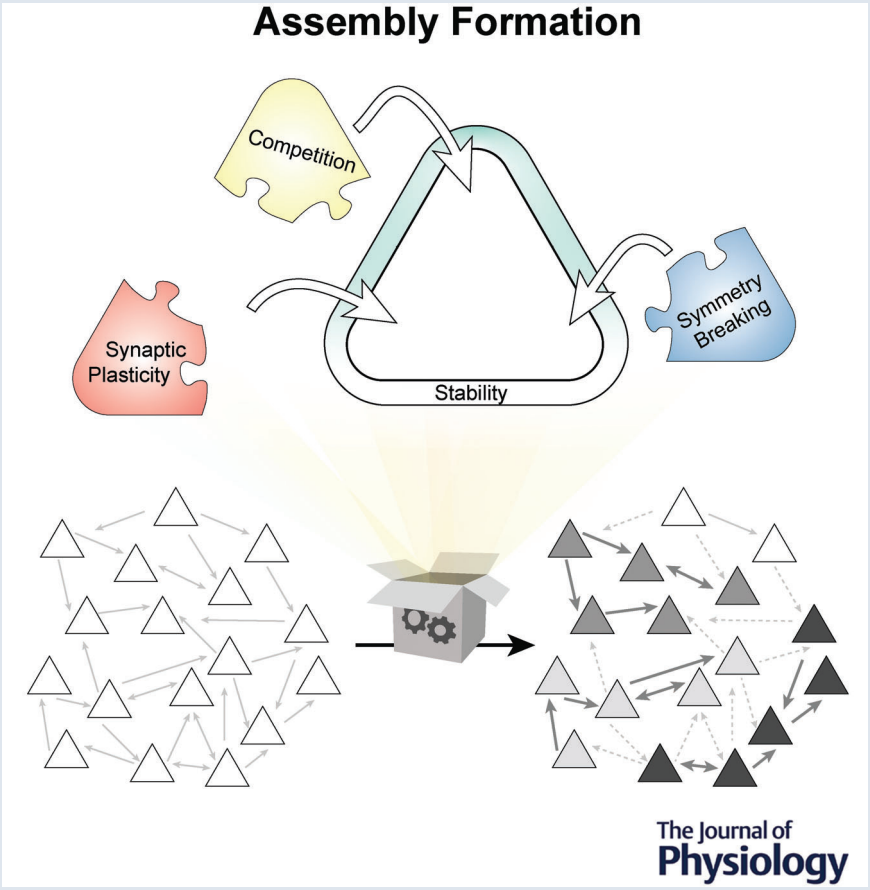
Formation and computational implications of assemblies in neural circuits

Christoph Miehl[1,2], Sebastian Onasch[1,2], Dylan Festa[1,2] and Julijana Gjorgjieva[1,2]

[1]*Computation in Neural Circuits, Max Planck Institute for Brain Research, Frankfurt, Germany*
[2]*School of Life Sciences, Technical University of Munich, Freising, Germany*

Handling Editors: Katalin Toth & Michael Okun

The peer review history is available in the Supporting Information section of this article (https://doi.org/10.1113/JP282750#support-information-section).



Abstract In the brain, patterns of neural activity represent sensory information and store it in non-random synaptic connectivity. A prominent theoretical hypothesis states that assemblies, groups of neurons that are strongly connected to each other, are the key computational units underlying perception and memory formation. Compatible with these hypothesised assemblies, experiments have revealed groups of neurons that display synchronous activity,

C. Miehl and S. Onasch contributed equally to this work.

DOI: 10.1113/JP282750

either spontaneously or upon stimulus presentation, and exhibit behavioural relevance. While it remains unclear how assemblies form in the brain, theoretical work has vastly contributed to the understanding of various interacting mechanisms in this process. Here, we review the recent theoretical literature on assembly formation by categorising the involved mechanisms into four components: synaptic plasticity, symmetry breaking, competition and stability. We highlight different approaches and assumptions behind assembly formation and discuss recent ideas of assemblies as the key computational unit in the brain.

(Received 20 May 2022; accepted after revision 22 August 2022; first published online 6 September 2022)
Corresponding author Julijana Gjorgjieva: School of Life Sciences, Technical University of Munich, Maximum-von-Imhof-Forum 3, Freising 85354, Germany. Email: gjorgjieva@tum.de

Abstract figure legend Assembly formation. Assemblies are groups of strongly connected neurons formed by the interaction of multiple mechanisms and with vast computational implications. Four interacting components are thought to drive assembly formation: synaptic plasticity, symmetry breaking, competition and stability.

Introduction

Originating from the ideas of Lorente de Nó (1938) and Donald O. Hebb (1949), a prominent theoretical hypothesis proposed that groups of neurons instead of single neurons are the basic unit of perceptive integration (Buzsáki, 2010; Eichenbaum, 2018; Huyck & Passmore, 2013; Yuste, 2015). It is now widely accepted that groups of neurons that display synchronous activity represent key computational units and are often referred to as 'assemblies', 'ensembles' or 'engrams'. These terms are not always used consistently in the literature; therefore, in this review, we propose the following disambiguation. An 'ensemble' refers to multiple neurons that express a certain degree of synchronous activity without any hypothesis about their connectivity. An 'assembly,' on the other hand, is defined as a group of neurons that has stronger or denser synaptic connections (called the within-assembly weights) among the neurons that constitute it, as opposed to the weights going into or coming out of the assembly (called the across-assembly weights), without necessarily having highly synchronous activity. Finally, an 'engram' describes multiple ensembles or assemblies interconnected and spread across multiple layers and even brain areas.

Although it is often hypothesised that highly synchronous activity within a group of neurons (ensemble) follows from strong synaptic connectivity among these neurons (assembly), distinguishing between activity and connectivity is important. Experimental studies usually investigate ensembles. This is despite the development of new recording techniques which have made it possible to image and manipulate the activity of groups of neurons and link them to behaviour (Carrillo-Reid et al., 2017; Wenzel & Hamm, 2021), and is due to the challenge to directly measure synaptic connectivity between specific neurons experimentally. Despite recent efforts to show experimentally that an ensemble also consists of strong, connected neurons (e.g. Alejandre-García et al., 2022), a clear link is still missing.

Due to the readily accessible information about connectivity and activity in *in silico* network models, theoretical studies can bridge the gap between activity and connectivity, seeking a mechanistic understanding of assembly formation and stability. An early example is Hebb's suggestion that long-term synaptic plasticity mechanisms favour the formation of assemblies among neurons that activate synchronously (Hebb, 1949). In this review, we examine contemporary literature on

Christoph Miehl and **Sebastian Onasch** are PhD students in the research group of Professor Julijana Gjorgjieva at the Max Planck Institute for Brain Research in Frankfurt and the School of Life Sciences at the Technical University in Munich. They joined the Gjorgjieva group during their master degrees and became fascinated by the idea of applying mathematical tools and computational approaches to neuroscience questions. Their interests lie within computational modelling of neuronal microcircuits inspired and constrained by experimental data.

assembly formation in recurrent networks, outlining the key components that, in addition to Hebb's basic principle where 'cells that fire together wire together' (Shatz, 1992), allow for the formation and stability of assembly structures. We identify four components: synaptic plasticity, symmetry breaking, competition and stability (Fig. 1*A*), and highlight how computational studies use different assumptions combining these four components in relation to experimental literature. Finally, we discuss recent ideas of how the concept of assemblies is useful in understanding how brains might 'compute' and we point towards open challenges and possible future research directions. Before diving into the computational perspective, we briefly review experimental findings on ensembles and assemblies.

Experimental evidence for ensembles, assemblies and their formation

Evidence of ensembles in neural activity. Ensembles of neurons have been identified by their temporally coordinated activity patterns in many species and across areas. These patterns appear upon stimulus presentation, such as in the mouse hippocampus (Harris et al., 2003) and visual cortex (Miller et al., 2014), the ferret visual cortex (Berkes et al., 2011) and the zebrafish (Romano et al., 2015). In addition, coordinated neural activity has been measured during spontaneous activity and even in *in vitro* preparations (Cossart et al., 2003; MacLean et al., 2005; Mao et al., 2001), hinting at the possibility that these patterns are generated by underlying local network structures (i.e. assemblies) rather than by common feedforward inputs. Therefore, one proposal supported by experimental evidence is that ensembles observed during spontaneous activity define the realm of possible activity patterns during stimulus presentation (Kenet et al., 2003; Luczak et al., 2009; Malvache et al., 2016; Miller et al., 2014; but see Avitan et al., 2021; Stringer et al., 2019).

The activation of neural ensembles not only correlates with stimulus presentation but also plays an important role in expressing specific behaviours. Precise optogenetic manipulation of ensembles induces, enhances or impairs expressed behaviour, for example in the hippocampus

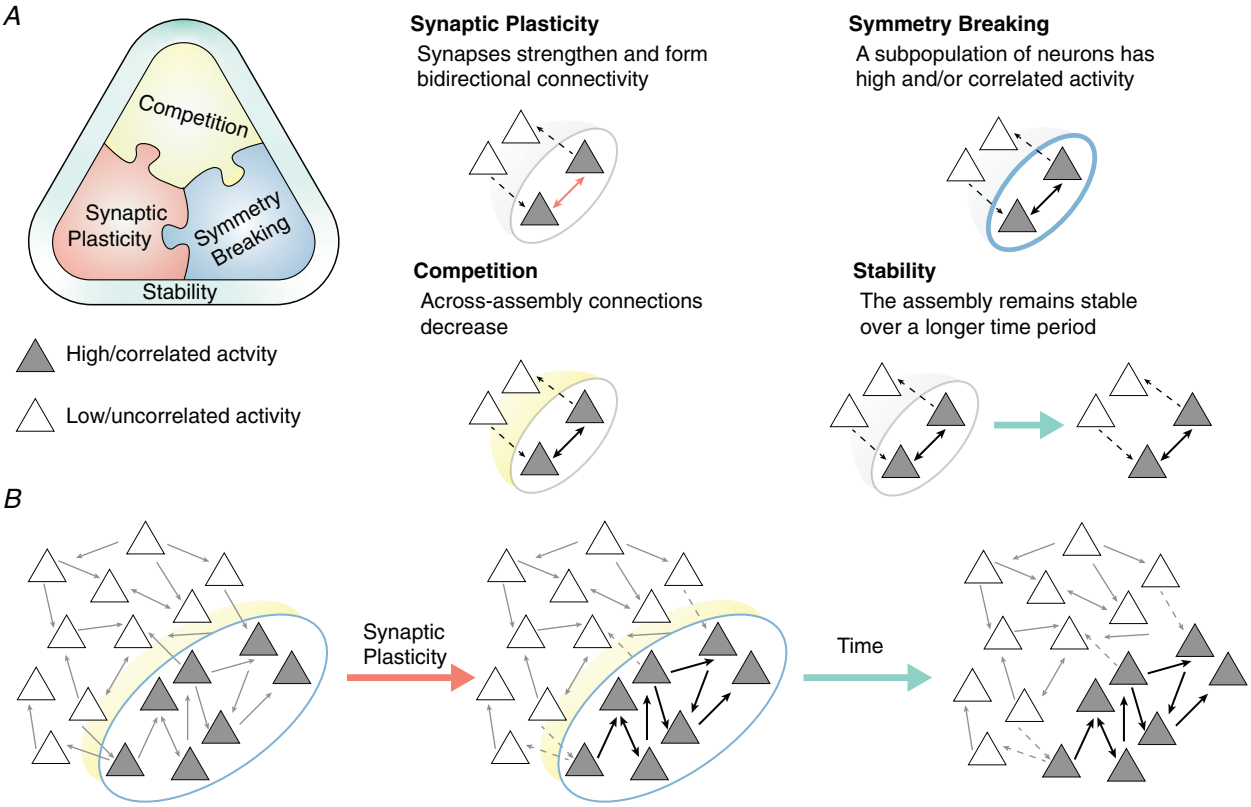


Figure 1. The basic building blocks of assembly formation
A, assembly formation is based on four key components in computational models: synaptic plasticity, symmetry breaking, competition and stability. *B*, due to symmetry breaking, a subpopulation of neurons fires at a high rate and/or with highly correlated activity compared to the remaining neurons. Synaptic plasticity promotes mutual connections within the assembly, while a competition mechanism decreases the across-assembly weights. The newly formed assembly structure is stable over time.

(Liu et al., 2012), the visual cortex (Carrillo-Reid et al., 2019; Marshel et al., 2019) and the orbitofrontal cortex (Jennings et al., 2019). Animals can also be trained to directly report the activation of a cell ensemble composed of fewer than 20 neurons in the barrel cortex and the olfactory bulb, even in the absence of a sensory stimulus (Dalgleish et al., 2020; Gill et al., 2020). However, it is unclear what generates the highly synchronous activity of the neurons within an ensemble: for example, whether the ensembles that lead to specific behaviours are strongly recurrently connected into assemblies or are strongly driven by common input. Additionally, it is unclear to what extent ensemble configuration and its function are preserved over longer time periods, such as days or months. Some recent work has shown that the cortical representation of natural images (Deitch et al., 2021) and odours (Schoonover et al., 2021) slowly shifts over time, a phenomenon termed 'representational drift' (Rule et al., 2019). Representational drift depends on the stimulus type (Marks & Goard, 2021) and is area-dependent, with representations in motor areas being especially stable (Jensen et al., 2022).

Evidence of assemblies in neural connectivity. Direct measures of neural connectivity based on electrophysiology, optogenetic stimulation and electron microscopy often reveal that network structure is far from random. In particular, bidirectional connections occur much more frequently compared to what would be expected if connections were random (Campagnola et al., 2022; Guzman et al., 2016; Jouhanneau et al., 2015; Song et al., 2005; Turner et al., 2022; but see Lefort et al., 2009). Although bidirectional connectivity does not uniquely define an assembly, it is consistent with the idea that neurons in the same assembly should have a denser and stronger mutual connectivity compared to the rest of the population.

Other studies have revealed that neurons are more likely to be connected when they share common neighbours (Perin et al., 2011; Turner et al., 2022), i.e. common input (Yoshimura et al., 2005) or common postsynaptic targets (Brown & Hestrin, 2009). Furthermore, highly correlated cells (Cossell et al., 2015; Ko et al., 2011) or cells that are tuned to the same stimuli have a higher probability of being connected (Ko et al., 2011; Lee et al., 2016; Rossi et al., 2020; Wertz et al., 2015). In the same vein, cells stemming from the same progenitor show similar selectivity for visual stimuli (Li et al., 2012; Ohtsuki et al., 2012) and display high connection probability as well as an increase in reciprocal connectivity (Tarusawa et al., 2016; Yu et al., 2009; but see Cadwell et al., 2020). The fact that functionally related neurons are more likely to be connected suggests that these groups of neurons can be considered assemblies.

Synaptic plasticity as a mechanism for assembly formation. It is widely believed that long-term synaptic plasticity is the underlying mechanism behind connectivity structure formation, including assembly formation (Abbott & Nelson, 2000; Brea & Gerstner, 2016; Feldman, 2009; Magee & Grienberger, 2020; Suvrathan, 2019). Multiple experimental studies have shown that the long-term plasticity of a synapse depends on firing rates and exact spike timing (Maffei, 2018). Long-term potentiation can be evoked by high firing rates (Kirkwood et al., 1996) or pairs of spikes which fire in a causal manner, whereby a postsynaptic spike follows a presynaptic spike (Bi & Poo, 1998; Markram et al., 1997). Therefore, groups of neurons firing at high rate and in a correlated manner (ensembles) should form strong synaptic connections among each other, forming an assembly.

However, experimentally it has been difficult to link synaptic plasticity directly to assembly formation. Recent findings in the mouse barrel cortex have suggested that the probability of neurons firing together increases when repeatedly activating them *in vivo*, tying this effect to long-lasting connectivity changes (Kim et al., 2016). Furthermore, fear memories could be artificially induced by stimulating neurons in the hippocampal region CA3 with a protocol that probably induces within-layer synaptic plasticity (Oishi et al., 2019).

A couple of experimental studies have attempted to imprint assemblies by repeatedly evoked spiking patterns in a selected subgroup of neurons (Carrillo-Reid et al., 2016; Zhang et al., 2020). While this increased correlated firing and spontaneous reactivations of the stimulated subgroup of neurons over days, the underlying mechanism has remained unclear. Long-term synaptic plasticity that leads to stronger recurrent inputs is one possibility (Zhang et al., 2020), but a change in the neuron's intrinsic excitability is also possible (Alejandre-García et al., 2022; Debanne et al., 2019; Zhang & Linden, 2003). Increased intrinsic excitability also plays an important role in memory formation. Studies in the hippocampus and amygdala have shown that cells with high intrinsic excitability during memory formation are likely to be part of the newly formed ensembles that correlate with the learned fear memories (Cai et al., 2016; Rashid et al., 2016). However, while recent studies suggest that memory formation requires changes in feedforward synapses between those ensembles (Abdou et al., 2018; Nabavi et al., 2014), it is unclear whether the same is true for recurrent synapses, i.e. whether memory formation relies on assembly formation.

Therefore, we conclude that despite the abundance of experimental data on the existence and formation of ensembles and assemblies, a clear link between how highly correlated activity and strong recurrent connectivity is still missing.

Formation of assemblies

To complement these experimental findings regarding the existence and relevance of neural ensembles and assemblies, computational studies have proven extremely useful in developing models linking synaptic plasticity to the formation of assemblies. We propose that the problem of stable assembly formation in neural circuits can be understood through four fundamental components (Fig. 1*A*).

The first component, synaptic plasticity, promotes assemblies by strengthening bidirectional connections of within-assembly neurons. The second component, symmetry breaking, represents a form of bias that influences neural activity, and thus can be used by synaptic plasticity to refine and shape assemblies. The third component, competition, strengthens synaptic connections within an assembly and weakens connections across assemblies. Lastly, stability prescribes that learned assemblies remain resilient to fluctuations and keep their connectivity structure intact over time.

Synaptic plasticity. To describe synaptic plasticity, computational models have defined phenomenological descriptions of the interactions between pre- and postsynaptic spikes in the form of learning rules. We first consider one of the best studied rules: Hebbian pairwise spike timing-dependent plasticity (STDP), where a presynaptic spike followed by a postsynaptic spike leads to long-term potentiation (LTP), whereas a postsynaptic spike followed by a presynaptic spike leads to long-term depression (Bi & Poo, 1998). The magnitude of the synaptic change depends on the time difference between the spikes, increasing as the events occur closer in time (Fig. 2*A*). At the circuit level, this rule does not favour the formation of neural assemblies, but rather it disrupts them (Ravid Tannenbaum & Burak, 2016), as it tends to cancel bidirectional connections between neurons (Abbott & Nelson, 2000; Clopath et al., 2010; Song & Abbott, 2001; Song et al., 2000) including synaptic loops that involve multiple neurons (Kozloski & Cecchi, 2010). This is due to the potentiation/depression profile of the rule (Fig. 2*A*). For example, if neuron j spikes before neuron i the synaptic weight that goes from j to i, w_{ij}, will increase; however, the reciprocal connection from neuron i to neuron j, w_{ji}, will decrease (Fig. 2*A*, red vertical lines). Therefore, synaptic plasticity rules with an anti-symmetric profile, as in the Hebbian case, can produce a continuous competition between reciprocal connections, resulting in a 'winner-take-all' mechanism that leaves only one direction intact (Abbott & Nelson, 2000; Song et al., 2000; Song & Abbott, 2001).

To understand the effect of STDP rules on bidirectional connections more generally, computational models describe the dependency between synaptic weight changes, firing statistics and plasticity rule parameters. The firing rates of single neurons and the pairwise correlations between the neurons' spike trains together with the parameters of the STDP rule determine the mean weight change (Kempter et al., 1999; Fig. 2*B*). However, any change of synaptic weights can, in turn, produce changes in the neurons' firing rates and correlation structure. Decomposing synaptic plasticity into structural motifs is a widely used theoretical approach that captures this complex relationship (Hu et al., 2013, 2014; Jovanović & Rotter, 2016; Montangie et al., 2020; Ocker, Hu et al., 2017; Ocker, Josić et al., 2017; Pernice et al., 2011; Ravid Tannenbaum & Burak, 2016; Trousdale et al., 2012). The structural motif framework is based on two assumptions: first, synaptic plasticity occurs on a much slower timescale than the dynamics of neural firing; and second, the neural dynamics follows an approximately linear behaviour. When the first assumption holds, the (slow) weight update does not depend on specific realisations of neural firing, but is instead determined by mean firing rates and by the correlation structure between the neurons' spike trains (Kempter et al., 1999). The second assumption of linearity allows us to compute the full firing statistics analytically from the network weights (Jovanović et al., 2015). This leads to a self-consistent solution (Ocker et al., 2015), where at each iteration first the firing statistics are derived from fixed weights, and then the weights are updated following the interaction between those statistics and synaptic plasticity. While this framework applies to any synaptic plasticity rule, for simplicity here we explain it in the context of the pairwise STDP rule.

Structural motifs are defined as the connectivity paths that a given spike from neuron k in a network travels to neuron j and neuron i, consequently affecting the correlation between neuron j and i. Using this framework, we can formulate the mean synaptic weight change of a synapse from neuron j to neuron i, $\langle \dot{w}_{ij} \rangle$, as a sum of terms which depend on structural motifs with a certain 'order of interaction' (Montangie et al., 2020; Ravid Tannenbaum & Burak, 2016; Fig. 2*C*). The order of interaction refers to the total number of synapses a spike from any neuron in a network needs to travel to affect the mean connection strength $\langle w_{ij} \rangle$ between presynaptic neuron j and postsynaptic neuron i, indicated below with the number in the superscript:

$$\langle \dot{w}_{ij} \rangle = \langle \dot{w}_{ij} \rangle^{(0)} + \langle \dot{w}_{ij} \rangle^{(1)} + \langle \dot{w}_{ij} \rangle^{(2)} + \langle \dot{w}_{ij} \rangle^{(3)} + \cdots \quad (1)$$

Hence, each of the terms of different orders of interaction consists of the product of different activity statistics (rates, correlations) and motif coefficients that scale each term's contribution. The first term in eqn (1) represents the zero-order structural motif (order of interaction 0), also referred to as the rate motif. This term describes the

mean weight change as a function of only the pre- and postsynaptic firing rates (r_j, r_i):

$$\langle \dot{w}_{ij} \rangle^{(0)} = r_i r_j M_0 \tag{2}$$

where M_0 is the zero-order motif coefficient and can be calculated as the area under the plasticity rule (Fig. 2*A*). The second term in eqn (1) represents the first-order structural motif (order of interaction 1), describing how a spike in either the pre- or the postsynaptic neuron can affect $\langle w_{ij} \rangle$:

$$\langle \dot{w}_{ij} \rangle^{(1)} = r_j w_{ij} M_{1,0} + r_i w_{ji} M_{0,1}. \tag{3}$$

The first-order motif coefficients, $M_{1,0}$ and $M_{0,1}$, and in general other higher-order motif coefficients, can be calculated based on the shape of the excitatory postsynaptic current (EPSC) and the STDP parameters (Fig. 2*D*). As the order of the interaction increases, the contributions on the weight dynamics become smaller because more synapses are involved, and therefore higher orders of interaction are often truncated (Montangie et al., 2020; Ravid Tannenbaum & Burak, 2016).

Given the importance of bidirectional connections for the formation of assemblies, the above mathematical framework can be used to describe the mean weight

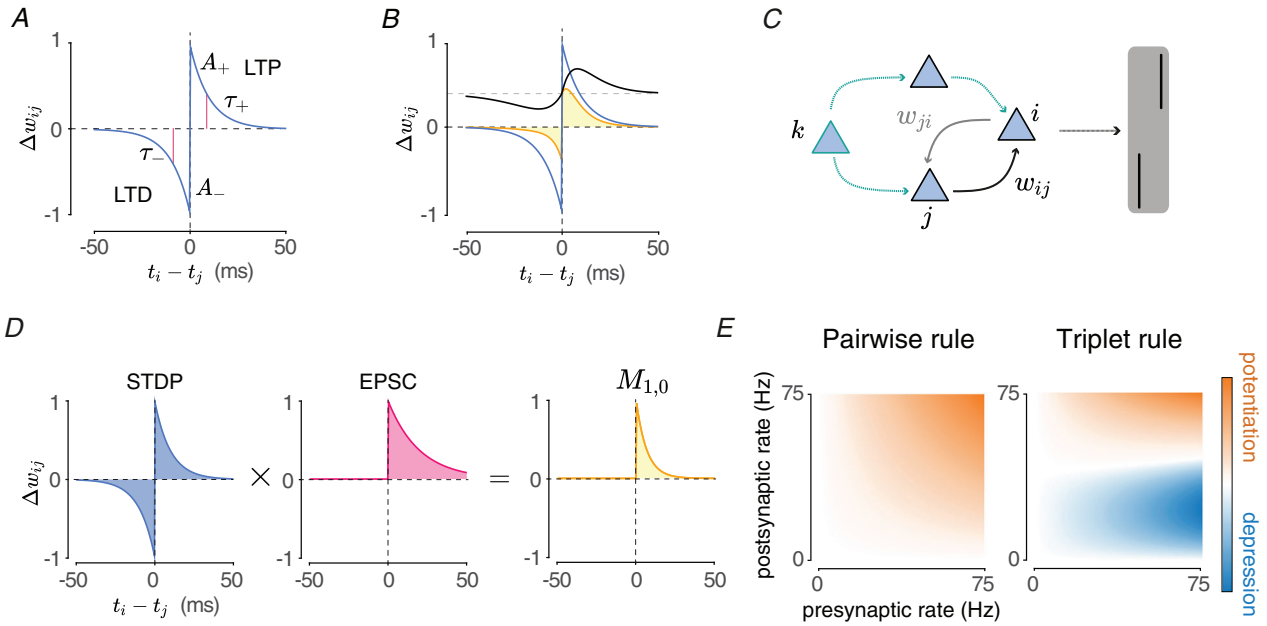


Figure 2. The effect of spike timing-dependent plasticity (STDP) on the formation of bidirectional connections
A, weight change of the synapse from neuron *j* to neuron *i*, Δw_{ij}, via the pairwise STDP rule as a function of the time difference between post- (t_i) and presynaptic spike (t_j), with long-term potentiation (LTP) parameters A_+ and τ_+ and long-term depression (LTD) parameters A_- and τ_-. Red vertical lines indicate weight change of reciprocal connections for one time difference of spikes in neuron *j* and *i*. The total area under the STDP rule is $A_+\tau_+ + A_-\tau_-$. *B*, the mean synaptic weight change (yellow area) can be computed as the area under the product of two curves: the pairwise STDP rule as in *A* (blue line) and the correlation density (black line). *C*, example of a structural motif with order of interaction 3 as one part of the decomposition of synaptic weight change ($\langle \dot{w}_{ij} \rangle^{(3)}$, see eqn 1). A spike in neuron *k* travels through one synapse to the presynaptic neuron *j* and through two synapses to the postsynaptic neuron *i*, influencing the correlation between neuron *j* and neuron *i* and therefore the weight change at synapse w_{ij} (adapted from Montangie et al., 2020). *D*, the structural motif with order of interaction 1 (also called the first-order motif) contributes to the mean weight change of synapse w_{ij} with two terms, $M_{1,0}$ and $M_{0,1}$ (see eqn 3). The first-order motif coefficient $M_{1,0}$ can be calculated by multiplication of the STDP area (blue) with the EPSC area (magenta). The EPSC here is defined as $E(t) = \exp(-t/\tau_e)$ for $t > 0$ with decay time constant τ_e. The first-order motif coefficients are $M_{1,0} = A_+ \tau_+ \tau_e/(\tau_+ + \tau_e)$ (yellow area under the curve in the right panel), and $M_{0,1} = A_- \tau_- \tau_e/(\tau_- + \tau_e)$ (not shown) (adapted from Montangie et al., 2020). *E*, weight change as a function of the pre- and postsynaptic firing rates for a pairwise STDP rule (left, see also *A*) with dominant potentiation, i.e. positive total area under the STDP rule $A_+\tau_+ + A_-\tau_- > 0$ (meaning that the zero-order motif coefficient is positive $M_0 > 0$), and the mean weight dynamics depend on pre- and postsynaptic firing rates. Weight change as a function of the pre- and postsynaptic firing rates for a triplet (right) STDP rule which has a rate-dependent zero-order contribution that depends non-linearly on the postsynaptic firing rate and depends linearly on the presynaptic rate. In both panels, the pre- and postsynaptic neurons fire independently. Potentiation (orange) and depression (blue) are normalised to their respective maximum value (adapted from Litwin-Kumar & Doiron, 2014).

dynamics for the reciprocal weights between neuron j and i as follows, considering only the zero- and first-order structural motifs (eqns 2 and 3):

$$\langle \dot{w}_{ij} \rangle^{(0,1)} = r_i r_j M_0 + r_j w_{ij} M_{1,0} + r_i w_{ji} M_{0,1}, \quad (4)$$

$$\langle \dot{w}_{ji} \rangle^{(0,1)} = r_i r_j M_0 + r_i w_{ji} M_{1,0} + r_j w_{ij} M_{0,1}. \quad (5)$$

Simplifying this for the case where the pairwise STDP plasticity rule has the same area of long-term depression (LTD) and LTP (i.e. $A_+ \tau_+ = -A_- \tau_-$, see Fig. 2*A*), the motif coefficients can be calculated as $M_0 = 0$ and $M_{0,1} = -M_{1,0}$, which we call simply M, yielding:

$$\langle \dot{w}_{ij} \rangle^{(0,1)} = \left(r_j w_{ij} - r_i w_{ji} \right) M, \quad (6)$$

$$\langle \dot{w}_{ji} \rangle^{(0,1)} = \left(r_i w_{ji} - r_j w_{ij} \right) M. \quad (7)$$

Therefore, it is easy to see that an increase of w_{ij} leads to a decrease of the reciprocal connection w_{ji} and vice versa. Furthermore, the weight changes depend on the weight strength, leading to a 'winner-take-all mechanism', where only one synaptic weight 'wins.' However, this is only true in the case of approximately equal areas of LTD and LTP and has been pointed out by multiple computational studies as the inability of the asymmetric pair-based STDP rule to generate bidirectional connections and, as a result, assemblies (Abbott & Nelson, 2000; Clopath et al., 2010; Song et al., 2000; Song & Abbott, 2001). For an STDP rule which has dominant potentiation, the zero-order motif coefficient is positive ($M_0 > 0$), and the mean weight dynamics depend on pre- and postsynaptic firing rates (Fig. 2*E*, left). Notably, this can lead to bidirectional connectivity (Babadi & Abbott, 2013).

Another way to promote bidirectional connections is using a symmetric STDP rule which has dominant depression (Manz & Memmesheimer, 2022) or without any LTD (Ravid Tannenbaum & Burak, 2016). In this setting, synaptic weights grow linearly with the pre- and postsynaptic rates, and further increase bidirectionally when neurons fire close in time, regardless of the firing order. Finally, a third possibility is to introduce synaptic delays (Babadi & Abbott, 2013; Gilson, Burkitt & van Hemmen et al., 2010; Ravid Tannenbaum & Burak, 2016). It is likely that a combination of multiples of these mechanisms operates in real biological circuits and here the computational models provide a principled mathematical investigation of each contribution.

While the pairwise STDP rule describes excitatory plasticity in the case of specific induction protocols, other plasticity rules have been proposed that can capture plasticity induced by more naturalistic induction protocols or directly resulting from well-defined molecular components. For example, to explain plasticity in response to increasing stimulation frequencies during classical pre–post pairing, Pfister and Gerstner (2006) proposed that triplets of spikes, in addition to pairs of spikes, contribute to synaptic plasticity. As a result, this triplet STDP rule is sensitive to higher-order firing statistics (Gjorgjieva et al., 2011). In the motif expansion framework, the rule adds a positive zero-order contribution that grows quadratically with the post-synaptic firing rate and linearly with the presynaptic rate (Montangie et al., 2020). This non-linear component generates a threshold between LTD and LTP that depends on presynaptic and postsynaptic firing rates (Litwin-Kumar & Doiron, 2014; Fig. 2*E*, right): when two neurons fire above threshold, both synaptic weights between them increase, promoting the bidirectionality needed for assemblies. Other computational models of assembly formation use similar non-linear dependencies of the weight change on the postsynaptic firing rate in the plasticity rules they implement, including the voltage-based STDP rule (Clopath et al., 2010; Ko et al., 2013; Miconi et al., 2016), the calcium rule (Graupner & Brunel, 2012) and the nearest-neighbour implementation of the pairwise STDP rule (Izhikevich & Desai, 2003; Izhikevich et al., 2004). The strong rate dependency in plasticity rules (Fig. 2*E*) justifies modelling neuronal dynamics based purely on the firing rate to study assembly formation (Eckmann & Gjorgjieva, 2022; Mackwood et al., 2021; Miehl & Gjorgjieva, 2022; Sadeh & Clopath, 2021).

Besides synaptic plasticity, assembly formation may also emerge from other mechanisms. We highlight structural plasticity, which refers to the activity-dependent pruning and sprouting of synapses (Gallinaro & Rotter, 2018; Gallinaro et al., 2022; Lu et al., 2019). A structural plasticity mechanism aiming to stabilise excitatory firing rates can lead to the formation of assemblies in which the number of connections, rather than connection strength, is increased among neurons within the assembly (Gallinaro et al., 2022).

In summary, theoretical frameworks have made important progress in explaining how the properties of the synaptic plasticity rule affect the formation of different connectivity structures, especially assemblies. We have focused on pair-based or triplet-based STDP rules, with the requirement that they should promote, and not hinder, the formation of bidirectional connections, a fundamental building block for the formation of assemblies. Bidirectional connections alone, however, do not guarantee an assembly-like structure. Consider, for example, the degenerate case of an all-to-all connected network, or a network with prominent bidirectional connectivity, but still no clear separation among groups of neurons. The two closely related principles of symmetry breaking and competition tackle this issue. As we discuss below, the former imposes an intrinsic or external bias that induces heterogeneity in the circuit and acts as a

learning signal. The latter provides a mechanism for consolidating within-assembly weights while reducing across-assembly weights.

Symmetry breaking. The most straightforward way to induce symmetry breaking for assembly formation is to train network weights via synaptic plasticity by structured external input. One possibility is to simultaneously stimulate neurons expected to be in one assembly with a high input rate, and neurons which are not part of the assembly with a low input rate (Fig. 3*A*). A synaptic plasticity rule dominated by firing rates will then potentiate bidirectional connections within the stimulated subpopulation (Clopath et al., 2010; Litwin-Kumar & Doiron, 2012, 2014; Schulz et al., 2021). Another possibility is to maintain constant firing rates across the neural population, but drive neurons expected to be in the same assembly by correlated inputs (Gilson et al., 2009a; Ocker & Doiron, 2019; Wu et al., 2020; Fig. 3*B*). Both of these assembly training paradigms are similar in that they impose a structure on the network via an external input. During the training protocol subpopulations of neurons (the future assemblies) are sequentially stimulated multiple times (Fig. 3*C*). In the correlation-based approach it is also possible to stimulate subpopulations of neurons in parallel (Ocker & Doiron, 2019).

Besides assuming an already selective external feedforward input performing symmetry breaking, input synapses can also be plastic and the external selectivity can be learned in parallel with the formation of assemblies (Clopath et al., 2010; Gilson et al., 2009a; Gilson, Burkitt, Grayden et al., 2010; Miconi et al., 2016; Zenke et al., 2015). This is in line with experimental studies suggesting that feedforward input becomes stimulus specific before strong recurrent connections form (Ko et al., 2011, 2013). The formation of feedforward selectivity could potentially be guided by recurrent gap junctions (as modelled by Crodelle & McLaughlin, 2021; Ko et al., 2013). These connections are observed during early development preferentially between cells that stem from the same progenitor and seem to play a crucial role in the formation of chemical synapses between these cells (Yu et al., 2009, 2012).

Experiments in the sensory deprived zebrafish larvae have shown that assemblies can also form without external input, only due to network-intrinsic activity (Pietri et al., 2017). Such assembly formation without structured input has been obtained in computational models with different forms of STDP based on the two frameworks discussed above: either mainly driven by the rate contribution (zero-order motif) (Babadi & Abbott, 2013; Burkitt et al., 2007; Gilson et al., 2009b; Ocker et al., 2015; Ocker & Doiron, 2019), or by higher-order motifs arising from internal correlation structure with the rate contribution minimised (Montangie et al., 2020; Ocker et al., 2015; Ocker & Doiron, 2019; Ravid Tannenbaum & Burak, 2016), or a combination thereof (Manz & Memmesheimer, 2022). When network-intrinsic correlations contribute significantly to assembly formation, symmetry breaking in

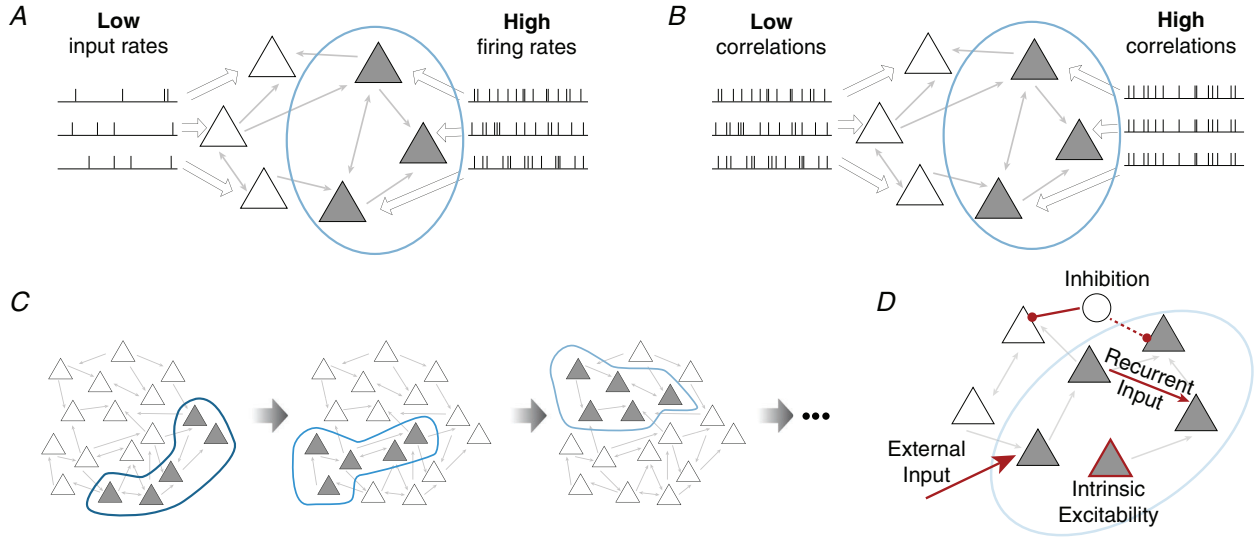


Figure 3. Symmetry breaking via external input
A, strong external input onto a subpopulation of neurons (blue ellipse) leads to high firing of the targeted neurons (grey triangles). *B*, same as *A*, but for strongly correlated external input. *C*, training protocol in which three distinct subpopulations of neurons (blue outlines) are stimulated sequentially. *D*, different mechanisms can contribute to symmetry breaking: structured external input, recurrent input that reflects existing structures, inhibition and changes in intrinsic excitability.

the network develops either due to random fluctuations in an otherwise symmetric network, or due to an initial bias in the connectivity matrix (Ocker & Doiron, 2019; Triplett et al., 2018).

Experimental studies have suggested additional mechanisms that might drive symmetry breaking: for example, neurons which become part of an engram during memory formation first have a higher excitability during the memory-encoding phase that seems to be partially due to cell-intrinsic mechanisms (Alejandre-García et al., 2022; Josselyn & Tonegawa, 2020). This highlights that the effect of symmetry breaking is probably a result of many interacting mechanisms: structured external input, recurrent input from already existing structures, intrinsic excitability, and also local inhibition or disinhibition (Fig. 3*D*).

In summary, the symmetry breaking mechanism is necessary to enable the potentiation of synapses by synaptic plasticity within the assembly. However, while successfully promoting weight potentiation within each assembly, symmetry breaking does not guarantee that weights across assemblies do not also increase, due to random rate fluctuations, or due to synaptic plasticity rules biased towards potentiation. Hence, the overall increase in synaptic strength might eventually lead to unstable dynamics, preventing the formation of desired assembly structure, as the assemblies tend to merge together. Below, we outline how competition can solve this problem.

Competition. Competition describes a mechanism to decrease across-assembly weights while increasing within-assembly weights. Competition therefore enables a clear separation of assembly from non-assembly neurons and prevents assemblies from merging. In some configurations, symmetry breaking driven by external inputs together with a plasticity rule is sufficient to induce competition between the synapses within and across assemblies. When a plasticity rule depends non-linearly on firing rates, such as the triplet STDP rule (Pfister & Gerstner, 2006), it is possible to choose the firing rates of external inputs such that within-assembly weights potentiate while across-assembly weights depress (Ocker & Doiron, 2019).

To conceptualise this, we consider a toy example with three neurons, of which only two belong to an assembly (Fig. 4*A*, red). We use a plasticity rule where potentiation and depression depend on firing rates, as for example the triplet STDP rule introduced in Fig. 2*E* (right). During the training protocol (see Fig. 3*C*), neurons within the assembly have high firing rates, while the outside neuron has low firing rates (Fig. 4*B*, left). The plasticity rule then leads to strong potentiation of within-assembly weights (Fig. 4*C*, left, purple star) and weak potentiation of weights into the assembly (Fig. 4*C*, left, green star). If the outside neuron joins another assembly at a later time, the situation reverses, with high rates outside the assembly (green neuron) and low rates within it (Fig. 4*B*, right). In this scenario, the weights into the assembly decrease more (Fig. 4*C*, right, green star) than the within-assembly weights (Fig. 4*C*, right, purple star). The weights out of the assembly always change in an opposite manner to the weights into the assembly when the pre- and postfiring rates are reversed (Fig. 4*C*, black star). Averaged throughout the whole training protocol, this leads to an increase in the within-assembly weights (Fig. 4*A*, purple arrow). In contrast, the weights onto the neuron outside the assembly decrease (Fig. 4*A*, black dashed arrow). However, the weights into the assembly only decrease if the connected outside neuron is part of a distinct assembly that is trained at a later time point (Fig. 4*A*, green dashed arrow). In this example, synaptic competition follows directly from the plasticity rule in combination with the symmetry breaking of the training protocol with different input firing rates. A pairwise symmetric STDP rule dominated by depression can also induce competition without such structured external input (Manz & Memmesheimer, 2022). Here, the relative contributions of the zero- and higher-order motifs regulate assembly size.

Many computational studies include additional mechanisms for more robust and flexible assembly formation. A widely used mechanism is synaptic weight normalisation (Fiete et al., 2010; Tetzlaff et al., 2011, 2013, 2015), which is often linked to synaptic scaling, suggesting that synaptic weights are down- (up-) regulated if the firing rates of the neurons are high (low) (Turrigiano et al., 1998; Turrigiano, 2008), and heterosynaptic plasticity, where the induction of potentiation (depression) in synapses is accompanied by depression (potentiation) at nearby synapses (Chistiakova et al., 2015; Field et al., 2020; Lynch et al., 1977).

Weight normalisation keeps the sum (or sum of squares) of all outgoing or incoming weights (or both) constant for each neuron. This mechanism, first introduced in the context of feedforward receptive-field formation (Miller & MacKay, 1994; Miller, 1996), induces synaptic competition. Any increase in a group of synaptic weights due to synaptic plasticity leads to a decrease in the remaining connections due to normalisation. To emphasise this point, we revisit the toy example mentioned above with the additional constraint that the sum of the incoming weights remains constant (Fig. 4*D*). Here, we consider subtractive normalisation that affects all incoming weights by subtracting an equal amount independent of their strength. This mechanism introduces competition by depressing the weights into the assembly despite the potentiation induced by synaptic plasticity (Fig. 4*D*, left) or potentiating in the within-assembly

weights despite the depression by synaptic plasticity (Fig. 4*D*, right). Subtractive normalisation results in a winner-take-all competition (Miller & MacKay, 1994), whereby at the fixed point, i.e. the point at which synaptic weights no longer change, the weight with the highest potentiation (or equivalently the lowest depression) rate will 'win' (Fig. 4*D*, yellow diamond). Including weight normalisation enables synaptic plasticity rules that do not introduce competition themselves – like symmetric pairwise STDP – to generate assembly structures (Ravid Tannenbaum & Burak, 2016) because the normalisation mechanism amplifies small asymmetries in the weight dynamics. At the same time, weight normalisation also enables more reliable assembly formation even in cases where it is not explicitly needed to generate assemblies (Litwin-Kumar & Doiron, 2014; Schulz et al., 2021; Wu et al., 2020; Zenke et al., 2015).

In contrast to subtractive normalisation, divisive normalisation induces less competition and does not lead to a winner-take-all mechanism, but to a stable fixed point where weights are proportional to their respective potentiation strengths (Miller & MacKay, 1994; Fig. 4*E*, yellow diamond; compare with Fig. 4*D*, left). A divisive effect is in line with the biological idea that synapses compete for molecular resources (Triesch et al., 2018), and it has been used, for example, in a model of assembly formation that can disambiguate input features (Eckmann & Gjorgjieva, 2022).

Metaplasticity, a dynamic change of the plasticity mechanism itself (Abraham, 2008), can also induce competition. A stimulus which results in a high postsynaptic firing rate leads to an increase in the LTD/LTP threshold, hence making it harder for a subsequent stimulus to induce LTP. A classic rate-based

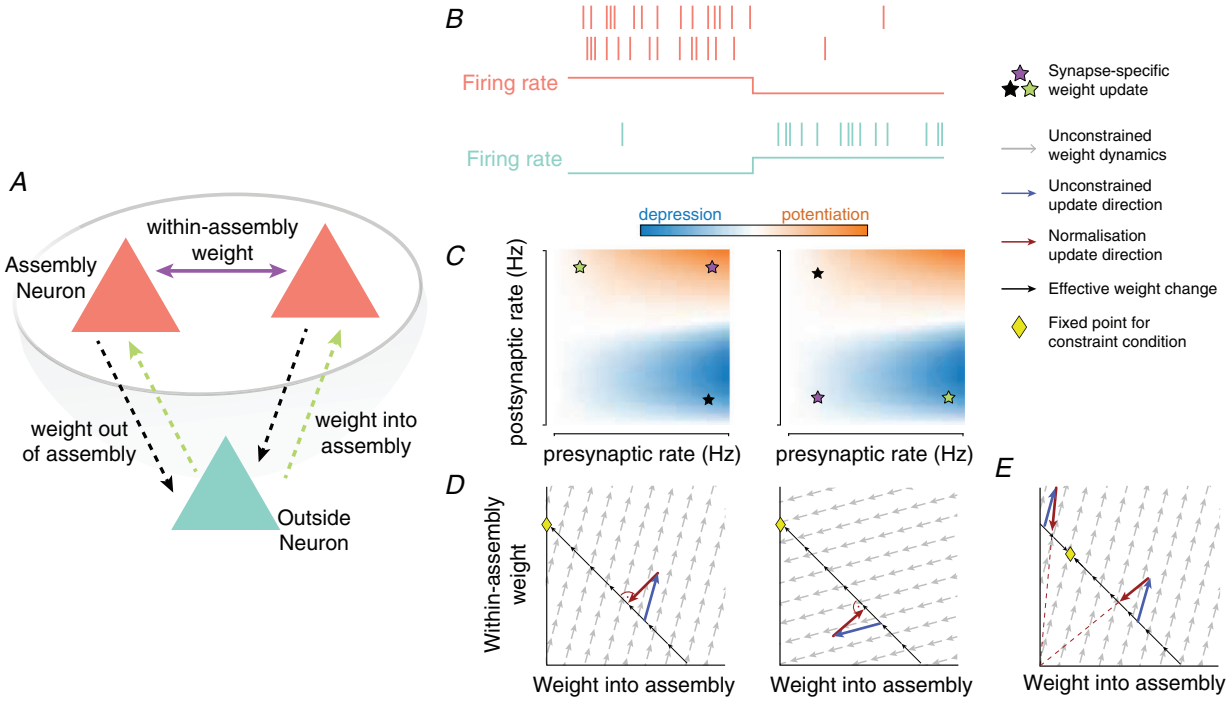


Figure 4. Competition between within- and across-assembly weights
A, a toy model schematic to explain the concept of competition. Two neurons (red) are part of an assembly, one neuron (mint) is not. The neurons are connected by synapses within the assembly (purple), into the assembly (green) and out of the assembly (black). The dotted line indicates the weights, which should decrease due to competition to form the assembly. *B*, firing rates and spikes for the three neurons in *A*, for two scenarios when training assemblies (left and right). Left: the within-assembly neurons have a high firing rate while the outside neuron fires with a low firing rate. Right: the firing rates of the within-assembly *vs*. outside-assembly neurons are reversed relative to the left (adapted from Ocker & Doiron, 2019). *C*, weight change as a function of the pre- and postsynaptic firing rates for the triplet STDP rule (as in Fig. 2*E*). The stars correspond to the pre- and postsynaptic rates in the scenarios sketched in *B*, with filling colour matching the scheme in *A* (adapted from Ocker & Doiron, 2019). *D*, phase plane of weight dynamics, showing weight into the assembly (*x*-axis, green arrow in *A*) *versus* within-assembly-weight (*y*-axis, purple arrow in *A*). The dynamics follow from the weight changes depicted in *C*, induced by the scenario in *B*. Grey arrows indicate the unconstrained weight dynamics, the blue arrow shows the weight change following the unconstrained synaptic plasticity dynamics, the red arrow shows the counteracting effect of the subtractive normalisation, the black arrow indicates the net weight change and the yellow diamond indicates the fixed point of the constrained weight evolution (adapted from Clopath et al., 2016). *E*, same scenario as in *D* (left) but with divisive normalisation.

plasticity rule implementing metaplasticity is the Bienenstock–Cooper–Munro (BCM) rule, which has been studied extensively in the context of competition in feedforward networks (Bienenstock et al., 1982; Yger & Gilson, 2015), and can be linked to the triplet STDP rule if one of the rule's parameters changes with ongoing activity (Gjorgjieva et al., 2011). Metaplasticity in the context of assembly formation is an important part of the voltage-based STDP rule (Clopath et al., 2010; Miconi et al., 2016), and adding metaplasticity to STDP ensures assembly formation over a broader range of parameters (Zenke et al., 2015).

Recent work has proposed that inhibition and the plasticity of inhibitory-to-excitatory synapses also play an important part in the competition between assemblies (Herpich & Tetzlaff, 2019; Lagzi et al., 2021; Miehl & Gjorgjieva, 2022; Sadeh & Clopath, 2021). Specifically, inhibitory plasticity can be linked to BCM-like metaplasticity and therefore can control the induction of LTD or LTP at excitatory synapses (Clopath et al., 2016; Miehl & Gjorgjieva, 2022).

In summary, multiple mechanisms to induce competition between within- and across-assembly weights have been suggested, including weight normalisation, metaplasticity and inhibitory plasticity. While all these mechanisms aid reliable formation of assemblies, the last outstanding question pertains to maintaining stable assemblies in the face of ongoing synaptic plasticity.

Stability of representations. Maintaining stable assemblies, or representations in general, faces two main challenges: first, how to control synaptic weight changes to prevent pathological firing rates in the network; and second, how to preserve the difference between within- and across-assembly weight strength so that the learned structure does not disappear over time (Fig. 5*A*).

Synaptic plasticity mechanisms lead to unbounded growth of the synaptic weights due to positive feedback between activity and plasticity, termed 'Hebbian runaway dynamics' (Turrigiano & Nelson, 2004). Solutions to this problem are weight normalisation (Fig. 4), metaplasticity mechanisms or applying upper bounds on the weights. Two types of upper bounds are often considered, 'hard' and 'soft'. Soft upper bounds implement 'weight-dependent' plasticity assuming that the weights change proportionally to the inverse of their strength (Gütig et al., 2003; Rossum et al., 2000; Rubin et al., 2001). Although soft bounds ensure stability, they can lead to unimodal weight distributions and, therefore, might counteract the formation of assemblies (Morrison et al., 2007). Similarly, metaplasticity and normalisation mechanisms might also stabilise weight growth at the expense of reducing competition (Yger & Gilson, 2015). Theoretical work has proposed that the key aspect to ensuring stable weight dynamics is the relative timescales of synaptic plasticity *versus* stabilising mechanisms (Zenke et al., 2013).

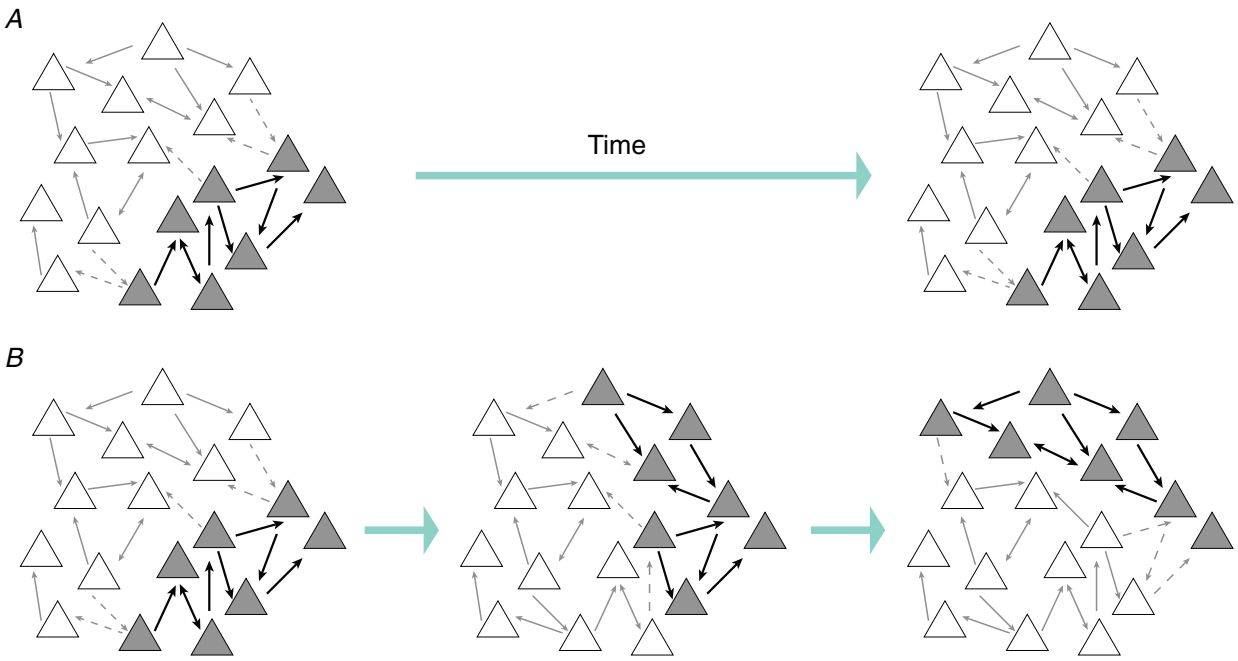


Figure 5. Stability of representations
A, the neurons of a given assembly remain the same over time, providing a stable representation of a sensory percept. *B*, the neurons of a given assembly change over time, leading to a representational drift.

When a network remains plastic after learning assemblies, learning additional assemblies or ongoing spontaneous activity alone can degrade the learned structure. Therefore, the network 'forgets' these representations (Fusi, 2017). It has been suggested that reactivation of the learned structure through the high activity of assembly neurons can reinforce the assembly structure (Fauth & van Rossum, 2019; Litwin-Kumar & Doiron, 2014). Essentially, the network transitions from learning due to external input to learning due to internal structure (see section Symmetry Breaking) and thus self-stabilises. Additional mechanisms, such as short-term plasticity, have also been suggested to keep the firing rates in the network in a healthy regime (Fauth & van Rossum, 2019; Hiratani & Fukai, 2014; Mongillo et al., 2005; Vasilaki & Giugliano, 2014; Zenke et al., 2015). An alternative to self-stabilisation via high firing rate reactivation is self-stabilisation via spiking statistics, for example when storing assemblies in a 'silent' fashion. In this framework, plasticity at inhibitory-to-excitatory synapses can maintain a stable balance of excitation and inhibition after learning, such that assembly neurons have similar firing rates to other neurons in the network (Barron et al., 2017; Ramaswami, 2014). Despite this, assembly neurons show higher correlations and spiking irregularities during spontaneous activity, which can stabilise and improve the long-term storage of the imprinted assemblies (Gallinaro & Clopath, 2021; Ocker & Doiron, 2019). The assemblies can then be accessed by disinhibitory mechanisms (Barron et al., 2016), or even be read out in seemingly quiet stages, by downstream neurons through synapses with short-term plasticity (Gallinaro & Clopath, 2021).

However, should the goal be to maintain stable assemblies at all? Experimental work suggests that assemblies may change over time, known as 'representational drift' (Rule et al., 2019). In computational models, ongoing synaptic plasticity after assembly formation results in single neurons dropping in and out of assemblies, leading to a drift of the assembly structure (Kossio et al., 2021; Manz & Memmesheimer, 2022; Raman & O'Leary, 2021; Triplett et al., 2018; Fig. 5*B*). Hence, an important question is how such unstable neuronal representations can lead to stable task performance (Mau et al., 2020; Rule et al., 2019). Current solutions to this problem propose that assemblies should be considered in conjunction with their readouts, for example by assuming that the connections from assemblies to downstream neurons can be also plastic. This leads to a constant readout despite changing assemblies (Kossio et al., 2021; Rule & O'Leary, 2022).

Another possible solution to the stability problem is to break free from the assumption that synaptic plasticity is 'on' at all times, continuously changing learned structures. To turn off, or gate, synaptic plasticity several mechanisms have been suggested, including inhibition of the inhibitory population, i.e. disinhibition as a gating signal (Froemke et al., 2007; Letzkus et al., 2011) or 'three-factor plasticity rules', where neuromodulators that convey a learning signal can regulate weight change in addition to the pre- and postsynaptic activity (Frémaux & Gerstner, 2016). Another mechanism is synaptic tagging and capture, where activity-dependent synaptic plasticity needs an additional stabilising internal signal to allow persistent connectivity changes (Luboeinski & Tetzlaff, 2021; Redondo & Morris, 2011).

In summary, multiple solutions have been proposed to ensure the stability of the learned assemblies: from self-stabilisation by activity (Litwin-Kumar & Doiron, 2014) or spiking statistics (Gallinaro & Clopath, 2021; Ocker & Doiron, 2019), to plastic downstream readouts despite drifting representation (Rule et al., 2019) and gating of synaptic plasticity.

Conclusions and outlook

Inspired by the wealth of experimental results on the existence and computational relevance of neural ensembles and assemblies, various computational models have investigated how synaptic plasticity mediates their formation and stability. In this review, we have identified four fundamental computational principles behind this process (Fig. 1). First, synaptic plasticity can lead to strong, bidirectional connectivity among neurons, but it needs to be accompanied by a second component, a symmetry breaking signal, to break up neurons into multiple assemblies. To prevent distinct assemblies from merging requires a third component, competition. Finally, the last component ensures the stability of assembly structures or of their representations.

While recent experimental work has made great advances in understanding synchronous activity (ensembles), a clear link between ensembles, assemblies and the formation of assemblies via synaptic plasticity mechanisms is still lacking. Computational models have proven indispensable in providing the missing link because they allow multiple mechanisms to be studied one at a time or in combination. An outstanding question that we did not address pertains to the functional and computational consequences of learned assemblies, which remains an important future direction. Nonetheless, we include a short overview on the functional relevance of assemblies.

The idea of assemblies, or ensembles, as the basic units of cognition has recently replaced the neuron-centric view (Buzsáki, 2010; Eichenbaum, 2018; Huyck & Passmore, 2013; Yuste, 2015). An emergent core assumption from this framework is that each ensemble represents a specific concept or feature, acting as the fundamental unit for

memory storage (Neves et al., 2008). One advantage of having strong recurrent connectivity (assembly) compared to only considering correlated rates (ensemble) is stimulus amplification (Peron et al., 2020), thus enabling weaker stimuli to elicit a recognisable response and to increase robustness whereby the malfunction or death of single neurons or synapses will not affect the represented concept. Consequently, an incomplete stimulus can be sufficient to evoke the complete assembly – a phenomenon named pattern completion, which is especially relevant for memory retrieval (Guzman et al., 2016). Moreover, recurrent interactions alone may be sufficient to keep an assembly active after stimulation, thus enabling the brain to decouple intrinsic activity from external stimulation and modify learned concepts independently from specific external inputs (Harris, 2005).

One can think of assemblies as the basis of any computation (Byrne & Huyck, 2010; Herpich & Tetzlaff, 2019; Ranhel, 2012), also referred to as 'assembly calculus' (Papadimitriou et al., 2020; Papadimitriou & Friederici, 2022). In this view, plastic changes of within- and across-assembly weights are abstracted in mathematically tractable basic operations. These basic operations describe how new assemblies are formed and how existing assemblies can be combined with each other. This framework allows us, for example, to build the full architecture of language syntax (Papadimitriou et al., 2020; Papadimitriou & Friederici, 2022). In general, assemblies can be combined in two distinct ways. First, chains of assemblies can form directed sequences, which has been suggested specifically in the hippocampus, and lead to reliable and even reversed reactivation after learning (Holtmaat & Caroni, 2016). For example, it has been shown that learning assemblies in a sequence can generate clock-like neuronal dynamics, enabling the learning of different spatiotemporal patterns (Maes et al., 2020, 2021). Second, assemblies can be associated with each other. Association models can be based on chaining models, where the association is encoded in the connection strength between the assemblies. Another option to encode associations is hierarchical models, where the associated concept is encoded in a newly formed assembly that only activates if both pre-existing assemblies are active (Pokorny et al., 2020).

Although most computational studies on assembly formation begin with an unstructured network from which assemblies and associations between assemblies are learned, this probably is not the case in the adult brain. There are two possibilities. First, assemblies might be formed in early development, and be later used as a 'backbone' to learn new sequences or associate new features (Holtmaat & Caroni, 2016). Cortical areas generate stereotypical structured 'spontaneous activity' during early development that is important for connectivity refinement (Richter & Gjorgjieva, 2017) and can drive assembly formation without structured external input (Loidolt et al., 2020; Montangie et al., 2020; Ravid Tannenbaum & Burak, 2016). Second, new assemblies might also form in the adult brain, where they should be integrated into an existing network structure without 'forgetting' previously learned representations and potentially allowing overlap between assemblies. This possibility is closely related to the existence of engrams, large-scale representations potentially spanning multiple brain regions. Experimental studies have identified various mechanisms of engram formation, such as plastic changes in synapses between regions, or modulated intrinsic excitability of recruited neurons (Holtmaat & Caroni, 2016; Josselyn & Tonegawa, 2020). Going one step further, Buzsáki (2010) suggested that assemblies should be defined from the perspective of a readout mechanism, as done, for example, in the context of drifting assemblies (Kossio et al., 2021; Rule & O'Leary, 2022). Future work needs to carefully study assembly formation and association in the context of existing network structures.

Once established, assembly structures shape ongoing network activity and generate distinct activity patterns often called discrete attractor dynamics (Aljadeff et al., 2021; Hopfield, 1982). These dynamics can also show sustained elevated activity independent of external cues which links to working memory tasks (Amit et al., 1994; Durstewitz et al., 2000). When multiple assemblies are embedded in a network, their activation can switch stochastically between different assemblies (Mazzucato et al., 2015). The resulting activity patterns are termed 'metastable dynamics', and closely relate to experimental findings (Abeles et al., 1995; La Camera et al., 2019). In the rat gustatory cortex, they have been linked to expectation: by reducing the stability of attractors via an external cue, the switching rate increases and stimuli can be detected faster (Mazzucato et al., 2019). Discrete attractor dynamics are also investigated in more abstract models called associative memory models and support the view of ensembles as computing units. As advanced significantly by Hopfield (1982), patterns of neuronal activation are imprinted in the network connectivity and can be activated even with incomplete stimuli, due to the attractor dynamics. Interestingly, the abstract assumptions correspond, especially as formulated for example by Tsodyks (1989), to the four components we identified when reviewing the literature on assembly formation: a Hebbian learning rule establishes bidirectional connections within an assembly based on well-defined memory patterns, which when imprinted, determine which neurons should be active, hence acting as a symmetry breaking mechanism. The abstractions in these models have enabled extensive theoretical results related to these attractor networks, including the theoretical limit for the number of patterns or concepts stored in a recurrent network (Amit et al., 1985; Gardner, 1988).

Various aspects in the models have also been made more biologically realistic, for example sparse coding (Amari, 1989; Gripon et al., 2016), asymmetric connections (Tsodyks, 1989) or context dependency (Podlaski et al., 2020). It is interesting to compare the abstracted learning rules of associative memory models with their more biologically plausible local and dynamic learning counterparts, as they are used in studies on assembly formation. These associative memory models have an advantage in the clear benchmarks they provide, such as the number of stored patterns, thus making it easier to compare different modelling approaches. Such a comparison is still an open research direction for assembly formation with more biologically inspired mechanisms. A formal integration of these two concepts will yield further insights, as initial efforts are already gaining some traction (Aljadeff et al., 2021).

In this review, we have focused on assemblies as a prominent type of connectivity structure described in many neural circuits in the brain. Beyond assemblies, at a smaller scale, specific connectivity motifs are either over- or under-represented than expected by chance (Song et al., 2005). At a larger scale, synfire chains, hub networks and other structures have been described in network neuroscience (Bassett & Sporns, 2017). Future work needs to show how these different, small- and large-scale, connectivity structures are related and how they can be learned via synaptic plasticity, leading to diverse activity dynamics and complex computations.

References

Abbott, L. F., & Nelson, S. B. (2000). Synaptic plasticity: Taming the beast. *Nature Neuroscience*, **3**(S11), 1178–1183.

Abdou, K., Shehata, M., Choko, K., Nishizono, H., Matsuo, M., Muramatsu, S. I., & Inokuchi, K. (2018). Synapse-specific representation of the identity of overlapping memory engrams. *Science*, **360**(6394), 1227–1231.

Abeles, M., Bergman, H., Gat, I., Meilijson, I., Seidemann, E., Tishby, N., & Vaadia, E. (1995). Cortical activity flips among quasi-stationary states. *Proceedings of the National Academy of Sciences of the United States of America*, **92**(19), 8616–8620.

Abraham, W. C. (2008). Metaplasticity: Tuning synapses and networks for plasticity. *Nature Reviews Neuroscience*, **9**(5), 387–387.

Alejandre-García, T., Kim, S., Pérez-Ortega, J., & Yuste, R. (2022). Intrinsic excitability mechanisms of neuronal ensemble formation. *eLife*, **11**, e77470.

Aljadeff, J., Gillett, M., Obilinovic, U. P., & Brunel, N. (2021). From synapse to network: Models of information storage and retrieval in neural circuits. *Current Opinion in Neurobiology*, **70**, 24–33.

Amari, S. (1989). Characteristics of sparsely encoded associative memory. *Neural Networks*, **2**(6), 451–457.

Amit, D. J., Brunel, N., & Tsodyks, M. V. (1994). Correlations of cortical Hebbian reverberations: Theory versus experiment. *Journal of Neuroscience*, **14**(11), 6435–6445.

Amit, D. J., Gutfreund, H., & Sompolinsky, H. (1985). Spin-glass models of neural networks. *Physical Review A*, **32**(2), 1007–1018.

Avitan, L., Pujic, Z., Mölter, J., Zhu, S., Sun, B., & Goodhill, G. J. (2021). Spontaneous and evoked activity patterns diverge over development. *eLife*, **10**, e61942.

Babadi, B., & Abbott, L. F. (2013). Pairwise analysis can account for network structures arising from spike-timing dependent plasticity. *Plos Computational Biology*, **9**(2), e1002906.

Barron, H. C., Vogels, T. P., Behrens, T. E., & Ramaswami, M. (2017). Inhibitory engrams in perception and memory. *Proceedings of the National Academy of Sciences of the United States of America*, **114**(26), 6666–6674.

Barron, H. C., Vogels, T. P., Emir, U. E., Makin, T. R., O'Shea, J., Clare, S., Jbabdi, S., Dolan, R. J., & Behrens, T. E. J. (2016). Unmasking latent inhibitory connections in human cortex to reveal dormant cortical memories. *Neuron*, **90**(1), 191–203.

Bassett, D. S., & Sporns, O. (2017). Network neuroscience. *Nature Neuroscience*, **20**(3), 353–364.

Berkes, P., Orbán, G., Lengyel, M., & Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, **331**(6013), 83–87.

Bi, G., & Poo, M. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, **18**(24), 10464–10472.

Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, **2**(1), 32–48.

Brea, J., & Gerstner, W. (2016). Does computational neuroscience need new synaptic learning paradigms? *Current Opinion in Behavioral Sciences*, **11**, 61–66.

Brown, S. P., & Hestrin, S. (2009). Intracortical circuits of pyramidal neurons reflect their long-range axonal targets. *Nature*, **457**(7233), 1133–1136.

Burkitt, A. N., Gilson, M., & van Hemmen, J. L. (2007). Spike-timing-dependent plasticity for neurons with recurrent connections. *Biological Cybernetics*, **96**(5), 533–546.

Buzsáki, G. (2010). Neural syntax: Cell assemblies, synapsembles, and readers. *Neuron*, **68**(3), 362–385.

Byrne, E., & Huyck, C. (2010). Processing with cell assemblies. *Neurocomputing*, **74**(1–3), 76–83.

Cadwell, C. R., Scala, F., Fahey, P. G., Kobak, D., Mulherkar, S., Sinz, F. H., Papadopoulos, S., Tan, Z. H., Johnsson, P., Hartmanis, L., Li, S., Cotton, R. J., Tolias, K. F., Sandberg, R., Berens, P., Jiang, X., & Tolias, A. S. (2020). Cell type composition and circuit organization of clonally related excitatory neurons in the juvenile mouse neocortex ed. West AE, Behrens TE, Hevner R & Fishell G. *eLife*, **9**, e52951.

Cai, D. J., Aharoni, D., Shuman, T., Shobe, J., Biane, J., Song, W., Wei, B., Veshkini, M., La-Vu, M., Lou, J., Flores, S. E., Kim, I., Sano, Y., Zhou, M., Baumgaertel, K., Lavi, A., Kamata, M., Tuszynski, M., Mayford, M., Golshani, P., & Silva, A. J. (2016). A shared neural ensemble links distinct contextual memories encoded close in time. *Nature*, **534**(7605), 115–118.

Campagnola, L., Seeman, S. C., Chartrand, T., Kim, L., Hoggarth, A., Gamlin, C., Ito, S., Trinh, J., Davoudian, P., Radaelli, C., Kim, M. H., Hage, T., Braun, T., Alfiler, L., Andrade, J., Bohn, P., Dalley, R., Henry, A., Kebede, S., …, Jarsky, T. (2022). Local connectivity and synaptic dynamics in mouse and human neocortex. *Science*, **375**(6585), eabj5861.

Carrillo-Reid, L., Han, S., Yang, W., Akrouh, A., & Yuste, R. (2019). Controlling visually guided behavior by holographic recalling of cortical ensembles. *Cell*, **178**(2), 447–457.e5.

Carrillo-Reid, L., Yang, W., Bando, Y., Peterka, D. S., & Yuste, R. (2016). Imprinting and recalling cortical ensembles. *Science*, **353**(6300), 691–694.

Carrillo-Reid, L., Yang, W., Miller, J.-E. K., Peterka, D. S., & Yuste, R. (2017). Imaging and optically manipulating neuronal ensembles. *Annual Review of Biophysics*, **46**(1), 271–293.

Chistiakova, M., Bannon, N., Chen, J.-Y., Bazhenov, M., & Volgushev, M. (2015). Homeostatic role of heterosynaptic plasticity: Models and experiments. *Frontiers in Computational Neuroscience*, **9**, 89.

Clopath, C., Büsing, L., Vasilaki, E., & Gerstner, W. (2010). Connectivity reflects coding: A model of voltage-based STDP with homeostasis. *Nature Neuroscience*, **13**(3), 344–352.

Clopath, C., Vogels, T. P., Froemke, R. C., & Sprekeler, H. (2016). Receptive field formation by interacting excitatory and inhibitory synaptic plasticity. *bioRxiv*. https://doi.org/10.1101/066589.

Cossart, R., Aronov, D., & Yuste, R. (2003). Attractor dynamics of network UP states in the neocortex. *Nature*, **423**(6937), 283–288.

Cossell, L., Iacaruso, M. F., Muir, D. R., Houlton, R., Sader, E. N., Ko, H., Hofer, S. B., & Mrsic-Flogel, T. D. (2015). Functional organization of excitatory synaptic strength in primary visual cortex. *Nature*, **518**(7539), 399–403.

Crodelle, J., & McLaughlin, D. W. (2021). Modeling the role of gap junctions between excitatory neurons in the developing visual cortex. *Plos Computational Biology*, **17**(7), e1007915.

Dalgleish, H. W. P., Russell, L. E., Packer, A. M., Roth, A., Gauld, O. M., Greenstreet, F., Thompson, E. J., & Häusser, M. (2020). How many neurons are sufficient for perception of cortical activity? *eLife*, **9**, e58889.

De Nó, R. L. (1938). Analysis of the activity of the chains of internuncial neurons. *Journal of Neurophysiology*, **1**(3), 207–244.

Debanne, D., Inglebert, Y., & Russier, M. (2019). Plasticity of intrinsic neuronal excitability. *Current Opinion in Neurobiology*, **54**, 73–82.

Deitch, D., Rubin, A., & Ziv, Y. (2021). Representational drift in the mouse visual cortex. *Current Biology*, **31**(19), 4327–4339.e6.e6.

Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Neurocomputational models of working memory. *Nature Neuroscience*, **3**(S11), 1184–1191.

Eckmann, S., & Gjorgjieva, J. (2022). Synapse-type-specific competitive Hebbian learning forms functional recurrent networks. *bioRxiv*. https://doi.org/10.1101/2022.03.11.483899.

Eichenbaum, H. (2018). Barlow versus Hebb: When is it time to abandon the notion of feature detectors and adopt the cell assembly as the unit of cognition? *Neuroscience Letters*, **680**, 88–93.

Fauth, M. J., & van Rossum, M. C. W. (2019). Self-organized reactivation maintains and reinforces memories despite synaptic turnover. *eLife*, **8**, e43717.

Feldman, D. E. (2009). Synaptic mechanisms for plasticity in neocortex. *Annual Review of Neuroscience*, **32**(1), 33–55.

Field, R. E., D'amour, J. A., Tremblay, R., Miehl, C., Rudy, B., Gjorgjieva, J., & Froemke, R. C. (2020). Heterosynaptic plasticity determines the set point for cortical excitatory-inhibitory balance. *Neuron*, **106**(5), 842–854.e4.

Fiete, I. R., Senn, W., Wang, C. Z. H., & Hahnloser, R. H. R. (2010). Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. *Neuron*, **65**(4), 563–576.

Frémaux, N., & Gerstner, W. (2016). Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules. *Frontiers in Neural Circuits*, **9**, 85.

Froemke, R. C., Merzenich, M. M., & Schreiner, C. E. (2007). A synaptic memory trace for cortical receptive field plasticity. *Nature*, **450**(7168), 425–429.

Fusi, S. (2017). Computational models of long term plasticity and memory. *arXiv*; https://doi.org/10.48550/arXiv.1706.04946.

Gallinaro, J. V., & Clopath, C. (2021). Memories in a network with excitatory and inhibitory plasticity are encoded in the spiking irregularity. *Plos Computational Biology*, **17**(11), e1009593.

Gallinaro, J. V., Gašparović, N., & Rotter, S. (2022). Homeostatic control of synaptic rewiring in recurrent networks induces the formation of stable memory engrams. *Plos Computational Biology*, **18**(2), e1009836.

Gallinaro, J. V., & Rotter, S. (2018). Associative properties of structural plasticity based on firing rate homeostasis in recurrent neuronal networks. *Science Reports*, **8**(1), 3754.

Gardner, E. (1988). The space of interactions in neural network models. *Journal of Physics A: Mathematical and General*, **21**(1), 257–270.

Gill, J. V., Lerman, G. M., Zhao, H., Stetler, B. J., Rinberg, D., & Shoham, S. (2020). Precise holographic manipulation of olfactory circuits reveals coding features determining perceptual detection. *Neuron*, **108**(2), 382–393.e5.

Gilson, M., Burkitt, A., & van Hemmen, J. L. (2010). STDP in recurrent neuronal networks. *Frontiers in Computational Neuroscience*, **4**, 23.

Gilson, M., Burkitt, A. N., Grayden, D. B., Thomas, D. A., & van Hemmen, J. L. (2009a). Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. II. Input selectivity—symmetry breaking. *Biological Cybernetics*, **101**(2), 103–114.

Gilson, M., Burkitt, A. N., Grayden, D. B., Thomas, D. A., & van Hemmen, J. L. (2009b). Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks III: Partially connected neurons driven by spontaneous activity. *Biological Cybernetics*, **101**(5–6), 411.

Gilson, M., Burkitt, A. N., Grayden, D. B., Thomas, D. A., & van Hemmen, J. L. (2010). Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks V: Self-organization schemes and weight dependence. *Biological Cybernetics*, **103**(5), 365–386.

Gjorgjieva, J., Clopath, C., Audet, J., & Pfister, J.-P. (2011). A triplet spike-timing–dependent plasticity model generalizes the Bienenstock–Cooper–Munro rule to higher-order spatiotemporal correlations. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(48), 19383–19388.

Graupner, M., & Brunel, N. (2012). Calcium-based plasticity model explains sensitivity of synaptic changes to spike pattern, rate, and dendritic location. *Pnas*, **109**(10), 3991–3996.

Gripon, V., Heusel, J., Löwe, M., & Vermet, F. (2016). A comparative study of sparse associative memories. *Journal of Statistical Physics*, **164**(1), 105–129.

Gütig, R., Aharonov, R., Rotter, S., & Sompolinsky, H. (2003). Learning input correlations through nonlinear temporally asymmetric hebbian plasticity. *Journal of Neuroscience*, **23**(9), 3697–3714.

Guzman, S. J., Schlögl, A., Frotscher, M., & Jonas, P. (2016). Synaptic mechanisms of pattern completion in the hippocampal CA3 network. *Science*, **353**(6304), 1117–1123.

Harris, K. D. (2005). Neural signatures of cell assembly organization. *Nature Reviews Neuroscience*, **6**(5), 399–407.

Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G., & Buzsáki, G. (2003). Organization of cell assemblies in the hippocampus. *Nature*, **424**(6948), 552–556.

Hebb, D. O. (1949). *The Organization of Behavior; a neuro-psychological theory*. Wiley, New York.

Herpich, J., & Tetzlaff, C. (2019). Principles underlying the input-dependent formation and organization of memories. *Network Neuroscience*, **3**(2), 606–634.

Hiratani, N., & Fukai, T. (2014). Interplay between short- and long-term plasticity in cell-assembly formation. *Plos One*, **9**(7), e101535.

Holtmaat, A., & Caroni, P. (2016). Functional and structural underpinnings of neuronal assembly formation in learning. *Nature Neuroscience*, **19**(12), 1553–1562.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, **79**(8), 2554–2558.

Hu, Y., Trousdale, J., Josić, K., & Shea-Brown, E. (2013). Motif statistics and spike correlations in neuronal networks. *Journal of Statistical Mechanics: Theory and Experiment*, **2013**(03), P03012.

Hu, Y., Trousdale, J., Josić, K., & Shea-Brown, E. (2014). Local paths to global coherence: Cutting networks down to size. *Physical Review E*, **89**(3), 032802.

Huyck, C. R., & Passmore, P. J. (2013). A review of cell assemblies. *Biological Cybernetics*, **107**(3), 263–288.

Izhikevich, E. M., & Desai, N. S. (2003). Relating STDP to BCM. *Neural Computation*, **15**(7), 1511–1523.

Izhikevich, E. M., Gally, J. A., & Edelman, G. M. (2004). Spike-timing dynamics of neuronal groups. *Cerebral Cortex*, **14**(8), 933–944.

Jennings, J. H., Kim, C. K., Marshel, J. H., Raffiee, M., Ye, L., Quirin, S., Pak, S., Ramakrishnan, C., & Deisseroth, K. (2019). Interacting neural ensembles in orbitofrontal cortex for social and feeding behaviour. *Nature*, **565**(7741), 645–649.

Jensen, K. T., Harpaz, N. K., Dhawale, A. K., Wolff, S. B. E., & Ölveczky, B. P. (2022). Long-term stability of neural activity in the motor system. *bioRxiv*. https://doi.org/10.1101/2021.10.27.465945.

Josselyn, S. A., & Tonegawa, S. (2020). Memory engrams: Recalling the past and imagining the future. *Science*, **367**(6473), eaaw4325.

Jouhanneau, J. S., Kremkow, J., Dorrn, A. L., & Poulet, J. F. A. (2015). In vivo monosynaptic excitatory transmission between layer 2 cortical pyramidal neurons. *Cell reports*, **13**(10), 2098–2106.

Jovanović, S., Hertz, J., & Rotter, S. (2015). Cumulants of Hawkes point processes. *Physical Review E*, **91**(4), 042802.

Jovanović, S., & Rotter, S. (2016). Interplay between graph topology and correlations of third order in spiking neuronal networks. *Plos Computational Biology*, **12**(6), e1004963.

Kempter, R., Gerstner, W., & van Hemmen, J. L. (1999). Hebbian learning and spiking neurons. *Physical Review E*, **59**(4), 4498–4514.

Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A., & Arieli, A. (2003). Spontaneously emerging cortical representations of visual attributes. *Nature*, **425**(6961), 954–956.

Kim, T., Oh, W. C., Choi, J. H., & Kwon, H.-B. (2016). Emergence of functional subnetworks in layer 2/3 cortex induced by sequential spikes in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, **113**(10), E1372–E1381.

Kirkwood, A., Rioult, M. G., & Bear, M. F. (1996). Experience-dependent modification of synaptic plasticity in visual cortex. *Nature*, **381**(6582), 526–528.

Ko, H., Cossell, L., Baragli, C., Antolik, J., Clopath, C., Hofer, S. B., & Mrsic-Flogel, T. D. (2013). The emergence of functional microcircuits in visual cortex. *Nature*, **496**(7443), 96–100.

Ko, H., Hofer, S. B., Pichler, B., Buchanan, K. A., Sjöström, P. J., & Mrsic-Flogel, T. D. (2011). Functional specificity of local synaptic connections in neocortical networks. *Nature*, **473**(7345), 87–91.

Kossio, Y. F. K., Goedeke, S., Klos, C., & Memmesheimer, R.-M. (2021). Drifting assemblies for persistent memory: Neuron transitions and unsupervised compensation. *Proceedings of the National Academy of Sciences of the United States of America*, **118**(46), e2023832118.

Kozloski, J., & Cecchi, G. A. (2010). A theory of loop formation and elimination by spike timing-dependent plasticity. *Frontiers in Neural Circuits*, **4**, 7.

La Camera, G., Fontanini, A., & Mazzucato, L. (2019). Cortical computations via metastable activity. *Current Opinion in Neurobiology*, **58**, 37–45.

Lagzi, F., Bustos, M. C., Oswald, A.-M., & Doiron, B. (2021). Assembly formation is stabilized by Parvalbumin neurons and accelerated by Somatostatin neurons. *bioRxiv*. https://doi.org/10.1101/2021.09.06.459211.

Lee, W. C. A., Bonin, V., Reed, M., Graham, B. J., Hood, G., Glattfelder, K., & Reid, R. C. (2016). Anatomy and function of an excitatory network in the visual cortex. *Nature*, **532**(7599), 370–374.

Lefort, S., Tomm, C., Floyd Sarria, J. C. F., & Petersen, C. C. H. (2009). The excitatory neuronal network of the C2 barrel column in mouse primary somatosensory cortex. *Neuron*, **61**(2), 301–316.

Letzkus, J. J., Wolff, S. B. E., Meyer, E. M. M., Tovote, P., Courtin, J., Herry, C., & Lüthi, A. (2011). A disinhibitory microcircuit for associative fear learning in the auditory cortex. *Nature*, **480**(7377), 331–335.

Li, Y., Lu, H., Cheng, P., Ge, S., Xu, H., Shi, S.-H., & Dan, Y. (2012). Clonally related visual cortical neurons show similar stimulus feature selectivity. *Nature*, **486**(7401), 118–121.

Litwin-Kumar, A., & Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature Neuroscience*, **15**(11), 1498–1505.

Litwin-Kumar, A., & Doiron, B. (2014). Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nature Communication*, **5**(1), 5319.

Liu, X., Ramirez, S., Pang, P. T., Puryear, C. B., Govindarajan, A., Deisseroth, K., & Tonegawa, S. (2012). Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature*, **484**(7394), 381–385.

Loidolt, M., Rudelt, L., & Priesemann, V. (2020). Sequence memory in recurrent neuronal network can develop without structured input. *bioRxiv*. https://doi.org/10.1101/2020.09.15.297580.

Lu, H., Gallinaro, J. V., & Rotter, S. (2019). Network remodeling induced by transcranial brain stimulation: A computational model of tDCS-triggered cell assembly formation. *Network Neuroscience*, **3**(4), 924–943.

Luboeinski, J., & Tetzlaff, C. (2021). Memory consolidation and improvement by synaptic tagging and capture in recurrent neural networks. *Communications Biology*, **4**(1), 1–17.

Luczak, A., Barthó, P., & Harris, K. D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron*, **62**(3), 413–425.

Lynch, G. S., Dunwiddie, T., & Gribkoff, V. (1977). Heterosynaptic depression: a postsynaptic correlate of long-term potentiation. *Nature*, **266**(5604), 737–739.

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

MacLean, J. N., Watson, B. O., Aaron, G. B., & Yuste, R. (2005). Internal dynamics determine the cortical response to thalamic stimulation. *Neuron*, **48**(5), 811–823.

Maes, A., Barahona, M., & Clopath, C. (2020). Learning spatiotemporal signals using a recurrent spiking network that discretizes time. *Plos Computational Biology*, **16**(1), e1007606.

Maes, A., Barahona, M., & Clopath, C. (2021). Learning compositional sequences with multiple time scales through a hierarchical network of spiking neurons. *Plos Computational Biology*, **17**(3), e1008866.

Maffei, A. (2018). Long-term potentiation and long-term depression. In *Oxford Research Encyclopedia of Neuroscience*.

Magee, J. C., & Grienberger, C. (2020). Synaptic plasticity forms and functions. *Annual Review of Neuroscience*, **43**(1), 95–117.

Malvache, A., Reichinnek, S., Villette, V., Haimerl, C., & Cossart, R. (2016). Awake hippocampal reactivations project onto orthogonal neuronal assemblies. *Science*, **353**(6305), 1280–1283.

Manz, P., & Memmesheimer, R.-M. (2022). Purely STDP-based assembly dynamics: Stability, learning, overlaps, drift and aging. *bioRxiv*. https://doi.org/10.1101/2022.06.20.496825.

Mao, B. Q., Hamzei-Sichani, F., Aronov, D., Froemke, R. C., & Yuste, R. (2001). Dynamics of spontaneous activity in neocortical slices. *Neuron*, **32**(5), 883–898.

Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science*, **275**(5297), 213–215.

Marks, T. D., & Goard, M. J. (2021). Stimulus-dependent representational drift in primary visual cortex. *Nature Communication*, **12**(1), 5169.

Marshel, J. H., Kim, Y. S., Machado, T. A., Quirin, S., Benson, B., Kadmon, J., Raja, C., Chibukhchyan, A., Ramakrishnan, C., Inoue, M., Shane, J. C., McKnight, D. J., Yoshizawa, S., Kato, H. E., Ganguli, S., & Deisseroth, K. (2019). Cortical layer-specific critical dynamics triggering perception. *Science*, **365**(6453), eaaw5202.

Mau, W., Hasselmo, M. E., & Cai, D. J. (2020). The brain in motion: How ensemble fluidity drives memory-updating and flexibility. *eLife*, **9**, e63550.

Mazzucato, L., Fontanini, A., & La Camera, G. (2015). Dynamics of multistable states during ongoing and evoked cortical activity. *Journal of Neuroscience*, **35**(21), 8214–8231.

Mazzucato, L., La Camera, G., & Fontanini, A. (2019). Expectation-induced modulation of metastable activity underlies faster coding of sensory stimuli. *Nature Neuroscience*, **22**(5), 787–796.

Miconi, T., McKinstry, J. L., & Edelman, G. M. (2016). Spontaneous emergence of fast attractor dynamics in a model of developing primary visual cortex. *Nature Communication*, **7**(1), 13208.

Miehl, C., & Gjorgjieva, J. (2022). Stability and learning in excitatory synapses by nonlinear inhibitory plasticity. *bioRxiv*. https://doi.org/10.1101/2022.03.28.486052.

Miller, J. E. K., Ayzenshtat, I., Carrillo-Reid, L., & Yuste, R. (2014). Visual stimuli recruit intrinsically generated cortical ensembles. *Proceedings of the National Academy of Sciences of the United States of America*, **111**(38), E4053–E4061.

Miller, K. D. (1996). Synaptic economics: Competition and cooperation in synaptic plasticity. *Neuron*, **17**(3), 371–374.

Miller, K. D., & MacKay, D. J. C. (1994). The role of constraints in hebbian learning. *Neural Computation*, **6**(1), 100–126.

Mongillo, G., Curti, E., Romani, S., & Amit, D. J. (2005). Learning in realistic networks of spiking neurons and spike-driven plastic synapses. *European Journal of Neuroscience*, **21**(11), 3143–3160.

Montangie, L., Miehl, C., & Gjorgjieva, J. (2020). Autonomous emergence of connectivity assemblies via spike triplet interactions. *Plos Computational Biology*, **16**(5), e1007835.

Morrison, A., Aertsen, A., & Diesmann, M. (2007). Spike-timing-dependent plasticity in balanced random networks. *Neural Computation*, **19**(6), 1437–1467.

Nabavi, S., Fox, R., Proulx, C. D., Lin, J. Y., Tsien, R. Y., & Malinow, R. (2014). Engineering a memory with LTD and LTP. *Nature*, **511**(7509), 348–352.

Neves, G., Cooke, S. F., & Bliss, T. V. P. (2008). Synaptic plasticity, memory and the hippocampus: A neural network approach to causality. *Nature Reviews Neuroscience*, **9**(1), 65–75.

Ocker, G. K., & Doiron, B. (2019). Training and spontaneous reinforcement of neuronal assemblies by spike timing plasticity. *Cerebral Cortex*, **29**(3), 937–951.

Ocker, G. K., Hu, Y., Buice, M. A., Doiron, B., Josić, K., Rosenbaum, R., & Shea-Brown, E. (2017). From the statistics of connectivity to the statistics of spike times in neuronal networks. *Current Opinion in Neurobiology*, **46**, 109–119.

Ocker, G. K., Josić, K., Shea-Brown, E., & Buice, M. A. (2017). Linking structure and activity in nonlinear spiking networks. *Plos Computational Biology*, **13**(6), e1005583.

Ocker, G. K., Litwin-Kumar, A., & Doiron, B. (2015). Self-organization of microcircuits in networks of spiking neurons with plastic synapses. *Plos Computational Biology*, **11**(8), e1004458.

Ohtsuki, G., Nishiyama, M., Yoshida, T., Murakami, T., Histed, M., Lois, C., & Ohki, K. (2012). Similarity of visual selectivity among clonally related neurons in visual cortex. *Neuron*, **75**(1), 65–72.

Oishi, N., Nomoto, M., Ohkawa, N., Saitoh, Y., Sano, Y., Tsujimura, S., Nishizono, H., Matsuo, M., Muramatsu, S. I., & Inokuchi, K. (2019). Artificial association of memory events by optogenetic stimulation of hippocampal CA3 cell ensembles. *Molecular Brain*, **12**(1), 1–10.

Papadimitriou, C. H., & Friederici, A. D. (2022). Bridging the gap between neurons and cognition through assemblies of neurons. *Neural Computation*, **34**(2), 291–306.

Papadimitriou, C. H., Vempala, S. S., Mitropolsky, D., Collins, M., & Maass, W. (2020). Brain computation by assemblies of neurons. *Pnas*, **117**(25), 14464–14472.

Perin, R., Berger, T. K., & Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. *Proceedings of the National Academy of Sciences of the United States of America*, **108**(13), 5419–5424.

Pernice, V., Staude, B., Cardanobile, S., & Rotter, S. (2011). How structure determines correlations in neuronal networks. *Plos Computational Biology*, **7**(5), e1002059.

Peron, S., Pancholi, R., Voelcker, B., Wittenbach, J. D., Ólafsdóttir, H. F., Freeman, J., & Svoboda, K. (2020). Recurrent interactions in local cortical circuits. *Nature*, **579**(7798), 256–259.

Pfister, J.-P., & Gerstner, W. (2006). Triplets of spikes in a model of spike timing-dependent plasticity. *Journal of Neuroscience*, **26**(38), 9673–9682.

Pietri, T., Romano, S. A., Pérez-Schuster, V., Boulanger-Weill, J., Candat, V., & Sumbre, G. (2017). The emergence of the spatial structure of tectal spontaneous activity is independent of visual inputs. *Cell reports*, **19**(5), 939–948.

Podlaski, W. F., Agnes, E. J., & Vogels, T. P. (2020). Context-modular memory networks support high-capacity, flexible, and robust associative memories. *bioRxiv*. https://doi.org/10.1101/2020.01.08.898528.

Pokorny, C., Ison, M. J., Rao, A., Legenstein, R., Papadimitriou, C., & Maass, W. (2020). STDP forms associations between memory traces in networks of spiking neurons. *Cerebral Cortex*, **30**(3), 952–968.

Raman, D. V., & O'Leary, T. (2021). Optimal plasticity for memory maintenance during ongoing synaptic change. *eLife*, **10**, e62912.

Ramaswami, M. (2014). Network plasticity in adaptive filtering and behavioral habituation. *Neuron*, **82**(6), 1216–1229.

Ranhel, J. (2012). Neural assembly computing. *IEEE Transactions on Neural Networks and Learning Systems*, **23**(6), 916–927.

Rashid, A. J., Yan, C., Mercaldo, V., Hsiang, H.-L. L., Park, S., Cole, C. J., De Cristofaro, A., Yu, J., Ramakrishnan, C., Lee, S. Y., Deisseroth, K., Frankland, P. W., & Josselyn, S. A. (2016). Competition between engrams influences fear memory formation and recall. *Science*, **353**(6297), 383–387.

Ravid Tannenbaum, N., & Burak, Y. (2016). Shaping neural circuits by high order synaptic interactions. *Plos Computational Biology*, **12**(8), e1005056.

Redondo, R. L., & Morris, R. G. M. (2011). Making memories last: the synaptic tagging and capture hypothesis. *Nature Reviews Neuroscience*, **12**(1), 17–30.

Richter, L. M., & Gjorgjieva, J. (2017). Understanding neural circuit development through theory and models. *Current Opinion in Neurobiology*, **46**, 39–47.

Romano, S. A., Pietri, T., Pérez-Schuster, V., Jouary, A., Haudrechy, M., & Sumbre, G. (2015). Spontaneous neuronal network dynamics reveal circuit's functional adaptations for behavior. *Neuron*, **85**(5), 1070–1085.

Rossi, L. F., Harris, K. D., & Carandini, M. (2020). Spatial connectivity matches direction selectivity in visual cortex. *Nature*, **588**(7839), 648–652.

van Rossum, M. C. W., Bi, G. Q., & Turrigiano, G. G. (2000). Stable hebbian learning from spike timing-dependent plasticity. *Journal of Neuroscience*, **20**, 8812–8821.

Rubin, J., Lee, D. D., & Sompolinsky, H. (2001). Equilibrium properties of temporally asymmetric hebbian plasticity. *Physical Review Letter*, **86**(2), 364–367.

Rule, M. E., & O'Leary, T. (2022). Self-healing codes: How stable neural populations can track continually reconfiguring neural representations. *Proceedings of the National Academy of Sciences of the United States of America*, **119**(7), e2106692119.

Rule, M. E., O'Leary, T., & Harvey, C. D. (2019). Causes and consequences of representational drift. *Current Opinion in Neurobiology*, **58**, 141–147.

Sadeh, S., & Clopath, C. (2021). Excitatory-inhibitory balance modulates the formation and dynamics of neuronal assemblies in cortical networks. *Science Advances*, **7**(45), eabg8411.

Schoonover, C. E., Ohashi, S. N., Axel, R., & Fink, A. J. P. (2021). Representational drift in primary olfactory cortex. *Nature*, **594**(7864), 541–546.

Schulz, A., Miehl, C., Berry II M. J., & Gjorgjieva J. (2021). The generation of cortical novelty responses through inhibitory plasticity. *eLife*, **10**, e65309.

Shatz, C. J. (1992). The developing brain. *Scientific American*, **267**(3), 60–67.

Song, S., & Abbott, L. F. (2001). Cortical development and remapping through spike timing-dependent plasticity. *Neuron*, **32**(2), 339–350.

Song, S., Miller, K. D., & Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature Neuroscience*, **3**(9), 919–926.

Song, S., Sjöström, P. J., Reigl, M., Nelson, S., & Chklovskii, D. B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *Plos Biology*, **3**(3), e68.

Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M., & Harris, K. D. (2019). Spontaneous behaviors drive multidimensional, brainwide activity. *Science*, **364**(6437), eaav7893.

Suvrathan, A. (2019). Beyond STDP—towards diverse and functionally relevant plasticity rules. *Current Opinion in Neurobiology*, **54**, 12–19.

Tarusawa, E., Sanbo, M., Okayama, A., Miyashita, T., Kitsukawa, T., Hirayama, T., Hirabayashi, T., Hasegawa, S., Kaneko, R., Toyoda, S., Kobayashi, T., Kato-Itoh, M., Nakauchi, H., Hirabayashi, M., Yagi, T., & Yoshimura, Y. (2016). Establishment of high reciprocal connectivity between clonal cortical neurons is regulated by the Dnmt3b DNA methyltransferase and clustered protocadherins. *Bmc Biology*, **14**(1), 103.

Tetzlaff, C., Dasgupta, S., Kulvicius, T., & Wörgötter, F. (2015). The use of hebbian cell assemblies for nonlinear computation. *Science Reports*, **5**, 1–14.

Tetzlaff, C., Kolodziejski, C., Timme, M., Tsodyks, M., & Wörgötter, F. (2013). Synaptic scaling enables dynamically distinct short- and long-term memory formation. *Plos Computational Biology*, **9**(10), e1003307.

Tetzlaff, C., Kolodziejski, C., Timme, M., & Wörgötter, F. (2011). Synaptic scaling in combination with many generic plasticity mechanisms stabilizes circuit connectivity. *Frontiers in Computational Neuroscience*, **5**, 47.

Triesch, J., Vo, A. D., & Hafner, A.-S. (2018). Competition for synaptic building blocks shapes synaptic plasticity. *eLife*, **7**, e37836.

Triplett, M. A., Avitan, L., & Goodhill, G. J. (2018). Emergence of spontaneous assembly activity in developing neural networks without afferent input. *Plos Computational Biology*, **14**(9), e1006421.

Trousdale, J., Hu, Y., Shea-Brown, E., & Josić, K. (2012). Impact of network structure and cellular response on spike time correlations. *Plos Computational Biology*, **8**(3), e1002408.

Tsodyks, M. V. (1989). Associative memory in neural networks with the hebbian learning rule. *Modern Physics Letters B*, **03**, 555–560.

Turner, N. L., Macrina, T., Bae, J. A., Yang, R., Wilson, A. M., Schneider-Mizell, C., Lee, K., Lu, R., Wu, J., Bodor, A. L., Bleckert, A. A., Brittain, D., Froudarakis, E., Dorkenwald, S., Collman, F., Kemnitz, N., Ih, D., Silversmith, W. M., Zung, J., …, & Seung H. S. (2022). Reconstruction of neocortex: Organelles, compartments, cells, circuits, and activity. *Cell*, **185**(6), 1082–1100.e24.

Turrigiano, G. G. (2008). The self-tuning neuron: Synaptic scaling of excitatory synapses. *Cell*, **135**(3), 422–435.

Turrigiano, G. G., Leslie, K. R., Desai, N. S., Rutherford, L. C., & Nelson, S. B. (1998). Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, **391**(6670), 892–896.

Turrigiano, G. G., & Nelson, S. B. (2004). Homeostatic plasticity in the developing nervous system. *Nature Reviews Neuroscience*, **5**(2), 97–107.

Vasilaki, E., & Giugliano, M. (2014). Emergence of connectivity motifs in networks of model neurons with short- and long-term plastic synapses. *Plos One*, **9**(1), e84626.

Wenzel, M., & Hamm, J. P. (2021). Identification and quantification of neuronal ensembles in optical imaging experiments. *Journal of Neuroscience Methods*, **351**, 109046.

Wertz, A., Trenholm, S., Yonehara, K., Hillier, D., Raics, Z., Leinweber, M., Szalay, G., Ghanem, A., Keller, G. B., Rózsa, B., Conzelmann, K. K., & Roska, B. (2015). Single-cell-initiated monosynaptic tracing reveals layer-specific cortical network modules. *Science*, **349**(6243), 70–74.

Wu, Y. K., Hengen, K. B., Turrigiano, G. G., & Gjorgjieva, J. (2020). Homeostatic mechanisms regulate distinct aspects of cortical circuit dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, **117**(39), 24514–24525.

Yger, P., & Gilson, M. (2015). Models of metaplasticity: A review of concepts. *Frontiers in Computational Neuroscience*, **9**, 138.

Yoshimura, Y., Dantzker, J. L., & Callaway, E. M. (2005). Excitatory cortical neurons form fine-scale functional networks. *Nature*, **433**(7028), 868–873.

Yu, Y.-C., Bultje, R. S., Wang, X., & Shi, S.-H. (2009). Specific synapses develop preferentially among sister excitatory neurons in the neocortex. *Nature*, **458**(7237), 501–504.

Yu, Y.-C., He, S., Chen, S., Fu, Y., Brown, K. N., Yao, X.-H., Ma, J., Gao, K. P., Sosinsky, G. E., Huang, K., & Shi, S.-H. (2012). Preferential electrical coupling regulates neocortical lineage-dependent microcircuit assembly. *Nature*, **486**(7401), 113–117.

Yuste, R. (2015). From the neuron doctrine to neural networks. *Nature Reviews Neuroscience*, **16**(8), 487–497.

Zenke, F., Agnes, E. J., & Gerstner, W. (2015). Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature Communication*, **6**(1), 6922.

Zenke, F., Hennequin, G., & Gerstner, W. (2013). Synaptic plasticity in neural networks needs homeostasis with a fast rate detector. *Plos Computational Biology*, **9**(11), e1003330.

Zhang, D., Yan, X., She, L., Wen, Y., & Poo, M. M. (2020). Global enhancement of cortical excitability following coactivation of large neuronal populations. *Proceedings of the National Academy of Sciences of the United States of America*, **117**(33), 20254–20264.

Zhang, W., & Linden, D. J. (2003). The other side of the engram: Experience-driven changes in neuronal intrinsic excitability. *Nature Reviews Neuroscience*, **4**(11), 885–900.

Additional information

Competing interests

None.

Author contributions

C.M., S.O. and J.G. conceptualised the manuscript, C.M. and S.O. wrote a first draft of the manuscript, and all authors revised the manuscript.

Funding

The authors thank the Max Planck Society for funding. S.O. is supported by the Joachim Herz Foundation. C.M. is supported by the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Centre (CRC) 1080 (awarded to J.G.). This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 804824).

Acknowledgements

The authors thank all members of the 'Computation in Neural Circuits' group for useful discussions and Alice Dauphin and Claudia Cusseddu for comments on the manuscript.

Open Access funding enabled and organized by Projekt DEAL.

Keywords

assembly, network model, neural circuits, synaptic plasticity

Supporting information

Additional supporting information can be found online in the Supporting Information section at the end of the HTML view of the article. Supporting information files available:

Peer Review History

VI. Autonomous emergence of connectivity assemblies via spike triplet interactions

Montangie, L., Miehl, C. & Gjorgjieva, J. Autonomous emergence of connectivity assemblies via spike triplet interactions. *PLoS Computational Biology* **16(5)**, e1007835 (2020). https://doi.org/10.1371/journal.pcbi.1007835

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Autonomous emergence of connectivity assemblies via spike triplet interactions

Lisandro Montangie[1], **Christoph Miehl**[1,2], **Julijana Gjorgjieva**[1,2]*

1 Computation in Neural Circuits Group, Max Planck Institute for Brain Research, Frankfurt, Germany, **2** Technical University of Munich, School of Life Sciences, Freising, Germany

* gjorgjieva@brain.mpg.de

Citation: Montangie L, Miehl C, Gjorgjieva J (2020) Autonomous emergence of connectivity assemblies via spike triplet interactions. PLoS Comput Biol 16(5): e1007835. https://doi.org/10.1371/journal.pcbi.1007835

Editor: Brent Doiron, University of Pittsburgh, UNITED STATES

Received: June 20, 2019

Accepted: March 31, 2020

Published: May 8, 2020

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files. All code is available at https://github.com/comp-neural-circuits/Autonomous-assembly-emergence-via-triplet-STDP.

Funding: LM, CM and JG thank the Max Planck Society for funding. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 804824). The funders had no role in

Abstract

Non-random connectivity can emerge without structured external input driven by activity-dependent mechanisms of synaptic plasticity based on precise spiking patterns. Here we analyze the emergence of global structures in recurrent networks based on a triplet model of spike timing dependent plasticity (STDP), which depends on the interactions of three precisely-timed spikes, and can describe plasticity experiments with varying spike frequency better than the classical pair-based STDP rule. We derive synaptic changes arising from correlations up to third-order and describe them as the sum of structural motifs, which determine how any spike in the network influences a given synaptic connection through possible connectivity paths. This motif expansion framework reveals novel structural motifs under the triplet STDP rule, which support the formation of bidirectional connections and ultimately the spontaneous emergence of global network structure in the form of self-connected groups of neurons, or assemblies. We propose that under triplet STDP assembly structure can emerge without the need for externally patterned inputs or assuming a symmetric pair-based STDP rule common in previous studies. The emergence of non-random network structure under triplet STDP occurs through internally-generated higher-order correlations, which are ubiquitous in natural stimuli and neuronal spiking activity, and important for coding. We further demonstrate how neuromodulatory mechanisms that modulate the shape of the triplet STDP rule or the synaptic transmission function differentially promote structural motifs underlying the emergence of assemblies, and quantify the differences using graph theoretic measures.

Author summary

Emergent non-random connectivity structures in different brain regions are tightly related to specific patterns of neural activity and support diverse brain functions. For instance, self-connected groups of neurons, known as assemblies, have been proposed to represent functional units in brain circuits and can emerge even without patterned external instruction. Here we investigate the emergence of non-random connectivity in recurrent networks using a particular plasticity rule, triplet STDP, which relies on the interaction of spike triplets and can capture higher-order statistical dependencies in

study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

neural activity. We derive the evolution of the synaptic strengths in the network and explore the conditions for the self-organization of connectivity into assemblies. We demonstrate key differences of the triplet STDP rule compared to the classical pair-based rule in terms of how assemblies are formed, including the realistic asymmetric shape and influence of novel connectivity motifs on network plasticity driven by higher-order correlations. Assembly formation depends on the specific shape of the STDP window and synaptic transmission function, pointing towards an important role of neuromodulatory signals on formation of intrinsically generated assemblies.

Introduction

The synaptic wiring between neurons—originally proposed as a mechanism for learning and memory—is sculpted by experience and has become a most relevant link between circuit structure and function [1]. The original formulation of Hebbian plasticity, whereby "cells that fire together, wire together" [2, 3], fostered the concept of 'cell assemblies' [4], defined as groups of neurons that are repeatedly co-activated leading to the strengthening of synaptic connectivity between individual neurons. This has suggested that activity-dependent synaptic plasticity, including both long-term potentiation and long-term depression, is a key mechanism for the emergence of assemblies in the organization of neural circuits [5–7]. These interconnected groups of neurons have become an important target for many theories of neural computation and associative memory [8–11]. Recent technological developments that enable multiple neurons to be simultaneously recorded have provided the much needed physiological evidence of assembly organization [12–15]. For instance, synaptically connected neurons tend to receive more common input than would be expected by chance, [12, 16–18] and cortical pyramidal neurons tend to be more strongly connected to neurons that share stimulus preference [13, 19, 20], providing evidence for clustered architecture. It has been proposed that this organization enables the cortex to intrinsically generate reverberating patterns of neural activity when representing different stimulus features [1, 21]. Thus, neuronal assemblies can be interpreted as the building blocks of cortical microcircuits which are differentially recruited during distinct functions, such as the binding of different features of a sensory stimulus [7, 17, 22]. In addition to cortical circuits, neuronal assemblies have also been observed in the optic tectum (a structure homologous to the superior colliculus in mammals [23]) in the developing zebrafish larva [24–27]. Experiments in sensory deprived larvae have demonstrated that the basic structure of spontaneous activity and functional connectivity emerges without intact retinal inputs, suggesting that neuronal assemblies are intrinsically generated in the tectum and not just the product of correlated external inputs [25–27]. This raises the important question of what drives the emergence of these clustered structures, and whether patterned external input is necessary.

To understand the emergence of such non-random connectivity, a growing body of theoretical and computational work has been developed to link connectivity architecture to the coordinated spiking activity of neurons, especially in recurrent networks [28–41]. These studies can be divided into two classes: those that examine the influence of externally structured input on activity-dependent refinement [42–47], and those that investigate the autonomous emergence of non-random connectivity in the absence of patterned external input, purely driven by emergent network interactions [5, 6, 48]. Specifically, assemblies in recurrent networks can be imprinted based on internally-generated network interactions [6] or through rate-based plasticity where inputs with higher firing rates to subsets of neurons strengthen

recurrent connections [49, 50]; assemblies can also be initially determined by externally patterned input but maintained by internal correlations [51].

Despite this success, all of these studies have assumed pair-based models of STDP, which induce plasticity based on the precise timing and order of a pair of pre- and postsynaptic spikes [52, 53]. Here, we consider a spike-based learning rule, "the triplet STDP model" [54], which uses sets of three spikes (triplets) to induce plasticity. Specifically, we focus on the 'minimal' triplet STDP model, where only potentiation depends on the interval between the pre- and postsynaptic spikes, and on the timing of the previous postsynaptic spike. This triplet learning rule has been shown to explain a variety of synaptic plasticity data [55, 56] significantly better than pair-based STDP [54]. We have previously shown a tight correspondence between the triplet STDP rule and the well-known Bienenstock-Cooper-Munro (BCM) synaptic learning rule, which maximizes the selectivity of the postsynaptic neuron, and thereby offers a possible explanation for experience-dependent cortical plasticity such as orientation and direction selectivity [57]. In addition, triplet STDP can also induce selectivity for input patterns consisting of up to third-order correlations, here referred to as higher-order correlations (HOCs). HOCs have been experimentally measured in several brain areas [58], and shown to account for a substantial amount of information transfer in sensory cortex [58–61]. HOCs are also important for characterizing the firing of a postsynaptic neuron [62, 63], circuit function and coding [64, 65], and the synchronous firing and the distribution of activity in a neuronal pool [66–69]. Here we investigated the functional significance of such HOCs for shaping recurrent network structure through synaptic plasticity.

First, we investigate how HOCs up to third order affect the development of connectivity in recurrent networks of Poisson spiking neurons in the absence of structured external stimuli, where the stochastic activity of each neuron is described by a mutually exciting Hawkes process [70]. Assuming a slow change of synaptic efficacies and fast spiking dynamics, we develop a formal analytical framework for the evolution of synaptic connections in the network based on the second- and third-order cumulants of spike timing interactions, which arise from assuming an STDP rule governed by pairs and triplets of spikes [54, 55]. The simplified neuronal model allows us to write exact and self-consistent equations for the synaptic change depending on the full network connectivity by taking into account non-local interactions between different neurons in the network and writing them as a sum of structural motifs of varying orders. We demonstrate differences to the classical pair-based STDP rule [52, 71] that ignores those HOCs, and compare the relative strength of the emergent structural motifs up to third-order induced by triplet STDP. Second, we examine the biological conditions which promote the formation of assembly structures of self-connected neurons without externally structured inputs under the triplet STDP rule. We find that this is achieved either by modulating the shape of the STDP function through neuromodulators or the shape of the evoked postsynaptic current (EPSC) and characterize changes in functional connectivity in terms of graph theoretic measures [25–27]. Third, we show that the novel structural motifs, and specifically 'loop' motifs, which follow from the triplet STDP rule, are crucial for the spontaneous emergence of assemblies. Finally, we compare them to assemblies generated via correlated external input.

Results

We present two main results: first, we derive a formal analytical framework for the evolution of synaptic weights depending on the second- and third-order cumulants of spike time interactions under the triplet STDP rule by expressing them as a sum of structural motifs; second, we

discuss the functional implications of this framework and present the biological conditions which promote the formation of assemblies without external instruction.

Average synaptic modification due to the interaction of pairs and triplets of spikes in recurrent networks

To study the autonomous emergence of assemblies in a recurrent network from a general form of STDP that includes the contribution of pairs and triplets of spikes to synaptic plasticity, we require a minimal theoretical representation of the network with plastic synapses driven by internal correlations in the spike timing statistics. In our model, structure is given by the connectivity matrix $\boldsymbol{W}$ between all excitatory neurons in the network ("all-to-all connectivity"), where the synaptic efficacy element W_{ij} denotes the connection strength between postsynaptic neuron i and presynaptic neuron j. The analytical description of the dynamics in recurrent networks can be dauntingly complex. On the one hand, to rigorously analyze the impact of STDP on the formation of functional structures it is indispensable to take into account the precise timing of action potentials or spikes. Therefore, models of neural activity that are based on rates cannot fulfill this criterion [72]. More elaborate models such as Hodgkin-Huxley with multiple ion channels [73] and even the simpler spiking leaky integrate-and-fire (LIF) models are much more accurate in reproducing the spiking dynamics of a population of neurons [74–76]. Although they are computationally tractable, to extract extensive and exact mathematical features from these models remains an elusive task. Under certain conditions of approximately asynchronous firing, the spiking statistics in networks of LIF neuron can be described by a linear theory [29]. Using this approach, here we make approximations for the spiking dynamics of each individual excitatory cell and treat each pre- and postsynaptic spike train as if they follow inhomogeneous Poisson statistics [6, 44, 52, 57].

In this model we assume that the probability of each neuron emitting an action potential at a certain time (the 'intensity' or mean activity) is proportional to the weighted sum of the preceding activity of all the other cells in the network and a constant, unstructured external input (Fig 1A). The activity of each neuron in this network is a stochastic process, also referred to as a 'mutually exciting point process' or a Hawkes process [70]. The availability of an exact expression for spike correlations in this model allows us to develop a precise theory for the synaptic efficacies' dynamics that are governed by different forms of STDP. To prevent runaway excitation, we also consider that the firing of excitatory neurons is modulated by the activity of a population of inhibitory neurons (Fig 1A). We assume that the total inhibitory input to each excitatory neuron is tuned in order to balance the sum of inhibitory efficacies with the sum of the excitatory ones (Methods) [6, 77–79].

Given the connectivity matrix $\boldsymbol{W}$ and assuming a slow learning rate (much slower than the dynamics of neural activity), the rate of change in the strength of synaptic efficacy $\langle \dot{W}_{ij} \rangle$ between postsynaptic neuron i and presynaptic neuron j, can be expressed in terms of the product of the time dependent cumulants of different orders and the STDP function, accordingly (Methods). Specifically, we consider STDP learning rules where plasticity depends on the timing and order of pairs and triplets of spikes, referred to as pair-based and triplet STDP. Initially, we make no assumptions about the shape of these learning rules keeping the framework general. The sign and magnitude of the net weight modification depends on the time interval between the firing of the pre- and postsynaptic neurons, and also on the relative spike times of individual pre- and postsynaptic neurons (Fig 1B). The exact expression for the evolution of the average (denoted by $\langle \cdot \rangle$) synaptic efficacy in the recurrent

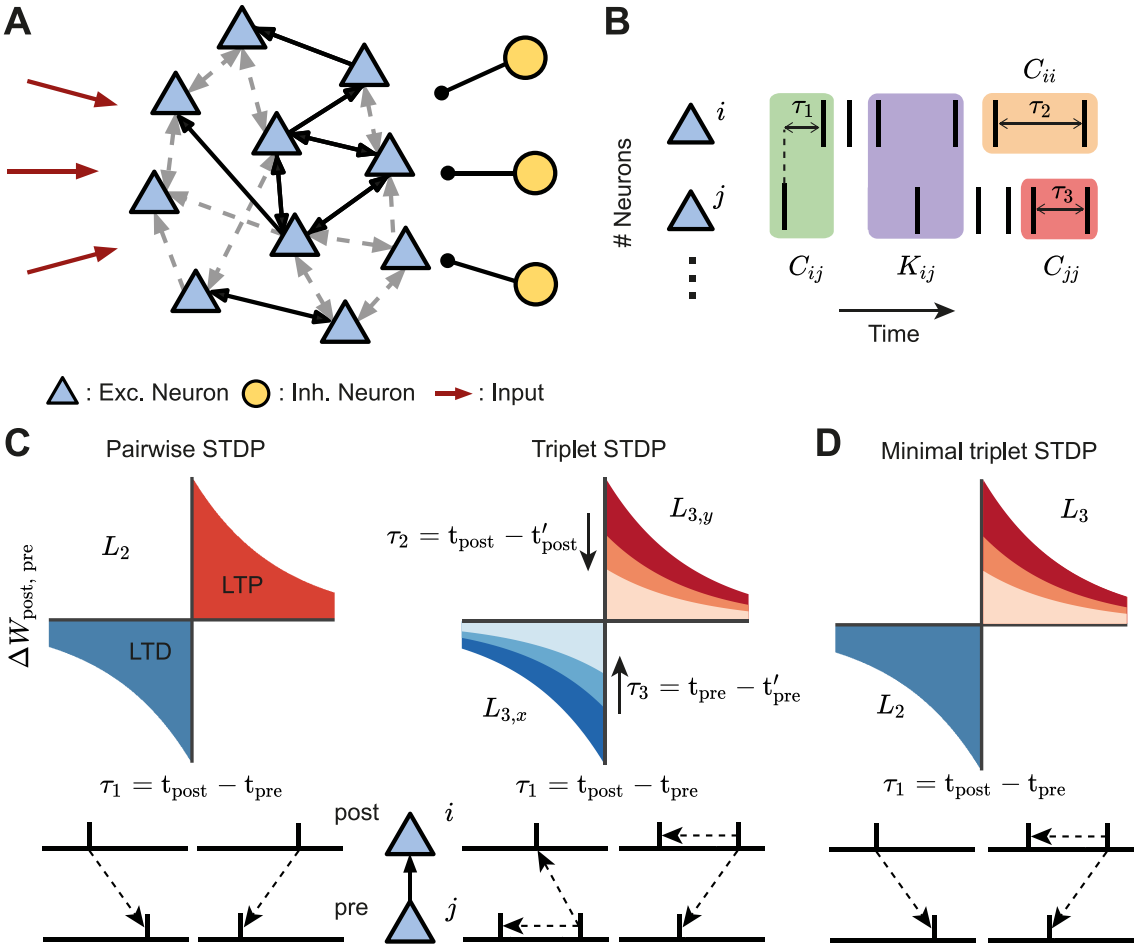


Fig 1. Framework set-up. A. A network of excitatory neurons (light blue triangles) fire stochastically, while their activity is driven by unstructured external input (red arrows) and modulated by a population of inhibitory neurons (yellow circles). Excitatory connections among the neurons can be weak (gray dashed arrows) or strong (black solid arrows), unidirectional or bidirectional. **B**. Cumulants of the spike trains (see Eq 1). The second-order cumulants C_{ij}, C_{ii} and C_{jj} are calculated based on the time difference between a pair of spikes (cross-covariance in green; auto-covariances in orange/red). The third-order cumulant K_{ij} is calculated based on the time differences between three spikes (purple). The spike triplets can be two post- and one presynaptic spikes, or one post- and two presynaptic spikes. The time differences are: τ_1 between a presynaptic spike and a postsynaptic spike, τ_2 between different postsynaptic spikes and τ_3 between different presynaptic spikes. **C**. STDP-induced plasticity by pairs and triplets of spikes. Left: An example of a classical pair-based STDP rule, with a learning window denoted by L_2. Potentiation is triggered by a postsynaptic following a presynaptic spike ($\tau_1 = t_{post} - t_{pre} > 0$), whereas if a presynaptic spike follows a postsynaptic spike ($\tau_1 = t_{post} - t_{pre} < 0$), depression is induced. The total potentiation (depression) is given by the red (blue) area under the curve. Right: Examples of triplet STDP rules denoted by $L_{3,y}$ and $L_{3,x}$. Potentiation (red) and depression (blue) are given by triplets of spikes: post-pre-post with a time difference $\tau_2 = t_{post} - t'_{post}$, and pre-post-pre with a time difference $\tau_3 = t_{pre} - t'_{pre}$, respectively. **D**. Minimal triplet STDP rule where potentiation depends on triplets of spikes L_3 and depression on pairs of spikes L_2.

https://doi.org/10.1371/journal.pcbi.1007835.g001

network due to STDP is

$$\langle \dot{W}_{ij}^{\text{STDP}} \rangle = \int_{-\infty}^{\infty} (C_{ij}(\tau_1) + r_i r_j) L_2(\tau_1) \mathrm{d}\tau_1$$
$$+ \iint_{-\infty}^{\infty} (K_{ij}(\tau_1, \tau_2) + r_i(C_{ij}(\tau_1) + C_{ij}(\tau_2 - \tau_1)) + r_j C_{ii}(\tau_2) + r_i^2 r_j) L_{3,y}(\tau_1, \tau_2) \mathrm{d}\tau_1 \, \mathrm{d}\tau_2 \qquad (1)$$
$$+ \iint_{-\infty}^{\infty} (K_{ij}(\tau_1, \tau_3) + r_j(C_{ij}(\tau_1) + C_{ij}(\tau_3 - \tau_1)) + r_i C_{jj}(\tau_3) + r_i r_j^2) L_{3,x}(-\tau_1, -\tau_3) \mathrm{d}\tau_1 \, \mathrm{d}\tau_3.$$

Here r_i and r_j denote the mean firing rates of neuron i and j, respectively; C_{ij} is the cross-covariance between neuron i and neuron j, with C_{ii} and C_{jj} being the auto-covariances (note that all of these covariance terms, C_{ij}, C_{ii} and C_{jj}, make up the second-order cumulant); and K_{ij} is the third-order cumulant between neuron i and neuron j. These quantities represent internal (i.e. not driven by external input) correlations in the network and are calculated as functions of the excitatory postsynaptic current (EPSC), and assumed to be identical for every pair of neurons. Both the second-order cumulants C and the third-order cumulants K are probability densities of pairs and triplets of spikes separated by the given time lapses τ accordingly (Fig 1B). τ_1 is the time difference between a spike emitted by the presynaptic neuron and one from the postsynaptic neuron, whereas τ_2 and τ_3 are the time intervals between different spikes from the postsynaptic neuron and the presynaptic neuron, respectively. The cumulant K_{ij} is calculated for both 'post-pre-post' or 'pre-post-pre' spike triplets and therefore depends on combinations of τ_1 and τ_2 or τ_3, according to each case.

The STDP functions that describe how potentiation or depression depend on the spike timing intervals are given by L_2 for pairs of spikes, and $L_{3,x}$ and $L_{3,y}$ for triplets of spikes. The subindices x and y correspond to the triplet sets 'pre-post-pre' and 'post-pre-post,' respectively. While Eq 1 can be calculated for any shape of the STDP function that depends on pairs and triplets of spikes, an illustrative example for these learning rules, commonly used in other studies based on fits to experimental data [54, 55, 71], is given in Fig 1C.

The average synaptic efficacy change (Eq 1) is sufficient to describe the plasticity dynamics when the learning rate is small relative to the spiking dynamics, and noise in the STDP dynamics, arising from random fluctuations, is averaged out. Furthermore, Eq 1 is combined with heterosynaptic competition [80] to restrict the amount of connections a neuron can make with the rest and prevent the dominance of a few (Methods). For the sake of simplicity, in the next steps we consider that triplets of spikes contribute only to potentiation and thus $L_{3,y}(\tau_1, \tau_2) = L_3(\tau_1, \tau_2)$ and $L_{3,x}(\tau_1, \tau_3) = 0$, for all τ_1 and τ_3, in agreement with the so-called 'minimal' triplet STDP rule [54] (Fig 1D). Nevertheless, if spike triplets would also be taken into account for depression, the derivation would be identical, with the corresponding modification to the variables involved. We can rewrite Eq (1) in the Fourier domain as

$$\begin{aligned}\langle \dot{W}_{ij}^{\mathrm{STDP}} \rangle &= \underbrace{r_i r_j \left(\tilde{L}_2(0) + r_i \tilde{L}_3(0,0) \right)}_{\text{Independent spikes}} \\ &+ \underbrace{\int_{-\infty}^{\infty} \left[\tilde{C}_{ij}(\omega_1)\tilde{L}_2(-\omega_1)\delta(\omega_2) + r_i(\tilde{C}_{ij}(\omega_1)\delta(\omega_2) + \tilde{C}_{ij}(\omega_2)\delta(\omega_1+\omega_2))\tilde{L}_3(-\omega_1,-\omega_2) \right] \mathrm{d}\omega_1 \, \mathrm{d}\omega_2}_{\text{Pre-post pairwise correlations } (*)} \\ &+ \underbrace{r_j \int_{-\infty}^{\infty} \tilde{C}_{ii}(\omega_2)\tilde{L}_3(-\omega_1,-\omega_2)\delta(\omega_1) d\omega_1 d\omega_2}_{\text{Post-post pairwise correlations } (**)} + \underbrace{\int_{-\infty}^{\infty} \tilde{K}_{ij}(\omega_1,\omega_2)\tilde{L}_3(-\omega_1,-\omega_2) \mathrm{d}\omega_1 \, \mathrm{d}\omega_2}_{\text{Post-pre-post triplet correlations } (***)} \end{aligned} \quad (2)$$

where we use the notation $\tilde{f}$ for the Fourier transform of a function f and δ is the Dirac delta function. It should be noted that Eq 2 is not the Fourier transform of Eq 1 but rather an equivalent expression of the latter. This comes about because we can express the integral of the product of two functions as the convolution of the Fourier transform of those functions, evaluated at zero.

This formulation of the previous equation allows us to clearly break down the contribution of spike interactions of different orders to the average synaptic efficacy in the recurrent network. The first term of Eq 2 considers the change in synaptic efficacy that is obtained from independent spiking and thus depends on the first-order cumulant (the mean firing rates) of the activity of both the pre- and postsynaptic neurons, r_j and r_i, respectively. As firing rates increase, 'chance' contributions to plasticity can occur. The second and third term account for the probability of observing changes to the mean synaptic efficacy due to pairwise correlations in the pre- and postsynaptic neurons. C_{ij} refers to the family of probabilities that generate pairwise cross-correlations (second-order cumulant) between neurons i and j, depending on spikes of other neurons in the network (Fig 1B, green). Accordingly, C_{ii} includes the family of probabilities that generate pairwise auto-correlations in the same neuron i due to the spiking activity of all other neurons in the network (Fig 1B, orange). Therefore, the second (*) and third (**) terms describe the total contribution of correlated spike pairs to plasticity through the pair-based STDP rule L_2 (Fig 1C, left) and the triplet STDP rule L_3 (Fig 1C, right). In the case of the latter, the first-order cumulant of the uncorrelated single postsynaptic neuron's spikes, r_i, is also included in the second term (*) and the first-order cumulant of the uncorrelated single presynaptic neuron's spikes, r_j, in the third term (**). The fourth term (***) describes the total contribution of correlated spike triplets (third-order cumulant) to plasticity. Thus, K_{ij} includes the family of probabilities for third-order correlations, where the relative spike timing interacts with the triplet STDP learning window L_3 to induce plasticity (Fig 1B, purple and Fig 1C, right).

In conclusion, we have derived an exact analytical expression for the average change in synaptic efficacy due to firing rates, pairwise and triplet correlations under a general STDP rule that includes pairwise and triplet spike interactions. The resulting cumulants of up to third order can depend in non-trivial ways on the full recurrent connectivity in the network.

Novel structural motifs emerge under triplet STDP compared to pair-based STDP

The calculation of the cumulants involved in the equation for the average weight dynamics (Eq 2) depends on the full network connectivity. Therefore, the second- and third-order cumulants in Eq 2 can be written as a sum over contributions from different structural motifs, following the convention of [6]. These structural motifs determine all possible connectivity paths that a given spike from a source neuron k travels to the postsynaptic neuron i or presynaptic neuron j, and as a consequence affects the synaptic weight W_{ij}. Thus, to calculate each term in Eq 2 we break down the second- and third-order cumulants C_{ij}, C_{ii} and K_{ij} into expressions that include the contribution of every spike propagated in the network through existing synaptic connections, taking into account the full recurrence in the network (Methods). These expressions consist of products of the corresponding synaptic efficacies along the two paths to the presynaptic and postsynaptic neuron, the firing rate of the source neuron and the motif coefficient functions M, which depend on the number of synapses along the two paths, the EPSC function, E, and the STDP learning rules, L_2 and L_3. The probability that neurons i and j jointly fire a spike is transiently modulated whenever a neuron anywhere in the network produces a spike. We can write the pairwise cross-covariance from Eq 2 as

$$(*) \equiv \sum_{k=1}^{N} r_k \sum_{\alpha,\beta} (\boldsymbol{W}^{\alpha})_{ik} (\boldsymbol{W}^{\beta})_{jk} (M^{pair}_{\alpha,\beta} + r_i M^{trip}_{\alpha,\beta}) \tag{3}$$

which combines the contribution of structural motifs from the pair-based and triplet STDP rules to a change in the connectivity matrix $\boldsymbol{W}$. The expression consists of sums over two

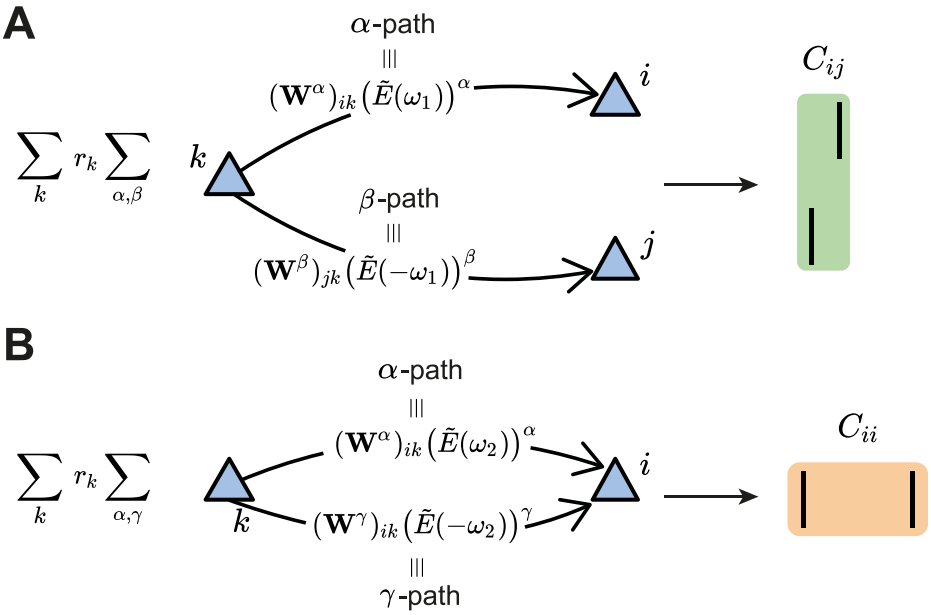


Fig 2. Second-order cumulant contributions to plasticity. A. The cross-covariance C_{ij} between the presynaptic neuron j and the postsynaptic neuron i is obtained by summing over all the possible α- and β-paths from every possible source neuron k in the network. Each path is calculated via the corresponding weights in the connectivity matrix and the EPSC function (see Eq 3). **B**. Same as **A** but for the auto-covariance C_{ii} of the postsynaptic neuron i (see Eq 4). In this case, γ is the second index to sum over the path from the source neuron k to the postsynaptic neuron i. It should be noted that the main difference between the α- and γ-path is given by the time dependence of the EPSC function, here represented in the Fourier domain for convenience.

https://doi.org/10.1371/journal.pcbi.1007835.g002

aspects to provide an intuitive description of the contribution of the pairwise cross-covariance C_{ij} between neurons i and j to plasticity of the connection W_{ij}. The first sum takes into account all spiking neurons in the network, while the second sum takes into account all possible 'paths' by which spikes originating from a 'source' neuron k affect the cross-covariance C_{ij}. Specifically, α and β constitute the 'path lengths' of synapses from source neuron k to the postsynaptic neuron i and the presynaptic neuron j, respectively (Fig 2A; see also [6]). We refer to the total path length of a motif, $\alpha + \beta$, as the 'order' of the motif.

The contribution of the pair-based STDP rule includes the motif coefficient functions, $M^{pair}_{\alpha,\beta}$, which are calculated in the Fourier domain (Eq 29 in Methods). The pairwise correlations between i and j also contribute to plasticity of W_{ij} based on the triplet STDP rule through the motif coefficient functions $M^{trip}_{\alpha,\beta}$ (Eq 30 in Methods). Examples of some motifs common for both the pair-based and the triplet STDP rule are provided in Fig 3A. Their contribution to plasticity through the EPSC function E and the STDP rules L_2 and L_3 is illustrated in Fig 3B.

In addition to the α and β path lengths, to derive the contribution of the triplet STDP rule to the average change in synaptic efficacy, we also introduced the γ-path so that now motifs have order $\alpha + \beta + \gamma$. γ is the synapse path length from the source neuron k to the postsynaptic neuron i, including a time delay relative to the α path from k to i, to account for the second postsynaptic spike of the triplet (Fig 3C and 3D). Thus, for the auto-covariance term in Eq 2, we obtain (Fig 2B)

$$(**) \equiv \sum_{k=1}^{N} r_k \sum_{\alpha,\gamma} (\boldsymbol{W}^{\alpha})_{ik} (\boldsymbol{W}^{\gamma})_{ik} r_j M^{trip}_{\alpha,\gamma} \tag{4}$$

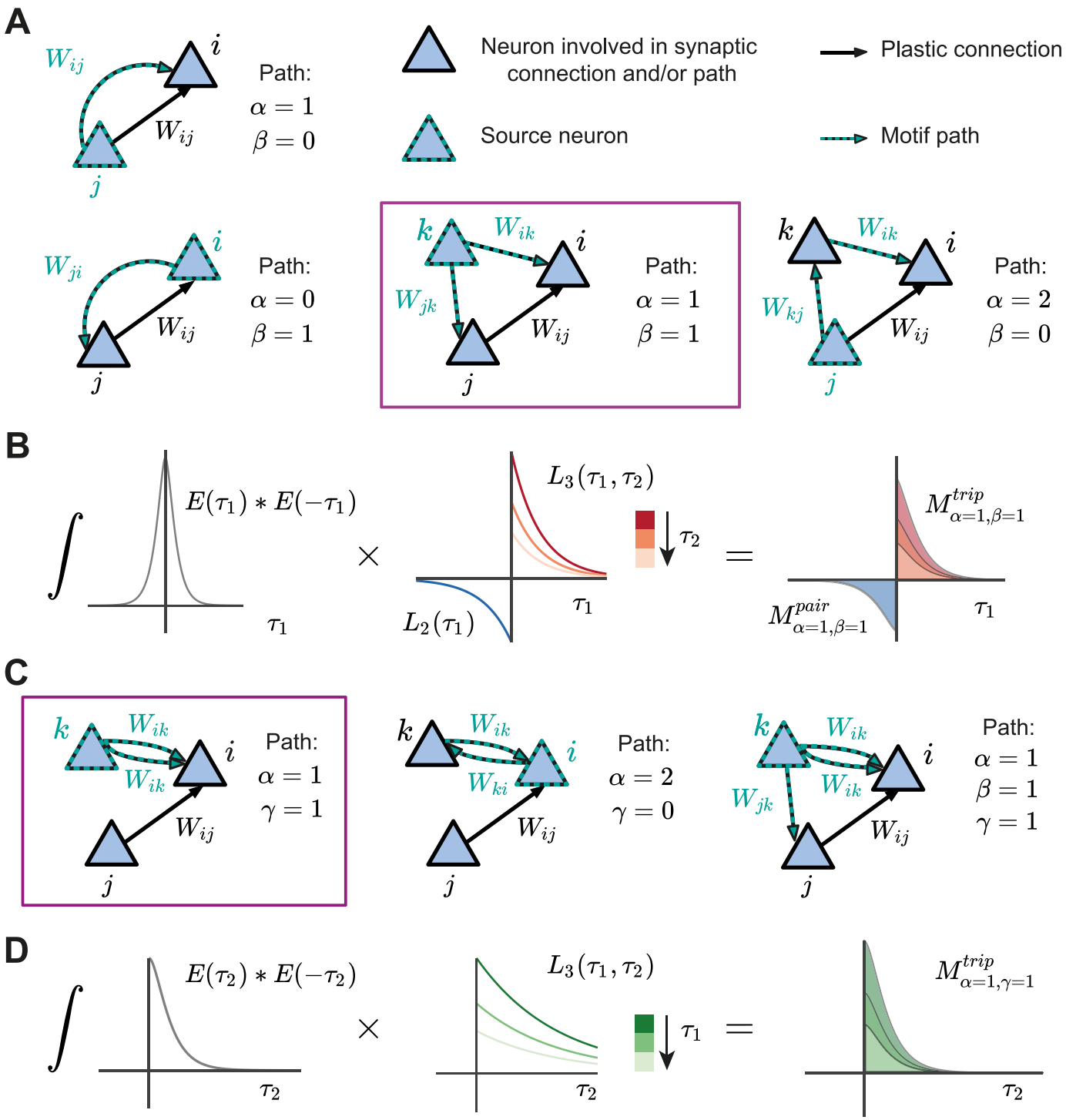


Fig 3. Structural motifs in the network under pair-based and triplet STDP. A. Examples of structural motifs common for both the pair-based and triplet STDP framework. Here α and β constitute the path lengths of synapses from the source neuron to the postsynaptic neuron i and the presynaptic neuron j. $\alpha = 1, \beta = 0$: Presynaptic neuron j projects to the postsynaptic neuron i. $\alpha = 0, \beta = 1$: Postsynaptic neuron i projects to the presynaptic neuron j. $\alpha = 1, \beta = 1$: Common input from source neuron k to presynaptic neuron j and postsynaptic neuron i. $\alpha = 2, \beta = 0$: Presynaptic neuron j projects to the postsynaptic neuron i through another neuron k in the network. **B.** Illustration of the calculation of the common input motif with $\alpha = 1$ and $\beta = 1$ framed in purple in **A** (there are also additional terms which are not illustrated). The motif coefficients $M_{\alpha=1,\beta=1}$ (right) are calculated as the total area under the curve resulting from the product of the convolution of the EPSC function E (left) and the STDP functions (pair-based L_2 and triplet L_3, middle). **C.** Examples of structural motifs found only in the triplet STDP framework, where γ denotes the

time-delayed path length from the source neuron to the postsynaptic neuron i. $\alpha = 1, \gamma = 1$: Source neuron k projects twice to postsynaptic neuron i with a different time delay. $\alpha = 2, \gamma = 0$: Feedback loop through another neuron k in the network (source and projecting neuron are the postsynaptic neuron i). $\alpha = 1, \beta = 1, \gamma = 1$: Source neuron k projects to the presynaptic neuron j and postsynaptic neuron i via all the three possible paths. **D**. Illustration of the calculation of the motif with $\alpha = 1$ and $\gamma = 1$ for the triplet STDP rule framed in purple in **C**, compare to **B**.

https://doi.org/10.1371/journal.pcbi.1007835.g003

where the motif coefficient function involving the triplet STDP rule is given in the Methods (Eq 31).

For third-order interactions, however, it is possible that the paths by which spikes are propagated branch out from a neuron other than the source neuron. Therefore, the third-order cumulant K_{ij} (Eq 2) is broken down into four sums:

$$
\begin{aligned}
(***) \equiv \sum_{k=1}^{N} r_k \sum_{\alpha,\beta,\gamma} & \overbrace{(\boldsymbol{W}^{\alpha})_{ik}(\boldsymbol{W}^{\gamma})_{ik}(\boldsymbol{W}^{\beta})_{jk} M^{trip}_{\alpha,\beta,\gamma}}^{\text{No branching (straight paths)}} \\
& \left.\begin{aligned}
&+\sum_{k,l=1}^{N} r_k \sum_{\alpha,\beta,\gamma,\zeta} (\boldsymbol{W}^{\zeta})_{lk}(\boldsymbol{W}^{\alpha})_{ik}(\boldsymbol{W}^{\beta})_{jl}(\boldsymbol{W}^{\gamma})_{il} M^{trip}_{(\alpha,\zeta),\beta,\gamma} \\
&+\sum_{k,l=1}^{N} r_k \sum_{\alpha,\beta,\gamma,\zeta} (\boldsymbol{W}^{\zeta})_{lk}(\boldsymbol{W}^{\alpha})_{il}(\boldsymbol{W}^{\beta})_{jk}(\boldsymbol{W}^{\gamma})_{il} M^{trip}_{\alpha,(\beta,\zeta),\gamma} \\
&+\sum_{k,l=1}^{N} r_k \sum_{\alpha,\beta,\gamma,\zeta} (\boldsymbol{W}^{\zeta})_{lk}(\boldsymbol{W}^{\alpha})_{il}(\boldsymbol{W}^{\beta})_{jl}(\boldsymbol{W}^{\gamma})_{ik} M^{trip}_{\alpha,\beta,(\gamma,\zeta).}
\end{aligned}\right\} \text{Branched paths}
\end{aligned}
\tag{5}
$$

The first term in Eq 5 sums over the paths to the presynaptic neuron j and postsynaptic neuron i from a source neuron k in the network that do not branch out. In other words, it considers that the 'distance' to each respective spike of the triplet is given by α, β and γ (Fig 4A). The remaining terms include the sum over possible branches in the network 'tree': $\zeta \geq 1$ is the synapse path length from the source neuron k to the neuron l that is the branching point (Fig 4B–4D). It should be noted that the branched paths all have a total path length of at least four (i.e. $\alpha + \beta + \gamma + \zeta \geq 4$) so that the motif order is at minimum four, since at least one synapse must be taken into account before the splitting of the path. The corresponding motif coefficients for the 'straight' triplet motif (Fig 4A, see Eq 32), and for the 'branching' motifs (Fig 4B–4D, see Eqs 33, 34 and 35) are provided in the Methods.

This analysis reveals novel motifs in the triplet STDP rule which have the potential to promote particular connectivity structures that are not possible with pair-based STDP [6] (Fig 3C). These include motifs which directly exclude the presynaptic neuron j but can still impact the synaptic weight, W_{ij} (Fig 3C, left and middle). This can be achieved, for example, through an additional neuron k that does not directly affect the weight W_{ij} but projects to the postsynaptic neuron i through the synaptic weight W_{ik} (Fig 3C, left and middle). Because these motifs exclude the presynaptic neuron j, they do not impact the pairwise cross-covariance term C_{ij} and do not have influence on the weight W_{ij} through pair-based STDP. For example, in the case when $\alpha = 2$ and $\gamma = 0$ (Fig 3C, middle), the postsynaptic neuron i is both the source neuron and the neuron involved in the path with the additional neuron k. We call this path involving the synaptic efficacies W_{ik} and W_{ki} a 'loop'. These loops involve a different neuron in addition to the pre- and postsynaptic neuron of the weight W_{ij}, and are a unique feature of incorporating spike triplets in the STDP rule. Loops include a neuron as both the source and target for the spike in the corresponding path, so that a 'loop' closes on itself. The direction

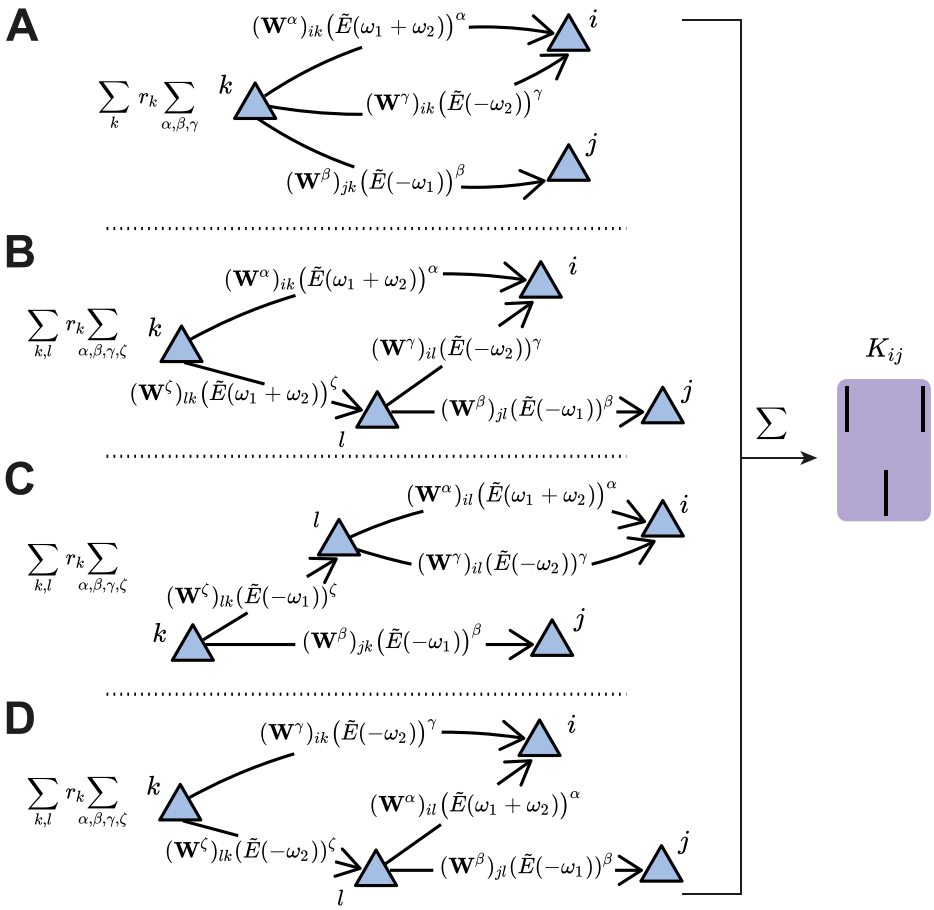


Fig 4. Third-order cumulant contributions to plasticity can be broken down into four terms. A. The first term contains all the α-, β- and γ-paths originating from the source neuron k to the spiking neurons i and j. **B-D**. The other terms take into account the possibility of an intermediate neuron l that acts as a new source neuron for two of the paths. These are referred to as 'branched paths', and the path length from the source neuron k to the intermediate neuron l is denoted with ζ. The branching describes the individual terms in Eq 5.

https://doi.org/10.1371/journal.pcbi.1007835.g004

of the edges are relevant for this definition. We propose that motifs with these 'loop' characteristics promote the formation of connections between clusters of neurons, and therefore assemblies.

To illustrate motifs of different orders and their relationship to cumulants of different orders, we depict all motifs up to third order arising from the expansion of the second- and the third-order cumulants (Fig 5). While it is clear that the full network connectivity through motifs of different orders from the cross-covariance C_{ij} influences plasticity under pair-based, as well as triplet STDP (Fig 5, first row), we also reveal novel motifs from the auto-covariance C_{ii} and the third-order cumulant K_{ij} that influence plasticity uniquely under triplet STDP (Fig 5, second and third row).

Taken together, our motif expansion framework reveals novel structural motifs under the triplet STDP rule that have the potential to form assemblies without structured external input. We next investigated the role of the different structural motifs (specifically the 'loop' motifs) on the emergence of assemblies under triplet STDP.

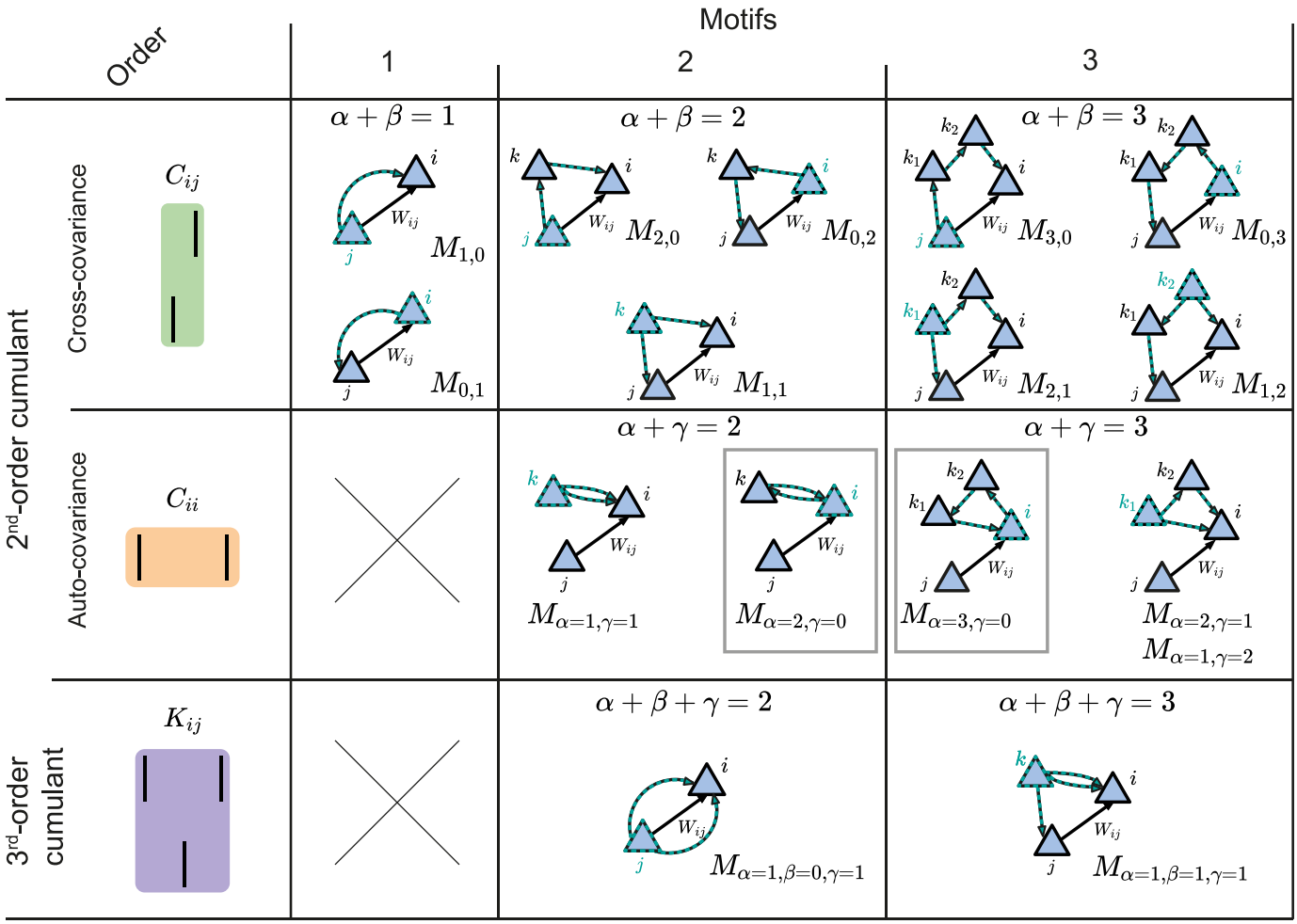


Fig 5. Second- and third-order cumulants can be described in terms of structural motifs that contribute to weight change. All motifs up to third order as they arise from the cross-covariance C_{ij} (top row and Eq 3), the auto-covariance C_{ii} (middle row and Eq 4) (both C_{ij} and C_{ii} together represent the second-order cumulant) or the third-order cumulant K_{ij} (bottom row and Eq 5). The gray boxes indicate the 'loop' motifs. The novel motifs which follow from the triplet STDP rule are those that include the path γ (second and third row).

https://doi.org/10.1371/journal.pcbi.1007835.g005

Modulation of the triplet STDP rule promotes the autonomous emergence of assemblies

So far, we considered general STDP rules that depend on the precise timing between pairs and triplets of spikes, without taking into account the exact dependence of potentiation or depression on these spikes. To further study the complex relationship between plasticity and network correlations, we considered a particular biologically identified STDP rule that relies on third-order interactions (Methods; Fig 1C). This rule has an asymmetric shape around the time lag of 0 (where pre- and postsynaptic spikes are coincident), similar to the classical pair-based STDP rule [71]. However, while synaptic depression is induced by the relative timing of pairs of presynaptic and postsynaptic spikes, the minimal triplet STDP model uses sets of three spikes to induce potentiation: the amount depends on the timing between pre- and postsynaptic spike pairs and in addition, on the timing between the current and the previous postsynaptic spike (Fig 1D). This *minimal* model successfully captures experimental data, where the

pairing frequency of pre- and postsynaptic spikes was varied, equally well compared to a full model that also uses triplets of spikes for depression [54].

Implementations of classical Hebbian learning, such as STDP, use joint pre- and postsynaptic activity to induce potentiation and depression, while neglecting other potential factors such as heterosynaptic plasticity [81], or the location of synaptic inputs on the dendritic tree [82]. However, recent experimental studies have highlighted an important role of neuromodulators in regulating plasticity across the brain [83–86], as they convey information about novelty or reward. Indeed, neuromodulators such as dopamine, acetylcholine and noradrenaline, but also brain-derived neurotrophic factor (BDNF) and gamma-aminobutyric acid (GABA), can predominantly act via two mechanisms: by reshaping the learning window for STDP or by regulating neuronal activity at the level of synaptic transmission [84, 86]. Therefore, we next investigated how neuromodulation of synaptic plasticity affects recurrently connected networks considering that pairwise and triplet spike interactions determine plasticity. We assume that the shape of the STDP function can be modulated via the modulation parameter η_- which preserves the overall level of depression by trading off the depression learning rate A_- and the depression time constant τ_- (Methods; Fig 6A). Such a modification of the learning rule has been observed in the lateral amygdala due to the action of dopamine via D2 receptors [85, 87], or in rat visual cortex slices with the activation of both the noradrenaline pathway through β-adrenergic receptors and the acetylcholine pathway through M_1-muscarinic receptors [84, 86, 88]. A similar modulation parameter could similarly be included for potentiation.

To determine contributions to plasticity arising due to internal network correlations and not just differences in neuronal firing rates [5], we consider the case in which the plasticity rule is balanced, such that $\tilde{L}_2(0) + r_i\tilde{L}_3(0,0) = 0$. We use this condition to calculate all motif coefficients, $M_{\alpha,\beta}$, that arise from the cross-covariance C_{ij} (Eqs 47–56 in Methods). We consider only motifs up to third-order in the evolution of the weights (Eq 2) since higher-than-third-order motif contributions are negligible (S1 Fig). Thus, we no longer include the branched path motifs of Eq 5 as they are higher-than-third-order motifs (Fig 4B–4D). This leaves us with a handful of motifs which arise from the second-order cumulant (consisting of the cross-covariance C_{ij} and the auto-covariance C_{ii}) and the third-order cumulant K_{ij} (Methods; Fig 5). This simplification allows us to study the spontaneous emergence of assemblies under the triplet STDP rule based on both the triplet rule contributions to the cross-covariance C_{ij} (Eq 3, Fig 5, top row) and the influence of the novel branching structures that follow from the auto-covariance C_{ii} (Eq 4, Fig 5, second row, including the loop motifs in the gray boxes) and the third-order cumulant K_{ij} (Eq 5, Fig 5, third row).

To systematically study how the dependence of these up to third-order motif coefficients on the shape of the STDP rule affects connectivity structure in the network, we visualized the connectivity matrices obtained by integrating the motif expansion up to third-order (Eqs 42–46) numerically, using experimentally-fitted parameters for the triplet STDP rule and the EPSC function (Table 1). Specifically, we investigated the emergence of global network structures, or assemblies, as a function of the modulation parameter η_-. This parameter has a direct influence on the motifs which follow from the cross-covariance C_{ij} (Fig 5) and the LTD window in the minimal triplet STDP rule (see Eqs 49–58 and S2 Text). A key requirement for the emergence of assemblies is the formation of bidirectional or reciprocal connections among groups of neurons. We compare the reciprocal connections of the first-order motif contributions to gain intuition:

$$\begin{aligned}\langle \dot{W}_{ij}^{\mathrm{STDP}} \rangle^{(1)} &= r_j W_{ij} M_{1,0} + r_i W_{ji} M_{0,1},\\ \langle \dot{W}_{ji}^{\mathrm{STDP}} \rangle^{(1)} &= r_i W_{ji} M_{1,0} + r_j W_{ij} M_{0,1}.\end{aligned} \tag{6}$$

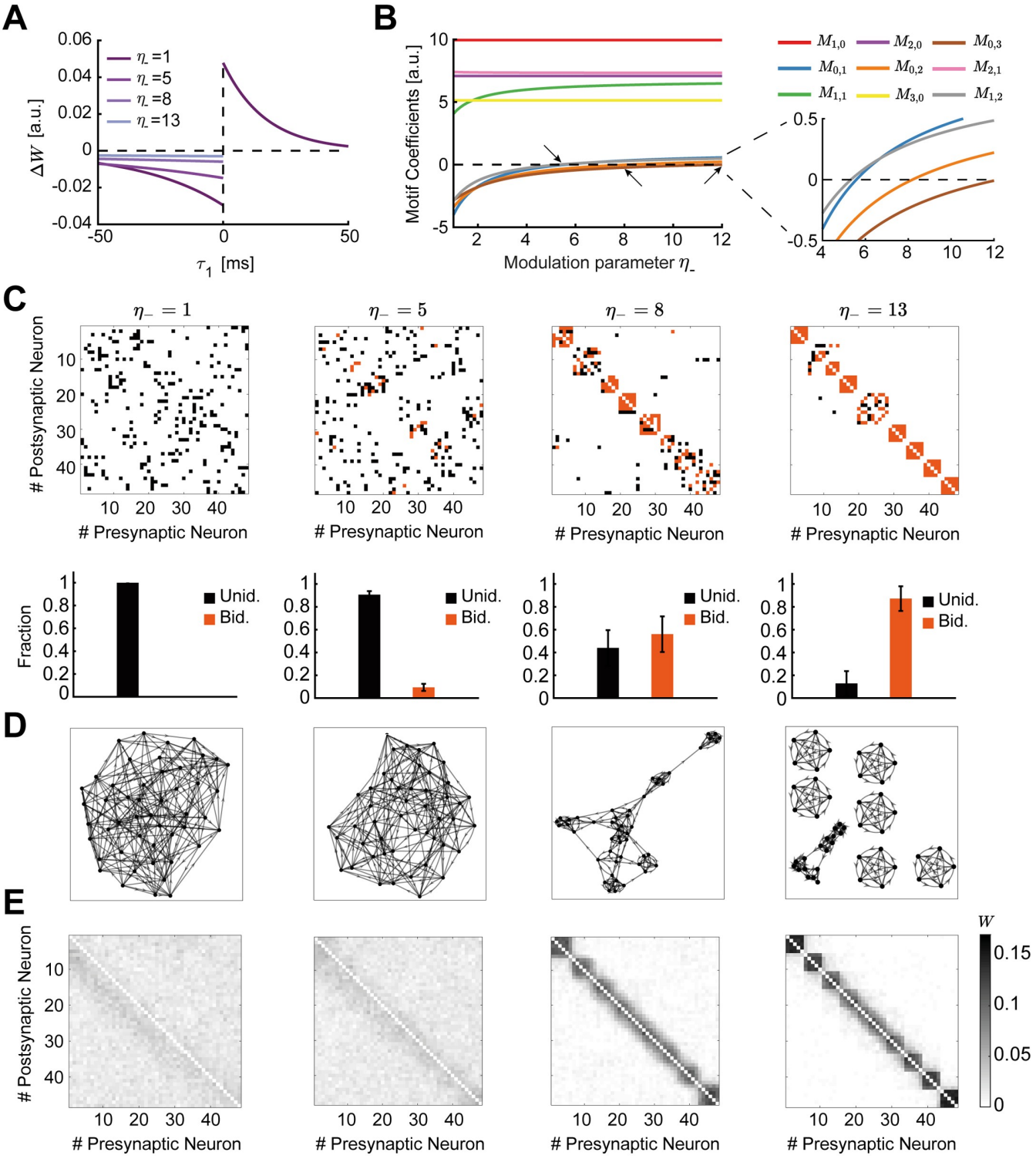


Fig 6. Spontaneous emergence of assemblies via modulation of the triplet STDP rule. A. The shape of the STDP function changes as a function of the modulation parameter η_-, which preserves the overall level of depression by trading off the depression learning rate and the depression time constant. **B**. Motif coefficients as the modulation parameter η_- increases. Points of interest given by the crossovers of the strength of particular motifs are indicated by a small arrow. Inset: Amplified scale around zero. Motif coefficients including γ paths are not illustrated, since they are always constant and positive in η_- and do not provide a meaningful comparison to the other motifs. **C**. Top: Examples of connectivity matrices obtained with different values of η_- at steady state. Unidirectional

connections are shown in black, bidirectional connections in orange. Matrices are reordered using the k-means clustering algorithm (see Methods). Bottom: Mean fraction of unidirectional and bidirectional connections for 100 trials with different initial synaptic efficacies as a function of η_-. Error bars represent the standard error of the mean. **D**. Graphs of the connectivity matrices in C. **E**. Averaged connectivity matrices over 100 trials at steady state. Note that the tighter clusters emerging near the edges of the matrices are the result of the clustering algorithm but do not affect the quantification of connectivity.

https://doi.org/10.1371/journal.pcbi.1007835.g006

Since in the triplet STDP rule $M_{1,0} > 0$ (Fig 6B, red), the two bidirectional connections compete if $M_{0,1} < 0$, and potentiate if $M_{0,1} > 0$. Therefore, the sign of the motif coefficient $M_{0,1}$, which depends on the contribution from the triplet STDP rule, determines the formation of bidirectional connections. Indeed, increasing η_- supports the formation of bidirectional connections (Fig 6C) as the motif coefficient $M_{0,1}$ changes sign (Fig 6B, blue, see inset). In contrast, as previously shown, the classical pair-based STDP rule is unable to support the formation of assemblies and bidirectional connections due to its asymmetric shape [89, 90], although under certain conditions (dominant potentiation) it can promote bidirectional connections [51, 91]. Under the asymmetric pair-based STDP rule, $M_{1,0}>$ and $M_{0,1} < 0$ result in competition between the two reciprocal connections. To autonomously generate self-connected assemblies without structured network input requires a symmetric pair-based STDP rule (which is not biologically motivated) and a sufficiently large synaptic latency [6]. In this case, the prominence of the common input motif driven by the $M_{1,1}$ motif coefficient over all other motif coefficients in the network supports assembly formation [6].

Under the triplet STDP rule, small increases in η_- increase the motif coefficient $M_{1,1}$, resulting in the formation of bidirectional connections and assemblies, similarly to the symmetric pair-based STDP rule. However, despite its asymmetric shape, the triplet STDP rule can robustly generate bidirectional connections and assemblies even when the $M_{1,1}$ motif coefficient has already saturated and other motif coefficients dominate (Fig 6C–6E), upon further increases in η_-. This is because higher-order structural motifs also contribute to the formation of bidirectional connections and assemblies. To understand this, we consider the motif contributions of feedforward motif coefficients—the motifs for which the α-path is longer than the β-path, $M_{1,0}$, $M_{2,1}$, $M_{2,0}$ and $M_{3,0}$—and reciprocal motif coefficients, where the β-path is longer than the α-path, $M_{0,1}$, $M_{1,2}$, $M_{0,2}$ and $M_{0,3}$. Given the asymmetry of triplet STDP, the feedforward motif coefficients are stronger. The reciprocal motifs, $M_{0,1}$, $M_{1,2}$, $M_{0,2}$ and $M_{0,3}$ play an important role in the formation of bidirectional connections as they change sign from negative to positive with increasing η_- (Fig 6B). A positive contribution from all motifs supports the

Table 1. Parameter values for figures. ⋆ denotes that values are provided in the figures.

Symbol	Description	Fig 6	Fig 7	Fig 8	Fig 9	Fig 10	Fig 11
N	Number of neurons	48					
$\boldsymbol{\mu}$	External input firing rate	150 Hz					
w_{max}	Upper bound for each individual weight	0.17					
W_{max}	Upper bound for total row/ column weight	0.85					
A_-	Depression learning rate	0.01					
τ_-	Depression time constant	33.7 ms [54]					
τ_+	Potentiation time constant	16.8 ms [54]					
τ_y	Second potentiation time constant	114 ms [54]					
η_-	Depression modulation parameter	⋆			1		⋆
τ_ε	First membrane time constant	5 ms					
τ_ι	Second membrane time constant	5 ms			⋆	5 ms	
ν	Scaling parameter of learning rate	3.5×10^{-4}					
ψ	Heterosynaptic competition scaling parameter	0.7					

https://doi.org/10.1371/journal.pcbi.1007835.t001

robust formation of bidirectional connections in the network as the competition between reciprocal connections decreases. Together with the strong common input motif driven by the $M_{1,1}$ motif coefficient, this leads to the robust emergence of assemblies. In this scenario, η_- controls the competition between feedforward (W_{ji}) and reciprocal connections (W_{ij}), with large η_- enabling the potentiation of both. This is not possible under the classical asymmetric pair-based STDP rule as previously discussed.

In summary, we find that the spontaneous formation of self-connected assemblies depends on the modulation parameter η_-, which influences most of the motifs arising from the cross-covariance C_{ij}. Furthermore, self-connected assemblies can be formed under triplet STDP even when motifs other than the common input motif $M_{1,1}$ dominate. This occurs despite the asymmetric shape of the triplet STDP rule, in contrast to pair-based STDP which requires a symmetric shape to promote $M_{1,1}$. Importantly, the dependence of assembly formation on the specific form of the STDP window points towards an important role of neuromodulatory signals on formation of intrinsically generated assemblies.

Characterizing emergent assembly structures

To determine the conditions on the learning rule and EPSC properties for the emergence of self-connected assemblies, it is convenient to represent the functional organization of the network given a connectivity matrix as a mathematical graph. In our context, graphs are composed of a set of nodes or neurons with pairs of them joined by edges or synaptic efficacies. The resulting graphs can be described by standard metrics, whose dependence on the shape of the learning rule and the EPSC function might yield insight into the emergent structures during circuit organization driven by spontaneous activity. We focused on common quantities for describing graph structure, including the clustering coefficient, the global efficiency and the modularity [92, 93], used previously in experimental systems like the zebrafish tectum and the mammalian cortex [25, 94].

The clustering coefficient quantifies the existence of densely interconnected groups of nodes in the graph [95]. It represents a measure of segregation, based on counting the number of connection triangles around a node (Methods). In this manner, it reflects the prevalence of clustered connectivity around individual nodes by calculating the fraction of neighbors of that particular node that are also neighbors of each other. As a result, the mean clustering coefficient of a network determines the prevalence of three-neuron-clusters in the network architecture. We find that as the modulation parameter η_- increases, the mean clustering coefficient also increases until it reaches a plateau (Fig 7A). Ensuring that the motif coefficients $M_{0,1}$ and $M_{1,2}$ are positive is sufficient for the formation of clusters beyond the critical value of $\eta_- \approx 5$ (Fig 6B and 6C), where the clustering coefficient begins to increase (Fig 7A). The value of η_- at which the clustering coefficient saturates corresponds to the emergence of more robust assemblies where all the motif coefficients are positive (Fig 6B and 6C). Although strong bidirectional connections are localized within clusters, connections from one cluster to some others still exist globally. This is different to the clustering enabled by strong symmetric interactions in which the motif $M_{1,1}$ dominates, considered previously by a symmetric pair-based STDP rule [6], where the clusters would be unconnected (i.e. isolated from each other) and the clustering coefficient would be much higher.

Complementary to the clustering coefficient, the global efficiency is a measure of functional integration, which determines how easily nodes can communicate between each other through sequences of edges [96]. Consequently, the lengths of the paths estimate the potential for the flow of information between nodes, with shorter paths denoting stronger capacity for integration. Then, global efficiency is defined as the average inverse shortest path length of the

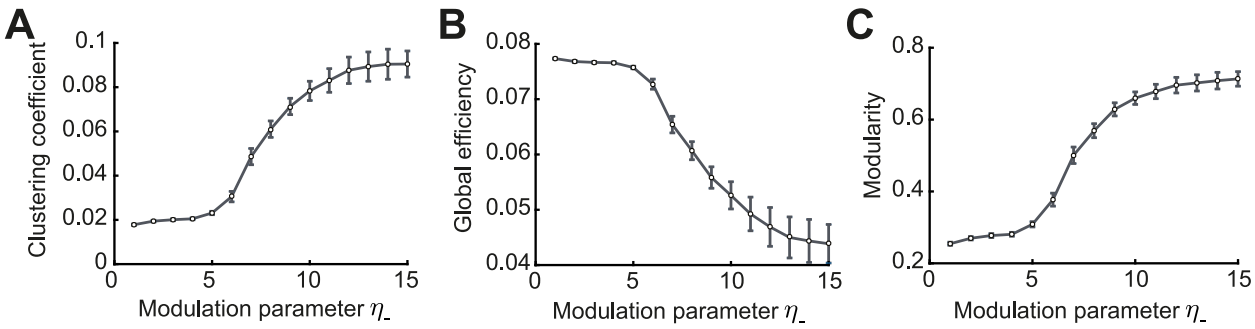


Fig 7. Graph measures of the stable network configuration. A. Mean clustering coefficient versus the modulation parameter η_-. **B**. Mean global efficiency versus the modulation parameter η_-. **C**. Mean modularity versus the modulation parameter η_-. All results are calculated from 100 trials at steady state connectivity. Error bars represent the standard error of the mean.

https://doi.org/10.1371/journal.pcbi.1007835.g007

network (Methods). In comparison to the clustering coefficient, this quantity initially remains approximately constant and then decreases until the point at which strong assemblies emerge autonomously since network structure no longer varies with the parameter η_- (Fig 7B). We find that as for the clustering coefficient, the value of η_- for which the motif coefficients $M_{0,1}$ and $M_{1,2}$ become positive ($\eta_- \approx 5$) constitutes a landmark for the formation of assemblies, after which global efficiency significantly decreases.

Finally, modularity is a graph theoretic measure that describes how strongly a network can be divided into modules, by comparing the relative strengths of connections within and outside modules to the case when the network has randomly chosen weights [93, 97, 98]. Recently, it was shown that even in models with rate-based dynamics, increasing modularity amplifies the recurrent excitation within assemblies evoking spontaneous activation [48]. With increasing η_-, modularity increases until strong assemblies are formed in a similar fashion as the clustering coefficient (Fig 7C). Interestingly, the critical value of $\eta_- \approx 5$ where assemblies begin to form robustly is consistent with experimental evidence of the shape of STDP where the time constant for depression has been found to be approximately 5 times longer than for potentiation [55, 99].

Contribution of the novel structural motifs under triplet STDP on assembly formation

So far we demonstrated that the spontaneous emergence of assemblies via modulation of triplet STDP depends on the interaction of different motifs that primarily arise from the second-order cross-covariance C_{ij} (Figs 5 and 6), which is also present under pair-based STDP. However, whether the novel structural motifs that are unique to triplet STDP (Figs 2B and 3–5) play a role remains unclear. We hypothesize that the 'loop' motifs, which do not appear for the pair-based STDP rule (Fig 5; gray box) are important for assembly formation.

To investigate the implications of these novel 'loop' motifs, we compare the three graph measures in four different scenarios: Using the motifs (1) only from the cross-covariance C_{ij} (Fig 5, top row); (2) from all cumulants (C_{ij}, C_{ii} and K_{ij}) without the 'loop' terms (Fig 5 all except the gray boxes in the second row); (3) from the cross-covariance plus the two additional 'loop' terms (Fig 5, top row plus the gray boxes in the second row); and (4) from all cumulants (Fig 5, all). We find that cases (1) and (2) have worse performance in all three graph measures compared to cases (3) and (4) (Fig 8). Adding the third-order cumulant K_{ij} and the 'non-loop' terms from the second-order auto-covariance C_{ii} (case 4) even worsens the graph measures. We find that the third-order cumulant K_{ij} alone has almost no influence on the spontaneous

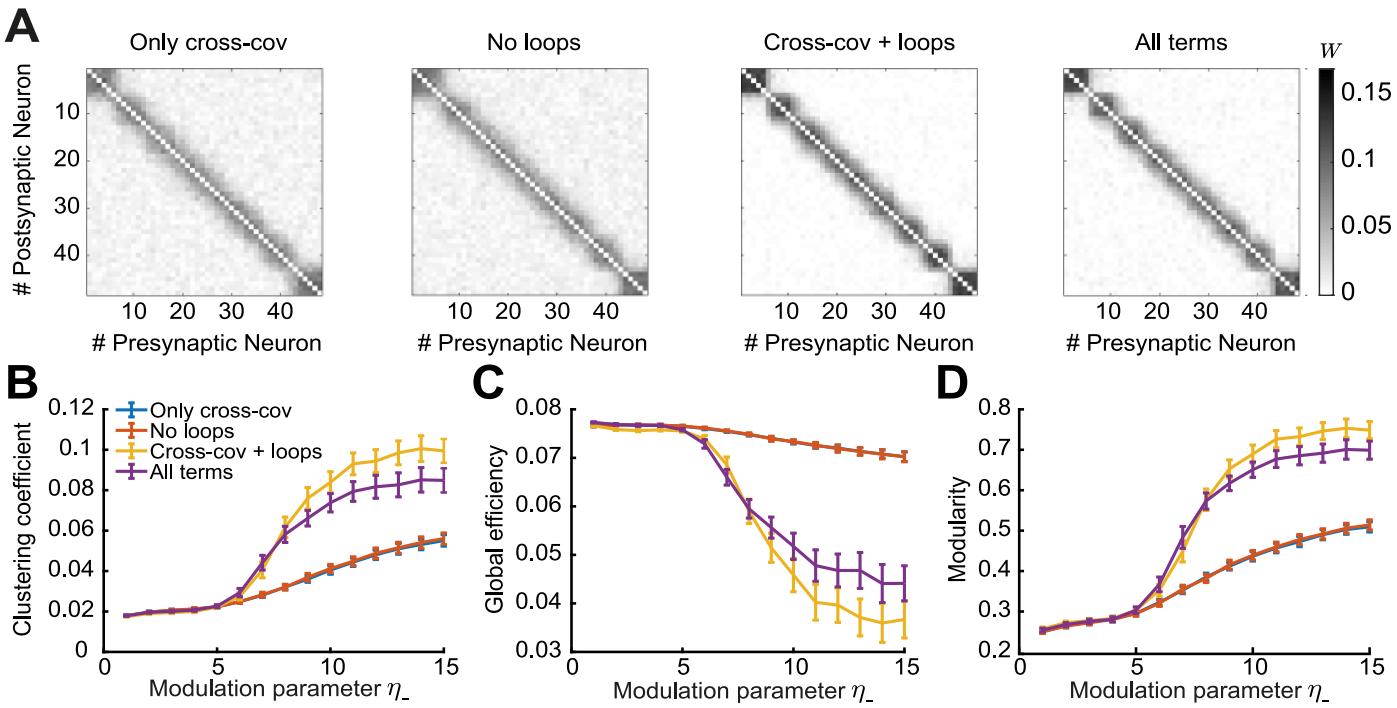


Fig 8. Spontaneous emergence of assemblies for four different motif combinations. Considering only motifs related to the cross-covariance C_{ij} (blue), from all cumulants (C_{ij}, C_{ii} and K_{ij}) without the 'loop' terms (red), from the cross-covariance C_{ij} plus the 'loop' terms (yellow) and from all cumulants (purple). **A**. Averaged connectivity matrices over 100 trials at steady state for four different motif combinations and modulation parameter $\eta_- = 13$. Matrices are reordered using the k-means clustering algorithm (see Methods). **B**. Mean clustering coefficient versus the modulation parameter η_-. **C**. Mean global efficiency versus the modulation parameter η_-. **D**. Mean modularity versus the modulation parameter η_-. All results are calculated from 100 trials at steady state connectivity. Error bars represent the standard error of the mean.

https://doi.org/10.1371/journal.pcbi.1007835.g008

emergence of assemblies (S2 Fig), since its contribution to the weight change is small, as shown before [100]. We conclude that the additional 'loop' terms, which arise as novel structural motifs from the triplet STDP rule (Fig 5), have a significant contribution to spontaneous assembly formation.

The triplet STDP rule and the EPSC together modulate network structure

The spontaneous emergence of assemblies discussed so far requires a relatively high value of the STDP modulation parameter η_-, raising the issue of biological plausibility. Although several experimental studies on induction of STDP indeed find longer depression than potentiation time constants [55, 99], we demonstrate an alternative mechanisms for the assembly formation by regulating the synaptic transmission of action potentials between neurons through the shape of the EPSC function. In this case, the strength of internally generated correlations can be changed independently of the STDP functions, L_2 and L_3. We investigated how the rise of the EPSC function modulated by delay of the spike transmission in the synapse, τ_t (Fig 9A), shapes motif coefficients (Methods; Fig 9B).

The parameter τ_t has a prominent impact on plasticity in the network. Even small shifts in the peak of the EPSC function by a few milliseconds have a strong impact on the cumulants of different orders, as reflected in the values of the motif coefficients (Fig 9A and 9B). Different to the modulation with the parameter η_-, the parameter τ_t affects all motif coefficients. However, the influence of τ_t on the auto-covariance C_{ii} and the third-order cumulant K_{ij} is negligible. Therefore, although the main cumulant driving plasticity is the second-order cross-covariance

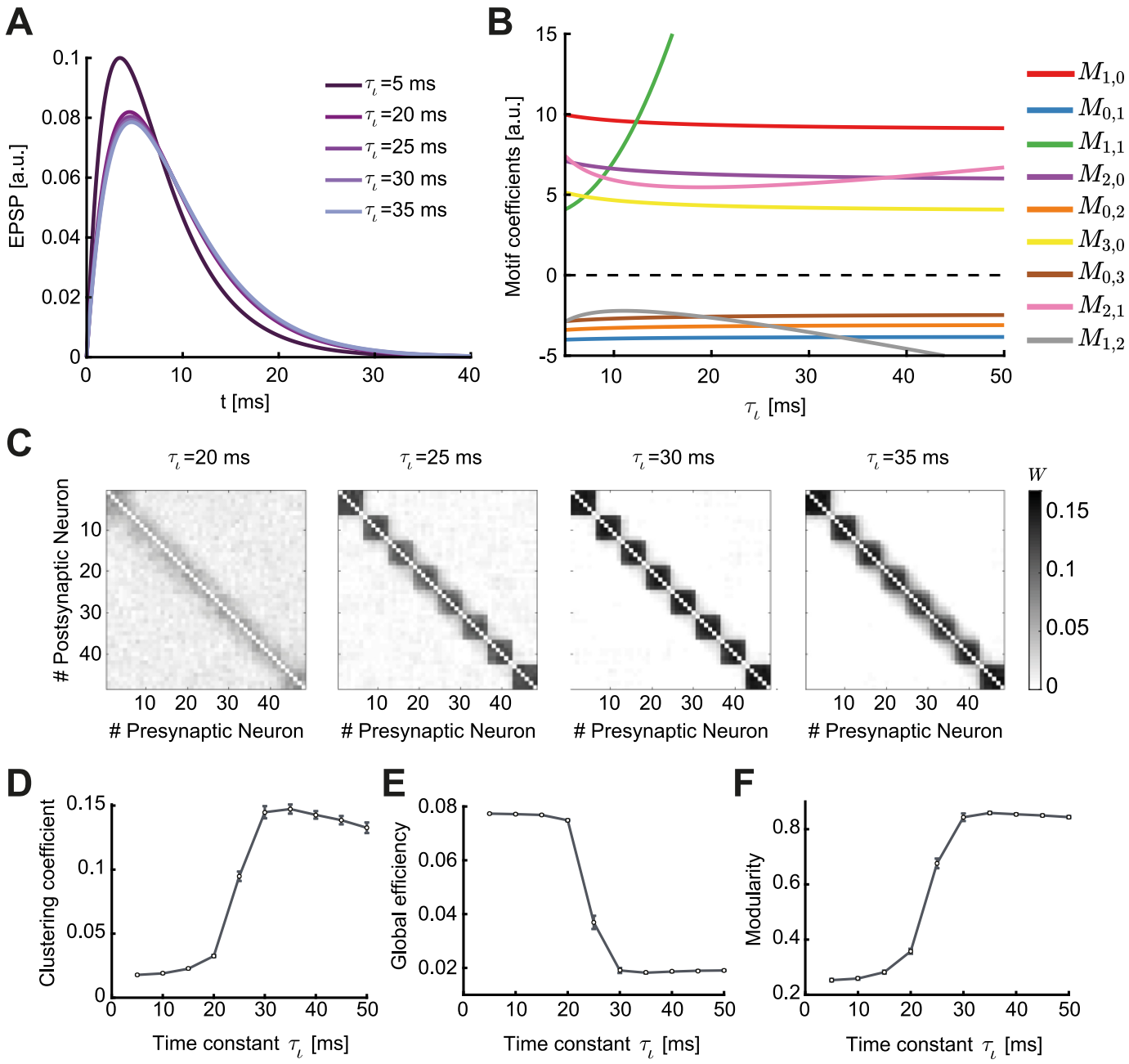


Fig 9. Spontaneous emergence of assemblies due to the modulation of synaptic transmission. A. Varying the time constant τ_ι changes the shape of the EPSC function, shifting its peak by a few milliseconds. **B**. Relative value of the motif coefficients as a function of τ_ι. While the common input motif $M_{1,1}$ rapidly assumes dominance, the motif coefficient $M_{1,2}$ crosses over in strength with the feedback motifs $M_{0,1}$, $M_{0,2}$ and $M_{0,3}$. **C**. Averaged connectivity matrices over 100 trials at steady state and different values of the time constant τ_ι. Matrices are reordered using the k-means clustering algorithm (see Methods). **D**. Mean clustering coefficient versus the time constant τ_ι. **E**. Mean global efficiency versus the time constant τ_ι. **F**. Mean modularity versus the the time constant τ_ι. All results are calculated from 100 trials at steady state connectivity. Error bars represent the standard error of the mean.

https://doi.org/10.1371/journal.pcbi.1007835.g009

C_{ij}, which exists even under pair-based STDP (Fig 5), assemblies easily form under the triplet STDP rule (Fig 9C). The common input motif $M_{1,1}$ abruptly assumes dominance over all others as τ_ι increases (Fig 9B). However, we observed that the reciprocal motif coefficients $M_{0,1}$, $M_{1,2}$, $M_{0,2}$ and $M_{0,3}$ remain negative for all values of τ_ι, in contrast to when we modulated the

STDP learning rule (Fig 6B). This tells us that assemblies in the network are spontaneously formed in a different fashion (by promoting the potentiation of reciprocal connections in each cluster due to the common input motif, $M_{1,1}$) than when modulating the STDP rule through η_-. In fact, assemblies emerge for minor modulations in τ_t (Fig 9C).

These differences in assembly formation become apparent when we consider the mean clustering coefficient, the global efficiency and the modularity as functions of τ_t (Fig 9D–9F): the three measures reflect the connectivity matrices as $M_{1,2}$ crosses the motif coefficients $M_{0,1}$, $M_{0,2}$ and $M_{0,3}$, in the case when $M_{1,1}$ is already large. When the motif coefficient $M_{1,2}$ becomes more negative than $M_{0,3}$ ($\tau_t \approx 20$ ms), bidirectional connections are strongly promoted and assemblies robustly form. Even for $\tau_t > 20$ ms, where the EPSC function does not change significantly (Fig 9A), one sees noticeable changes in the 'tightness' of the assemblies as observed in the averaged connectivity matrices (Fig 9C). Interestingly, as $M_{1,2}$ decreases below $M_{0,2}$ ($\tau_t \approx 25$ ms), the value of the clustering coefficient (≈ 0.1) and the modularity (≈ 0.7) correspond to the values where the clustering coefficient, the modularity, and the global efficiency saturate when modulating the STDP function (compare Figs 7 and 9D–9F). This means that the network structure is very similar (compare Fig 6E, right, with Fig 9C, second from left). Nevertheless, further increasing τ_t leads to more refined assemblies (Fig 9C, third from left) when $M_{1,2} < M_{0,2}$. However, for $\tau_t \gtrsim 35$ ms where $M_{1,2} < M_{0,1}$, the clustering coefficient slightly decreases (Fig 9D) suggesting the existence of optimal regions in the parameter space of τ_t to obtain the 'tightest' assemblies.

Taken together, our analytical framework enables us to interpret changes in the motif coefficients as changes in the connectivity structure in terms of the formation of self-organized assemblies. Modifying either the shape of the learning rule, or the shape of the EPSC function, can achieve this, however, with different consequences on the nature of the formed structures as demonstrated by the graph theoretic measures.

Comparison with assemblies generated via external correlated input

Until now, we sought to understand the mechanisms that contribute to the autonomous emergence of assemblies in neural circuits without any structured external input. Yet, the training of assemblies and plasticity of recurrent connections has been more frequently studied when these networks are driven by structured external input, both in simulations [49, 89] and analytically [42–45, 51]. Significant experimental evidence also exists for the emergence of functional connectivity underlying feature selectivity in the visual cortex around the time of eye opening, which is presumably influenced by structured visual input through the open eyes [14]. Therefore, we wanted to compare the formation of assemblies without structured external input under the triplet STDP rule to that with structured external input. To investigate spatiotemporal input patterns in our framework, we studied the overall mean impact of an external pairwise correlated input. This was implemented by assuming that the driving signal, which could for instance represent retinal input in the optic tectum or visual cortex, is correlated for a pair of neurons in the network, so that the structure of the input is represented as common input to that particular pair of neurons.

We write the covariance as a sum of the internal correlation and a novel term that conveys the external structured activity as common input [40]:

$$\tilde{\boldsymbol{C}}(\omega) = \tilde{\boldsymbol{C}}^{\text{int}}(\omega) + \left(\mathbb{I} - \tilde{E}(\omega)\boldsymbol{W}\right)^{-1} \tilde{E}(\omega)\tilde{\boldsymbol{C}}^{\text{ext}}(\omega)\tilde{E}(-\omega)\left(\mathbb{I} - \tilde{E}(-\omega)\boldsymbol{W}^{T}\right)^{-1}. \tag{7}$$

Here, $\boldsymbol{C}^{\text{int}}$ denotes the covariance matrix (see Eq 22) and $\boldsymbol{C}^{\text{ext}}$ is the covariance matrix of the external input. We model the input signal as a correlated pattern that promotes the joint activity of pairs of neurons that belong to a certain assembly.

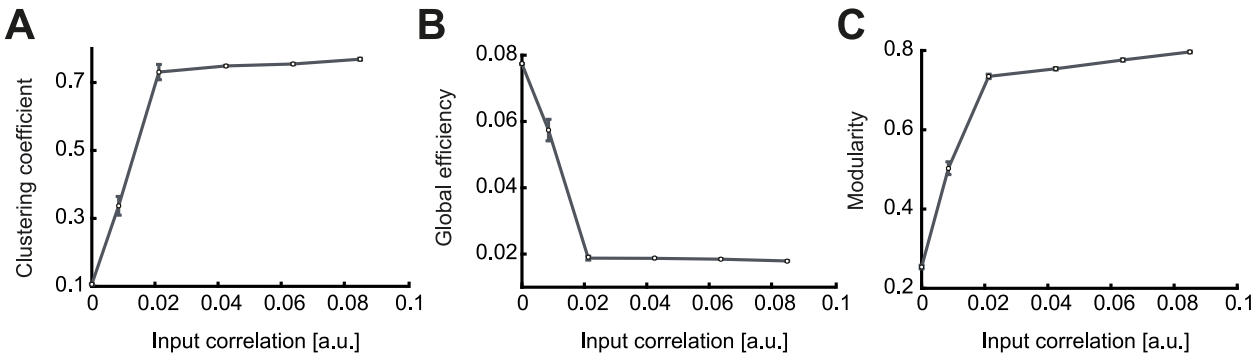


Fig 10. Emergence of assemblies in the presence of structured external input. A. Mean clustering coefficient versus the pairwise correlation coefficient of the input pattern. The strength of the correlation was provided as ratios (0.01, 0.05, 0.125, 0.25, 0.375 and 0.5) of the possible maximum weight of each individual synaptic connection w_{max}. **B**. Mean global efficiency versus the pairwise correlation coefficient of the input pattern. **C**. Mean modularity versus the pairwise correlation coefficient of the input pattern. The rapid increase of the clustering coefficient and the modularity combined with a decrease of the global efficiency is a feature of robust assembly formation. Sufficiently strong correlations in the external signal generate tight assemblies. All results are calculated from 100 trials at steady state connectivity. Error bars represent the standard error of the mean.

https://doi.org/10.1371/journal.pcbi.1007835.g010

Using the standard parameters of the minimal triplet model (Table 1; Fig 9C, $\eta_- = 1$) assembly formation is difficult when the feedforward motif coefficients dominate (the motifs for which the α-path is longer than the β-path). However, a significantly stronger external correlation relative to internally generated network correlations can promote the common input motif, $M_{1,1}$, and support assembly formation. As a function of the external correlation matrix, we quantify the structure of the resulting self-connected assemblies of neurons via the same graph measures used previously (Fig 10). The tight assemblies observed for the modulation of the STDP and the EPSC functions (Figs 6E and 9C) can now be formed for values of correlation strength one order of magnitude smaller than the synaptic upper bound.

Disrupting the balance between potentiation and depression affects assembly formation

We considered an STDP rule that is balanced in the total potentiation and depression, because disrupting this balance by increasing some firing rater over others favors the particular circuit motifs affected by those rates, as shown before [5, 45, 101]. When the balance is disrupted, the firing rate contribution to plasticity from chance spike coincidences dominates over internal correlations. When the zero-order term of the motif expansion (Eq 42) is non-zero, the mean change in the synaptic efficacies has a term that only depends on the firing rates. In this case, the firing rates of the pre- and postsynaptic neurons are the main drivers of network structure. This means that the overall impact of motifs in the network is diminished [5]. We explored the possible departures from balance through the inclusion of a perturbation parameter δ that can be either positive or negative and we scaled this parameter in proportion to the learning rate (Methods).

Therefore, to study the sensitivity of the emergence of network structure to perturbations on the depression vs. potentiation balance we consider that the zero-order 'rate' motif is different from zero. We find that departures from the balanced regime impact plasticity significantly. In the case of a depression dominated imbalance, $\delta < 0$, all connections depress no matter the strength of the modulation through η_-, even for small absolute value of $\delta = -0.0001$. In the case of potentiation, $\delta > 0$, one might expect that all synaptic efficacies will just saturate; however, due to heterosynaptic competition, some network structure still forms when δ is small (Fig 11). If the perturbation is sufficiently strong, the autonomous emergence of

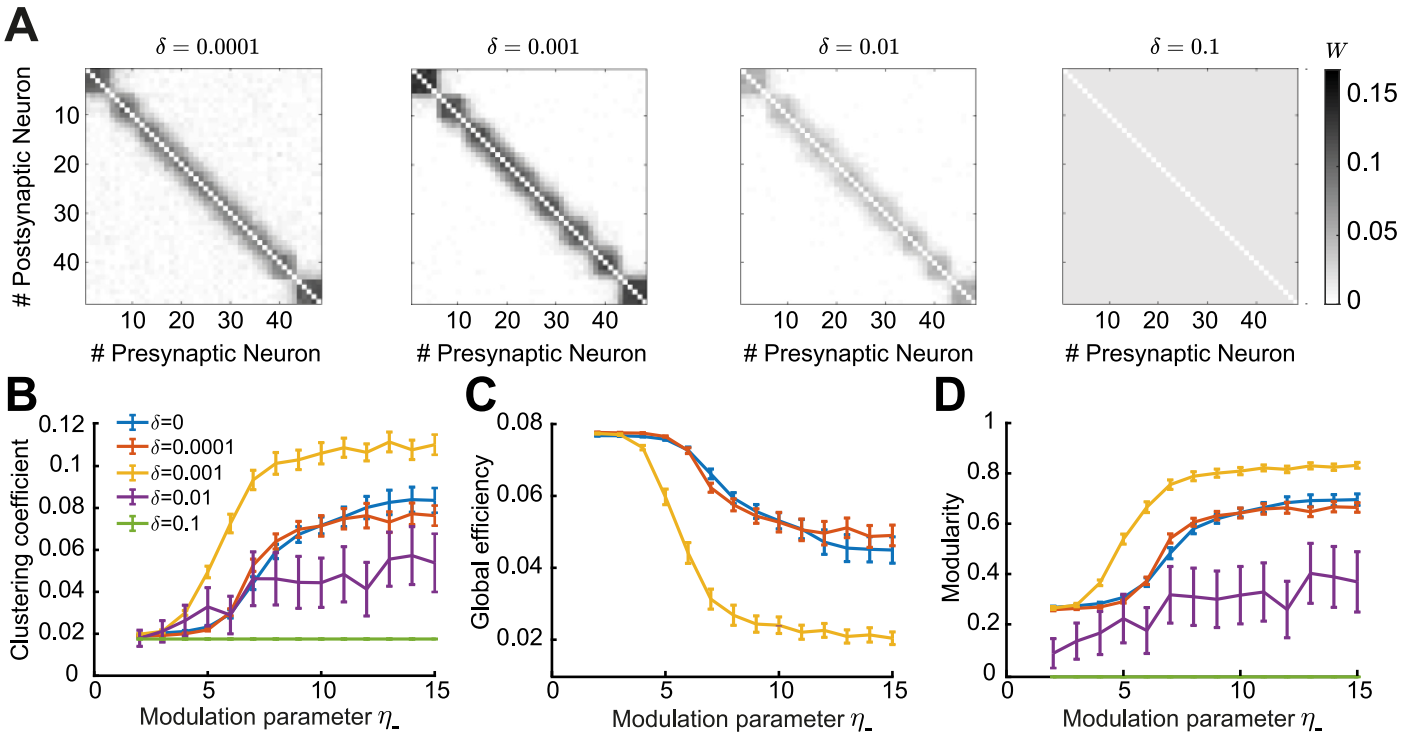


Fig 11. Impact of perturbations in the balance of potentiation and depression of the triplet STDP rule. A. Averaged connectivity matrices over 100 trials at steady state for the four different cases of the perturbation parameter δ and modulation parameter $\eta_- = 13$. **B**. Mean clustering coefficient versus the modulation parameter η_-. **C**. Mean global efficiency versus the modulation parameter η_-. We removed the cases $\delta = [0.01, 0.1]$ here since the global efficiency cannot be computed for weight matrices where all entries are identical. **D**. Mean modularity versus the modulation parameter η_-. All results are calculated from 100 trials at steady state connectivity. Error bars represent the standard error of the mean.

https://doi.org/10.1371/journal.pcbi.1007835.g011

assemblies by increasing the parameter η_- is disrupted. This is also evidenced when computing the graph measures for the resulting network structures (Fig 11B–11D). In summary, we find that considering an unbalanced STDP rule where either depression or potentiation dominates, prevents the autonomous emergence of assemblies.

Discussion

We developed a self-consistent theoretical framework to study the impact of HOCs, specifically up to third order, on the plasticity of recurrent networks by using the triplet STDP rule. We derived the dependence of the drift in synaptic efficacy on network structure, taking into account contributions from structural motifs of different orders, and demonstrated the emergence of global network structures i.e. assemblies, from these local motifs. Based on recent work on the calculation of beyond second-order cumulants of mutually exciting Hawkes processes [37, 102], we broke down the spike interactions (including pairs and triplets of spikes) to include the influence of spikes from any source neuron in the network on the firing of the pre- and postsynaptic neurons via paths of different length thus taking into account the full network recurrence (Figs 2 and 4). We characterized structural motifs that arise from these spike interactions, including novel motifs arising due to triplet STDP, and analyzed their impact on the internal up to third-order correlation structure and plasticity in the network through the motif coefficients (Figs 3 and 5). While linearization of neuronal dynamics was required for this approach, it is a common technique used to approximate the dynamics of more realistic biophysical neurons [5, 34]. We found that motif contributions to plasticity

from the second-order cross-covariance C_{ij} support assembly formation under triplet STDP. However, since these same motifs exist also under pair-based STDP, we wondered if the novel motifs unique to triplet STDP are important for assembly formation. Indeed, we showed that several novel motifs and specifically the 'loop' motifs, which emerge under the triplet STDP, have an important contribution to the formation of assemblies (Fig 8).

We investigated the contribution of up to third-order structural motifs on assembly formation using an asymmetric minimal triplet STDP rule, in which depression is induced by pairs of spikes and, conversely, potentiation is induced by triplets of spikes (Fig 1D). This rule has been shown to describe plasticity experiments that the classical STDP rule, based on pairs of spikes, has failed to capture; for instance, plasticity experiments in which the pairing frequency during plasticity induction was varied [54, 55]. As such, the triplet STDP rule is sensitive to third-order correlations, here referred to as HOCs. HOCs have not only been measured in the brain, but also shown to play an important role in visual coding and representing experimental data [58, 59, 103, 104]. HOCs are ubiquitous in sensory stimuli, such as natural stimuli and speech signals [105, 106]. These correlations have been previously utilized in learning rules, such as the BCM rule, to extract the independent components or features in natural images resulting in simple cell receptive fields as seen in V1 [105, 107–109]. Because of its mapping to the BCM rule [57], we can interpret the triplet STDP rule as a method for performing similar computations.

Modulating either the STDP rule (Fig 6) or the EPSC function (Fig 9) enabled the spontaneous formation of self-connected assemblies without the need for externally patterned inputs [49–51] or assuming a symmetric pair-based STDP rule [6]. We quantified the nature of the emergent assemblies using three graph theoretic measures used to characterize spontaneous assemblies in the tectum of zebrafish larvae [25]. Directly comparing the values of these measures between the experimental data and our model results is difficult given inhomogeneities in the size of biological network assemblies and a multitude of mechanisms that shape their formation during development. Yet, comparing how these measures change as a function of the STDP rule or the EPSC kernel in our model could offer insights into how modulating plasticity and synaptic transmission affect network structure through spontaneous activity under minimal assumptions (Fig 7). Interestingly, the final assemblies formed by modulating the EPSC function were more consistent across networks with different initial connectivity than the assemblies generated through the modification of the STDP function. This could be seen by the 'tighter' structures in the average connectivity matrices (Figs 6E and 9C), and the higher values of graph measures (Figs 6 and 9D–9F). The ultimate connectivity structure was determined by the relative strength of motifs which were regulated differently by each modulatory mechanism. In particular, modifying the EPSC function reinforced the influence of the common input motif (driven by the motif coefficient $M_{1,1}$) over all others (Fig 9B). In comparison, the modulation of the STDP rule by extending the time constant for depression over potentiation reduced the competition between reciprocal connections by maintaining a strong feedforward drive (driven by the feedforward motif coefficients $M_{1,0}$, $M_{2,1}$, $M_{2,0}$ and $M_{3,0}$) and making the corresponding reciprocal motif coefficients ($M_{0,1}$, $M_{1,2}$, $M_{0,2}$ and $M_{0,3}$) positive. Therefore, assembly formation was driven by the strengthening of reciprocal connections, even though the $M_{1,1}$ coefficient was still strong (Fig 6B). Although experimental evidence exists for a longer time constant for depression over potentiation in STDP [55, 99], the much longer values of the STDP modulation parameter η_- needed for our results raise the question of whether this mechanism is biologically plausible. This might make the modulation of the EPSC function under triplet STDP more suitable for explaining the autonomous emergence of self-connected assemblies. It is probably the case that both mechanisms are used in biological circuits. Studying the effects of neuromodulation, which can alter the shape of STDP or the synaptic transmission function, on the plasticity of connections in many brain regions is

possible with recent advances in experimental techniques [84–86]. Understanding the consequences of changing the properties of the underlying plasticity mechanisms on network dynamics can further elucidate the process of learning and memory storage in recurrent networks found everywhere in the brain [84–86].

Applying external correlated input led to the emergence of self-organized assemblies (Fig 10) that were similar to the assemblies from changing the EPSC function. Consequently, we propose that the mechanisms that promote the formation of assemblies can be diverse in different circuits depending on the nature of the plasticity rules, synaptic transmission (EPSC function) or the structure of external input that dominate in these circuits.

Our framework enabled us to derive global connectivity structures that emerge in recurrent networks such as assemblies, which have been abundantly observed in experimental data. Connectivity matrices of large recurrent networks are generally difficult to assay experimentally, requiring multiple cells to be patched simultaneously [110], although recent developments in the field of connectomics offer potential for these matrices to be obtained in the future [111, 112]. However, a good experimental determinate of assemblies may be derived from functional interactions among neurons, inferred from physiological experiments that simultaneously record the activity of a large number of neurons. While it is clear that neuronal activity exhibits structure in response to sensory input, assemblies are present even during spontaneous activity and have similar spatial organization [21, 25, 26]. This has suggested that these self-organized assemblies are biologically relevant for the processing of information in these networks and for the representation of sensory stimulus attributes [21]. In the rodent visual cortex, a given stimulus, of the form of a natural scene or an orientated grating, consistently activates a specific assembly [21]. On the behavioral scale, recent experiments suggest that functional circuit connectivity may be intrinsically adapted to respond preferentially to stimuli of biological relevance for the survival of the animal, such as catching prey or avoiding predators [24, 27].

Our analytical approach offers a precise description of how synaptic plasticity shapes connectivity in recurrent networks driven by spontaneous activity (though we also considered the role of structured external input). Such spontaneous activity is especially common during early postnatal development, where it activates neural networks before the onset of sensory experience and the maturation of sensory organs. In the rodent visual system, for instance, eye opening only occurs during the second postnatal week of development [113]. Prior to this, spontaneous patterns of activity propagate throughout the entire visual system, including the retina, thalamus and cortex [114], which are known to instruct different aspects of circuit organization [115]. Interestingly, during very early postnatal development of somatosensory cortex in rodents (postnatal day 4), spontaneous activity exhibits a highly correlated state consisting of cell assemblies where multiple neurons show correlated activity [116]. By the second postnatal week this spontaneous activity transitions to a much more decorrelated state that lacks a clear spatial structure. A similar sparsification of spontaneous activity during development is also observed in the visual cortex, though lacking the spatial structure observed in the somatosensory cortex [117]. Since these two studies argue that over development functional connectivity becomes more desynchronized, this framework is more consistent with our analysis of the depression window of the STDP rule becoming smaller over development (Fig 6). This broadening of the depression window in early development is consistent with a previously described burst-timing-dependent plasticity where the temporal integration of activity occurs over much longer timescales on the order of several hundred milliseconds than in adulthood [115, 118, 119].

Assembly formation has been the goal of many other previous works, typically instructed by externally structured input in recurrent network models with balanced excitation and inhibition [42–47]. These assemblies exhibit attractor dynamics which have been argued to serve as the substrate of different computations, such as predictive coding through the spontaneous

retrieval of evoked response patterns [49, 50, 120]. We investigated the generation of assemblies through triplet STDP driven by higher-order correlations generated internally in the network. Other works have also studied the emergence of non-random structure in the absence of structured external input [5, 6, 48]; our work takes a similar approach of incorporating the full recurrence in the network through the expansion into structural motifs as [6]. As it becomes evident from these studies, the investigation of STDP in recurrent networks for unsupervised learning involves a lot of parameters and additional mechanisms (including short-term plasticity, heterosynaptic plasticity and inhibitory plasticity) which make the identification of general rules difficult. Nevertheless, the precise theoretical description of triplet STDP in recurrent networks provided by our framework highlights a set of novel motifs absent in the case of pair-based STDP that promote assembly formation, in the process highlighting an important role of HOCs in the generation of global network structure from local motifs.

Methods

Network dynamics

The time dependent activity of a neuron i is given by a stochastic realization of an inhomogeneous Poisson process [70], with expectation value

$$\lambda_i(t) = \mu_i + \sum_{k=1}^{N} W_{ik}[E * S_k](t), \tag{8}$$

where μ_i is the external input firing rate, $\boldsymbol{W}$ is the synaptic weight matrix, $S(t)$ is the spike train and $E(t)$ is the EPSC function, which we assume to be identical for all N neurons. Then, the product $\boldsymbol{W}E(t)$ is referred to as the interaction kernel. The operator '$*$' corresponds to the convolution operation. In all plasticity simulations, the connectivity weight matrix is divided into an excitatory and an inhibitory component, such that the effective connectivity matrix is calculated as $\boldsymbol{W}^{\text{eff}} = \boldsymbol{W} - \boldsymbol{W}^{\text{inh}}$. The inhibitory weight matrix $\boldsymbol{W}^{\text{inh}}$ is updated to balance the excitatory (see section 'Additional plasticity mechanisms besides STDP'). For simplicity in notation, we refer to $\boldsymbol{W}^{\text{eff}}$ as $\boldsymbol{W}$ in the manuscript.

Averaged synaptic efficacy dynamics for pair-based and triplet STDP rules

Plasticity of the connectivity matrix $\boldsymbol{W}$ is determined by pair-based and triplet STDP rules. We assume 'all-to-all' interactions between spikes, where each postsynaptic spike interacts with every previous pre- and postsynaptic spike and vice-versa [52, 121–123].

Plasticity due to the pair-based STDP can be expressed as:

$$\dot{W}_{ij}^{\text{pair STDP}}(t) = \int_{-\infty}^{\infty} S_i(t) S_j(t-\tau_1) L_2(\tau_1) \mathrm{d}\tau_1 \tag{9}$$

and plasticity due to the triplet STDP rule as:

$$\begin{aligned} \dot{W}_{ij}^{\text{triplet STDP}}(t) = & \iint_{-\infty}^{\infty} S_i(t) S_j(t-\tau_1) S_i(t-\tau_2) L_{3,y}(\tau_1, \tau_2) \mathrm{d}\tau_1 \, \mathrm{d}\tau_2 \\ & + \iint_{-\infty}^{\infty} S_j(t) S_i(t-\tau_1) S_j(t-\tau_3) L_{3,x}(-\tau_1, -\tau_3) \mathrm{d}\tau_1 \, \mathrm{d}\tau_3 \end{aligned} \tag{10}$$

L_2 corresponds to the pair-based STDP rule and L_3 to the triplet STDP rule. The additional

subscripts x and y denote that the triplets which contribute to plasticity are two pre- and one postsynaptic spikes and one pre- and two postsynaptic spikes, respectively. τ_1 is the time difference between the spikes of the pre- and the postsynaptic neuron. τ_2 is the time difference between two postsynaptic spikes and τ_3 is the time difference between two presynaptic spikes (Fig 1B). It should be highlighted that this derivation is independent of the specific shape of the STDP functions.

Assuming slow learning in comparison to neuronal dynamics and that pairs and triplets of spikes between the pre- and postsynaptic neurons are relevant to plasticity [54, 57], the mean evolution of the synaptic efficacies due to STDP is given by

$$\begin{aligned}\langle \dot{W}_{ij}^{\mathrm{STDP}}(t)\rangle = &\int_{-\infty}^{\infty}\langle S_i(t)S_j(t-\tau_1)\rangle L_2(\tau_1)\mathrm{d}\tau_1 + \iint_{-\infty}^{\infty}\langle S_i(t)S_j(t-\tau_1)S_i(t-\tau_2)\rangle L_{3,y}(\tau_1,\tau_2)\mathrm{d}\tau_1\,\mathrm{d}\tau_2 \\ &+\iint_{-\infty}^{\infty}\langle S_j(t)S_i(t-\tau_1)S_j(t-\tau_3)\rangle L_{3,x}(-\tau_1,-\tau_3)\mathrm{d}\tau_1\,\mathrm{d}\tau_3\end{aligned} \tag{11}$$

where $\langle\cdot\rangle$ denotes averaging over different realizations of the Poisson neuronal dynamics for different connectivity.

We define the mean rates of the pre- (j) and postsynaptic neuron (i) as r_j and r_i. We consider both to be stationary at equilibrium. The second-order correlation between the pre- and postsynaptic neurons with time delay τ_1 is $\langle S_i(t)S_j(t-\tau_1)\rangle$ and we define the covariance matrix (second-order cumulant) C (Fig 1B) as

$$C_{ij}(\tau_1) = \langle S_i(t)S_j(t-\tau_1)\rangle - r_i r_j. \tag{12}$$

Note that [6, 37] use a different convention for signs.

The third-order correlation between the triplet of spikes 'post-pre-post' with time delays between the pre and one post τ_1 and between the two post τ_2 is $\langle S_i(t)S_j(t-\tau_1)S_i(t-\tau_2)\rangle$ and we define the third-order cumulant as [57]

$$K_{ij}(\tau_1,\tau_2) = \langle S_i(t)S_j(t-\tau_1)S_i(t-\tau_2)\rangle - r_i(C_{ij}(\tau_1)+C_{ij}(\tau_2-\tau_1)) - r_j C_{ii}(\tau_2) - r_i^2 r_j. \tag{13}$$

Analogously, for the 'pre-post-pre' third-order correlation $\langle S_j(t)S_i(t-\tau_1)S_j(t-\tau_3)\rangle$, we can define the third-order cumulant

$$K_{ij}(\tau_1,\tau_3) = \langle S_j(t)S_i(t-\tau_1)S_j(t-\tau_3)\rangle - r_j(C_{ij}(\tau_1)+C_{ij}(\tau_3-\tau_1)) - r_i C_{jj}(\tau_3) - r_i r_j^2. \tag{14}$$

With these definitions, Eq 11 becomes Eq 1 of the main text.

Calculation of cumulants

The definition of cumulants in the Fourier space is imperative for the derivation of our results. Assuming stationarity, the expected firing rate (i.e. the first order cumulant) vector $\boldsymbol{r} = \langle \boldsymbol{\lambda}(t)\rangle$ no longer depends on time and can be written as

$$\boldsymbol{r} = \left(\mathbb{I} - \tilde{E}(0)\boldsymbol{W}\right)^{-1}\boldsymbol{\mu}, \tag{15}$$

where $\tilde{E}(0) = \mathcal{F}[E(t)]|_{t=0}$ denotes the Fourier transform of the EPSC function evaluated at

zero. For all the calculations, we define the Fourier transform as

$$\mathcal{F}[f(\tau)] = \tilde{f}(\omega) = \int_{-\infty}^{\infty} f(\tau) e^{-j\omega\tau} \mathrm{d}\tau. \tag{16}$$

The second-order cumulant, consisting of the cross- and auto-covariance, can be calculated in the time domain as [37, 102]

$$C_{ij}(\tau_1) = \sum_{k=1}^{N} r_k \int_{-\infty}^{\infty} R_{ik}(u) R_{jk}(u - \tau_1) \mathrm{d}u, \tag{17}$$

where $\boldsymbol{R}(t) = \sum_{n\geq 0} \boldsymbol{G}^{*n}(t)$ is defined as a 'convolution power series' [37, 102] of the interaction kernel $\boldsymbol{G}(t) = \boldsymbol{W}E(t)$, with

$$\boldsymbol{G}^{*n}(t) = \begin{cases} \mathbb{I}\delta(t), & \text{if } n = 0 \\ \int_{-\infty}^{t} \boldsymbol{G}^{*(n-1)}(t-s)\boldsymbol{G}(s)\mathrm{d}s = \int_{-\infty}^{t} (\boldsymbol{W}E(t-s))^{*(n-1)} \boldsymbol{W}E(s)\mathrm{d}s, & \text{if } n \geq 1. \end{cases} \tag{18}$$

Since $\boldsymbol{W}$ does not depend on the integration domain, the convolution in the operation ^{*n} is calculated on E, while for $\boldsymbol{W}$ it becomes a power operation. Formally, the computation of each element $R_{mn}(t)$ consists of calculating the probability of a spike from neuron m at time t given that neuron n fired at time 0. Therefore, in Eq 17 we write the covariance for the spike trains of neurons i and j as the probability of a pair of spikes in neurons i and j at a time lag τ_1 given that neuron k fired, where k can be *any* neuron in the network. This representation provides a convenient formalism for representing causality of spiking events in our model. Then, considering the definition of 'path lengths' α and β from the source neuron k to the postsynaptic neuron i and the presynaptic neuron j (Fig 2A), we can rewrite Eq 17 as

$$C_{ij}(\tau_1) = \sum_{\alpha,\beta} \int_{-\infty}^{\infty} E^{*\alpha}(u) E^{*\beta}(u - \tau_1) \mathrm{d}u \sum_{k=1}^{N} r_k (\boldsymbol{W}^{\alpha})_{ik} (\boldsymbol{W}^{\beta})_{jk}. \tag{19}$$

Here, $E^{*\kappa}$ denotes a series of convolutions of the EPSC function

$$E^{*\kappa}(t) = \underbrace{E(t) * E(t) * \ldots * E(t)}_{\kappa \text{ terms}}. \tag{20}$$

For the auto-covariance C_{ii} for path lengths α and γ from the source neuron k to the postsynaptic neuron i (Fig 2B), we analogously obtain

$$C_{ii}(\tau_2) = \sum_{\alpha,\gamma} \int_{-\infty}^{\infty} E^{*\alpha}(u) E^{*\gamma}(u - \tau_2) \mathrm{d}u \sum_{k=1}^{N} r_k (\boldsymbol{W}^{\alpha})_{ik} (\boldsymbol{W}^{\gamma})_{ik}. \tag{21}$$

Since each R function consists of the convolution of the EPSC functions, then its Fourier transform is the product of the Fourier transforms of each of those functions, which simplifies calculations. Therefore, the cross-covariance C_{ij} in the frequency domain (i.e. the Fourier transform of Eq 17) is given by (detailed derivation in S1 Text)

$$\tilde{C}_{ij}(\omega) = \sum_{k=1}^{N} r_k \tilde{R}_{ik}(\omega) \tilde{R}_{jk}(-\omega), \tag{22}$$

and, finally we obtain the expression

$$\tilde{C}_{ij}(\omega) = \sum_{\alpha,\beta} \tilde{E}^{\alpha}(\omega)\, \tilde{E}^{\beta}(-\omega) \sum_{k=1}^{N} r_k (\boldsymbol{W}^{\alpha})_{ik} (\boldsymbol{W}^{\beta})_{jk}. \tag{23}$$

It should be noted that Eq 23 was also derived in previous works using a different approach [6, 70]. However, for the third-order cumulant K_{ij} (Fig 4) that same approach is not possible. Therefore, it is convenient to write K_{ij} in the time domain in terms of the previously defined $\boldsymbol{R}$ [37, 102] as

$$\begin{aligned} K_{ij}(\tau_1,\tau_2) \;=& \sum_{k=1}^{N} r_k \int_{-\infty}^{\infty} R_{ik}(u) R_{jk}(u-\tau_1) R_{ik}(u-\tau_2)\mathrm{d}u \\ &+ \sum_{k,l=1}^{N} r_k \iint_{-\infty}^{\infty} R_{ik}(u) R_{jl}(v-\tau_1) R_{il}(v-\tau_2) \Psi_{lk}(v-u)\mathrm{d}v\,\mathrm{d}u \\ &+ \sum_{k,l=1}^{N} r_k \iint_{-\infty}^{\infty} R_{jk}(u-\tau_1) R_{il}(v) R_{il}(v-\tau_2) \Psi_{lk}(v-u)\mathrm{d}v\,\mathrm{d}u \\ &+ \sum_{k,l=1}^{N} r_k \iint_{-\infty}^{\infty} R_{ik}(u-\tau_2) R_{il}(v) R_{jl}(v-\tau_1) \Psi_{lk}(v-u)\mathrm{d}v\,\mathrm{d}u, \end{aligned} \tag{24}$$

where additionally

$$\boldsymbol{\Psi}(t) = \boldsymbol{R}(t) - \mathbb{I}\,\delta(t) = \sum_{n\geq 1} \boldsymbol{G}^{*n}(t). \tag{25}$$

In Eq 24, $\Psi_{lk}(v-u)$ is the probability density of the event that a spike from neuron k at a time $v-u=0$ causes a neuron l (different from neuron k) to emit a spike at a time $v-u\neq 0$, after at least one synaptic connection. The function $\boldsymbol{\Psi}$ is necessary in Eq 24 to take into account the branching structures in the calculation of K_{ij} (Fig 4B–4D). In addition to α, β and γ, ζ is the path length from the source neuron k to the neuron l where the synaptic connection path branches out and is equal to or larger than one. Then, replacing both the R and Ψ functions by their corresponding definitions in terms of the connectivity matrix $\boldsymbol{W}$ and EPSC function $E(t)$ yields

$$\begin{aligned} K_{ij}(\tau_1,\tau_2) \;=& \sum_{\alpha,\beta,\gamma} \int_{-\infty}^{\infty} E^{*\alpha}(u) E^{*\beta}(u-\tau_1) E^{*\gamma}(u-\tau_2)\mathrm{d}u \sum_{k=1}^{N} r_k (\boldsymbol{W}^{\alpha})_{ik} (\boldsymbol{W}^{\gamma})_{ik} (\boldsymbol{W}^{\beta})_{jk} \\ &+ \sum_{\alpha,\beta,\gamma} \sum_{\zeta\geq 1} \iint_{-\infty}^{\infty} E^{*\alpha}(u) E^{*\beta}(v-\tau_1) E^{*\gamma}(v-\tau_2) E^{*\zeta}(v-u)\mathrm{d}v\,\mathrm{d}u \sum_{k,l}^{N} r_k (\boldsymbol{W}^{\alpha})_{ik} (\boldsymbol{W}^{\beta})_{jl} (\boldsymbol{W}^{\gamma})_{il} (\boldsymbol{W}^{\zeta})_{lk} \\ &+ \sum_{\alpha,\beta,\gamma} \sum_{\zeta\geq 1} \iint_{-\infty}^{\infty} E^{*\beta}(u-\tau_1) E^{*\alpha}(v) E^{*\gamma}(v-\tau_2) E^{*\zeta}(v-u)\mathrm{d}v\,\mathrm{d}u \sum_{k,l}^{N} r_k (\boldsymbol{W}^{\beta})_{jk} (\boldsymbol{W}^{\alpha})_{il} (\boldsymbol{W}^{\gamma})_{il} (\boldsymbol{W}^{\zeta})_{lk} \\ &+ \sum_{\alpha,\beta,\gamma} \sum_{\zeta\geq 1} \iint_{-\infty}^{\infty} E^{*\gamma}(u-\tau_2) E^{*\alpha}(v) E^{*\beta}(v-\tau_1) E^{*\zeta}(v-u)\mathrm{d}v\,\mathrm{d}u \sum_{k,l}^{N} r_k (\boldsymbol{W}^{\gamma})_{ik} (\boldsymbol{W}^{\alpha})_{il} (\boldsymbol{W}^{\beta})_{jl} (\boldsymbol{W}^{\zeta})_{lk}. \end{aligned} \tag{26}$$

As with the second-order cumulant, we can calculate the Fourier transform of the third-order cumulant K_{ij} from Eq 24 as (detailed derivation in S1 Text)

$$\begin{aligned}\tilde{K}_{ij}(\omega_1,\omega_2) &= \sum_{k=1}^{N} r_k \tilde{R}_{ik}(\omega_1+\omega_2)\tilde{R}_{jk}(-\omega_1)\tilde{R}_{ik}(-\omega_2) \\ &+\sum_{k,l=1}^{N} r_k \tilde{R}_{ik}(\omega_1+\omega_2)\tilde{R}_{jl}(-\omega_1)\tilde{R}_{il}(-\omega_2)\tilde{\Psi}_{lk}(\omega_1+\omega_2) \\ &+\sum_{k,l=1}^{N} r_k \tilde{R}_{il}(\omega_1+\omega_2)\tilde{R}_{jk}(-\omega_1)\tilde{R}_{il}(-\omega_2)\tilde{\Psi}_{lk}(-\omega_1) \\ &+\sum_{k,l=1}^{N} r_k \tilde{R}_{il}(\omega_1+\omega_2)\tilde{R}_{jl}(-\omega_1)\tilde{R}_{ik}(-\omega_2)\tilde{\Psi}_{lk}(-\omega_2).\end{aligned} \tag{27}$$

Finally, we obtain the third-order cumulant K_{ij} in the Fourier domain in terms of the connectivity matrix $\boldsymbol{W}$, the EPSC function $E(t)$, and the path lengths α, β, γ and ζ as

$$\begin{aligned}K_{ij}(\omega_1,\omega_2) &= \sum_{\alpha,\beta,\gamma}\tilde{E}^{\alpha}(\omega_1+\omega_2)\,\tilde{E}^{\beta}(-\omega_1)\,\tilde{E}^{\gamma}(-\omega_2)\sum_{k=1}^{N} r_k (\boldsymbol{W}^{\alpha})_{ik}(\boldsymbol{W}^{\gamma})_{ik}(\boldsymbol{W}^{\beta})_{jk} \\ &+\sum_{\alpha,\beta,\gamma}\sum_{\zeta\geq 1}\tilde{E}^{\alpha+\zeta}(\omega_1+\omega_2)\,\tilde{E}^{\beta}(-\omega_1)\,\tilde{E}^{\gamma}(-\omega_2)\sum_{k,l=1}^{N} r_k (\boldsymbol{W}^{\zeta})_{lk}(\boldsymbol{W}^{\alpha})_{ik}(\boldsymbol{W}^{\beta})_{jl}(\boldsymbol{W}^{\gamma})_{il} \\ &+\sum_{\alpha,\beta,\gamma}\sum_{\zeta\geq 1}\tilde{E}^{\alpha}(\omega_1+\omega_2)\,\tilde{E}^{\beta+\zeta}(-\omega_1)\tilde{E}^{\gamma}(-\omega_2)\sum_{k,l=1}^{N} r_k (\boldsymbol{W}^{\zeta})_{lk}(\boldsymbol{W}^{\alpha})_{il}(\boldsymbol{W}^{\beta})_{jk}(\boldsymbol{W}^{\gamma})_{il} \\ &+\sum_{\alpha,\beta,\gamma}\sum_{\zeta\geq 1}\tilde{E}^{\alpha}(\omega_1+\omega_2)\,\tilde{E}^{\beta}(-\omega_1)\,\tilde{E}^{\gamma+\zeta}(-\omega_2)\sum_{k,l=1}^{N} r_k (\boldsymbol{W}^{\zeta})_{lk}(\boldsymbol{W}^{\alpha})_{il}(\boldsymbol{W}^{\beta})_{jl}(\boldsymbol{W}^{\gamma})_{ik}.\end{aligned} \tag{28}$$

Calculation of motif coefficients

Extending the work of [6], who artificially tuned the values of motif coefficients to investigate the consequences on the network structures, we derive them analytically as a function of the STDP rule and the EPSC function. To obtain the expression for the motif coefficients necessary for Eqs 3, 4 and 5, we first need to insert Eqs 23 and 28, i.e. the definitions of the second- and third-order cumulants in the frequency domain, in Eq 2. Then, it can easily be seen that it is possible to separate the part that depends on the products of the connectivity matrix $\boldsymbol{W}$ from the rest. This way we define the motif coefficients as the integral of the products of the Fourier transforms of the STDP functions and the EPSC functions, considering the appropriate path lengths α, β, γ and ζ (Fig 3B and 3D). In particular, for the motif coefficients in Eq 3 we derive

$$M^{pair}_{\alpha,\beta} = \int_{-\infty}^{\infty}\tilde{E}^{\alpha}(\omega_1)\,\tilde{E}^{\beta}(-\omega_1)\,\tilde{L}_2(-\omega_1)\mathrm{d}\omega_1. \tag{29}$$

and

$$M^{trip}_{\alpha,\beta} = \iint_{-\infty}^{\infty}\Big(\tilde{E}^{\alpha}(\omega_1)\,\tilde{E}^{\beta}(-\omega_1)\,\delta(\omega_2)+\tilde{E}^{\alpha}(\omega_2)\,\tilde{E}^{\beta}(-\omega_2)\,\delta(\omega_1+\omega_2)\Big)\tilde{L}_3(-\omega_1,-\omega_2)\mathrm{d}\omega_1\,\mathrm{d}\omega_2. \tag{30}$$

We note that this definition combines motif coefficients where α is the index corresponding to paths to the postsynaptic neuron, regardless of which of the two postsynaptic spike of the spike triplet it refers to (Fig 1C). For the motif coefficient in Eq 4 we derive

$$M^{trip}_{\alpha,\gamma} = \iint_{-\infty}^{\infty} \tilde{E}^{\alpha}(\omega_2)\,\tilde{E}^{\gamma}(-\omega_2)\,\tilde{L}_3(-\omega_1,-\omega_2)\,\delta(\omega_1)\mathrm{d}\omega_1\,\mathrm{d}\omega_2. \tag{31}$$

Lastly, for the 'straight' triplet motif (Fig 4A) in Eq 5 we get:

$$M^{trip}_{\alpha,\beta,\gamma} = \iint_{-\infty}^{\infty} \tilde{E}^{\alpha}(\omega_1+\omega_2)\,\tilde{E}^{\beta}(-\omega_1)\,\tilde{E}^{\gamma}(-\omega_2)\,\tilde{L}_3(-\omega_1,-\omega_2)\,\mathrm{d}\omega_1\,\mathrm{d}\omega_2 \tag{32}$$

while for the 'branching' motifs (Fig 4B–4D) in Eq 5:

$$M^{trip}_{(\alpha,\zeta),\beta,\gamma} = \iint_{-\infty}^{\infty} \tilde{E}^{\alpha+\zeta}(\omega_1+\omega_2)\,\tilde{E}^{\beta}(-\omega_1)\,\tilde{E}^{\gamma}(-\omega_2)\,\tilde{L}_3(-\omega_1,-\omega_2)\,\mathrm{d}\omega_1\,\mathrm{d}\omega_2, \tag{33}$$

$$M^{trip}_{\alpha,(\beta,\zeta),\gamma} = \iint_{-\infty}^{\infty} \tilde{E}^{\alpha}(\omega_1+\omega_2)\,\tilde{E}^{\beta+\zeta}(-\omega_1)\,\tilde{E}^{\gamma}(-\omega_2)\,\tilde{L}_3(-\omega_1,-\omega_2)\,\mathrm{d}\omega_1\,\mathrm{d}\omega_2, \tag{34}$$

and

$$M^{trip}_{\alpha,\beta,(\gamma,\zeta)} = \iint_{-\infty}^{\infty} \tilde{E}^{\alpha}(\omega_1+\omega_2)\,\tilde{E}^{\beta}(-\omega_1)\,\tilde{E}^{\gamma+\zeta}(-\omega_2)\,\tilde{L}_3(-\omega_1,-\omega_2)\,\mathrm{d}\omega_1\,\mathrm{d}\omega_2. \tag{35}$$

These expressions give us a concise representation of how the spiking activity interacts with network structure to impact plasticity.

Synaptic dynamics

To calculate the values for the motif coefficients in Eqs 29–35, we define the EPSC function $E(t)$ as

$$E(t) = \begin{cases} \frac{\tau_\varepsilon+\tau_\iota}{\tau_\varepsilon^2} e^{\frac{-t}{\tau_\varepsilon}}\left(1-e^{\frac{-t}{\tau_\iota}}\right) & \text{if } t \geq 0 \\ 0 & \text{if } t < 0. \end{cases} \tag{36}$$

This function depends on two time constants τ_ε and τ_ι that define the onset and decay of the increase in the membrane potential with each spike. In particular, when $\tau_\iota \to 0$ the current is instantaneous and decays exponentially. The function is normalized to have an integral equal to 1, so that on average the number of postsynaptic spikes with the arrival of a presynaptic spike scales with the same order of magnitude as the synaptic efficacy. Its Fourier transform is

$$\tilde{E}(\omega) = \left(1+\frac{\tau_\iota}{\tau_\varepsilon}\right)\frac{1-j\tau_\varepsilon\omega}{1+\tau_\varepsilon^2\omega^2} - \frac{\tau_\iota}{\tau_\varepsilon}\frac{1-j\frac{\tau_\varepsilon\tau_\iota}{\tau_\varepsilon+\tau_\iota}\omega}{1+\left(\frac{\tau_\varepsilon\tau_\iota}{\tau_\varepsilon+\tau_\iota}\right)^2\omega^2}. \tag{37}$$

With respect to the choice of STDP function, we consider the minimal triplet STDP rule [54, 57] that consists of the pair-based STDP function for depression and of a triplet STDP

function for potentiation (Fig 1C). Furthermore, we introduce a 'modulation parameter' η_- to model the reshaping of the depression window of the STDP function via modulatory effects.

The depression window of the STDP function can be written as

$$L_2(\tau_1) = \begin{cases} -\frac{A_-}{\eta_-} e^{\frac{\tau_1}{\eta_- \tau_-}} & \text{if } \tau_1 < 0 \\ 0 & \text{otherwise,} \end{cases} \tag{38}$$

where $\tau_1 = t_{post} - t_{pre}$ denotes the time difference between a post- and a presynaptic spike, A_- is the depression learning rate, τ_- is the depression time constant and η_- is the depression modulation parameter. The potentiation window of the STDP function depends on the timing of spike triplets (t_{pre}, t_{post}, t'_{post})

$$L_3(\tau_1, \tau_2) = \begin{cases} A_+ e^{-\frac{\tau_1}{\tau_+}} e^{-\frac{\tau_2}{\tau_y}} & \text{if } \tau_1 \geq 0, \tau_2 \geq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{39}$$

where again $\tau_1 = t_{post} - t_{pre}$ denotes the time difference between a post- and a presynaptic spike and $\tau_2 = t_{post} - t'_{post}$ is the time difference between the two postsynaptic spikes; A_+ is the potentiation learning rate, τ_+ is the potentiation time constant and τ_y is the second potentiation time constant.

While the 'A' parameters scale the amplitude of weight changes, the 'τ' coefficients determine how synchronous pre- and post-synaptic spikes must be to drive plasticity. The η_- parameter enables the modification of the shape of the STDP function. This additional parameter does not affect the total depression and potentiation in the rule and one can easily recover the 'standard' expressions for $\eta_- \to 1$.

The Fourier transforms for these two functions are

$$\tilde{L}_2(\omega_1) = -A_- \tau_- \frac{1 + j\eta_- \tau_- \omega_1}{1 + \eta_-^2 \tau_-^2 \omega_1^2}, \tag{40}$$

and

$$\tilde{L}_3(\omega_1, \omega_2) = A_+ \tau_+ \tau_y \frac{1 - j\tau_+ \omega_1}{1 + \tau_+^2 \omega_1^2} \frac{1 - j\tau_y \omega_2}{1 + \tau_y^2 \omega_2^2}. \tag{41}$$

Motif expansion up to third-order

After including Eqs 3, 4 and 5 into Eq 2, we can rewrite it in terms of the order of interactions in which they contribute to the averaged synaptic modification:

$$\langle \dot{W}_{ij} \rangle = \langle \dot{W}_{ij} \rangle^{(0)} + \langle \dot{W}_{ij} \rangle^{(1)} + \langle \dot{W}_{ij} \rangle^{(2)} + \langle \dot{W}_{ij} \rangle^{(3)} + \dots \tag{42}$$

Assuming non-zero mean rates, no self-excitation (i.e. $W_{ii} = 0$) and that terms of order higher than three can be disregarded in comparison to lower order ones, the terms of Eq 42 are

$$\langle \dot{W}_{ij} \rangle^{(0)} = r_i r_j M_0 \tag{43}$$

for the zeroth-order contributions,

$$\langle \dot{W}_{ij} \rangle^{(1)} = r_j W_{ij} M_{1,0} + r_i W_{ji} M_{0,1} \tag{44}$$

for the first-order contributions,

$$\begin{aligned}\langle \dot{W}_{ij}\rangle^{(2)} &= \sum_{k\neq i,j} r_k W_{ik} W_{jk} M_{1,1} + r_j (\boldsymbol{W}^2)_{ij} M_{2,0} + r_i (\boldsymbol{W}^2)_{ji} M_{0,2} \\ &+ \sum_{k\neq i,j} r_i W_{ik} W_{ki} r_j M^{trip}_{\alpha=2,\gamma=0} + \sum_{k\neq i,j} r_k W^2_{ik} r_j M^{trip}_{\alpha=1,\gamma=1} \\ &+ r_j W^2_{ij} M^{trip}_{\alpha=1,\beta=0,\gamma=1},\end{aligned} \tag{45}$$

for the second-order contributions, and finally

$$\begin{aligned}\langle \dot{W}_{ij}\rangle^{(3)} &= \sum_{k\neq i,j} r_k (\boldsymbol{W}^2)_{ik} W_{jk} M_{2,1} + \sum_{k\neq i,j} r_k W_{ik} (\boldsymbol{W}^2)_{jk} M_{1,2} + r_j (\boldsymbol{W}^3)_{ij} M_{3,0} \\ &+ r_i (\boldsymbol{W}^3)_{ji} M_{0,3} + \sum_{k\neq i,j} r_k (\boldsymbol{W}^2)_{ik} W_{ik} r_j (M^{trip}_{\alpha=2,\gamma=1} + M^{trip}_{\alpha=1,\gamma=2}) \\ &+ r_j^2 (\boldsymbol{W}^3)_{ii} M^{trip}_{\alpha=3,\gamma=0} + \sum_{k\neq i,j} r_k W^2_{ik} W_{jk} M^{trip}_{\alpha=1,\beta=1,\gamma=1}\end{aligned} \tag{46}$$

for the third-order contributions. Examples and illustrations of these motifs are given in Figs 3 and 5. For conciseness, we grouped motif coefficients arising from the pair-based STDP rule and from the triplet STDP rule that shared values of α and β and relabeled them as

$$M_0 = \tilde{L}_2(0) + r_i \tilde{L}_3(0,0), \tag{47}$$

$$M_{1,0} = r_i M^{trip}_{\alpha=1,\beta=0}, \tag{48}$$

$$M_{0,1} = M^{pair}_{\alpha=0,\beta=1} + r_i M^{trip}_{\alpha=0,\beta=1}, \tag{49}$$

$$M_{2,0} = r_i M^{trip}_{\alpha=2,\beta=0}, \tag{50}$$

$$M_{0,2} = M^{pair}_{\alpha=0,\beta=2} + r_i M^{trip}_{\alpha=0,\beta=2}, \tag{51}$$

$$M_{1,1} = M^{pair}_{\alpha=1,\beta=1} + r_i M^{trip}_{\alpha=1,\beta=1}, \tag{52}$$

$$M_{3,0} = r_i M^{trip}_{\alpha=3,\beta=0}, \tag{53}$$

$$M_{0,3} = M^{pair}_{\alpha=0,\beta=3} + r_i M^{trip}_{\alpha=0,\beta=3}, \tag{54}$$

$$M_{2,1} = M^{pair}_{\alpha=2,\beta=1} + r_i M^{trip}_{\alpha=2,\beta=1}, \tag{55}$$

$$M_{1,2} = M^{pair}_{\alpha=1,\beta=2} + r_i M^{trip}_{\alpha=1,\beta=2}. \tag{56}$$

Using the defined functions for the EPSC (Eq 37) and STDP functions (Eqs 40 and 41) we calculated the motif coefficients in Eqs 29–32, and consequently Eqs 47–56. Note, we excluded the higher-than-third-order motif coefficients in Eqs 33–35. The truncated approximation of motifs up to third-order is valid, since the difference of the weight change from the full contribution and the weight change from up to third-order motif truncation is very small (S1 Fig).

These quantities represent the strength of contributions of each particular combination of paths from the source neuron to the pre- and postsynaptic neurons involved in the synaptic connection. In principle, also the motifs $M_{\alpha=0,\gamma=2}$ and $M_{\alpha=0,\gamma=3}$ from the auto-covariance and the motifs $M_{\alpha=2,\beta=0,\gamma=1}$ and $M_{\alpha=1,\beta=0,\gamma=2}$ from the third-order cumulant would need to be considered, however we find that the contribution of these motifs is zero and therefore we did not include them in our analysis. Since we assume a learning rule balanced in potentiation and depression,

$$M_0 = 0, \tag{57}$$

and thus

$$-A_-\tau_- + r_i A_+ \tau_+ \tau_y = 0, \tag{58}$$

which is independent of the modulation parameter η_-, this allows us to rewrite these motifs so that they are independent of the mean firing rate of the postsynaptic neuron. Since the firing rate r_i is not fixed, it should be noted that this assumption implies that the amplitude of the LTP window A_+ adjusts to balance the learning rule, similar to metaplasticity [57]. However, we verify that the firing rates in the system are relatively stable over the time of the simulation (S5 Fig) and therefore A_+ does not vary much. We analyze the evolution of these quantities in the main text, because they involve only α and β paths and remain constant throughout the numerical integration, in contrast to the motif coefficients in Eqs 45 and 46 which involve both α and γ paths, and which have an additional rate dependence. The expressions for the motif coefficients defined by Eqs 47–56 in terms of the EPSC and STDP functions' parameters are given in S2 Text.

Perturbation of the zero-order motif

We consider a small perturbation δ to the zero-order (or rate) motif

$$M_0 = \pm\delta. \tag{59}$$

A minus sign indicates that the balance is tilted towards depression and, conversely, a plus sign conveys a potentiation-dominated regime. Then, given the minimal triplet STDP rule, we obtain that

$$-A_-\tau_- + r_i A_+ \tau_+ \tau_y = \pm\delta. \tag{60}$$

Since the firing rate contribution to plasticity is now different from zero, chance spike coincidences impact the averaged evolution of the synaptic efficacies as follows

$$\langle \dot{W}_{ij} \rangle = r_i r_j (\pm\delta) + \cdots \tag{61}$$

In this new scenario, the motif coefficients calculated from the triplet rule (equations given in S2 Text) now scale with $(A_- \tau_- \pm \delta)$. All weights depress to zero for $\delta < 0$ because we impose a lower bound on the weights at 0, and do not include any growth terms independent of synaptic potentiation.

Additional plasticity mechanisms besides STDP

Although we did not formally model inhibitory plasticity, we assume that the overall effect of the inhibitory population on the synaptic efficacies among excitatory neurons is to balance the network activity. Thus, the sum of inhibitory synapses into each neuron is dynamically adjusted to match the sum of the excitatory synaptic efficacies, such that each element of the

inhibitory connectivity matrix is equal to the average of the excitatory input as

$$\boldsymbol{W}^{\text{inh}} = \begin{bmatrix} w_1^{\text{inh}} \\ \vdots \\ w_N^{\text{inh}} \end{bmatrix} [1 \cdots 1] - \boldsymbol{D}^{\text{inh}} \tag{62}$$

where

$$w_l^{\text{inh}} = \frac{1}{N-1} \sum_k W_{lk} \tag{63}$$

is the value of each row element and

$$\boldsymbol{D}^{\text{inh}} = \begin{bmatrix} w_1^{\text{inh}} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & w_N^{\text{inh}} \end{bmatrix} \tag{64}$$

is a diagonal matrix to take into account there is no self-connectivity. Then, the effective connectivity weight matrix is calculated as $\boldsymbol{W}^{\text{eff}} = \boldsymbol{W} - \boldsymbol{W}^{\text{inh}}$. The inhibitory connections are fast and updated in each integration step. As mentioned earlier, we refer to $\boldsymbol{W}^{\text{eff}}$ as $\boldsymbol{W}$ in the manuscript. It should be noted that deviations from this perfect balance between excitation and inhibition, modeled with an inhibitory multiplicative factor δ_{inh} which scales the overall inhibitory inputs to deviate from a perfect balance, do not affect the emergence of network structure (S4 Fig). Furthermore, we find that the formation of network structure does somewhat depend on the input rate μ_i but not on heterogeneity in the input firing rates (S3 Fig). The input rates effectively determine the mean firing rates of the network throughout the whole simulation (S5 Fig).

We also implement heterosynaptic competition based on previous work [6, 80] as an additional mechanism for the plasticity dynamics to restrict the maximum number of strong connections a neuron can make, and thus keep the spectral radius of the connectivity matrix lower than one. The total synaptic input and output of each neuron is limited: the sum of the inbound (afferent) connections to each postsynaptic neuron i and the sum of outbound (efferent) excitatory synaptic efficacies from each presynaptic neuron j have an upper bound $W_{\max}$. The plasticity due to heterosynaptic competition can be written as

$$\langle \dot{W}_{ij}^{\text{hc}} \rangle = \Big(W_{\max} - \sum_k W_{ki} \Big) H\Big(\sum_k W_{ki} - W_{\max} \Big) + \Big(W_{\max} - \sum_k W_{jk} \Big) H\Big(\sum_k W_{jk} - W_{\max} \Big), \tag{65}$$

where H is the Heaviside function. Imposing an upper bound $w_{\max}$ for each synaptic efficacy restricts the possible number of connections a neuron can make to $\frac{W_{\max}}{w_{\max}}$. Therefore, the average amount of plasticity is the sum of the change due to STDP based on Eq 2 and heterosynaptic competition based on Eq 65.

$$\langle \dot{W}_{ij} \rangle = \nu \Big(\langle \dot{W}_{ij}^{\text{STDP}} \rangle + \psi \langle \dot{W}_{ij}^{\text{hc}} \rangle \Big). \tag{66}$$

Here, the learning rate scale ν ensures that the synaptic efficacy increments in each integration step are small. The relative contribution of heterosynaptic competition to overall plasticity is determined by the heterosynaptic competition term ψ. The values for these parameters can be

Table 2. Parameter values for supplementary figures. ⋆ denotes that values are provided in the figures.

Symbol	Description	S1 Fig	S2 Fig	S3 Fig	S4 Fig	S5 Fig
N	Number of neurons	12	48			
$\boldsymbol{\mu}$	External input firing rate	150 Hz		⋆	150 Hz	⋆
w_{max}	Upper bound for each individual weight	0.17				
W_{max}	Upper bound for total row/ column weight	0.85				
A_-	Depression learning rate	0.01				
τ_-	Depression time constant	33.7 ms [54]				
τ_+	Potentiation time constant	16.8 ms [54]				
τ_y	Second potentiation time constant	114 ms [54]				
η_-	Depression modulation parameter	13	⋆	13		
τ_e	First membrane time constant	5 ms				
τ_i	Second membrane time constant	5 ms				
ν	Scaling parameter of learning rate	3.5×10^{-4}				
ψ	Heterosynaptic competition scaling parameter	0.7				

https://doi.org/10.1371/journal.pcbi.1007835.t002

found in Table 1. The values for the parameters in the Supplementary Figures can be found in Table 2.

Numerical integration of connectivity matrices

To generate the different connectivity matrices in each Figure, we integrate Eq 2 numerically. The plasticity dynamics are implemented using the Euler method with an adaptive time step. The maximal amount that a weight can change in each integration step is 0.00035. Although the weight evolution is deterministic and determined by the plasticity parameters (Table 1), final connectivity matrices depend on initial connectivity matrices. The initial connection weights are chosen independently from a uniform distribution between 0 and $W_{max}/N \times 0.001$, and each one of these initial conditions corresponds to a different "trial". The numerical integration for each initial condition is continued until the network connectivity achieves a steady state (no longer changes).

Averaged ordered connectivity matrices

The connectivity matrices resulting from integrating Eq 2 numerically are ordered to reflect the graph structure of the network [6] (Fig 6C). K-means classification groups neurons that share similar connectivity using a squared Euclidean distance. We then reorder the connectivity matrix based on the groups identified by the k-means clustering. Since the structures studied depend on initial conditions, despite the deterministic nature of our approach, we average the rearranged synaptic efficacy matrix over many trials with different (but random and weak) initial connectivity to obtain the most likely connectivity (Figs 6E and 9C). Assemblies on the edges of the connectivity matrices have sharper edges due to an artifact created by the ordering algorithm, but this does not affect results.

Network analysis

We calculated graph theoretic measures for directed networks using algorithms of the Brain Connectivity Toolbox [124] from http://www.brain-connectivity-toolbox.net. All graph measures were calculated at the steady state and increased during a simulation as network organization improved.

Clustering coefficient. For each connectivity matrix we computed the clustering coefficient [95]. For node i, this is

$$C_i = \frac{\text{number of complete triplets}}{\text{number of all possible triplets}}. \tag{67}$$

The number of complete triplets is obtained from the product of the corresponding edges of the node (from the adjancency matrix), and the total number of triplets depends on network size. Then, the average of the clustering coefficients of all the vertices N is given by [92]

$$\bar{C} = \frac{1}{N}\sum_{i=1}^{N} C_i. \tag{68}$$

Global efficiency. The efficiency in the communication between nodes i and j can be defined to be inversely proportional to the shortest distance. The average efficiency of a network is calculated as [96]

$$E = \frac{1}{N(N-1)}\sum_{i\neq j}\frac{1}{d(i,j)} \tag{69}$$

where N denotes the nodes in the network and $d(i, j)$ is the length of the shortest path between a node i and a different node j. As an alternative to the average path length, the global efficiency of a network is defined as

$$E_{glob} = \frac{E}{E(\text{full network})} \tag{70}$$

where the efficiency is scaled by an ideal graph where all the possible edges exist (i.e. full network). The difference between these measures is that the first measure quantifies the efficiency in a network where only one packet of information is being moved through it and the global measure quantifies the efficiency where all the vertices are exchanging packets of information with each other [96].

Modularity. The modularity Q of a connectivity matrix is a measure of the strength of its division into clusters or modules. Formally, modularity can be calculated as [93]

$$Q = \frac{1}{2m}\sum_{vw}\left(A_{vw} - \frac{k_v k_w}{2m}\right)\frac{s_v s_w + 1}{2} \tag{71}$$

where $\boldsymbol{A}$ is the adjancency matrix of the graph, k is the node degree, v and w are the nodes' indices and s is a variable that determines if the node belongs to a community or not. Modularity is the non-randomly distributed proportion of the edges that belong to the given cluster in a graph. It is positive if the number of edges within groups exceeds the number expected at random and depends on the chosen method for community detection.

The first algorithm we used for community detection, referred to as 'spectral clustering' algorithm, is based on the fact that modularity of a network is closely related to the structure of the eigenvalue spectrum of its weight matrix [93, 125, 126], high modularity means more strongly embedded communities. This is reflected in the spectra of the connectivity matrices as the separation of eigenvalues into a group with most eigenvalues and another of outliers, the number of which is often used to estimate the number of communities present in the network [93, 125, 126].

The second algorithm, called the Louvain method [98, 124] is a greedy optimization method. First, smaller cliques are found by optimizing modularity locally on all nodes, then each small-sized community is grouped into one node and the first step is repeated. The complete modularity is then calculated by maximizing this value over all the divisions of the network into clusters [98, 124]. We did not find any relevant differences between the Louvain method [98] and the spectral clustering algorithm [93, 125, 126], which were used to define community structure (see Fig 7C).

Supporting information

S1 Fig. The truncated weight change including motifs up to third-order provides a good match for the full weight change including motifs of all orders. The total weight change is calculated either by including motifs up to third-order, or using the full contribution of all motifs based on the integral of Eq 2 (main text) in Fourier space. Each dot represents the respective weight change calculated with the truncated (abscissa) and the full version (ordinate) starting with a random set of initial connectivity weights. The axis are normalized to the maximum weight change. Parameters used are the ones used in the manuscript, except for $N = 12$ (to speed up calculations) and $\eta_- = 13$.
(TIF)

S2 Fig. Spontaneous emergence of assemblies does not depend on the third-order cumulant. Considering only (up to third order) motifs related to the second-order cross-covariance C_{ij} (blue) leads to generally worse graph measures compared to the case when adding motifs from the second-order auto-covariance C_{ii} (red) and the case where all motifs are considered (yellow). **A**. Mean clustering coefficient versus the modulation parameter η_-. **B**. Mean global efficiency versus the modulation parameter η_-. **C**. Mean modularity versus the modulation parameter η_-. All results are calculated from 100 trials at steady state connectivity. Error bars represent the standard error of the mean.
(TIF)

S3 Fig. Network structure at steady state is sensitive to the external input firing rate, but not to heterogeneities in the input firing rates. A. Mean clustering coefficient, mean global efficiency and modularity versus the external input firing rate. Assembly formation breaks down for very large input firing rates. **B**. Mean clustering coefficient, mean global efficiency and modularity versus the standard deviation of firing rate distribution, σ, to introduce heterogeneity in the external input firing rates. Varying σ preserves assembly formation as can be seen from the different graph measures. The mean external input rate was chosen to be 150 Hz. The modulation parameter used is $\eta_- = 13$ and all other parameters are taken as in the main text. Note that the abscissa is logarithmic.
(TIF)

S4 Fig. Departures from the balance of excitation and inhibition do not effect assembly formation. Mean clustering coefficient, mean global efficiency and modularity versus the inhibitory multiplicative factor $\delta_{\text{inh}} = \{0.5, 0.8, 1.2, 1.5\}$, which scales the overall inhibitory matrix (see Methods). Increasing or decreasing δ_{inh} does not disrupt assembly formation as can be seen from the comparison of the different graph measures. The modulation parameter used is $\eta_- = 13$ and all other parameters are taken as in the main text.
(TIF)

S5 Fig. The network converges to steady firing rates during ongoing plasticity. The different curves indicate different external input firing rates and in each case the network converges

to the same rate as the external input firing rate. The modulation parameter used is $\eta_- = 13$ and all other parameters are taken as in the main text.
(TIF)

S1 Text. Fourier transform of the second- and third-order cumulants.
(PDF)

S2 Text. Calculation of motif coefficients up to third-order.
(PDF)

Acknowledgments

We thank members of the "Computation in Neural Circuits" group for useful discussions and Leonidas Richter for reading the manuscript. We thank Neta Ravid Tannenbaum and Yoram Burak for sharing the code used to obtain the connectivity matrices, and to Lilach Avitan and Jan Mölter for sharing further information regarding the methods used for community detection in their publications.

Author Contributions

Conceptualization: Lisandro Montangie, Julijana Gjorgjieva.

Formal analysis: Lisandro Montangie, Christoph Miehl, Julijana Gjorgjieva.

Funding acquisition: Julijana Gjorgjieva.

Investigation: Lisandro Montangie, Christoph Miehl.

Methodology: Lisandro Montangie, Christoph Miehl, Julijana Gjorgjieva.

Software: Lisandro Montangie, Christoph Miehl.

Supervision: Julijana Gjorgjieva.

Visualization: Lisandro Montangie, Christoph Miehl, Julijana Gjorgjieva.

Writing – original draft: Lisandro Montangie, Julijana Gjorgjieva.

Writing – review & editing: Lisandro Montangie, Christoph Miehl, Julijana Gjorgjieva.

References

1. Hebb DO. The Organization of Behavior; a neuropsychological theory. New York: Wiley. 1949.
2. Löwel S, Singer W. Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. Science. 1992; 255(5041):209–212.
3. Shatz CJ. The Developing Brain. Sci Am. 1992; 267(3):60–67. https://doi.org/10.1038/scientificamerican0992-60 PMID: 1502524
4. Singer W. Neuronal representations, assemblies and temporal coherence. Prog Brain Res. 1993; 95:461–474. https://doi.org/10.1016/s0079-6123(08)60388-x PMID: 8493353
5. Ocker GK, Litwin-Kumar A, Doiron B. Self-organization of microcircuits in networks of spiking neurons with plastic synapses. PLoS Comput Biol. 2015; 11(8):e1004458. https://doi.org/10.1371/journal.pcbi.1004458 PMID: 26291697
6. Ravid Tannenbaum N, Burak Y. Shaping neural circuits by high order synaptic interactions. PLoS Comput Bio. 2016; 12(8):e1005056.
7. Carrillo-Reid L, Yang W, Bando Y, Peterka DS, Yuste R. Imprinting and recalling cortical ensembles. Science. 2016; 353(6300):691–694. https://doi.org/10.1126/science.aaf7560 PMID: 27516599
8. Harris KD. Neural signatures of cell assembly organization. Nat Rev Neurosci. 2005; 6:399–407. https://doi.org/10.1038/nrn1669 PMID: 15861182

9. Neves G, Cooke SF, Bliss TVP. Synaptic plasticity, memory and the hippocampus: a neural network approach to causality. Nat Rev Neurosci. 2008; 9:65–75. https://doi.org/10.1038/nrn2303 PMID: 18094707
10. Lansner A. Associative memory models: from the cell-assembly theory to biophysically detailed cortex simulations. Trends Neurosci. 2009; 32(3):178–186. https://doi.org/10.1016/j.tins.2008.12.002
11. Buzsáki G. Neural syntax: Cell assemblies, synapsembles, and readers. Neuron. 2010; 68(3):362–385. https://doi.org/10.1016/j.neuron.2010.09.023 PMID: 21040841
12. Perin R, Berger TK, Markram H. A synaptic organizing principle for cortical neuronal groups. Proc Natl Acad Sci USA. 2011; 108(13):5419–5424. https://doi.org/10.1073/pnas.1016051108 PMID: 21383177
13. Ko H, Hofer SB, Pichler B, Buchanan KA, Sjöström PJ, Mrsic-Flogel TD. Functional specificity of local synaptic connections in neocortical networks. Nature. 2011; 473(7345):87–91. https://doi.org/10.1038/nature09880 PMID: 21478872
14. Ko H, Cossell L, Baragli C, Antolik J, Clopath C, Hofer SB, Mrsic-Flogel TD. The emergence of functional microcircuits in visual cortex. Nature. 2013; 496(7443):96–100. https://doi.org/10.1038/nature12015 PMID: 23552948
15. Holtmaat A, Caroni P. Functional and structural underpinnings of neuronal assembly formation in learning. Nat Neurosci. 2016; 19(12):1553–1562. https://doi.org/10.1038/nn.4418 PMID: 27749830
16. Yoshimura Y, Dantzker JLM, Callaway EM. Excitatory cortical neurons form fine-scale functional networks. Nature. 2005; 433(7028):868–873. https://doi.org/10.1038/nature03252 PMID: 15729343
17. Kampa BM, Letzkus JJ, Stuart GJ. Cortical feed-forward networks for binding different streams of sensory information. Nat Neurosci. 2006; 9(12):1472–1473. https://doi.org/10.1038/nn1798 PMID: 17099707
18. Miner D, Triesch J. Plasticity-driven self-organization under topological constraints accounts for non-random features of cortical synaptic wiring. PLoS Comput Bio. 2016; 12(2):e1004759.
19. Cossell L, Iacaruso MF, Muir DR, Houlton R, Sader EN, Ko H, Hofer SB, Mrsic-Flogel TD. Functional organization of excitatory synaptic strength in primary visual cortex. Nature. 2015; 518(7539):399–403. https://doi.org/10.1038/nature14182 PMID: 25652823
20. Lee WC, Bonin V, Reed M, Graham BJ, Hood G, Glattfelder K, Reid RC. Anatomy and function of an excitatory network in the visual cortex. Nature. 2016; 532(7599):370–374. https://doi.org/10.1038/nature17192 PMID: 27018655
21. Miller JK, Ayzenshtat I, Carrillo-Reid L, Yuste R. Visual stimuli recruit intrinsically generated cortical ensembles. Proc Natl Acad Sci USA. 2014; 111(38):E4053–E4061. https://doi.org/10.1073/pnas.1406077111 PMID: 25201983
22. Kruskal PB, Li L, MacLean JN. Circuit reactivation dynamically regulates synaptic plasticity in neocortex. Nat Commun. 2013; 4:2474.
23. Gebhardt C, Baier H, Del Bene F. Direction selectivity in the visual system of the zebrafish larva. Front Neural Circuits. 2013; 7:111. https://doi.org/10.3389/fncir.2013.00111 PMID: 23785314
24. Romano SA, Pietri T, Pérez-Schuster V, Jouary A, Haudrechy M, Sumbre G. Spontaneous neuronal network dynamics reveal circuit's functional adaptations for behavior. Neuron. 2015; 85(5):1070–1085. https://doi.org/10.1016/j.neuron.2015.01.027 PMID: 25704948
25. Avitan L, Pujic Z, Mölter J, Poll MVD, Sun B, Teng H, Amor R, Scott EK, Goodhill GJ. Spontaneous activity in the zebrafish tectum reorganizes over development and is influenced by visual experience. Curr Biol. 2017; 27(16):2407–2419.e4. https://doi.org/10.1016/j.cub.2017.06.056 PMID: 28781054
26. Pietri T, Romano SA, Pérez-Schuster V, Boulanger-Weill J, Candat V, Sumbre G. The emergence of the spatial structure of tectal spontaneous activity is independent of visual inputs. Cell Rep. 2017; 19(5):939–948. https://doi.org/10.1016/j.celrep.2017.04.015 PMID: 28467907
27. Marachlian E, Avitan L, Goodhill GJ, Sumbre G. Principles of functional circuit connectivity: Insights from spontaneous activity in the zebrafish optic tectum. Front in Neural Circuits. 2018; 12:46.
28. Kriener B, Helias M, Aertsen A, Rotter S. Correlations in spiking neuronal networks with distance dependent connections. J Comput Neurosci. 2009; 27(2):177–200. https://doi.org/10.1007/s10827-008-0135-1 PMID: 19568923
29. Pernice V, Staude B, Cardanobile S, Rotter S. How structure determines correlations in neuronal networks. PLoS Comput Bio. 2011; 7(5):e1002059.
30. Zhao L, Beverlin B, Netoff T, Nykamp DQ. Synchronization from second order network connectivity statistics. Front Comput Neurosci. 2011; 5:28. https://doi.org/10.3389/fncom.2011.00028 PMID: 21779239
31. Roxin A. The role of degree distribution in shaping the dynamics in networks of sparsely connected spiking neurons. Front Comput Neurosci. 2011; 5:8. https://doi.org/10.3389/fncom.2011.00008 PMID: 21556129

32. Gaiteri C, Rubin JE. The Interaction of Intrinsic Dynamics and Network Topology in Determining Network Burst Synchrony. Front Comput Neurosci. 2011; 5:10. https://doi.org/10.3389/fncom.2011.00010 PMID: 21373358

33. Litwin-Kumar A, Doiron B. Slow dynamics and high variability in balanced cortical networks with clustered connections. Nat Neurosci. 2012; 15(11):1498–1505. https://doi.org/10.1038/nn.3220 PMID: 23001062

34. Trousdale J, Hu Y, Shea-Brown E, Josić K. Impact of network structure and cellular response on spike time correlations. PLoS Comput Bio. 2012; 8(3):e1002408.

35. Pernice V, Deger M, Cardanobile S, Rotter S. The relevance of network micro-structure for neural dynamics. Front Comput Neurosci. 2013; 7:72. https://doi.org/10.3389/fncom.2013.00072 PMID: 23761758

36. Helias M, Tetzlaff T, Diesmann M. The correlation structure of local neuronal networks intrinsically results from recurrent dynamics. PLoS Comput Bio. 2014; 10(1):e1003428.

37. Jovanović S, Rotter S. Interplay between graph topology and correlations of third order in spiking neuronal networks. PLoS Comput Biol. 2016; 12(6):e1004963. https://doi.org/10.1371/journal.pcbi.1004963 PMID: 27271768

38. Hu Y, Trousdale J, Josić K, Shea-Brown E. Motif statistics and spike correlations in neuronal networks. J Stat Mech. 2013; P03012.

39. Hu Y, Trousdale J, Josić K, Shea-Brown E. Local paths to global coherence: Cutting networks down to size. Phys Rev E Stat Nonlin Soft Matter Phys. 2014; 89(3):032802. https://doi.org/10.1103/PhysRevE.89.032802 PMID: 24730894

40. Ocker GK, Hu Y, Buice MA, Doiron B, Josić K, Rosenbaum R, Shea-Brown E. From the statistics of connectivity to the statistics of spike times in neuronal networks. Curr Opin Neurobiol. 2017; 46:109–119. https://doi.org/10.1016/j.conb.2017.07.011 PMID: 28863386

41. Dechery JB, MacLean JN. Functional triplet motifs underlie accurate predictions of single-trial responses in populations of tuned and untuned V1 neurons. PLoS Comput Biol. 2018; 14(5):e1006153. https://doi.org/10.1371/journal.pcbi.1006153 PMID: 29727448

42. Gilson M, Burkitt AN, Grayden DB, Thomas DA, van Hemmen JL. Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. I. Input selectivity–strengthening correlated input pathways. Biol Cybern. 2009; 101(2):81–102. https://doi.org/10.1007/s00422-009-0319-4 PMID: 19536560

43. Gilson M, Burkitt AN, Grayden DB, Thomas DA, van Hemmen JL. Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. II. Input selectivity—symmetry breaking. Biol Cybern. 2009; 101(2):103–114. https://doi.org/10.1007/s00422-009-0320-y PMID: 19536559

44. Gilson M, Burkitt AN, Grayden DB, Thomas DA, van Hemmen JL. Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks IV: structuring synaptic pathways among recurrent connections. Biol Cybern. 2009; 101(5-6):427–444. https://doi.org/10.1007/s00422-009-0346-1 PMID: 19937070

45. Burkitt AN, Gilson M, van Hemmen JL. Spike-timing-dependent plasticity for neurons with recurrent connections. Biol Cybern. 2007; 96(5):533–546. https://doi.org/10.1007/s00422-007-0148-2 PMID: 17415586

46. Hiratani N, Fukai T. Interplay between Short- and Long-Term Plasticity in Cell-Assembly Formation. PLoS One. 2014; 9(7):e101535. https://doi.org/10.1371/journal.pone.0101535 PMID: 25007209

47. Vasilaki E, Giugliano M. Emergence of Connectivity Motifs in Networks of Model Neurons with Short- and Long-Term Plastic Synapses. PLoS One. 2014; 9(1):e84626. https://doi.org/10.1371/journal.pone.0084626 PMID: 24454735

48. Triplett MA, Avitan L, Goodhill GJ. Emergence of spontaneous assembly activity in developing neural networks without afferent input. PLoS Comput Biol. 2018; 14(9):e1006421. https://doi.org/10.1371/journal.pcbi.1006421 PMID: 30265665

49. Litwin-Kumar A, Doiron B. Formation and maintenance of neuronal assemblies through synaptic plasticity. Nat Commun. 2014; 5:5319. https://doi.org/10.1038/ncomms6319 PMID: 25395015

50. Zenke F, Agnes EJ, Gerstner W. Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. Nat Commun. 2015; 6:6922. https://doi.org/10.1038/ncomms7922 PMID: 25897632

51. Ocker GK, Doiron B. Training and spontaneous reinforcement of neuronal assemblies by spike timing plasticity. Cereb Cortex. 2019; 29(3):937–951. https://doi.org/10.1093/cercor/bhy001 PMID: 29415191

52. Kempter R, Gerstner W, van Hemmen JL. Hebbian learning and spiking neurons. Physical Review E. 1999; 59(4):4498–4514.

53. Markram H, Gerstner W, Sjöström PJ. A history of spike-timing-dependent plasticity. Front Synaptic Neurosci. 2011; 3:4. https://doi.org/10.3389/fnsyn.2011.00004 PMID: 22007168

54. Pfister JP, Gerstner W. Triplets of spikes in a model of spike timing-dependent plasticity. J Neurosci. 2006; 26(38):9673–9682. https://doi.org/10.1523/JNEUROSCI.1425-06.2006 PMID: 16988038

55. Sjöström PJ, Turrigiano GG, Nelson SB. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. Neuron. 2001; 32(6):1149–1164. https://doi.org/10.1016/s0896-6273(01)00542-6 PMID: 11754844

56. Wang HX, Gerkin RC, Nauen DW, Bi GQ. Coactivation and timing-dependent integration of synaptic potentiation and depression. Nat Neurosci. 2005; 8(2):187–193. https://doi.org/10.1038/nn1387 PMID: 15657596

57. Gjorgjieva J, Clopath C, Audet J, Pfister JP. A triplet spike-timing-dependent plasticity model generalizes the Bienenstock-Cooper-Munro rule to higher-order spatiotemporal correlations. Proc Natl Acad Sci USA. 2011; 108(48):19383–19388. https://doi.org/10.1073/pnas.1105933108 PMID: 22080608

58. Ohiorhenuan IE, Mechler F, Purpura KP, Schmid AM, Hu Q, Victor JD. Sparse coding and high-order correlations in fine-scale cortical networks. Nature. 2010; 466(7306):617–621. https://doi.org/10.1038/nature09178 PMID: 20601940

59. Montani F, Ince RAA, Senatore R, Arabzadeh E, Diamond ME, Panzeri S. The impact of high-order interactions on the rate of synchronous discharge and information transmission in somatosensory cortex. Philos Trans A Math Phys Eng Sci. 2009; 367(1901):3297–3310. https://doi.org/10.1098/rsta.2009.0082 PMID: 19620125

60. Ohiorhenuan IE, Victor JD. Information-geometric measure of 3-neuron firing patterns characterizes scale-dependence in cortical networks. J Comput Neurosci. 2010; 30(1):125–141. https://doi.org/10.1007/s10827-010-0257-0 PMID: 20635129

61. Yu S, Yang H, Nakahara H, Santos GS, Nikolic D, Plenz D. Higher-order interactions characterized in cortical activity. J Neuroscience. 2011; 31(48):17514–17526.

62. Bohte SM, Spekreijse H, Roelfsema PR. The effects of pair-wise and higher order correlations on the firing rate of a post-synaptic neuron. Neural Comput. 2000; 12(1):153–179. https://doi.org/10.1162/089976600300015934 PMID: 10636937

63. Zylberberg J, Shea-Brown E. Input nonlinearities can shape beyond-pairwise correlations and improve information transmission by neural populations. Phys Rev E Stat Nonlin Soft Matter Phys. 2015; 92(6):062707. https://doi.org/10.1103/PhysRevE.92.062707 PMID: 26764727

64. Barreiro AK, Gjorgjieva J, Rieke F, Shea-Brown E. When do microcircuits produce beyond-pairwise correlations? Front Comput Neurosci. 2014; 8:10. https://doi.org/10.3389/fncom.2014.00010 PMID: 24567715

65. Cayco-Gajic NA, Zylberberg J, Shea-Brown E. Triplet correlations among similarly tuned cells impact population coding. Front Comput Neurosci. 2015; 9:57. https://doi.org/10.3389/fncom.2015.00057 PMID: 26042024

66. Amari S, Nakahara H, Wu S, Sakai Y. Synchronous firing and higher-order interactions in neuron pool. Neural Comput. 2003; 15(1):127–142. https://doi.org/10.1162/089976603321043720 PMID: 12590822

67. Macke JH, Opper M, Bethge M. Common input explains higher-order correlations and entropy in a simple model of neural population activity. Phys Rev Lett. 2011; 106(20):208102. https://doi.org/10.1103/PhysRevLett.106.208102 PMID: 21668265

68. Montangie L, Montani F. Quantifying higher-order correlations in a neuronal pool. Physica A: Statistical Mechanics and its Applications. 2015; 421:388–400.

69. Montangie L, Montani F. Common inputs in subthreshold membrane potential: The role of quiescent states in neuronal activity. Phys Rev E. 2018; 97(6-1):060302. https://doi.org/10.1103/PhysRevE.97.060302 PMID: 30011540

70. Hawkes AG. Spectra of some self-exciting and mutually exciting point processes. Biometrika. 1971; 58(1):83–90.

71. Bi GQ, Poo MM. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. J Neurosci. 1998; 18(24):10464–10472. https://doi.org/10.1523/JNEUROSCI.18-24-10464.1998 PMID: 9852584

72. Brette R. Philosophy of the spike: Rate-based vs. spike-based theories of the brain. Front Syst Neurosci. 2015; 9:151. https://doi.org/10.3389/fnsys.2015.00151 PMID: 26617496

73. Gjorgjieva J, Drion G, Marder E. Computational implications of biophysical diversity and multiple timescales in neurons and synapses for circuit performance. Curr Opin Neurobiol. 2016; 37:44–52. https://doi.org/10.1016/j.conb.2015.12.008 PMID: 26774694

74. Hansel D, Mato G. Asynchronous states and the emergence of synchrony in large networks of interacting excitatory and inhibitory neurons. Neural Comput. 2003; 15(1):1–56. https://doi.org/10.1162/089976603321043685 PMID: 12590818

75. Cessac B, Viéville T. On dynamics of integrate-and-fire neural networks with conductance based synapses. Frontiers in Computational Neuroscience. 2008; 2:2. https://doi.org/10.3389/neuro.10.002.2008 PMID: 18946532

76. Renart A, de la Rocha J, Bartho P, Hollender L, Parga N, Reyes A, Harris KD. The asynchronous state in cortical circuits. Science. 2010; 327(5965):587–590. https://doi.org/10.1126/science.1179850 PMID: 20110507

77. Vogels TP, Sprekeler H, Zenke F, Clopath C, Gerstner W. Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. Science. 2011; 334(6062):1569–1573. https://doi.org/10.1126/science.1211095 PMID: 22075724

78. Luz Y, Shamir M. Balancing feed-forward excitation and inhibition via Hebbian inhibitory synaptic plasticity. PLoS Comput Biol. 2012; 8(1):e1002334. https://doi.org/10.1371/journal.pcbi.1002334 PMID: 22291583

79. Rubin R, Abbott LF, Sompolinsky H. Balanced excitation and inhibition are required for high-capacity, noise-robust neuronal selectivity. Proc Natl Acad Sci USA. 2017; 114(44):E9366–E9375. https://doi.org/10.1073/pnas.1705841114 PMID: 29042519

80. Fiete IR, Senn W, Wang CZH, Hahnloser RHR. Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. Neuron. 2010; 65(4):563–576. https://doi.org/10.1016/j.neuron.2010.02.003 PMID: 20188660

81. Field R, D'amour J, Tremblay R, Miehl C, Rudy B, Gjorgjieva J, Froemke RC. Heterosynaptic plasticity determines the set-point for cortical excitatory-inhibitory balance. Neuron. 2020; 106:1–13.

82. Froemke RC, Poo M, Dan Y. Spike-timing-dependent synaptic plasticity depends on dendritic location. Nature. 2005; 434(7030):221–225. https://doi.org/10.1038/nature03366 PMID: 15759002

83. Cassenaer S, Laurent G. Conditional modulation of spike-timing-dependent plasticity for olfactory learning. Nature. 2012; 482(7383):47–52. https://doi.org/10.1038/nature10776 PMID: 22278062

84. Frémaux N, Gerstner W. Neuromodulated spike-timing-dependent plasticity, and theory of three-factor learning rules. Front Neural Circuits. 2016; 9:85. https://doi.org/10.3389/fncir.2015.00085 PMID: 26834568

85. Pedrosa V, Clopath C. The role of neuromodulators in cortical plasticity. A computational perspective. Front Synaptic Neurosci. 2017; 8:38. https://doi.org/10.3389/fnsyn.2016.00038 PMID: 28119596

86. Foncelle A, Mendes A, Jędrzejewska-Szmek J, Valtcheva S, Berry H, Blackwell KT, Venance L. Modulation of spike-timing dependent plasticity: Towards the inclusion of a third factor in computational models. Front Comput Neurosci. 2018; 12:49. https://doi.org/10.3389/fncom.2018.00049 PMID: 30018546

87. Bissière S, Humeau Y, Lüthi A. Dopamine gates LTP induction in lateral amygdala by suppressing feedforward inhibition. Nat Neurosci. 2003; 6(6):587–592. https://doi.org/10.1038/nn1058 PMID: 12740581

88. Seol GH, Ziburkus J, Huang S, Song L, Kim IT, Takamiya K, Huganir RL, Lee HK, Kirkwood A. Neuromodulators control the polarity of spike-timing-dependent synaptic plasticity. Neuron. 2007; 55(6):919–929. https://doi.org/10.1016/j.neuron.2007.08.013 PMID: 17880895

89. Clopath C, Büsing L, Vasilaki E, Gerstner W. Connectivity reflects coding: a model of voltage-based STDP with homeostasis. Nat Neurosci. 2010; 13(3):344–352. https://doi.org/10.1038/nn.2479 PMID: 20098420

90. Gjorgjieva J, Evers JF, Eglen SJ. Homeostatic activity-dependent tuning of recurrent networks for robust propagation of activity. J Neurosci. 2016; 36(13):3722–3734. https://doi.org/10.1523/JNEUROSCI.2511-15.2016 PMID: 27030758

91. Babadi B, Abbott LF. Pairwise analysis can account for network structures arising from Spike-Timing Dependent Plasticity. PLoS Comput Biol. 2013; 9(2):e1002906. https://doi.org/10.1371/journal.pcbi.1002906 PMID: 23436986

92. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature. 1998; 393(6684):440–442. https://doi.org/10.1038/30918 PMID: 9623998

93. Newman MEJ. Modularity and community structure in networks. Proc Natl Acad Sci USA. 2006; 103(23):8577–8582. https://doi.org/10.1073/pnas.0601602103 PMID: 16723398

94. Betzel R, Wood KC, Angeloni C, Geffen MN, Bassett DS. Stability of spontaneous, correlated activity in mouse auditory cortex. PLoS Comput Biol. 2019; 15(12):e1007360. https://doi.org/10.1371/journal.pcbi.1007360 PMID: 31815941

95. Fagiolo G. Clustering in complex directed networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2007; 76(2):026107. https://doi.org/10.1103/PhysRevE.76.026107 PMID: 17930104

96. Latora V, Marchiori M. Efficient behavior of small-world networks. Phys Rev Lett. 2001; 87 (19):198701. https://doi.org/10.1103/PhysRevLett.87.198701 PMID: 11690461

97. Reichardt J, Bornholdt S. Statistical mechanics of community detection. Phys Rev E Stat Nonlin Soft Matter Phys. 2006; 74(1):016110. https://doi.org/10.1103/PhysRevE.74.016110 PMID: 16907154

98. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. 2008; 10:P10008.

99. Feldmann DE. Timing-Based LTP and LTD at Vertical Inputs to Layer II/III Pyramidal Cells in Rat Barrel Cortex. Neuron. 2000; 27(1):45–56.

100. Graupner M, Wallisch P, Ostojic S. Natural Firing Patterns Imply Low Sensitivity of Synaptic Plasticity to Spike Timing Compared with Firing Rate. J Neurosci. 2016; 36(44):11238–11258. https://doi.org/10.1523/JNEUROSCI.0104-16.2016 PMID: 27807166

101. Gilson M, Burkitt AN, Grayden DB, Thomas DA, van Hemmen JL. Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks III: Partially connected neurons driven by spontaneous activity. Biol Cybern. 2009; 101(5-6):411–426. https://doi.org/10.1007/s00422-009-0343-4 PMID: 19937071

102. Jovanović S, Hertz J, Rotter S. Cumulants of Hawkes point processes. Phys Rev E Stat Nonlin Soft Matter Phys. 2015; 91(4):042802. https://doi.org/10.1103/PhysRevE.91.042802 PMID: 25974542

103. Luczak A, Barthó P, Marguet SL, Buzsaki G, Harris KD. Sequential structure of neocortical spontaneous activity in vivo. Proc Natl Acad Sci USA. 2007; 104(1):347–352. https://doi.org/10.1073/pnas.0605643104 PMID: 17185420

104. Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP. Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature. 2008; 454(7207):995–999. https://doi.org/10.1038/nature07140 PMID: 18650810

105. Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature. 1996; 381(6583):607–609. https://doi.org/10.1038/381607a0 PMID: 8637596

106. Simoncelli EP, Olhausen BA. Natural Image Statistics and Neural Representation. Annu Rev Neurosci. 2001; 24:1193–1216. https://doi.org/10.1146/annurev.neuro.24.1.1193 PMID: 11520932

107. Bell AJ, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. Neural Comp. 1995; 7(6):1129–1159.

108. Intrator N, Cooper LN. Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions. Neural Networks. 1992; 5(1):3–17.

109. Blais BS, Intrator N, Shouval H, Cooper L. Receptive field formation in natural scene environments. Comparison of single-cell learning rules. Neural Comp. 1998; 10(7):1797–1813.

110. Song S, Sjöström PJ, Reigl M, Nelson S, Chklovskii DB. Highly nonrandom features of synaptic connectivity in local cortical circuits. PLoS Biol. 2005; 3(10):e350.

111. Lichtman J, Helmstaedter M. Connectomics at the cutting edge: Challenges and opportunities in high-resolution brain mapping. Science. 2014; 346(6209):651–651.

112. Swanson LW, Lichtman JW. From Cajal to Connectome and Beyond. Ann Rev Neurosci. 2016; 39:197–216. https://doi.org/10.1146/annurev-neuro-071714-033954 PMID: 27442070

113. Blankenship AG, Feller MB. Mechanisms underlying spontaneous patterned activity in developing neural circuits. Nat Rev Neurosci. 2010; 11(1):18–29. https://doi.org/10.1038/nrn2759 PMID: 19953103

114. Ackman JB, Burbridge TH, Crair MC. Retinal waves coordinate patterned activity throughout the developing visual system. Nature. 2012; 490(7419):219–225. https://doi.org/10.1038/nature11529 PMID: 23060192

115. Richter LM, Gjorgjieva J. Understanding neural circuit development through theory and models. Curr Opin in Neurobiol. 2017; 46:39–47.

116. Golshani P, Gonçalves JT, Khoshkhoo S, Mostany R, Smirnakis S, Portera-Cailliau C. Internally Mediated Developmental Desynchronization of Neocortical Network Activity. J Neurosci. 2009; 29 (35):10890–10899. https://doi.org/10.1523/JNEUROSCI.2012-09.2009 PMID: 19726647

117. Rochefort NL, Garaschuk O, Milos RI, Narushima M, Marandi N, Pichler B, Kovalchuk Y, Konnerth A. Sparsification of neuronal activity in the visual cortex at eye-opening. Proc Natl Acad Sci USA. 2009; 106(35):15049–15054. https://doi.org/10.1073/pnas.0907660106 PMID: 19706480

118. Butts DA, Kanold PO, Shatz CJ. A Burst-Based "Hebbian" Learning Rule at Retinogeniculate Synapses Links Retinal Waves to Activity-Dependent Refinement. PLoS Biol. 2007; 5(3):e61. https://doi.org/10.1371/journal.pbio.0050061 PMID: 17341130

119. Winnubst J, Cheyne JE, Niculescu D, Lohmann C. Spontaneous Activity Drives Local Synaptic Plasticity In Vivo. Neuron. 2015; 87(2):399–410. https://doi.org/10.1016/j.neuron.2015.06.029 PMID: 26182421

120. Miconi T, McKinstry JL, Edelman GM. Spontaneous emergence of fast attractor dynamics in a model of developing primary visual cortex. Nat Commun. 2016; 7:13208. https://doi.org/10.1038/ncomms13208 PMID: 27796298

121. Gerstner W, Kempter R, van Hemmen JL, Wagner H. A neuronal learning rule for sub-millisecond temporal coding. Nature. 1996; 383(6595):76–78. https://doi.org/10.1038/383076a0 PMID: 8779718

122. Kistler WM, van Hemmen JL. Modeling synaptic plasticity in conjunction with the timing of pre- and postsynaptic action potentials. Neural Comput. 2000; 12(2):385–405. https://doi.org/10.1162/089976600300015844 PMID: 10636948

123. Song S, Miller KD, Abbott LF. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. Nat Neurosci. 2000; 3(9):919–926. https://doi.org/10.1038/78829 PMID: 10966623

124. Rubinov M, Sporns O. Complex network measures of brain connectivity: Uses and interpretations. Neuroimage. 2010; 52(3):1059–1069. https://doi.org/10.1016/j.neuroimage.2009.10.003 PMID: 19819337

125. Leicht EA, Newman MEJ. Community structure in directed networks. Phys Rev Lett. 2008; 100(11):118703. https://doi.org/10.1103/PhysRevLett.100.118703 PMID: 18517839

126. Nadakuditi RR, Newman MEJ. Spectra of random graphs with arbitrary expected degrees. Phys Rev E Stat Nonlin Soft Matter Phys. 2013; 87(1):012803. https://doi.org/10.1103/PhysRevE.87.012803 PMID: 23410384