Homeostatic and functional implications of interneuron plasticity

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)
im Fach Biologie

eingereicht an der
Lebenswissenschaftlichen Fakultät
Humboldt-Universität zu Berlin

von

Owen John Mackwood
M.Sc.

Präsidentin der Humboldt-Universität zu Berlin:
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät:
Prof. Dr. Bernhard Grimm

Gutachter:
1. Prof. Dr. Henning Sprekeler
2. Prof. Dr. Richard Kempter
3. Dr. Timothy O'Leary

Eingereicht am 26. Juni 2018
Tag der mündlichen Prüfung: 21. September 2018

For Maja.

Acknowledgements

During the course of my doctoral studies, many people played key roles in directly helping, or simply inspiring me. Above all, I thank my supervisor and scientific mentor Henning Sprekeler. Over the course of this research his guidance, support, and patience have proven inexhaustible. In moments when I felt uninspired or hopeless, even a brief conversation about our work invariably renewed my excitement and motivated me to redouble my efforts. No matter my state of mind, his enthusiasm for science would remind me of why I had undertaken this effort. I now understand why in German a doctoral supervisor is known as a *Doktorvater*.

Of course this work was supported in other ways, often through conversation and friendship. In our lab, many people have proved excellent comrades-in-arms, including Cathrin Bunkelmann, our postdocs Loreen Hertäg, David Higgins, Richard Naud, Filip Vercruysse, and fellow doctoral students Joram Keijser, Laura Naumann, and Mathias Schmerling. A special mention is reserved for Simon Weber, the first person to join the lab after myself. Having gone through this multi-year process together, we have developed a unique friendship that has been truly rewarding.

My long-standing connection to the Bernstein Center for Computational Neuroscience has proven important to my completion of this thesis. The administrative support from Margret Franke and Robert Martin has been invaluable. I also want to thank past administrators and teachers throughout the masters program, all of whom laid the foundation for this doctoral research.

The willingness to volunteer their time and energy to review this thesis earns Richard Kempter and Timothy O'Leary my deep gratitude. I also extend my sincerest thanks to Michael Brecht and Matthew Larkum for agreeing to be part of my defence committee.

Without my family, none of this would have been possible. All of my parents, including those I had at birth and those that entered my life later, have always offered their unconditional love and support. My extended family has also played an important role, with Peter Sharpe in particular providing encouragement and generous material support.

I want to extend a special thanks to my friends, who are too plentiful to list here. I will single out Melanie Conrad and Lyle Crawford, who have served as great inspirations and always motivate me to strive for greater intellectual heights.

Most importantly, I thank Maja Svartåker. She has been more loving and tolerant through this process than I could have possibly hoped. She has kept me healthy and sane, and I am eternally indebted to her.

Abstract

Preserving brain function despite ongoing changes inside the organism, and out in the world, necessitates homeostatic mechanisms. Inhibitory interneurons play a key role in both computation and homeostasis within the brain. However, it remains unclear if there is a mechanism that can account for both of these properties. This thesis therefore aims to determine the homeostatic capabilities of such interneurons and elucidate the resulting computational consequences, using analytical and numerical techniques.

The central hypothesis of this thesis is that some interneurons slowly modulate their firing rates to maintain the long-term activity of excitatory neurons at a homeostatic set-point. Thus we begin with a normative approach, deriving a plasticity rule that regulates the activity of interneurons to minimise network-wide deviations from that set-point. In the interest of biological plausibility we also provide two approximations, both of which make each interneuron responsive to the excitatory population it inhibits, and show that all three variants exhibit comparable though distinct homeostatic capabilities. We contrast this normative approach by characterising the homeostatic properties of rules which instead alter the activity of an interneuron when the neurons that drive it deviate from the set-point. Those rules induce a competition between neurons, causing network activity to become sparse.

In the second part of this thesis, we investigate how one of the approximate rules affects computational properties of sensory cortex. We show that it can account for several experimentally reported results, including co-tuning of excitatory and inhibitory currents, and the development of excitatory-inhibitory cell assemblies.

In summation, this thesis provides new insight into how regulating interneuron activity can be homeostatic for neuronal networks, and reveals potential implications for development and preservation of brain function.

Zusammenfassung

Die Erhaltung der Gehirnfunktion trotz Veränderungen im Organismus und dessen Umwelt erfordert homöostatische Mechanismen. Inhibitorische Interneurone spielen eine Schlüsselrolle bei Berechnungen und Homöostase im Gehirn. Es ist jedoch unklar, welcher Mechanismus diese Eigenschaften erzeugen kann. Diese Arbeit hat das Ziel, die homöostatischen Fähigkeiten solcher Interneurone zu bestimmen und die daraus resultierenden funktionellen Konsequenzen mit analytischen und numerischen Techniken zu ergründen.

Die zentrale Hypothese dieser Arbeit ist, dass Interneurone ihre Feuerraten modulieren, um langfristig die Aktivität exzitatorischer Neurone bei einem homöostatischen Sollwert zu halten. Wir beginnen mit einem normativen Ansatz und leiten eine Plastizitätsregel her, welche die Aktivität von Interneuronen regelt, um netzwerkweite Abweichungen vom Sollwert zu minimieren. Um die biologische Plausibilität zu erhöhen, liefern wir zwei Approximationen, bei denen jede Interneurone auf die exzitatorische Population reagiert, die sie inhibiert und zeigen, dass alle drei Varianten vergleichbare aber unterschiedliche homöostatische Fähigkeiten haben. Wir kontrastieren den normativen Ansatz mit Regeln, welche die Aktivität einer Interneurone verändern, wenn die Neuronen, die sie treiben, vom Sollwert abweichen. Diese Regeln erzeugen Konkurrenz zwischen Neuronen und führen daher zu zerstreuter Netzwerkaktivität.

Im zweiten Teil dieser Arbeit untersuchen wir, wie eine der approximierten Regeln die funktionellen Eigenschaften des sensorischen Kortex beeinflusst. Wir zeigen, dass sie mehrere experimentell Beobachtungen erklären kann, inklusive des Ko-Tunings von exzitatorischen und inhibitorischen Strömen und der Entwicklung von Zellverbänden.

Zusammenfassend liefert diese Arbeit neue Erkenntnisse darüber, wie die Regulierung der Interneuron-Aktivität für neuronale Netzwerke homöostatisch sein kann, und zeigt mögliche Auswirkungen auf die Entwicklung und Erhaltung der Gehirnfunktion auf.

Contents

1 Introduction to the thesis

"Plus ça change, plus c'est la même chose."
— Jean-Baptiste Alphonse Karr, Les Guêpes

It is remarkable that animals can change so drastically, but nevertheless tend to remain in a stable equilibrium. This adaptability is often separated into two largely distinct processes: learning and homeostasis. Learning is how we adjust our behaviours, ideally becoming more successful in our interactions with the world. Homeostasis, on the other hand, is the preservation of the organism in a healthy state. At first glance, learning seems the more interesting process, with homeostasis trailing along afterwards simply doing the house cleaning. But if one carefully considers the nature of homeostasis, it becomes apparent that homeostasis is inextricably intertwined with learning. And those two processes operate hand-in-glove within the brain.

Within the mammalian brain, neurons are constantly buffeted by changes both originating from the outside world, and arising in the brain itself. Adapting to those changes requires both immediate response, relying on transforming sensory inputs into action, and long term responses that rely upon learning and concomitant homeostatic processes. The ability to make long lasting adaptive changes is broadly known as plasticity, and virtually every functional aspect of a neuron is plastic in some fashion. The complexity of such plasticity is breathtaking, with the rules dictating its effect varying over developmental phases, across brain regions, neuron type, and for synapses—the connections between neurons—it even depends on the type of neuron on either side of the synapse.

The broadest classification of the neurons in the brain is according to the effect they have on the cells they communicate with. The two largest classes are excitatory neurons and inhibitory neurons, so named because they excite or inhibit the cells that receive their synaptic output. The central role of excitatory neurons in all aspects of brain function has long been appreciated and extensively studied. Inhibitory cells were historically not nearly as well studied, and their role was poorly understood. It is only in recent decades that investigators have begun to piece together a comprehensive picture of how inhibitory neurons contribute to brain function, including sensory processing.

Whatever their role, how do inhibitory neurons come to develop the functional properties they need? And how do they contribute to homeostasis in the brain? Although there are some hints within the published literature, these are both essentially open questions. This thesis moves towards an answer by investigating a type of plasticity that regulates inhibitory neuron activity, in a manner that should be homeostatic for excitatory cells. We use mathematical and computational techniques to design and then explore the effects of such plasticity, both on the homeostasis of excitatory activity, as well as on sensory processing.

Structure of the thesis

The thesis is structured as follows:

In Chapter 2, we first introduce the many roles that inhibitory neurons play in the brain. We then describe what is currently known about the plasticity of inhibitory activity, including homeostatic and other forms of plasticity.

In Chapter 3, the aims of this thesis are stated.

The methods in Chapter 4 contain a comprehensive description of the computational models used to obtain our results, and the mathematical tools used to analyse them.

The results of our investigation are presented in Chapter 5, with the homeostatic capabilities of inhibitory neurons described in Section 5.1. There we show that to be effective homeostatic controllers of excitatory firing rates, interneurons should be responsive to the activity of cells that they inhibit, whereas if they are responsive to cell providing excitatory drive to them it induces a competition between those cells. We then describe the effect that one such plasticity rule of inhibitory activity has on sensory networks in Section 5.2, wherein we replicate certain experimental results.

These results are discussed at length, and placed in the broader context of existing research in Chapter 6.

Finally, Chapter 7 contains the mathematical derivation of the plasticity rule that serves as the foundation of much of the work presented here.

2 Background

Within the mammalian neocortex, neuronal networks are largely comprised of excitatory cells which release the neurotransmitter glutamate, and inhibitory cells which release γ-amino-butyric acid (GABA). Historically, excitatory principal cells were the focus of much of the research conducted, mainly for two reasons: First, they were easier to study. Their cell bodies and axonal processes—which carries their output to other cells—are larger than those of inhibitory interneurons, and they outnumber them at least 4-to-1, making them more accessible to the electrode-based recording techniques used since the birth of modern neuroscience (Graham and Gerard, 1946; Renshaw et al., 1940). Second, because principal cells have long-range axons, their activity was held to be the output of the local circuit, and therefore the computational result of interest.

Inhibitory neurons (INs) typically have short axons that predominantly target the local population—thus the common name inhibitory *interneuron*, and it was thought their primary role was providing negative feedback to excitatory cells, to control runaway activity. Despite so little being known about their function, researchers dating back as far as Ramón y Cajal (1923) had suspected that interneurons are what enable cortex to manifest such extraordinary computational capabilities.

2.1 The role of inhibitory neurons in cortex

One challenge in trying to understand the role of INs is their diversity. While excitatory cells typically have a highly stereotyped morphology (hence the terms *pyramidal* and *stellate* cell), inhibitory neurons exhibit a more diverse range of shapes and projections. This variability, first reported in the ground breaking work of Cajal (1899), was discovered by viewing Golgi stained cells through a light microscope (though his studies predated the discovery that interneurons are often inhibitory).

As experimental techniques advanced, more methods for characterising INs became available. Electron microscopy permitted detailed mapping of synaptic connectivity (Peters et al., 1976; Somogyi et al., 1982), and thereby enabled an early classification according to targets of their synapses: Interneurons can be soma-targeting, forming synapses on the cell body, basal dendrites, or axon initial segment, all of which can control the output of cells. Or they can be dendrite-targeting, forming synapses either on the dendritic shaft—which allows them to control integration of nearby inputs—or on dendritic spines (thorn-like protuberances) to affect input only on that spine. This distinction immediately pointed to unique functional roles for certain INs.

Although the electrophysiological measurement of neuronal intrinsic properties has a long history, it was the advent of experiments employing brain slices that removed barriers to high quality recordings (McCormick et al., 1985). These experiments revealed markedly different responses to current injection, permitting further classification. Interneurons can be fast-spiking (FS), regular-spiking (RS), late-spiking (LS) amongst

other categories (Markram et al., 2004). All of these differences have computational consequences that are not yet fully understood, and remain the topic of ongoing research.

More recently genetic and molecular markers enabled yet further classification. Various calcium-binding proteins, such as parvalbumin (PV), calbindin (CB), and calretinin (CR), as well as neuropeptides such as vasoactive intestinal peptide (VIP), somatostatin (SOM) and cholecystokinin (CCK), amongst others are expressed within INs of different morphologies and firing properties. These neurochemical indicators are especially important, as they bring to bear powerful experimental tools. Those tools have enabled the visualisation and functional manipulation of specific INs within living tissue.

Although the aforementioned methods of classification would seem a boon for investigators, it has created problems in defining what a "class" of interneuron is. Some have even argued that there are no well-defined groups of interneurons, but rather a continuum of traits (Parra et al., 1998), which would make ascribing any function that generalises to a subset of interneurons seem a fruitless endeavour. There is some reason to be concerned, as none of the traits uniquely predict others—e.g. the calcium-binding protein CR has been reported in cells that exhibit different morphologies, synaptic target domains and firing properties (Kepecs and Fishell, 2014).

To illustrate this complexity, consider the morphological categories of INs, which include basket cells (soma- and proximal dendrite-targeting), bouquet and bipolar cells (dendrite-targeting), chandelier cells (axon-targeting), to name just a few (Markram et al., 2004). Of these categories, basket cells are the most common, with three main subclasses defined by their dendritic and axonal morphologies: large basket cells, small basket cells and nest basket cells. Large basket cells (LBCs) have extensive axonal arborizations that provide lateral inhibition to other cortical layers and columns (the proposed computational unit of the cortex; Lorente de Nó, 1949). They can express CB, PV or other markers, though never VIP. Small basket cells (SBCs) have axonal arborizations that rarely extend outside a layer or column, and express VIP. Nest basket cells (NBCs) can appear to have a mixture of LBC and SBC morphologies, and express PV or CB amongst other markers, but never VIP.

Contemplating the full complexity of interneuron classification can be discouraging, but fortunately there is an emerging consensus on what constitutes a broadly applicable categorization for INs. The three major families are PV, SOM, and 5-hydroxytryptamine-receptor expressing interneurons, with the latter including VIP expressing INs (Harris and Mrsic-Flogel, 2013). Those three groups account for nearly 100% of neocortical INs (Rudy et al., 2011). Surely the most widely investigated class of interneurons are PV-positive cells. These cells are the largest population of inhibitory neurons and although the data are conflicted, may represent 40-50% of non-pyramidal cells (Kubota, 2014; Rudy et al., 2011). Along with SOM and VIP expressing INs, they are one of the essential components of the canonical neocortical microcircuit.

Inhibitory motifs and the canonical neocortical microcircuit

Before describing the canonical neocortical microcircuit, it is important to establish the most basic connection motifs between inhibitory and excitatory cells. Inhibition

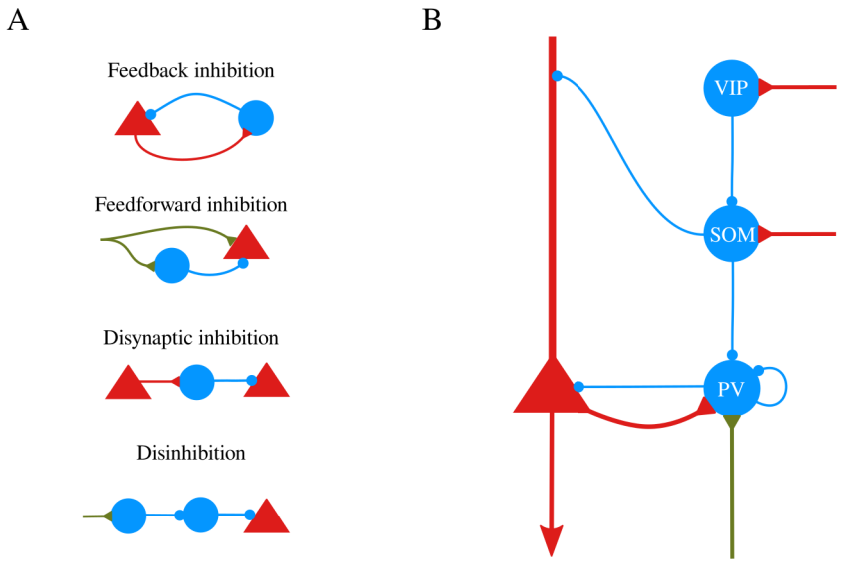

Figure 2.1: **Motifs and the canonical microcircuit.** Red triangles are excitatory principal cells, blue circles are inhibitory neurons. Olive lines indicate afferent excitatory connections. **A)** Inhibitory connection motifs. **B)** The canonical cortical microcircuit. See the text for a description of both.

can operate in a feedback, feedforward, disynaptic, or disinhibitory manner (Fino et al., 2013). These modes are illustrated by connectivity motifs in Fig 2.1A. *Feedback inhibition* occurs when excitatory cells are reciprocally connected with local INs (this is also known as recurrent inhibition). Due to the excitatory-to-inhibitory connection, IN firing rates rise and fall with excitatory rates, serving to counteract the destabilising positive feedback of recurrent excitatory-to-excitatory connections (Douglas et al., 1995). *Feedforward inhibition* occurs when afferent excitatory inputs from another population drive both excitatory and inhibitory cells in the local population. The INs also project onto the local excitatory cells, and thus the afferent drive is effectively both excitatory and inhibitory. *Disynaptic inhibition* occurs when an excitatory cell drives local INs that are not reciprocally connected with it, but that provide inhibition to other excitatory cells. Finally, *disinhibition* occurs when an excitatory cell is inhibited by an IN, which is in turn inhibited by yet another IN. When the latter interneuron fires it suppresses activity in the former IN which relieves the excitatory cell from inhibition, thus disinhibiting it.

All of these inhibitory motifs are present in what is known as the canonical microcircuit (Fig 2.1B). The name results from its pervasiveness in neocortex, with variations on it being found throughout both the sensory and association cortices (Creutzfeldt, 1977; Douglas and Martin, 2004). The microcircuit as described by Harris and Mrsic-Flogel (2013) is comprised of excitatory principal neurons (PNs), and inhibitory cells from the three major families of INs (for the sake of simplicity some types of interneuron in the microcircuit are excluded here). PV fast-spiking interneurons target the soma and proximal dendrites of PNs. They receive strong excitatory input from cortex and thalamus, and are recurrently connected amongst themselves. Interneurons expressing

SOM (SOMs) provide inhibition to other interneurons, as well as the distal dendrites of PNs. They receive most of their input from local PNs and VIPs, with little thalamic drive. The last component that we will consider here are the VIPs (interneurons expressing VIP). Those interneurons preferentially target SOMs and receive both thalamic and cortical drive.

Inspecting the connectivity of the microcircuit, PV INs are part of a feedforward inhibition circuit. As already stated, like PNs they receive afferent inputs from thalamus—known as thalamocortical projections—and other cortical populations, known as corticocortical projections. Recurrent inhibition is also provided by the PV population, which shares dense reciprocal connections with local PNs. It is the SOMs on the other hand which are involved in the mediation of disynaptic inhibition between neighbouring PNs (Berger et al., 2009; Silberberg and Markram, 2007). And disinhibition is present thanks to the VIPs, which can inhibit the SOMs, reducing their inhibitory input to the dendrites of PNs (Pi et al., 2013).

With this minimalist picture of cortical connectivity and inhibitory motifs in hand, we can now begin to address the question how INs contribute to computation.

How do inhibitory neurons contribute to computation?

Given the diversity of inhibitory interneurons and the complexity of the cortical microcircuit, it should come as little surprise that there are a multitude of roles ascribed to INs. First, they are essential for the generation of synchronous rhythms occurring during both awake behaviour and sleep (Steriade et al., 1993), with such oscillations believed to be important for coordination of neuronal activity—though the ability for INs to synchronise populations of neurons can unfortunately result in epileptiform activity when inhibition is dysregulated (Dudek and Sutula, 2007).

Feedforward inhibition can reduce latency jitter—variability of response time—in the spike responses of PNs to their afferent inputs (Pouille and Scanziani, 2001). The inhibitory microcircuit is also important for associating "bottom-up" inputs (from lower order cortex and primary thalamus) and "top-down" inputs (from higher order cortex and thalamus) that target the opposing poles of PNs (Larkum, 2013). And feedback inhibition can decorrelate the responses of PNs, maximizing the information content of spike trains (Averbeck et al., 2006; Tetzlaff et al., 2012).

Inhibition may play a key role in a phenomenon known as gain control. Neurons typically fire at a rate proportional to their input current, and the constant of that proportionality is known as the gain. Within cortex, if both excitatory and inhibitory background inputs are increased, the gain of PNs are modulated in a divisive manner (Chance et al., 2002). For sensory neurons, this can preserve their dynamic range despite large variations of afferent input intensity (Pouille et al., 2009).

A balance of excitation and inhibition

Some of the most important properties of cortical activity, and therefore processing, are closely linked to a balance between excitation and inhibition. In sensory cortical regions, changes in stimulus features lead to changes of excitation that are closely matched by inhibition (Wehr and Zador, 2003), and even during spontaneous activity fluctuations in excitation are reliably tracked by inhibition (Haider et al., 2006). Crucial

insight into the consequences of excitatory-inhibitory (E/I) balance comes from the landmark theoretical analysis of van Vreeswijk and Sompolinsky (1996). They showed that such a state yields chaotic dynamics—potentially explaining the asynchronous-irregular activity observed in cortical PNs—within a network that responds linearly to inputs despite highly non-linear neurons. And those balanced networks can respond to changes of input with a latency shorter than the response time of individual cells. Brunel (2000) demonstrated that, depending on the strength of inhibition and afferent drive, networks can exhibit different regimes of dynamical behaviour. This includes synchronous or asynchronous population activity, with regular or irregular spiking of individual cells.

Later, new theoretical insight to the balanced state was provided by Murphy and Miller (2009), who proposed that an amplification of specific input patterns arises when feedback inhibition stabilizes strong recurrent excitation, especially true in the presence of structured connectivity (Hennequin et al., 2012). Such balanced amplification can enhance reliable signal transmission in biologically constrained models of cortex (Joglekar et al., 2017), and posits a critical role for lateral (i.e. local disynaptic) inhibition.

Stimulus responses

A common property of many PNs in sensory cortices is that they respond preferentially to certain stimulus features. The seminal account of this was from Hubel and Wiesel (1959), wherein they reported that cells in the primary visual cortex of cats were responsive to moving bar stimuli, with individual cells firing at higher rates for bars of a particular orientation. That a cell spikes selectively in response to a stimulus feature is a direct consequence of the stimulus selectivity of the synaptic currents onto that cell.

The excitatory synaptic current onto a PN in sensory cortex is tuned to prefer a particular stimulus, being large for similar and small for dissimilar stimuli. Absent significant inhibitory current, the excitatory current will depolarise the cell membrane potential above the spiking threshold, generating action potentials in response to most stimuli (Fig 2.2, left).

The picture changes in the presence of inhibitory currents. If inhibition is much less selective for the stimulus feature than excitation, it has the effect of shifting the membrane potential downward by a near-constant amount (Fig 2.2, right). This narrows the range of stimuli for which the PN membrane potential exceeds its firing threshold, and makes the spiking response more selective to particular stimuli. This is known as response sharpening, and is believed to be one of the essential functions of inhibition within sensory cortices.

Inhibitory currents are sometimes much less stimulus selective—if at all—than the excitatory drive, but in some cases inhibition can be just as well-tuned as excitation. We now consider where this difference is observed and what might distinguish the two cases.

Stimulus selectivity and E/I balance in sensory cortex

When the total inhibitory currents onto a neuron exhibit less stimulus tuning than excitation, but on average the two are proportional, it is known as *global* balance. If excitatory and inhibitory currents arising from each individual subset of presynaptic

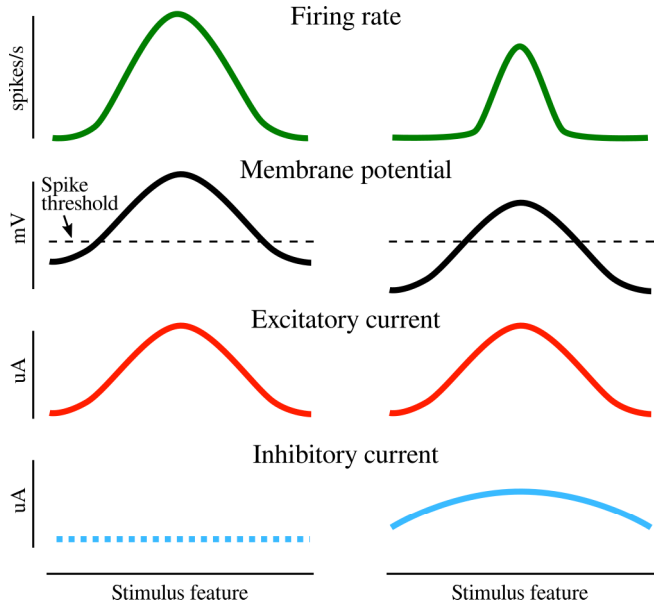

Figure 2.2: **Inhibition sharpens stimulus tuning in PNs.** The firing rate, membrane potential, excitatory and inhibitory currents measured in a principal cell within sensory cortex. Left: un-tuned background inhibition. Right: Moderately co-tuned synaptic inhibition. See text for full description. Modified from Isaacson and Scanziani (2011).

inputs are proportional, and are consequently well co-tuned for stimulus features, it is known as *detailed* balance (Vogels and Abbott, 2009). It has been shown computationally that global and detailed balance have differential effects on sensory processing, with global balance causing the cell to become responsive only to its preferred stimulus while making it largely insensitive to stimulus intensity (Vogels et al., 2011). In contrast, detailed balance produces sparse firing irrespective of stimulus, where spikes are produced due to transients in the afferent inputs, with broadly stimulus-selective responses graded to intensity (Vogels et al., 2011).

Whether global or detailed balance is at work within particular areas of cortex remains a contentious issue. We now consider two cases: the auditory and visual cortex of rodents.

Auditory cortex

Excitatory neurons within the primary auditory cortex (A1) of rodents typically emit spikes preferentially to certain stimuli—in this case, auditory tones of a particular frequency. Those stimulus selective cells are spatially distributed across A1 in a tonotopic fashion, meaning that cells with similar frequency preferences tend to be physically near one another (Horikawa et al., 1988). Such a tonotopic arrangement emerges over the course of development (Zhang et al., 2001). It is this tonotopic arrangement that Harris and Mrsic-Flogel (2013) claim produces the co-tuning of excitatory and inhibitory currents measured in neurons within A1 (Fig 2.3, right). The logic of their argument is straightforward: if neighbouring PNs have similar stimulus preferences,

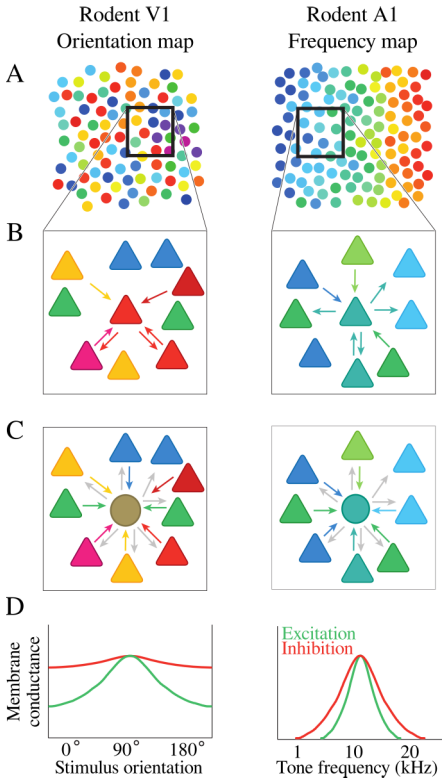

Figure 2.3: **Co-tuning in sensory cortex depends on topography.** Depicts the hypothesised origin of excitatory-inhibitory current co-tuning in both rodent primary visual cortex (left), and primary auditory cortex (right). **A)** Spatial distribution of stimulus preference in V1 and A1. **B)** Despite salt-and-pepper arrangement of V1, excitatory inputs to principal cells are biased towards cells with similar tuning. In A1 nearby principal cells have similar tuning. **C)** Interneurons of V1 receive excitatory input from cells with random preferred orientation. In A1 interneurons receive excitation from cells with similar frequency preference. **D)** The salt-and-pepper arrangement in conjunction with unspecific excitatory-to-inhibitory connections produces weak tuning of inhibition. The tonotopic arrangement of A1 produces strong co-tuning of excitation and inhibition. Modified with permission from Harris and Mrsic-Flogel (2013).

and most of them provide excitatory drive to local INs (Fino et al., 2013; Fino and Yuste, 2011), then those interneurons will be driven to respond to the same stimuli. It follows that if INs within the local population fire preferentially at the same stimulus as PNs, synaptic inhibition originating from the former will exhibit roughly the same stimulus preference as synaptic excitation from the latter. Experimental data seems to support this view. Li et al. (2014a), as well as Moore and Wehr (2013) report that PV cells exhibit responses nearly as selective as PNs, and Froemke et al. (2007) show that in adult rats, the stimulus tuning of inhibitory currents are highly correlated with excitatory currents.

Visual cortex

In contrast to auditory cortex, the primary visual cortex (V1) of rodents does not exhibit the same orderly arrangement of cells according to their stimulus preferences—in this case the orientation of a bar within their receptive field. Rodent V1 rather has a "salt and pepper" topography, whereby neighbouring cells can have drastically different orientation preferences. Harris and Mrsic-Flogel (2013) propose that such heterogeneity of stimulus preference would have clear consequences. Since cortex is densely connected, and with neighbouring PNs preferring different orientations, the excitatory currents onto inhibitory cells would exhibit no clear stimulus preference. This should cause interneurons to be essentially non-selective to orientation, and consequently inhibitory currents would exhibit very little co-tuning with excitation, as depicted in Fig 2.3 (left).

Investigators have published data that they claim either supports (Hofer et al., 2011; Kerlin et al., 2010; Zariwala et al., 2011) or contradicts (Runyan et al., 2010; Tan et al., 2011) this hypothesis. A close inspection of the data reveals that the difference between the two populations are not as substantial as their claims would imply. As those studies show, INs reliably exhibit modest levels of orientation selectivity, which on average is not drastically lower than what they report for PNs. Consequently, one of the questions addressed in this thesis then is: Given the salt and pepper topography of V1, how does the observed degree of stimulus selectivity of interneurons develop, and how is it maintained?

2.2 Homeostasis and the plasticity of inhibitory activity

Having established the central role that GABAergic cells play in shaping the functional response of neuronal circuits, we now consider how they develop and maintain the properties that enable them to fulfill their role. Mammals are dynamic organisms, changing throughout their lifetime due to ageing, injury, and adaptation to their environment. On the other hand, sensitivity to large deviations of their internal state—e.g. their temperature, pH, and the activity of their nervous system—means that mechanisms to detect and control those deviations are essential. This property of maintaining certain parameters within an operational, and thus healthy range, is known as homeostasis (Cannon, 1932).

One parameter that appears homeostatically regulated is the firing rate of neurons within the brain (Hengen et al., 2013; Keck et al., 2013). Since this is a single observable quantity, it might seem *prima facie* that a single mechanism could account for its regulation. The reality is considerably more complicated, with a plethora of mechanisms simultaneously operating on separate control variables that can interact with each other. Homeostatic mechanisms currently known to affect neurons of the mammalian central nervous system include

- changes to the intrinsic excitability of cells (Desai et al., 1999b),
- postsynaptic changes to the effectiveness of synapses, across an entire cell (Turrigiano et al., 1998) or parts of it (Yu and Goda, 2009),

- presynaptic changes that can alter quantal size, or release probability (Branco et al., 2008),
- regulation of the excitation-inhibition ratio (Maffei et al., 2004), and
- metaplastic changes to the threshold for induction of long-term potentiation (LTP) or depression (LTD) of synapses (Kirkwood et al., 1996).

Many of these mechanisms respond to different signals—e.g. long term changes to membrane polarization (O'Leary et al., 2010), or Ca+ concentration levels (Turrigiano, 2012)—and at different time scales. Some require chronic changes in activity, such as those produced by injury or certain experimental protocols (Rutherford et al., 1998). Others may be responsible for rapid compensatory responses, such as metaplasticity which may be necessary to stabilize networks in the presence of correlation-based synaptic plasticity (Zenke et al., 2017).

The role of inhibitory interneuron activity in homeostasis

In the face of this daunting complexity, there is yet another dimension: the role of GABAergic interneuron activity. On one hand, it seems likely that interneurons have regulatory processes that are homeostatic for their own activity (Desai et al., 1999a). But on the other hand, if this is true in the sense that those processes simply maintain interneuron firing rates at some set-point, it seems maladaptive for total circuit function. Arguably, they should regulate their activity—in fact, they might be more adaptable than excitatory cells within the adult brain (Kameyama et al., 2010)—so as to preserve the functional properties of the system. And because they serve a role distinct from that of excitatory cells, the regulation of their activity should likewise have a distinct form.

Some direct evidence of this has been observed *in vivo* for several different animal models. Hengen et al. (2013) show that changes to FS population activity follow a different time course than that of the PN population in response to monocular deprivation (MD). Those putative INs reduce their firing rate in the first 24 hours following deprivation, while the PN population rate remains unchanged. On the second day, FS rates recover and PN rates drop, before finally on the third day both cell types return to their baseline rates.

When looking at the rates of individual neurons, Barnes et al. (2015) report that for the GABAergic interneurons exhibiting activity 72 hours after monocular enucleation, their population rate does not recover to baseline, while the population of PNs that are active does. Finally, Kuhlman et al. (2013) show that L2/3 PV INs respond to 1-day MD by reducing their firing rate by about half, while PNs in the same layer actually increase their firing.

Taken together these data point to regulatory processes that act on inhibitory cell firing rates in a manner distinct from that of excitatory cells. The question is now whether the corresponding synaptic signature of these ostensibly homeostatic processes are known.

Study	Methods	Firing rates	Plasticity
Rutherford et al., 1998	Juvenile rat visual cortex cultures; 2d incubation in TTX, BDNF, or both; Bipolar interneurons;	TTX alone yields 10x PN rates after washout, but only 2x in INs; TTX+BDNF no change in PNs, small increase in INs; BDNF alone PN unchanged, IN 4x increase;	BDNF exposure significantly increases EPSC amplitude in PN-IN pair recording;
Bartley et al., 2008	Juvenile mouse somatosensory cortex slice culture; 5d blockade TTX; paired recordings of coupled PN-IN pairs, (PV & SOM)	N/A	Both PV & SOM INs had increased membrane excitability and increased excitatory synaptic drive. Other changes (STP, uIPSC size) lead to decreased PV uIPSCs, but balance out to maintain normal SOM;
Doyle et al., 2010	Dissociated embryonic mouse cortical culture; Multipolar GABAergic neurons; Increase Glut synaptic activity;	N/A	VGLUT2 and Narp induction by exc transmission should increase Glut transmission at AMPA receptors, due to greater Glut in vesicles;

Table 2.1: Summary of *in vitro* experiments studying homeostatic response of INs

Synaptic correlates of regulated interneuron activity

Unfortunately, the data on how such changes to IN activity might relate to synaptic changes are limited and conflicting. One of the first studies of homeostatic changes to glutamatergic synapses onto inhibitory cells was from Rutherford et al. (1998). They showed that in dissociated cortical cell cultures incubated in the presence of brain-derived neurotrophic factor (BDNF)—a compound secreted by PNs during elevated activity—the amplitude of excitatory postsynaptic currents (EPSCs) onto bipolar INs increases. Again in cortical cultures, Doyle et al. (2010) observed changes that should result in an increase of excitatory synaptic drive onto multipolar GABAergic interneurons following a prolonged increase of glutamatergic synaptic activity. Within hippocampal cell cultures, Chang et al. (2010) showed that mini-EPSC amplitude increased in PV interneurons following chronic enhancement of activity, and decreased following activity blockade.

While in those experiments the chronic manipulation of neuronal activity occurred in culture or slice preparations, there are a number where the manipulation was *in vivo*. The first that we are aware of is from Keck et al. (2011), who recorded from mouse V1 following retinal lesion. They report a rapid loss of glutamatergic synapses onto the subset of inhibitory cells that carry dendritic spines (whereas most INs lack spines). Takesian et al. (2013) induced hearing loss in juvenile gerbils, and later observed decreased EPSC amplitude onto the FS INs of auditory cortex. Finally, in addition to the data on IN firing rates already mentioned, Kuhlman et al. (2013) also report that excitatory drive onto those L2/3 PV INs is reduced by 70%, with a concomitant reduction in spontaneous and evoked EPSC frequency.

The aforementioned results appear in keeping with the hypothesis that the activity of certain INs might be regulated to be homeostatic for PN firing rates. In contrast Bartley et al. (2008) used TTX to block activity for 5 days in a cortical slice culture, and observed a rapid increase in excitatory synaptic drive onto PV and SOM expressing INs—though inhibitory neurotransmission from the PV cells decreased due to loss of GABAergic synapses. In an experiment from Maffei et al. (2006), who worked with an acute slice preparation taken from rat visual cortex, they report that prior visual deprivation potentiates synapses from pyramidal to inhibitory FS cells. Such an increase of synaptic excitation onto INs under these protocols appears to be homeostatic for IN, not PN firing rates.

To the best of our knowledge, that is a nearly exhaustive list of studies in which the homeostatic properties of excitatory-to-inhibitory synapses are studied, by chronically manipulating the activity of most or all of the network. There is an additional body of literature that investigates the plastic properties of excitatory synapses on GABAergic cells, using stimulation protocols that only manipulate the activity of the cells sharing a synapse (Kullmann et al., 2012). The results reported depend on the interneuron type, and brain area. Inhibitory cells in amygdala (Mahanty and Sah, 1998), striatum (Fino et al., 2008), somatosensory cortex (Lu et al., 2007), visual cortex (Sarihi et al., 2008), and hippocampus (Kullmann and Lamsa, 2007; Le Roux et al., 2013; Nissen et al., 2010) have all been reported to have plastic excitatory synapses.

It is important to reiterate that this latter set of results were not obtained using experimental protocols explicitly designed to test for homeostatic properties. Nonetheless, it is interesting to consider the homeostatic properties (or lack thereof) of the various

Study	Methods	Firing rates	Plasticity
Maffei et al., 2006	Juvenile mice; 3d visual deprivation (P18-P21); coronal slice; L4 FS IN;	N/A	EPSP amplitude onto FS increases. IPSC onto PN increases;
Keck et al., 2011	Adult mice; in vivo; Focal retinal lesion; 2-photon imaging; L1+2/3; INs w/ spines: 91% NPY, 20% CR, 5% SOM;	N/A	INs rapidly lose spines after lesion (begins 6hrs after, but 48hrs for bilateral lesion); GABA synapses take 24hrs to respond; No recovery after 2 months; Control: Dendritic branch tips stable, spine density stable;
Hengen et al., 2013	Juvenile freely behaving rats; MD; Multi-electrode recording; putative FS and RS units;	RSU firing in deprived hemisphere reduced (from ~5 to ~3 Hz) 2d after lid suture; pFS rate drops by 1/3 1d after, recovers on day 2.	N/A
			Continued on following page...

Kuhlman et al., 2013	Alert juvenile mice; MD; ODP; in vivo; L2/3; Glut uncaging via laser scanning photo stimulation (LSPS)	Monocular response after 1d: PYR 2x (equivalent to binocular stim), PV 1/2 (for both open and formerly closed eye); after 3d PV response to deprived eye remain weak, but OD shifts to open eye	E-E connections unchanged; PV show no intrinsic changes; L4+5 exc drive onto L2/3 PV 70% reduced, evoked+spontaneous EPSC freq. reduced, but amplitude unchanged
Takesian et al., 2013	Juvenile gerbils; thalamocortical brain slice; medial geniculate-evoked PSPs in FS or LTS measured	Thalamic input generated spikes in half as many FS cells after hearing loss versus control.	Hearing loss affects FS intrinsic properties; EPSPs onto FS reduced, LTS increased. Less STD onto FS, greater onto LTS.
Barnes et al., 2015	Adult behaving mice; Ca^{+} imaging of L2/3 following retinal enucleation; Compared PN vs GABA expressing INs;	~50% PN, ~60% IN remain inactive after 72h; Pop. rate of remaining cells recover in PNs, but is ~0.6 of baseline for INs; INs corr. w/ PNs at baseline more likely to recover if PN recovers;	Recovery of activity in PNs unlikely to be due to synaptic scaling; Reduction in mIPSC frequency but not amplitude was insufficient to account for recovery;

Table 2.2: Summary of *in vivo* experiments studying homeostatic response of INs

plastic behaviours reported, as done by Kullmann and Lamsa (2011). How is it that we might infer homeostatic properties? According to a simple model of the homeostasis of inhibition proposed by Wenner (2011), if excitatory activity rises we should expect that both the glutamatergic synapses onto inhibitory interneurons, as well as GABAergic synapses onto excitatory cells to strengthen. And the reverse should hold in the presence of a decrease in excitatory activity.

How well do these results comport with the simple model of Wenner? All of those experiments use protocols that increase the activity of the presynaptic excitatory neuron, so we should expect that if plasticity is homeostatic for that cell, the synapse should always increase in strength. Only Lu et al. (2007) contradict the simple model, showing a decrease in synaptic strength when pre- and postsynaptic firing are correlated. Mahanty and Sah (1998), as well as Sarihi et al. (2008) report that correlated activity leads to increases in synaptic strength—a form of plasticity typically referred to as Hebbian. Fino et al. (2008) report a stricter requirement, because in their case simple correlation is not enough, rather it requires presynaptic spikes to precede postsynaptic activity. In contrast, Kullmann and Lamsa (2007), as well as Le Roux et al. (2013) show that potentiation can require the interneuron to be hyperpolarised when the excitatory cell is firing, a form of plasticity sometimes called anti-Hebbian.

Although many of those results are compatible with the simple model of Wenner (2011), it remains an open question—that this thesis addresses—whether they are in fact homeostatic for excitatory firing rates.

Cell autonomous or network level regulation?

Much homeostatic work is done by bathing a cell culture in TTX (to block all network activity), or GABA antagonists such as bicuculline (to increase network activity). Even the aforementioned *in vivo* work relies upon gross manipulation of network-level activity by retinal lesioning, enucleation, or some other form of visual deprivation. These interventions alter both spiking activity and neurotransmission across the network. Hence, it is conceivable that either changes in network activity, single-cell firing, or synaptic transmission underlie homeostatic processes.

Increasing spiking activity of a single principal cell in hippocampus can trigger a compensatory increase of inhibition (Peng et al., 2010). However, when activity is blocked by modifying its membrane ion channels, no changes are observed (Hartman et al., 2006). So in some cases, deviations in single cell activity might not trigger homeostatic changes, but rather changes in activity or neurotransmission on the network level may be required.

If interneuron firing rates are responsive to network-level chronic deviations (or homeostatic "errors") in activity, what signal would permit detection of that error? Volume transmission might be adequate for the task. Volume transmission is defined as communication along a channel without clearly identifiable "wiring", e.g. axons and dendrites (Agnati et al., 1995). It often involves the transmission of signals by secreted factors released by a neuron (neurotrophins, cytokines, and pentraxins) that diffuse in the extracellular space and are detected some distance away by other neurons or glial cells (occasionally at great distances, see Kuczewski et al., 2010).

There is a clear role for brain-derived neurotrophic factor (BDNF) as a channel of

volume transmission (Huang et al., 1999; Kuczewski et al., 2010; Rutherford et al., 1997). It is released as a function of excitatory cell activity and is detected by tyrosine receptor kinase (Trk) receptors on both PNs and interneurons (Lu, 2003). It is this signal that is thought to be involved in the homeostatic regulation of glutamatergic synapses on inhibitory cells (Rutherford et al., 1998).

Other volume transmission signals implicated in homeostatic plasticity of synapses include neuronal activity-regulated pentraxin (Narp), nitric oxide (NO), tumour-necrosis factor-alpha (TNF-α), and endocannibinoids (Chang et al., 2010; Hardingham et al., 2013; Kim and Alger, 2010; Stellwagen and Malenka, 2006). Some of these molecules are unlikely to diffuse far in the extracellular space at concentrations high enough to be physiologically relevant (Hardingham et al., 2013; Syková and Nicholson, 2008), and so are more likely to be detected locally.

Nevertheless, their effect may be broader due to retrograde transmission along the axon of some neurons, where their signal is integrated centrally (Ginty and Segal, 2002; Zweifel et al., 2005). If present in inhibitory interneurons, this sort of "retroaxonal" transmission would permit them to respond to the activity of their entire postsynaptic population. Admittedly the slow propagation of retrograde signals—which might take hundreds of seconds to reach the soma (Maday et al., 2014)—introduces substantial latency, permitting only slow responses to chronic changes of activity (possibly necessitating additional mechanisms to counter rapid changes, e.g. the destabilising effect of Hebbian plasticity; Zenke et al., 2013). Another network-level sensor, extracellular TNF-α concentration is modulated by glial cells in response to population activity, and affects cell-surface expression of excitatory alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic (AMPA) receptors (Stellwagen and Malenka, 2006). Though this effect has not yet been explicitly shown in inhibitory interneurons, it does represent another potential mechanism for regulating interneuron activity in response to network-level deviations from from the homeostatic set-point.

Summary

Collectively, these data suggest it is plausible that some INs regulate their activity to homeostatically control PN firing rates. Glutamatergic synapses onto several IN types exhibit plastic changes in response to chronic manipulation of network activity that appear compatible with such a role. Importantly, it is unclear whether those synapses are changing in response to the activity of individual PNs, or populations of them. In the former case, a conventional homosynaptic plasticity might suffice to explain the observed behaviour. In the latter case, there are a variety of candidate mechanisms to potentially explain how INs could sense such network-level changes. In either case it remains an open question what effect such putative homeostatic plasticities would have on the network. That is precisely what is addressed in the remainder of this thesis.

3 Aims and main questions of this thesis

The central aim of this thesis is to contribute to the understanding of how interneurons are involved in the homeostatic regulation of excitatory activity, and what might cause the development and maintenance of their functional properties within sensory cortex. In the previous chapter, we laid out a litany of roles for inhibition within cortex, and some of the various forms of plasticity relevant for its regulation.

While homeostatic plasticity of synapses and intrinsic neuronal properties have been well characterised for many components of cortical networks, how plasticity of interneuron activity is involved in network homeostasis has not received as much attention. It is therefore the central subject of this thesis, and we endeavour to answer the following questions:

1. If IN activity were regulated—by plasticity of excitatory synapses onto them or of their intrinsic properties—to homeostatically control the firing rates of excitatory cells, how should this be accomplished?

2. What effect on a network do putatively homeostatic forms of plasticity have, when they act via the excitatory synapses onto inhibitory neurons?

3. If IN activity *is* regulated to be homeostatic for the excitatory population, what are the consequences for the operation of networks within sensory cortex?

We study these questions because while there exists published research claiming that some of the observed plasticity of IN activity is homeostatic for the network, it is a topic that remains underexplored. And despite these claims, we know of no research that has established that if INs are to be homeostatic controllers for the network, *how* they should regulate their activity. This is precisely why we begin with a normative approach, using mathematical analysis to determine what is required for interneurons to accomplish this hypothesised task. Furthermore, one of the great challenges in understanding homeostasis is that existing experimental protocols engage many mechanisms simultaneously, making it difficult to accurately determine the true effect of any one of them. A computational approach permits us to isolate a single mechanism, and determine its effect without having the complexity of a real biological system confounding the result. And that is the ultimate aim of this thesis: To show in what manner interneuron activity on its own can—even in principle—be regulated so as to be homeostatic for the brain.

4 Methods

We use a combination of mathematical and computational techniques to study whether different forms of interneuron plasticity can homeostatically control the activity of excitatory cells. In the following, we first outline the methodology, including the network model, the different plasticity rules, and the data analysis methods we use to characterize the effects of the plasticity on network activity. All simulations are performed using custom software written in the C programming language, which integrates the excitatory and inhibitory neuronal dynamics using the forward Euler method, while stochastic differential equations are integrated using Euler–Maruyama. Convergence of all simulations is ensured by verifying several measures of network activity, including population rate and error, obtain a steady state value.

4.1 Homeostatic network model

We study homeostatic plasticity in a rate-based network of N^{E} excitatory and N^{I} inhibitory neurons. Those neurons are arranged on a ring to study consequences of distance-dependent connectivity, and to avoid boundary effects (Fig 4.1). The dynamics of the network are determined by integrating a system of differential equations for the membrane potential of excitatory cells and inhibitory cells (h_j^{E} and h_i^{I} respectively)

$$\tau_{\mathrm{E}} \dot{h}_j^{\mathrm{E}} = -h_j^{\mathrm{E}} + \sum_{k=1}^{N^{\mathrm{E}}} W_{jk}^{\mathrm{E}\leftarrow\mathrm{E}} r_k^{\mathrm{E}} - \sum_{i=1}^{N^{\mathrm{I}}} W_{ji}^{\mathrm{E}\leftarrow\mathrm{I}} r_i^{\mathrm{I}} + w^{\mathrm{in}} r_j^{\mathrm{E}\leftarrow\mathrm{in}} \tag{4.1a}$$

$$\tau_{\mathrm{I}} \dot{h}_i^{\mathrm{I}} = -h_i^{\mathrm{I}} + \sum_{j=1}^{N^{\mathrm{E}}} W_{ij}^{\mathrm{I}\leftarrow\mathrm{E}} r_j^{\mathrm{E}} - \sum_{k=1}^{N^{\mathrm{I}}} W_{ik}^{\mathrm{I}\leftarrow\mathrm{I}} r_k^{\mathrm{I}} + w^{\mathrm{in}} r_i^{\mathrm{I}\leftarrow\mathrm{in}}. \tag{4.1b}$$

Here, $\dot{h}$ denotes the temporal derivative of h (which is unitless), $W_{ij}^{b\leftarrow a}$ denotes the synaptic weight from the j-th neuron in population a to the i-th neuron in population b (for $a, b \in \{\mathrm{E}, \mathrm{I}\}$ indicating the excitatory or inhibitory population), w^{in} denotes the synaptic strength of external input, and τ_{E} and τ_{I} are the excitatory and inhibitory membrane time constants. The firing rates r_j^{E} and r_i^{I} are determined by passing the membrane potential of a given cell through a rectifying non-linearity $s^+(x, \alpha) = \alpha^{-1} \ln(1 + e^{\alpha x})$, ensuring non-negative firing rates, as in

$$r_j^{\mathrm{E}} = s^+(h_j^{\mathrm{E}}, \alpha_r)$$
$$r_i^{\mathrm{I}} = g_i s^+(h_i^{\mathrm{I}} - \theta_i, \alpha_r).$$

The parameter α_r controls the sharpness of the firing threshold. For interneurons, there are also gain g_i and firing-threshold θ_i parameters, for which we derive intrinsic

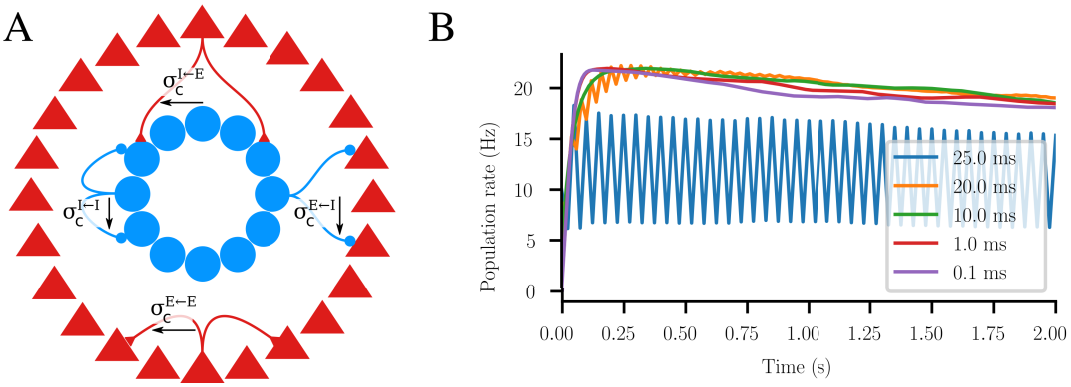

Figure 4.1: **Network schematic and numerical stability. A)** The neurons of the simulated network are arranged on a ring, and are connected up to a spatial distance of approximately $\sigma_c^{a\leftarrow b}$, from presynaptic population b to postsynaptic population a. See the text for details. **B)** Population firing rates, resulting from simulating identical networks each with a different value for the numerical integration time step Δt. Values range from 25 ms (equal to the membrane time constant of the inhibitory neurons τ_I), to 0.1 ms. The two largest values of Δt produce oscillations, while values ≤ 10 ms result in smooth integration. Differences in the particular time series is a consequence of drawing new Gaussian random variables for the external inputs on each time step, making identical results impossible for our simulator. All remaining results use $\Delta t = 10$ ms.

homeostatic plasticity rules in the Appendix (in Chapter 7). For the simulation results in Chapter 5 they are held constant at $g_i = 1$ and $\theta_i = 0$.

Inputs $r_j^{E\leftarrow in}$ and $r_i^{I\leftarrow in}$ are independent Ornstein-Uhlenbeck processes with means μ_{ou}^E and μ_{ou}^I respectively, standard deviation σ_{ou}, and time constant τ_{ou}. Those inputs stand in for background activity, e.g. inputs from a large network that the modelled network would be embedded within—in the diffusion limit of synaptic noise (Lánský, 1984). The mean values are chosen such that excitatory cells need to recruit inhibition to reach the homeostatic set-point, and that inhibitory cells rely upon drive from the excitatory population to exhibit any significant activity. This ensures that the plasticity mechanisms being studied are actually engaged. The values of the inputs at time t are determined by integrating the stochastic differential equations

$$\dot{r}_j^{a\leftarrow in} = \frac{\mu_{ou}^a - r_j^{a\leftarrow in}}{\tau_{ou}} + \sqrt{\frac{2\sigma_{ou}^2}{\tau_{ou}}}\xi(t),$$

where ξ is a zero-mean Gaussian white noise having units $s^{-1/2}$ and an autocorrelation function $\langle \xi(t)\xi(s)\rangle = \delta(t-s)$.

Connectivity

Synaptic weight matrices are constructed to form a ring network with a circular distance dependence, and defined as $W_{ij}^{a\leftarrow b} = \Delta_{ij}^{a\leftarrow b} w_{ij}^{a\leftarrow b}$, where $w_{ij}^{a\leftarrow b}$ is the synaptic weight from the j-th neuron in population b onto the i-th neuron in population a. $\Delta^{a\leftarrow b}$ is an

N^{E}	256	N^{I}	64	Number of exc. & inh. neurons.
τ_{E}	50 ms	τ_{I}	25 ms	Membrane time constants.
α_r	10	$r_{\max}$	200 Hz	Threshold sharpness & max. rate.
x_j^{E}	$2\pi j/N^{\mathrm{E}}$ rad	x_i^{I}	$2\pi i/N^{\mathrm{I}}$ rad	Location of exc. cell j and inh. cell i.
$\sigma_c^{a\leftarrow\mathrm{E}}$	$3\pi/8$ rad	$\sigma_c^{a\leftarrow\mathrm{I}}$	$\pi/4$ rad	Connection distances.
$p^{\mathrm{I}\leftarrow\mathrm{E}}$	0.25	$p^{\mathrm{E}\leftarrow\mathrm{I}}$	0.25	Connection probabilities.
$p^{\mathrm{E}\leftarrow\mathrm{E}}$	0.25	$p^{\mathrm{I}\leftarrow\mathrm{I}}$	0.25	
$\mu^{\mathrm{I}\leftarrow\mathrm{E}}$	$p^{\mathrm{I}\leftarrow\mathrm{E}}N^{\mathrm{E}}$	$\mu^{\mathrm{E}\leftarrow\mathrm{I}}$	$p^{\mathrm{E}\leftarrow\mathrm{I}}N^{\mathrm{I}}$	Mean in-degree of connection matrices.
$\mu^{\mathrm{E}\leftarrow\mathrm{E}}$	$p^{\mathrm{E}\leftarrow\mathrm{E}}(N^{\mathrm{E}}-1)$	$\mu^{\mathrm{I}\leftarrow\mathrm{I}}$	$p^{\mathrm{I}\leftarrow\mathrm{I}}(N^{\mathrm{I}}-1)$	
$\mathrm{CV}^{\mathrm{I}\leftarrow\mathrm{E}}$	0.35	$\mathrm{CV}^{\mathrm{E}\leftarrow\mathrm{I}}$	0.31	Coefficient-of-variation of in-degree.
$\mathrm{CV}^{\mathrm{E}\leftarrow\mathrm{E}}$	0.27	$\mathrm{CV}^{\mathrm{I}\leftarrow\mathrm{I}}$	0.38	
$w^{\mathrm{I}\leftarrow\mathrm{E}}$	$2/\mu^{\mathrm{I}\leftarrow\mathrm{E}}$s	$w^{\mathrm{E}\leftarrow\mathrm{I}}$	$2/\mu^{\mathrm{E}\leftarrow\mathrm{I}}$ s	Default synaptic weights. Units of s required for dimensional consistency.
$w^{\mathrm{E}\leftarrow\mathrm{E}}$	$2/\mu^{\mathrm{E}\leftarrow\mathrm{E}}$s	$w^{\mathrm{I}\leftarrow\mathrm{I}}$	$1/\mu^{\mathrm{I}\leftarrow\mathrm{I}}$s	
w^{in}	1 s			
$\mu_{\mathrm{ou}}^{\mathrm{E}}$	10 Hz	$\mu_{\mathrm{ou}}^{\mathrm{I}}$	0 Hz	Ornstein-Uhlenbeck parameters.
σ_{ou}	5 Hz	τ_{ou}	5 s	
T	2×10^5 s	Δt	10 ms	Simulation time & step size.
ρ_0	6 Hz	τ_{slow}	600 s	Target rate & slow time constant.
η^w	2×10^{-8}	$w_{\max}^{\mathrm{I}\leftarrow\mathrm{E}}$	$10w^{\mathrm{I}\leftarrow\mathrm{E}}$	Learning rate & max. weight.
α_p	10			Sharpness of synaptic rectification.

Table 4.1: **Network model parameters.** For explanations see the text.

adjacency matrix, with value 1 if neurons i and j are connected, otherwise 0. Each adjacency matrix has an in-degree of mean $\mu^{a\leftarrow b}$ and coefficient of variation $\mathrm{CV}^{a\leftarrow b}$, with values taken from Landau et al. (2016). The probability of a given connection $\Delta_{ij}^{a\leftarrow b}$ being present is

$$p\left(\Delta_{ij}^{a\leftarrow b}\right) \propto p^{\mathrm{vM}}\left(x_i^a \middle| \mu = x_j^b,\ \kappa^{-1/2} = \sigma_c^{a\leftarrow b}\right)$$

Here, x_i^a denotes the location of the i-th neuron in population a on the ring, scaled between $-\pi$ and π, and $p^{\mathrm{vM}}(x \mid \mu, \kappa)$ is the von Mises density function, a circular analogue of the Gaussian density. Its parameters μ and κ control the location and width of its peak, respectively. The connectivity from population b to population a decays on a spatial scale $\sigma_c^{a\leftarrow b}$ $\left(\hat{=}\kappa^{-1/2}\right)$. The amplitude of p^{vM} is scaled so as to reach an overall connection probability $p^{a\leftarrow b}$ between the two populations. Autapses are not permitted in the network, and thus $p(\Delta_{ii}^{a\leftarrow a}) \equiv 0$.

4.2 Sensory cortex network model

In Section 5.2, we study the effects of different forms of homeostatic plasticity on the stimulus responses of cells in sensory cortices. In those simulations the dynamics of the network are determined by numerically integrating a slightly modified version of

the previous network model (Eq 4.1):

$$\tau_{\mathrm{E}}\dot{h}_j^{\mathrm{E}} = -h_j^{\mathrm{E}} - \sum_{i=1}^{N^{\mathrm{I}}} W_{ji}^{\mathrm{E}\leftarrow\mathrm{I}} r_i^{\mathrm{I}} + w^{\mathrm{in}} r_j^{\mathrm{E}\leftarrow\mathrm{in}} + \sum_{l=1}^{N^{\mathrm{A}}} W_{jl}^{\mathrm{E}\leftarrow\mathrm{A}} r_l^{\mathrm{A}}(q(t)) \tag{4.3a}$$

$$\tau_{\mathrm{I}}\dot{h}_i^{\mathrm{I}} = -h_i^{\mathrm{I}} + \sum_{j=1}^{N^{\mathrm{E}}} W_{ij}^{\mathrm{I}\leftarrow\mathrm{E}} r_j^{\mathrm{E}} - \sum_{k=1}^{N^{\mathrm{I}}} W_{ik}^{\mathrm{I}\leftarrow\mathrm{I}} r_k^{\mathrm{I}} + w^{\mathrm{in}} r_i^{\mathrm{I}\leftarrow\mathrm{in}} + \sum_{l=1}^{N^{\mathrm{A}}} W_{il}^{\mathrm{I}\leftarrow\mathrm{A}} r_l^{\mathrm{A}}(q(t)). \tag{4.3b}$$

The dynamics are as before, except the excitatory recurrence is absent (see below), and there is a population of N^{A} topographically-ordered afferent cells providing input to both populations of neurons. The rates of the afferent cells r_l^{A} are dependent on the current stimulus $-\pi \leq q(t) < \pi$, which represents the orientation or rescaled frequency of a stimulus (for visual and auditory stimuli respectively). For stimulus $q(t)$ the afferent cells fire at a rate defined by

$$r_l^{\mathrm{A}}(q(t)) = r^{\mathrm{vM}}\left(q(t)\middle|\mu = x_l^{\mathrm{A}},\ \kappa^{-1/2} = \sigma^{\mathrm{A}}\right) \tag{4.4}$$

where r^{vM} is the von Mises density function, scaled in amplitude so that its peak value is $r_{\max}^{\mathrm{A}}$. For clarity, when reporting results for the auditory networks, we convert the stimulus values to frequencies. All network parameters are the same as in Table 4.1, except for those listed here in Table 4.2.

N^{A}	256	N^{S}	16	Number of afferent cells, stimuli.
$\sigma^{\mathrm{E}\leftarrow\mathrm{A}}$	$\pi/4$	$\sigma^{\mathrm{I}\leftarrow\mathrm{A}}$	$\pi/16$	Connection distances.
σ^{A}	$\pi/16$	$r_{\max}^{\mathrm{A}}$	10 Hz	Afferent rate parameters.
$p^{\mathrm{I}\leftarrow\mathrm{E}}$	.2	$p^{\mathrm{E}\leftarrow\mathrm{I}}$	.2	Connection probabilities.
$p^{\mathrm{I}\leftarrow\mathrm{A}}$	.2	$p^{\mathrm{I}\leftarrow\mathrm{I}}$	.2	
ρ_0	1 Hz	τ_{slow}	10 ms	Retroaxonal target & slow time constant.
Θ	2 Hz	τ_{bcm}	30 s	BCM target & slow time constant.
η^w	8×10^{-8}	η^{I}	5×10^{-8}	Learning rates.
$w_{\max}^{\mathrm{I}\leftarrow b}$	$10w^{\mathrm{I}\leftarrow b}$	$w_{\max}^{\mathrm{E}\leftarrow\mathrm{I}}$	$10w^{\mathrm{E}\leftarrow\mathrm{I}}$	Max. weight for plastic synapses.

Table 4.2: **Sensory network model parameters.** For explanations see the text. Parameters not specified here remain the same as in Table 4.1. Retroaxonal plasticity parameters are modified due to interaction with BCM plasticity, and time scale of stimulus protocol.

Connectivity in the sensory networks

The afferent-to-interneuron connections, as well as those that are also present in the homeostatic network model, are constructed as before. To impose stimulus tuning on cells in the excitatory population, the afferent-to-excitatory connections are computed as a dense matrix, which causes subpopulations of excitatory cells to respond to similar stimuli. This is intended to replace the excitatory-to-excitatory recurrent connections, which would contain strongly connected assemblies of principal cells that are selective

for similar stimuli. Each row of the connection matrix is filled with values corresponding to a von Mises function (with $\mu = x_j^{\mathrm{E}}$ and $\kappa^{-1/2} = \sigma^{\mathrm{E}\leftarrow\mathrm{A}}$ for the j-th row), but as opposed to using it as the connection probability—as for the other connections—the values of the function directly determine the weight matrix.

Experimental protocol for sensory networks

The stimulus tuning of the neurons in the sensory networks are obtained by measuring the network response to various stimuli. The stimuli are intended to mimic the afferent input that would arise from pure auditory tones, or visual gratings of different orientations. Although the stimulus space q is continuous, we use only N^{S} values q_k, distributed over the stimulus space in equal intervals.

During the experimental protocol the network is sequentially exposed to 300,000 stimulus sets. A set is 10 seconds long, and stimulates the network with each of the $N^{\mathrm{S}} = 16$ stimuli q_k for 0.5 seconds, intermixed with 4 half-second periods of no afferent drive (the background state). The order of stimulus presentation is randomised in each set.

Plasticity is disabled for the first 3000 sets to measure the initial response to the stimuli, and for the last 3000 sets while measuring how the response has changed due to plasticity. In both phases, we evaluate the stimulus-dependent response of cells by measuring the firing rate of every cell in response to each stimulus q_k, averaged over the period being measured. The result is the receptive fields of the excitatory cells $r_j^{\mathrm{E}}(q_k)$ and inhibitory cells $r_i^{\mathrm{I}}(q_k)$, before and after learning.

In total, the experimental protocol requires 50,000 minutes, or slightly more than a month, of simulated time.

4.3 Synaptic plasticity models

The main focus of the thesis rests upon understanding the effect of different forms of interneuron plasticity on the homeostasis of excitatory activity, and the stimulus tuning of sensory neurons. The simulation results are only for synaptic forms of plasticity. Because synapses in excitatory-inhibitory networks must maintain their sign to satisfy Dale's principle (Dale, 1935), we rewrite plastic synaptic weights as a non-linear rectifying function of an underlying parameter V that is without sign constraints. For example, when synapses from population b to population a are plastic, we set $W_{ij}^{a\leftarrow b} = \Delta_{ij}^{a\leftarrow b} s^+(V_{ij}^{a\leftarrow b}, \alpha_p)$, where $V_{ij}^{a\leftarrow b}$ is the unconstrained parameter, and the rectification function $s^+(x, \alpha)$ is as previously defined. That parameter $V_{ij}^{a\leftarrow b}$ is then updated according to one of the following rules.

Heterosynaptic plasticity rules

One principled approach to deriving homeostatic plasticity rules is to perform gradient descent on an objective function $\mathcal{E}$ that measures the squared deviation of the excitatory

firing rates from a homeostatic target value ρ_0:

$$\mathcal{E}\left(\boldsymbol{r}^{\mathrm{E}}\right) = \frac{1}{2}\sum_{l=1}^{N^{\mathrm{E}}}\left(r_l^{\mathrm{E}} - \rho_0\right)^2$$

Here, $\boldsymbol{r}^{\mathrm{E}}$ denotes the vector of excitatory firing rates. Although the objective offered above measures error in the firing rates, the objective we employ for the derivation of the gradient uses the membrane potentials $\boldsymbol{h}^{\mathrm{E}}$. Despite this difference, following that gradient has the effect of reducing error in the firing rates—which are simply the rectified potentials. We use the potential because it is linear in membrane current, whereas the firing rate is strictly non-negative and its gradient disappears for hyperpolarized cells. The detailed derivation of the gradient rules is rather technical and can be found in Chapter 7.

These derived rules are highly non-local and hence biologically implausible, so we developed several approximate learning rules possessing greater plausibility. We refer to the synaptic variant of these plasticity rules as heterosynaptic plasticity, due to the fact that the activity of a single excitatory cell can affect many synapses on the same inhibitory neuron. In those rules, the variables $V_{ij}^{\mathrm{I}\leftarrow b}$ (that parametrize the excitatory synaptic weight onto the interneurons before rectification), are updated according to

$$\Delta V_{ij}^{\mathrm{I}\leftarrow b} \propto \left[\sum_{k=1}^{N^{\mathrm{E}}}\left(\left\langle r_k^{\mathrm{E}}\right\rangle_t - \rho_0\right)K_{ki}\right]\frac{\partial r_i^{\mathrm{I}}}{\partial h_i^{\mathrm{I}}}\frac{\partial W_{ij}^{\mathrm{I}\leftarrow b}}{\partial V_{ij}^{\mathrm{I}\leftarrow b}}r_j^b, \tag{4.5}$$

where x_i is the i-th component of the vector expression $\boldsymbol{x}$. The presynaptic population b is the afferent population A for the auditory network, or the excitatory population E for all other simulations. The partial derivatives arise from the derivation of the gradient, with the firing rate term causing only synapses onto active interneurons to change, and the synaptic weight term producing an effective lower bound on the value of $V_{ij}^{\mathrm{I}\leftarrow b}$. The constant of proportionality is the learning rate η^w. K is a matrix that takes different forms depending on the specific approximation of the gradient-based rule. We studied three different variants:

- $K = \Delta^{\mathrm{E}\leftarrow\mathrm{I}}$ for the retroaxonal rule, (4.5a)
- $K = \left\langle\Delta^{\mathrm{E}\leftarrow\mathrm{I}}\right\rangle$ for the diffusion-like rule, and (4.5c)
- $K^\intercal = \Delta^{\mathrm{I}\leftarrow\mathrm{E}}$ for the anterograde rule. (4.5b)

For the diffusion-like rule, the expectation $\langle\cdot\rangle$ is the ensemble average, and thus a Toeplitz matrix filled with the von Mises density function used to compute the inhibitory-to-excitatory connections. The justification for these variants is discussed in Chapter 5. The time expectation $\left\langle r_k^{\mathrm{E}}\right\rangle_t$ denotes a running average of the firing rate of excitatory cell k

$$\left\langle r_k^{\mathrm{E}}(t)\right\rangle_t = r_k^{\mathrm{E}}(0)e^{-t/\tau_{\mathrm{slow}}} + \frac{1}{\tau_{\mathrm{slow}}}\int_0^t \mathrm{d}t'\, r_k^{\mathrm{E}}(t')e^{(t'-t)/\tau_{\mathrm{slow}}}, \tag{4.6}$$

where τ_{slow} is the time scale of the exponentially decaying averaging window.

Homosynaptic plasticity rules

In addition to the gradient-based heterosynaptic rules, we also investigate the effect of simpler, homosynaptic forms of synaptic plasticity (see Section 5.1.3). Their form—at first glance—would be consistent with a homeostatic role, in that they have a set-point where the excitatory rates are at the homeostatic target:

$$\Delta V_{ij}^{\mathrm{I}\leftarrow\mathrm{E}} \propto \left(\left\langle r_j^{\mathrm{E}}\right\rangle_t - \rho_0\right) r_i^{\mathrm{post}} \frac{\partial W_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}}{\partial V_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}}. \tag{4.7}$$

Again, we study three variants: $r_i^{\mathrm{post}} = r_i^{\mathrm{I}}$ leads to a Hebbian rule (Eq 4.7a), while $r_i^{\mathrm{post}} = r_{\mathrm{ref}}(r_i^{\mathrm{I}} + r_{\mathrm{ref}})^{-1}$ is an anti-Hebbian rule (Eq 4.7b), and $r_i^{\mathrm{post}} = 1$ is a non-Hebbian rule (Eq 4.7c). For the anti-Hebbian rule, the asymptotic approach of $r_i^{\mathrm{post}} \to 0$ as $r_i^{\mathrm{I}} \to \infty$ is dictated by the constant $r_{\mathrm{ref}} = 25$ Hz (while $r_i^{\mathrm{post}} = 1$ when $r_i^{\mathrm{I}} = 0$).

We also contrast (in Section 5.1) or combine (in Section 5.2) plasticity of excitatory connections onto interneurons with plasticity of inhibitory connections onto excitatory cells. When plastic, inhibitory synapses evolve according to

$$\Delta V_{ji}^{\mathrm{E}\leftarrow\mathrm{I}} \propto \left(\left\langle r_j^{\mathrm{E}}\right\rangle_t - \rho_0\right) \frac{\partial W_{ji}^{\mathrm{E}\leftarrow\mathrm{I}}}{\partial V_{ji}^{\mathrm{E}\leftarrow\mathrm{I}}} r_i^{\mathrm{I}}, \tag{4.8}$$

obtained by simplifying the gradient derived in the Appendix (Eq 7.10) by assuming weak synapses, as done for the heterosynaptic rules. The constant of proportionality is the learning rate η^I. Note that this rule (Eq 4.8) is very similar to that of Vogels et al. (2011), although it is obtained through a novel derivation.

In the sensory cortex simulations, for excitatory synapses onto interneurons, we contrast the aforementioned retroaxonal rule (Eq 4.5a) with the classic Bienenstock-Cooper-Munro (BCM) rule (Bienenstock et al., 1982), slightly modified to account for positivity of weights:

$$\Delta V_{ij}^{\mathrm{I}\leftarrow b} \propto \left(r_i^{\mathrm{I}} - \frac{\left\langle r_i^{\mathrm{I}}\right\rangle_t^2}{\Theta}\right) r_i^{\mathrm{I}} \frac{\partial W_{ij}^{\mathrm{I}\leftarrow b}}{\partial V_{ij}^{\mathrm{I}\leftarrow b}} r_j^b, \tag{4.9}$$

where the time expectation of the post-synaptic firing rate is estimated as in Eq 4.6, except with time constant τ_{BCM}. The parameter Θ serves as an approximate set-point for the firing rate of the postsynaptic inhibitory cell. The presynaptic population b is dictated by the network configuration (afferent population A for the auditory cortex model, excitatory population E for the visual cortex model).

Structural plasticity in sensory cortex

When the retroaxonal rule (Eq 4.5a) is combined with inhibitory synaptic plasticity (Eq 4.8), the propagation kernel of the former is replaced with

$$K_{ji} = s^+(W_{ji}^{\mathrm{E}\leftarrow\mathrm{I}} - 0.1, 10^3)\Delta_{ji}^{\mathrm{E}\leftarrow\mathrm{I}}.$$

This is similar to previous uses of s^+, except that there is a very sharp cut-off (with $\alpha = 10^3$), ensuring $K_{ji} = 0$ whenever $W_{ji}^{\mathrm{E}\leftarrow\mathrm{I}} < 0.1$. The effect of this is to cause interneurons to become unresponsive to the rate of postsynaptic excitatory cells that have become decoupled by inhibitory synaptic plasticity. Aside from this exception, there is no structural plasticity for synapses.

4.4 Measures

Total error

To evaluate the homeostatic properties of the various plasticity rules, one of the primary measures we use is the root-mean-square error (RMSE) with respect to the homeostatic set-point ρ_0. We refer to this in the text as the "total error" and it is computed as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N^{\mathrm{E}}} \sum_j (r_j^{\mathrm{E}} - \rho_0)^2}. \tag{4.10}$$

This is essentially the objective function of the gradient-based rules, and thus total error plays a central role in this thesis.

Quiescent cells

When evaluating the homosynaptic and anterograde rules in Section 5.1.3 we often observe very sparse network activity. To quantify this effect, we measure the fraction of excitatory cells in each network that are deemed inactive for a period of time. To decide which cells are inactive, we first compute the average firing rate of each cell over the course of the 55 hour experiment in 15 minute bins (this binning is done of necessity, to keep the amount of data manageable). Then, for the last 10 hours of the simulation, we find the bin with the maximum firing rate of each cell $r_j^{\max}$. If that maximum rate is $r_j^{\max} < (10 \times 60 \times 60)^{-1}$ Hz, i.e. the cell would be expected to fire less than one spike per 10 hours in any of the 15 minute bins—then cell j is declared quiescent.

Stimulus selectivity index

An important measure of the receptive field of a cell in sensory areas is its stimulus selectivity index. This measures the degree to which a cell i fires preferentially in response to a single stimulus. It is defined as

$$\mathrm{SI} = \frac{\sqrt{\sum_k^{N^{\mathrm{S}}} r_i(q_k)\sin(q_k) + r_i(q_k)\cos(q_k)}}{\sum_k^{N^{\mathrm{S}}} r_i(q_k)}. \tag{4.11}$$

Goodness of fit

To determine the preferred stimulus of each cell, we use a linear regression to fit a von Mises density function $\hat{r}_i(q_k)$ to the receptive field $r_i(q_k)$ of each cell. To evaluate how

well the stimulus tuning of the cell is described by this ansatz, we define the "goodness of fit" for each cell to be the coefficient of determination

$$R^2 \equiv 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \tag{4.12}$$

wherein the residual sum of squares is

$$SS_{\text{res}} = \sum_{k}^{N^{\text{S}}} \left(r_i(q_k) - \hat{r}_i(q_k)\right)^2,$$

and the total sum of squares is

$$SS_{\text{tot}} = \sum_{k}^{N^{\text{S}}} \left(r_i(q_k) - \widetilde{r_i}\right)^2$$

with $\widetilde{r_i} = \frac{1}{N^{\text{S}}} \sum_k^{N^{\text{S}}} r_i(q_k)$, the average response of cell i across all stimuli.

4.5 Limits of interneuron-based homeostatic plasticity

In order to gain a first understanding of the limitations of a homeostatic rule operating via interneuron activity, it is illustrative to consider a system in which an "ideal rule" can set inhibitory firing rates with arbitrary precision, so long as they remain non-negative. The two constraints facing us are a fixed inhibitory-to-excitatory synaptic weight matrix, and input to the system that is non-negative and contains some spatial heterogeneities.

We consider the steady-state of the system, and assume that an ideal rule could bring the firing rate of every excitatory cell to the target rate ρ_0, assuming that the set-point is such that $\boldsymbol{r}^{\text{E}}(\rho_0) = \rho_0$. This is achieved by taking the dynamics (Eq 4.1) to be zero, and then setting the excitatory membrane potentials to ρ_0:

$$0 \stackrel{!}{=} \tau_{\text{E}} \dot{\boldsymbol{h}}^{\text{E}} = (W^{\text{E}\leftarrow\text{E}} - \mathbb{I})\rho_0 - W^{\text{E}\leftarrow\text{I}}\boldsymbol{r}^{\text{I}} + w^{\text{in}}\boldsymbol{r}^{\text{E}\leftarrow\text{in}}. \tag{4.13}$$

By defining the $w^{\text{in}}\boldsymbol{r}^{\text{E}\leftarrow\text{in}} + (W^{\text{E}\leftarrow\text{E}} - \mathbb{I})\rho_0 \equiv \Delta\mathbf{I}$ and rearranging, we are left with a condition for the inhibitory firing rates:

$$\Delta\mathbf{I} = W^{\text{E}\leftarrow\text{I}}\boldsymbol{r}^{\text{I}}. \tag{4.14}$$

Here, $\Delta\mathbf{I}$ is the vector of inhibitory currents the excitatory neurons would require to reach the target rate.

For the sake of argument, we shall assume that there are as many inhibitory cells as excitatory. This one-to-one relationship between the two populations represents a "minimal scenario" for any homeostatic rule that attempts to bring all excitatory cells to the homeostatic target rate. In the presence of significant input heterogeneities, a smaller inhibitory population (in the absence of inhibitory plasticity) is unlikely to be able to provide adequate degrees of freedom to achieve the target in every excitatory cell.

If the two populations are of equal size, we can study a simple, but illustrative case where each inhibitory cell has a preferred excitatory partner (corresponding to the diagonal values of the weight matrix being $W_{ii}^{\mathrm{E}\leftarrow\mathrm{I}} = 1$) while non-preferred cells have a inhibitory synaptic weight of $c < 1$ as in

$$W^{\mathrm{E}\leftarrow\mathrm{I}} = \begin{pmatrix} 1 & c & \cdots & c \\ c & 1 & \cdots & c \\ \vdots & \vdots & \ddots & \vdots \\ c & c & \cdots & 1 \end{pmatrix}.$$

If $c = 0$ the ideal homeostatic rule could fine tune the firing rate of every inhibitory cell to meet the needs of its postsynaptic excitatory cell. But what happens if $c > 0$? To determine the effect of non-zero cross-coupling, we decompose the inhibitory weight matrix into

$$W^{\mathrm{E}\leftarrow\mathrm{I}} = c\mathbb{1}\mathbb{1}^\intercal + (1-c)\mathbb{I} \tag{4.15}$$

where $\mathbb{1}$ is the 1 vector, and $\mathbb{I}$ is the identity matrix. The first eigenvector is $\mathbf{v}_1 = \mathbb{1}/\sqrt{N}$ with eigenvalue

$$\lambda_1 = \frac{N + (1-c)}{\sqrt{N}},$$

while all remaining eigenvectors are orthogonal to it $\mathbf{v}_{i>1} \perp \mathbb{1}$, with eigenvalues $\lambda_i = 1 - c$. Because this matrix is Hermitian, we can express its inverse as

$$\left(W^{\mathrm{E}\leftarrow\mathrm{I}}\right)^{-1} = \sum_i \frac{1}{\lambda_i}\mathbf{v}_i\mathbf{v}_i^\intercal.$$

Next we decompose $\Delta\mathbf{I}$ into two parts, one that corresponds to its mean μ_I, and another which corresponds to deviations from the mean for each term (the heterogeneity due to afferent input and recurrent excitation)

$$\Delta\mathbf{I} = \mu_I\mathbb{1} + \sigma_I\delta\mathbf{I},$$

where $\|\delta\mathbf{I}\| = 1$ and $\delta\mathbf{I}\mathbb{1}^\intercal = 0$. Finally, combining the above

$$\boldsymbol{r}^{\mathrm{I}} = \left(W^{\mathrm{E}\leftarrow\mathrm{I}}\right)^{-1}\Delta\mathbf{I} = \frac{\sqrt{N}\mu_I}{cN + (1-c)} + \frac{\sigma_I}{1-c}\sum_{i\neq 1}\mathbf{v}_i\mathbf{v}_i^\intercal\delta\mathbf{I}.$$

Because $\sum_{i\neq 1}\mathbf{v}_i\mathbf{v}_i^\intercal\delta\mathbf{I} = \delta\mathbf{I}$ ($\delta\mathbf{I}$ is orthogonal to $\mathbb{1}$, and the remaining vectors form a complete orthonormal basis for the remaining space), it follows that

$$\boldsymbol{r}^{\mathrm{I}} = \left(W^{\mathrm{E}\leftarrow\mathrm{I}}\right)^{-1}\Delta\mathbf{I} = \frac{\sqrt{N}\mu_I}{cN + (1-c)} + \frac{\sigma_I}{1-c}\delta\mathbf{I},$$

which must be non-negative. If $c = 0$ then

$$\boldsymbol{r}^{\mathrm{I}} = \sqrt{N}\mu_I + \sigma_I\delta\mathbf{I}$$

and existence of a non-negative r^{I} is guaranteed so long as $\sqrt{N}\mu_I > \sigma_I$, which is to say

that the mean input should be larger than the maximum difference in inputs.

Increasing the cross-coupling c has the effect of amplifying the heterogeneity of the input σ_I, reducing the likelihood that there exists a set of non-negative firing rates for the inhibitory neurons that will satisfy the homeostasis target. In the limit of $1 - c = \epsilon$, inhibitory rates would have to become extremely large, both positive and negative for the excitatory cells to fire at the homeostatic target rate.

Precisely this limitation is encountered in the networks we simulate, due to the structure of connectivity. With that in mind, we proceed to characterize what is achievable with plasticity that regulates interneuron activity to be homeostatic for excitatory neurons in the following chapter.

5 Results

5.1 Homeostasis

To understand the ways in which regulating interneuron activity could be homeostatic for excitatory principal cell firing rates, we employ two complementary approaches. The first begins by choosing a measure of global network activity (the total deviation from a homeostatic set-point across all excitatory cells), and subsequently deriving a set of gradient-based plasticity rules that minimize this measure by modifying interneuron activity. We contrast the homeostatic capabilities of one of those gradient-based rules with that of inhibitory synaptic plasticity in Section 5.1.1. The gradient-based rules are non-biological, and so we find plausible approximations to them which modify the activity of each interneuron in response to the firing rates of its postsynaptic excitatory targets. We call this postsynaptic-population interneuron plasticity (Post-IP), and investigate its effect on network homeostasis in Section 5.1.2.

In the second approach we investigate plasticity rules that regulate the activity of an interneuron in response to deviations in the principal cells providing excitatory drive to it. We refer to this as presynaptic-population interneuron plasticity (Pre-IP; Section 5.1.3). The first approach is normative, and is intended to find optimal solutions. The second is descriptive, and elucidates the consequences of certain experimentally reported forms of interneuron plasticity.

5.1.1 Limited homeostasis via regulation of interneuron activity

How should interneuron activity be regulated to effectively control principal cell firing rates within a cortical network? We begin by deriving a gradient designed to minimize the network-wide total error in the activity of individual excitatory cells with respect to some homeostatic set-point. For reasons of readability, the detailed derivation of these rules is provided elsewhere (see Section 7.1 of the Appendix). Here, we merely state one result of that derivation, and discuss its features and consequences.

Following the gradient reduces error by changing excitatory-to-inhibitory synapses $W_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}$, or interneuron intrinsic properties (their firing threshold θ_i or gain g_i). By choosing one of those variables the gradient simplifies to a plasticity rule. The synaptic variant is

$$\Delta W_{ij}^{\mathrm{I}\leftarrow\mathrm{E}} \propto -\Big[\underbrace{\left(\left\langle \boldsymbol{h}^{\mathrm{E}}\right\rangle_t - \rho_0\right)^{\intercal}}_{\text{(i)}}\underbrace{\mathcal{W}^{-1}}_{\text{(ii)}}\underbrace{W^{\mathrm{E}\leftarrow\mathrm{I}}}_{\text{(iii)}}\underbrace{\mathcal{M}^{-1}}_{\text{(iv)}}\Big]_i \underbrace{\frac{\partial r_i^{\mathrm{I}}}{\partial h_i^{\mathrm{I}}}\frac{\partial W_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}}{\partial V_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}} r_j^{\mathrm{E}}}_{\text{(v)}} \tag{5.1}$$

where $\mathcal{W} = \mathbb{I} - W^{\mathrm{E}\leftarrow\mathrm{E}}\mathrm{diag}\left(\frac{\partial \boldsymbol{r}^{\mathrm{E}}}{\partial \boldsymbol{h}^{\mathrm{E}}}\right) + W^{\mathrm{E}\leftarrow\mathrm{I}}\mathcal{M}^{-1}\mathrm{diag}\left(\frac{\partial \boldsymbol{r}^{\mathrm{I}}}{\partial \boldsymbol{h}^{\mathrm{I}}}\right) W^{\mathrm{I}\leftarrow\mathrm{E}}\mathrm{diag}\left(\frac{\partial \boldsymbol{r}^{\mathrm{E}}}{\partial \boldsymbol{h}^{\mathrm{E}}}\right)$, and $\mathcal{M} = \mathbb{I} + \mathrm{diag}\left(\frac{\partial \boldsymbol{r}^{\mathrm{I}}}{\partial \boldsymbol{h}^{\mathrm{I}}}\right) W^{\mathrm{I}\leftarrow\mathrm{I}}$. In those expressions, $\mathbb{I}$ is the identity matrix, and $\mathrm{diag}(\cdot)$ is a

diagonal matrix containing the present activation of each neuron in either the excitatory or inhibitory population. Their values are ≈ 0 if the neuron is hyperpolarised, and ≈ 1 if it is depolarised. For the sake of simplicity, the rule is shown here to operate on the synaptic weights directly. In actual fact it modifies an underlying parameter $V_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}$, which is then rectified to $W_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}$.

In brief, the gradient-based rules consist of a product of five factors (i) - (v), and achieve their objective by doing the following (as illustrated in Fig 5.1A):

(i) Calculate the deviation of each excitatory principal cell from the homeostatic set-point, to determine whether to increase or decrease its inhibitory input.

(ii) Using detailed knowledge of recurrent connectivity and the activation of all cells $\mathcal{W}$, determine how to change the input to excitatory cells such that once propagated through the network, its deviation from the set-point will decrease.

(iii) Use the exact structure of inhibitory-to-excitatory connectivity $W^{\mathrm{E}\leftarrow\mathrm{I}}$ to decide how to change the activity of each interneuron such that it will produce the necessary changes for its postsynaptic excitatory population, as found in the previous step.

(iv) Evaluate how changes in the activity of the interneurons will interact within the inhibitory circuit $\mathcal{M}$ to determine how to achieve the desired change of rate in each interneuron.

(v) Finally, preferentially modify interneurons that are not hyperpolarised, and for the synaptic rule, using synapses that are strongly driven and non-zero.

It is important to emphasize that updating any single plastic parameter requires knowledge of every other variable in the network (synaptic weights, intrinsic variables, and membrane potentials). This is because the objective is to minimize the total error in the network, and the ultimate effect of changing one parameter can only be accounted for if the rule considers how all elements of the system interact.

With these gradient-based plasticity rules in hand, we begin by choosing the variant that modifies excitatory-to-inhibitory synapses (Eq 5.1), and contrast its effects with homeostatic plasticity of inhibitory synapses (Eq 4.8). We choose an experimental paradigm that mimics monocular deprivation experiments. This means simulating a network with either rule for about 28 hours so that plasticity brings firing rates to a steady-state distribution (the baseline). Then input to the network is reduced, and simulated another 28 hours to observe the final (adapted) rate distributions.

Inhibitory synaptic plasticity (iSP) modifies inhibitory synapses onto excitatory cells such that the postsynaptic cell fires at the homeostatic set-point. In our simulations it brings the entire population of principal cells to the homeostatic set-point (both in the baseline and adapted state; Fig 5.1B, left). This is thanks to the many degrees of freedom afforded by changing each inhibitory synapse made by a given interneuron separately, while leaving its own firing rate to be determined by other variables in the network.

By contrast, controlling principal cells by regulating the firing rates of interneurons using the plasticity of Eq 5.1 achieves a considerably coarser homeostasis (Fig 5.1B,

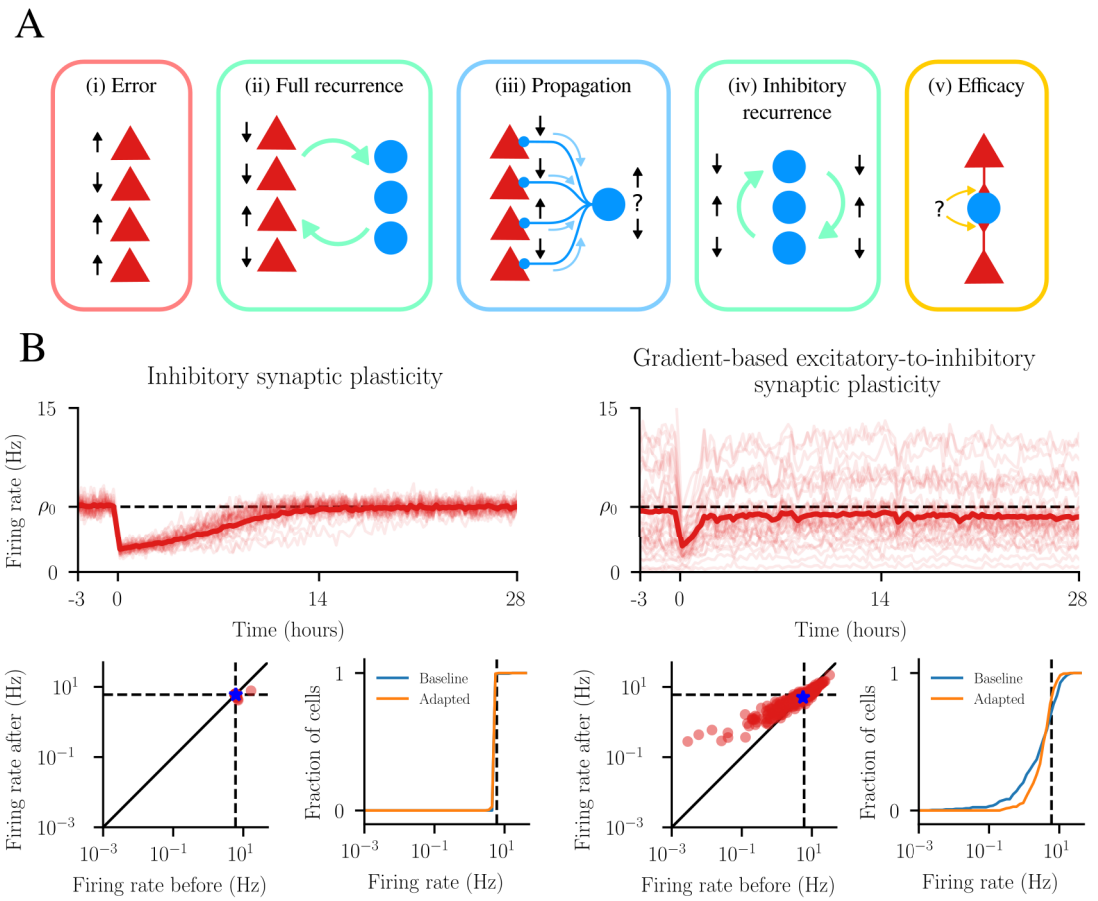

Figure 5.1: **Gradient-based plasticity of interneuron activity achieves limited control over error in excitatory firing rates compared to inhibitory synaptic plasticity. A)** Schematic of the steps the gradient-based rule follows to minimize the total squared-error of the membrane potential w.r.t. a homeostatic set-point ρ_0. For a description, see the text. **B)** Results from simulating input deprived networks with iSP (left) and the gradient-based synaptic plasticity (right). Top are the firing rates over the course of the experiment, dark lines are the population rate, light lines are individual cells. Deprivation of input occurs at hour 0. Bottom: scatter plots comparing the rates of individual cells at -3 hours (baseline) and 28 hours (adapted). Next to the scatter plots are their corresponding cumulative distributions. The blue star indicates the population firing rate. Black dashed lines are the target rate ρ_0.

right). The firing rate distribution prior to deprivation is broad, with cells firing above and below the set-point. Twenty-eight hours after the input deprivation, firing rates have reached a steady-state distribution different than that prior to the intervention. This inability to bring all cells to the set-point (both before and after deprivation) is because each interneuron inhibits many principal cells, and if plasticity can only modify interneuron rates, then no biologically feasible (i.e. non-negative) distribution of inhibitory activity can bring every cell in the excitatory population to the homeostatic set-point. Any plasticity using interneuron rates as a homeostatic controller will

inevitably be limited by the structure of inhibitory-to-excitatory connectivity, and heterogeneities within the network activity and inputs. We show this analytically for a simplified system in the Methods chapter (Section 4.5). Furthermore, excitatory cells outnumber their inhibitory counterparts by a factor of four, yielding degrees of freedom typically inadequate to perfectly control the excitatory firing rate distribution.

5.1.2 Post-ISP: Approximating the gradient

Although any plasticity regulating interneuron activity can only achieve limited control over excitatory rates, and the gradient-based rules are clearly non-biological, they remain useful. Since our objective is to ascertain how interneuron-based homeostatic plasticity might act within the brain, we must find approximations that are plausibly biological. In the gradient-based rule, the factors $\mathcal{W}^{-1}$ and $\mathcal{M}^{-1}$ pose a particular challenge for implementation because they include the state of every synapse and neuron in the network. As we have explained, they permit the rule to compensate for the interaction between neurons in the network. But if synapses are sufficiently weak (i.e. much less than one, as they are in our network), only monosynaptic interactions are essential for the gradient, and multisynaptic connections may be neglected.

To eliminate multisynaptic interactions from the rule, we perform a first-order geometric expansion of $\mathcal{W}^{-1}$ and $\mathcal{M}^{-1}$, and eliminate any terms that are not monosynaptic (for details see Section 7.2). In this manner $\mathcal{W}^{-1}$ and $\mathcal{M}^{-1}$ disappear from the rule, leaving behind only $W^{\mathrm{E}\leftarrow\mathrm{I}}$, which is factor (iii) of the previous section. That this factor remains is crucial, because the rule must transmit information about the firing rates of excitatory cells to the interneurons, so that they may alter their activity levels. This is precisely what the detailed inhibitory-to-excitatory connectivity $W^{\mathrm{E}\leftarrow\mathrm{I}}$ accomplishes.

That is not to say that only $W^{\mathrm{E}\leftarrow\mathrm{I}}$ can accomplish this task. In fact, it has the disadvantage that the rule must possess specific information about synaptic weights. For the sake of reducing the biological complexity of our approximation, we replace the synaptic weights with the inhibitory-to-excitatory connectivity $\Delta^{\mathrm{E}\leftarrow\mathrm{I}}$ (which indicates which synapses are present, but not their weight). We call this approximation the retroaxonal rule (Eq 4.5a). But our approximations need not end there. That retroaxonal rule requires detailed connectivity information, which might be still too onerous a requirement. We can next ask where, on average, connections might be? Averaging $\Delta^{\mathrm{E}\leftarrow\mathrm{I}}$ over many networks (the ensemble average) tells us exactly that. When we replace the inhibitory-to-excitatory connectivity with the ensemble average $\left\langle \Delta^{\mathrm{E}\leftarrow\mathrm{I}} \right\rangle$, we refer to the rule as diffusion-like (Eq 4.5c). In the following section we provide an explanation of how both the diffusion-like and retroaxonal rules operate.

Diffusion-like and retroaxonal plasticity

In the previous section, we offered two mathematical approximations to the gradient-based rule, for reasons of biological plausibility. So how can those two approximate rules be interpreted? They operate on distinct principles: One approximate rule responds to deviations from the set-point in all nearby principal cells, via a non-specific diffusion-like signal (Fig 5.2A, left). In contrast, the other approximate rule regulates the activity of each interneuron in response to deviations from the set-point in the

principal cells it forms inhibitory synapses onto, via retroaxonal signalling (Fig 5.2A, right).

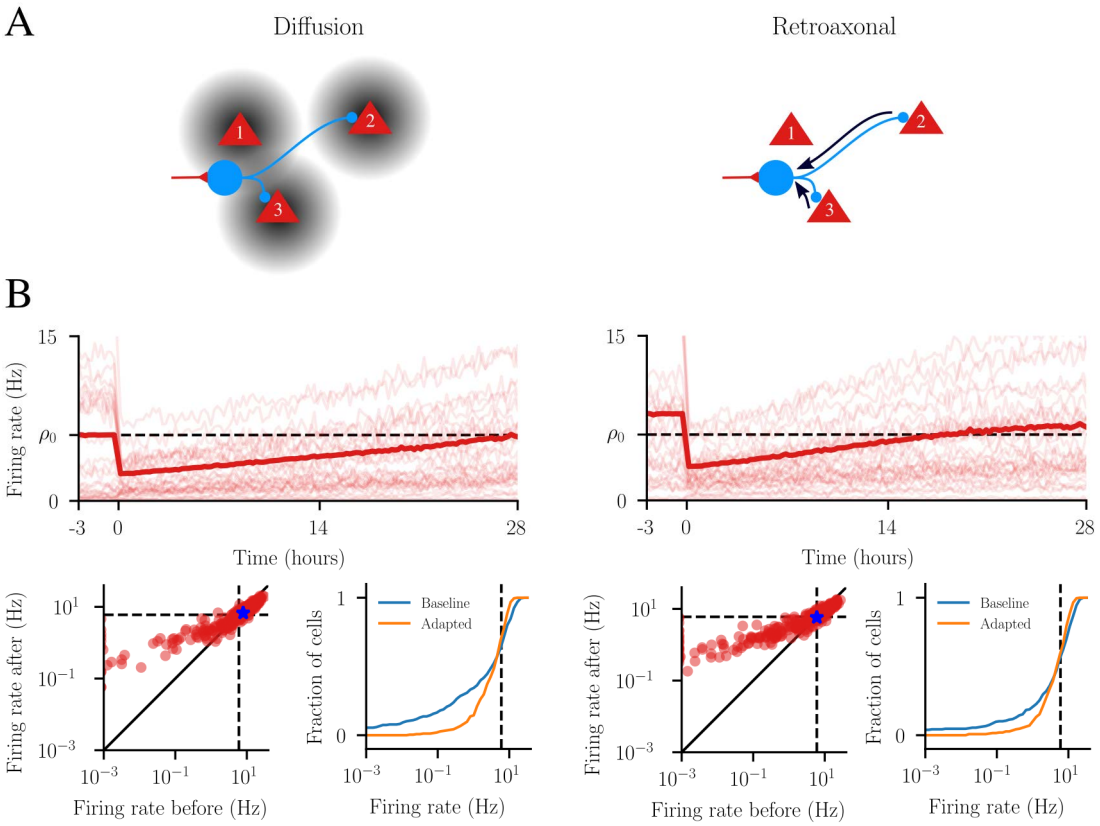

Figure 5.2: **Post-IP rules respond to input deprivation similarly to gradient-based rule. A)** Left: The diffusion-like rule (Eq 4.5c). Right: The retroaxonal rule (Eq 4.5a). See the text for a description of both. **B)** Plots are the same as Fig 5.1B, but with different plasticity.

To illustrate the difference between the two rules, consider the toy networks depicted in Fig 5.2A. For the diffusion-like rule (left; Eq 4.5c), there is one interneuron whose incoming excitatory synapse changes according to

$$\Delta w^{\mathrm{I}\leftarrow\mathrm{E}} \propto K_1(r_1^{\mathrm{E}} - \rho_0) + K_3(r_3^{\mathrm{E}} - \rho_0).$$

As that expression indicates, it is responsive to firing rate error in excitatory cells scaled by a constant K. The value of K is inversely related to the distance between the interneuron and the cell (as depicted by the grey diffusion cloud). So $K_2 = 0$ because cell 2 is too far away, and consequently despite being inhibited by the interneuron, the synapse will not be altered in response to any error in the rate of that cell. In the full network, the scalars $K_{ji} = \langle \Delta_{ji}^{\mathrm{E}\leftarrow\mathrm{I}} \rangle$ are equal the connection probability from a typical inhibitory cell i to a typical excitatory cell j, corresponding to the ensemble average $\langle \cdot \rangle$ of the connectivity matrix.

For the retroaxonal rule (right; Eq 4.5a), the synapse onto the interneuron changes

according to

$$\Delta w^{\mathrm{I}\leftarrow\mathrm{E}} \propto \Delta_2(r_2^{\mathrm{E}} - \rho_0) + \Delta_3(r_3^{\mathrm{E}} - \rho_0).$$

In this case the synapse onto the interneuron will respond to error only in cells it inhibits, where $\Delta_2 = \Delta_3 = 1$. In contrast with the diffusion-like rule, it ignores the nearby cell 1 because $\Delta_1 = 0$, as they are not connected. In the full network, these Δ coefficients are defined by the connectivity matrix $\Delta^{\mathrm{E}\leftarrow\mathrm{I}}$.

Now that we have a basic understanding of how these two rules work, we simulate them in networks using the same input deprivation protocol as before (Fig 5.2B). Both at the baseline and after adapting to reduced input, neither rule brings individual rates to the set-point, nor do they restore the distribution of excitatory firing rates. But qualitatively, both rules appear to behave like the full gradient-based plasticity. The question is, how robust is the resemblance, and in what sense are the rules homeostatic?

Retroaxonal and diffusion-like plasticity exhibit differential responses

At first glance, the diffusion-like and retroaxonal rules behave similarly to the full gradient rule. To determine whether this holds up when the simplifying assumption is violated, we co-vary the strength of excitatory and inhibitory synaptic weights across simulations (Fig 5.3A), such that locally generated excitatory drive onto principal cells is slightly less than the afferent excitation (weak synapses) or is about three times more (strong synapses). The inhibitory-to-inhibitory synapses are kept at their default value to avoid winner-take-all dynamics within the interneuron population.

Although the gradient rule achieves a lower total error in rates than the other rules (Fig 5.3A, left), their firing rate distributions are qualitatively similar, even for the strong synaptic weights (Fig 5.3A, middle). On the other hand, the three rules have markedly different effects on the population rate (Fig 5.3A, right). Regardless of synaptic weight, the retroaxonal plasticity brings the excitatory population to fire at a rate of 7-8 Hz, slightly higher than the set point $\rho_0 = 6$ Hz. We have shown analytically in Section 7.3 that this positive offset results from variability of inhibitory in-degree across the excitatory population. To confirm this we simulate networks with different coefficients of variation $\mathrm{CV}^{\mathrm{E}\leftarrow\mathrm{I}}_{\text{in-degree}}$ and as predicted by our analysis the population rate rises with the CV of the in-degree (Fig 5.3B, top). In that analysis (Section 7.3) we assume that the rates of excitatory cells are anti-correlated with their inhibitory in-degree. This assumption proves true in the simulations, with the expected negative correlation between rates and in-degree holding (Fig 5.3B, bottom). In contrast the diffusion-like plasticity brings the population rate precisely to the set-point, regardless of synaptic weight or inhibitory in-degree, once again in agreement with the same analysis.

The non-zero total error for the diffusion-like rule can be understood in the context of back-propagation. Lillicrap et al. (2016) have shown that the classical back-propagation algorithm can effectively learn even if the feedback matrix is not identical to the transpose of the feedforward matrix (as it is in classical back-propagation), so long as the forward and back matrices are sufficiently similar (measured by their angle). The greater their similarity, the lower the expected error. For the diffusion-like rule, in sparse networks the diffusion operator is over 70 degrees from the back-propagation matrix $\Delta^{\mathrm{E}\leftarrow\mathrm{I}}$, and that angle is reduced as the connection probability increases towards the limit of dense local connectivity, achieving a value of 40 degrees (Fig 5.3C). There is

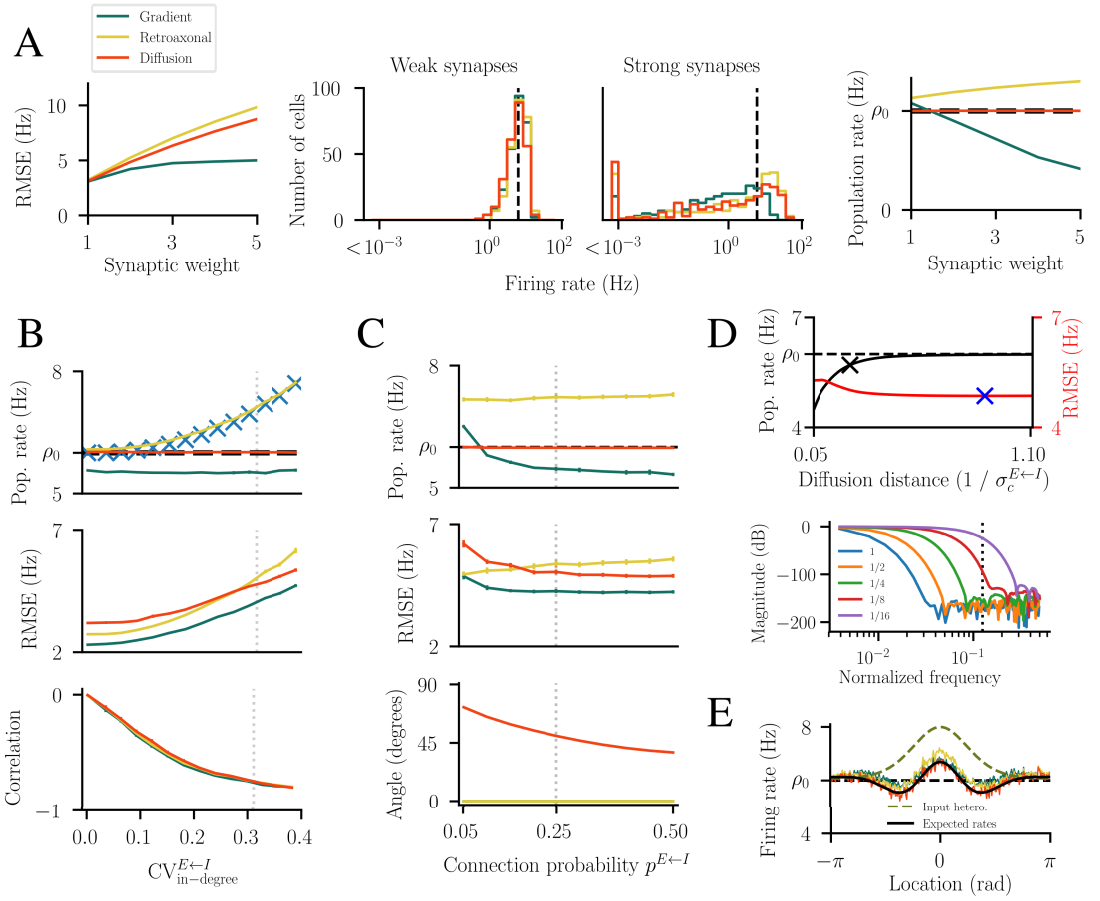

Figure 5.3: **Post-IP rules exhibit differential responses to variation in network parameters A)** Simulation results for the gradient-based, retroaxonal and diffusion-like plasticity in networks with different excitatory and inhibitory synaptic weight values. Left: Total error of all excitatory firing rates with respect to the homeostatic set-point. Middle two: Final distribution of excitatory firing rates for networks with weak and strong synapses. Right: Excitatory population firing rate. **B)** Results when varying $\mathrm{CV}^{\mathrm{E}\leftarrow\mathrm{I}}_{\text{in-degree}}$, the inhibitory-to-excitatory in-degree coefficient of variation. Top: Excitatory population firing rate. Blue crosses indicate analytic prediction. Middle: Total error of excitatory rates. Bottom: Correlation between the firing rate of an excitatory cell and its inhibitory in-degree. **C)** Results when varying connection probability $p^{\mathrm{E}\leftarrow\mathrm{I}}$. Top: Excitatory population firing rate. Middle: Total error of excitatory rates. Bottom: Angle between the diffusion operator $K = \langle \Delta^{\mathrm{E}\leftarrow\mathrm{I}} \rangle$ and the connection matrix $\Delta^{\mathrm{E}\leftarrow\mathrm{I}}$. **D)** Results for diffusion-like plasticity when the diffusion distance is varied from one-twentieth (0.05) of the connection distance $\sigma_c^{\mathrm{E}\leftarrow\mathrm{I}}$ to slightly larger. Top: Excitatory population firing rate (black) and total error (red). Black cross indicates where the rate is 95% of ρ_0, at 0.23 of the default distance. Blue cross indicates the diffusion distance yielding the minimum error, which occurs at 0.9 of the default. Bottom: Filter gain as a function of spatial frequency, for several diffusion distances (normalized to $\sigma_c^{\mathrm{E}\leftarrow\mathrm{I}}$). Dotted black line indicates the Nyquist frequency following N^{E} to N^{I} down-sampling. **E)** Spatial firing rate plots after plasticity has adapted to an input heterogeneity (dashed olive line). The solid black line is the expected response, as described in the text. Faint dashed lines in some plots indicate default network parameters.

a corresponding drop in the total error, while error for the retroaxonal rule rises slightly, despite it employing the back-propagation matrix. This non-zero error observed in the latter case is due in part to the constant offset from the target rate discussed above.

Diffusion as decimation of an error signal

Viewed from a signal processing perspective, the task of these rules is to take a spatial error signal (N^{E} error terms, or "samples") and somehow reduce it to a signal of N^{I} samples, which are then used to alter the firing rate of the interneurons. That reduction from N^{E} to N^{I} samples is known as down-sampling. Down-sampling a signal necessitates prior application of a low-pass filter to avoid signal aliasing—a phenomenon where power from high frequencies in the original signal appears as spurious signals at lower frequencies in the down-sampled signal. To avoid problematic aliasing, such a filter must block sufficient power from frequencies greater than the Nyquist frequency of the new signal (the Nyquist frequency is half the sample frequency). These two steps, applying an anti-aliasing filter and then down-sampling, when taken sequentially are known as decimation.

Filters can be applied to a periodic sampled signal by constructing a circulant convolution matrix and multiplying it with the signal vector (this comparison to circular convolution is valid for our system because it is a ring network, meaning it is periodic). A circulant convolution matrix is square, with each row containing the filter kernel, but shifted such that the centre of the kernel in row i is at column i. Returning to the case at hand, one can construct our diffusion operator $\left\langle \Delta^{\mathrm{E}\leftarrow\mathrm{I}} \right\rangle$ by first forming an $N^{\mathrm{E}} \times N^{\mathrm{E}}$ convolution matrix $\hat{K}$, where the low-pass filter kernel is the von Mises function used to generate synaptic connections:

$$\hat{K}_{ij} = p^{\mathrm{vM}}\left(x_i^{\mathrm{E}} \middle| \mu = x_j^{\mathrm{E}},\ \kappa^{-1/2} = \sigma_c^{\mathrm{E}\leftarrow\mathrm{I}}\right)$$

To complete its construction, we right-multiply it with an $N^{\mathrm{E}} \times N^{\mathrm{I}}$ down-sampling matrix D that is everywhere zero, except for in each column k where row $4(k-1)+1$ is 1 (owing to the down-sampling factor $N^{\mathrm{E}}/N^{\mathrm{I}} = 4$):

$$D_{jk} = \begin{cases} 1, & \text{if } 1 \equiv j (\mathrm{mod}\ 4) \\ 0, & \text{otherwise.} \end{cases}$$

That will reduce the square convolution matrix to an $N^{\mathrm{E}} \times N^{\mathrm{I}}$ matrix identical to the diffusion operator, as in

$$\left\langle \Delta^{\mathrm{E}\leftarrow\mathrm{I}} \right\rangle_{ik} = \sum_{j=1}^{N^{\mathrm{E}}} \hat{K}_{ij} D_{jk}.$$

These two matrices correspond to the two steps of decimation: first a filter is applied by the convolution matrix $\hat{K}$ resulting in a filtered vector of N^{E} terms, and then that vector is reduced to N^{I} terms by the down-sampling matrix D, which is then used to modify interneuron activity.

Filter design is a field in itself (Antoniou, 1993), and well outside the scope of this work. But it is plausible that our default diffusion operator implements a filter

that is more restrictive than needed to avoid problematic aliasing. It might be that shorter distances—allowing more of the signal at high spatial frequencies to alias—are sufficient to achieve the homeostatic set-point for the population.

We tested this hypothesis, and varied the diffusion distance (filter kernel width) from about one-twentieth of the default to slightly larger than it (Fig 5.3D, top). Any diffusion greater than one quarter of the default distance is sufficient to bring the population rate to >95% of ρ_0. And though error is at minimum near the default distance, it does not increase significantly for any diffusion distance at which the population rate remains close to the set-point. Computing Bode magnitude plots for diffusion kernels of several different widths further confirms our hypothesis that the default kernel is more restrictive than necessary from a signal processing perspective (a Bode magnitude plot depicts the output gain of a filter as a function of input frequency; Fig 5.3D, bottom). For diffusion distances greater than 1/4 of the default, the Nyquist frequency lies well within the stop-band of the corresponding filter. These results suggest that satisfactory error is achievable at diffusion distances well below that obtained by a naive approximation of the full gradient.

Spatial input heterogeneities

As a final comparison of the plasticity rules, we exposed networks to spatially localized input heterogeneities, where a subpopulation receives elevated input, and waited for plasticity to reach a steady state. While none of the rules completely eliminated the heterogeneity (Fig 5.3E), both approximate rules perform comparably to the gradient-based rule. This inability to completely eliminate spatial deviations from the set-point is unavoidable for such plasticity rules, due to fan-out from inhibitory cells. The sort of spatial control afforded by changing interneuron firing rates is akin to iteratively measuring the error and applying a spatial low-pass filter—the inhibitory-to-excitatory connection kernel—to the desired control signal. To demonstrate this effect, we show the result of applying this iterative filtering procedure to the error, starting with the heterogeneity as the initial error (Fig 5.3E, solid black line). It well predicts the behaviour of all three rules. We argue that spatially detailed control cannot be achieved, even in the case of the optimal gradient-based rule.

5.1.3 Pre-IP rules induce competition between excitatory cells

Up until now, we have sought to understand the homeostatic capabilities of interneurons starting from a normative perspective by deriving plasticity rules using a global objective. We now characterise the behaviour of several synaptic plasticity rules inspired by experimentally reported forms of excitatory-to-inhibitory synaptic plasticity.

We start from the assumption that the activity of interneurons should increase if excitatory cells are firing above the homeostatic set-point, and decrease if they are below it, as per Wenner (2011). The rules considered achieve this by modifying the strength of excitatory synapses onto interneurons in response to deviations in the presynaptic principal cells (thus the abbreviation, pre-IP). We investigate three homosynaptic rules and one heterosynaptic rule. The homosynaptic rules include

- a *Hebbian* rule: a synapse increases if the presynaptic principal cell fires above the

set-point, and decreases if below. The magnitude of change is in proportion to the postsynaptic interneuron rate (Eq 4.7a):

$$\Delta V_{ij}^{\mathrm{I}\leftarrow\mathrm{E}} \propto \left(\left\langle r_j^{\mathrm{E}}\right\rangle_t - \rho_0\right) r_i^{\mathrm{I}} \frac{\partial W_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}}{\partial V_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}}$$

- an *anti-Hebbian* rule: similar to the Hebbian rule, but weight change is inversely proportional to the postsynaptic rate (Eq 4.7b):

$$\Delta V_{ij}^{\mathrm{I}\leftarrow\mathrm{E}} \propto \left(\left\langle r_j^{\mathrm{E}}\right\rangle_t - \rho_0\right) \frac{r_{\mathrm{ref}}}{(r_i^{\mathrm{I}} + r_{\mathrm{ref}})} \frac{\partial W_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}}{\partial V_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}}$$

- a *non-Hebbian* rule: without any postsynaptic rate term (Eq 4.7c):

$$\Delta V_{ij}^{\mathrm{I}\leftarrow\mathrm{E}} \propto \left(\left\langle r_j^{\mathrm{E}}\right\rangle_t - \rho_0\right) \frac{\partial W_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}}{\partial V_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}}.$$

The heterosynaptic rule is an *anterograde* rule where all synapses onto an interneuron respond to the average deviation in the entire population presynaptic to the cell (Eq 4.5a). That rule is obtained by replacing the propagation matrix of the post-IP rules (e.g. $\Delta^{\mathrm{E}\leftarrow\mathrm{I}}$ for the retroaxonal rule) with the excitatory-to-inhibitory connectivity matrix $\Delta^{\mathrm{I}\leftarrow\mathrm{E}}$. This alters the character of the rule making it responsive to the cells providing excitatory drive, rather than those being inhibited.

Homeostasis for neither individual nor population rates

Using the same model network, we numerically simulated how these new rules respond to variations in synaptic weight. None of the rules reliably bring the population rate to the homeostatic set-point, with population rates decreasing as synaptic weights increase (Fig 5.4A, left). Simultaneously, the error and fraction of cells quiescent for hours at a time increase along with the synaptic weights (Fig 5.4A, middle and right respectively; see Section 4.4 for the definition of quiescence employed here). The Hebbian rule is most pathological in this regard, with nearly two in five cells exhibiting no activity whatsoever at the strongest synaptic weight. The total error for all three homosynaptic rules nearly saturates at its maximum value, given that for stronger weights most of the cells are firing below the target rate (Fig 5.4B). Only the anterograde rule avoids pushing nearly the entire population below the target, but does so with higher total error.

We hypothesize that the pre-IP rules induce a competition between excitatory cells in response to two widely observed features of biological neuronal networks: shared inhibition in the presence of heterogeneity, and disynaptic inhibition.

Shared inhibition

To understand the importance of shared inhibition in the competition between cells, it is instructive to consider connectivity patterns within this cortical-like network.

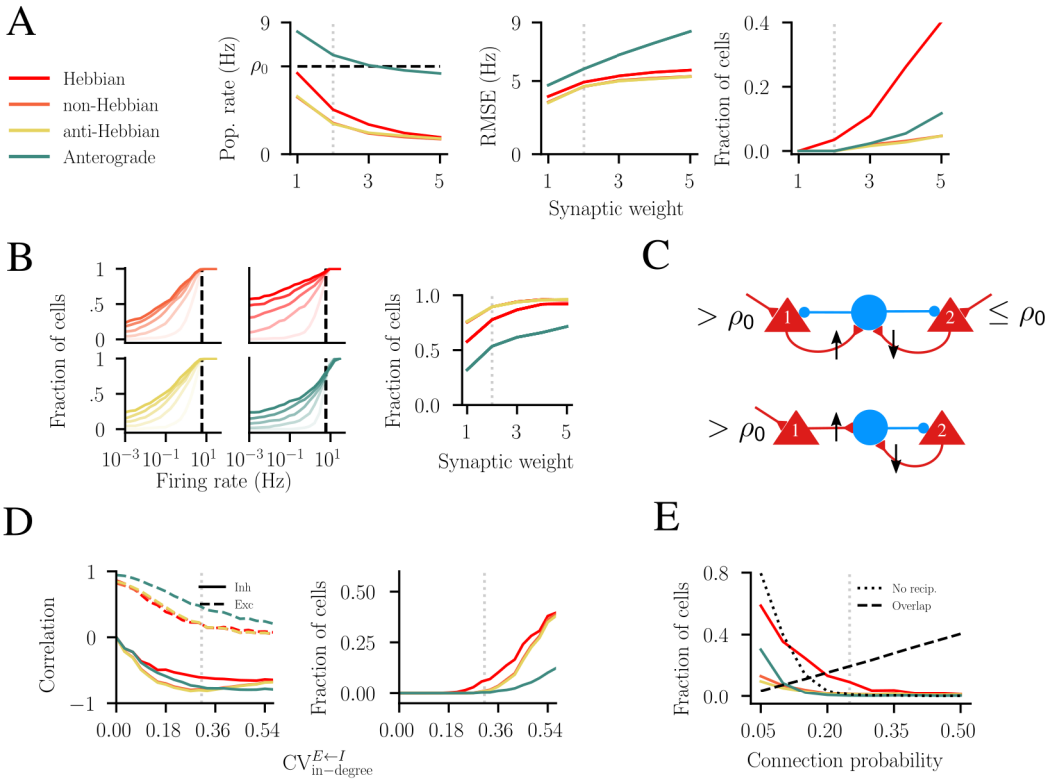

Figure 5.4: **Pre-IP rules induce competition.** Measures computed after convergence of networks with one of four pre-IP plasticity rules (for convergence criteria and definition of the measures see Section 4). For these results non- and anti-Hebbian plasticity produce nearly identical outcomes, and as such the former are often concealed behind the latter in the plots. **A)** Results from simulations of the pre-IP rules in networks with different co-varying excitatory and inhibitory synaptic weight. Left: Excitatory population firing rate. Middle: Total error of excitatory cell rates with respect to the homeostatic set-point ρ_0. Right: Fraction of excitatory cells quiescent for hours at a time. Criteria as per Section 4.4. **B)** Left: Firing rate cumulative distributions for various synaptic weights (darker lines are stronger weights). Right: Fraction of cells with firing rates below the target rate. **C)** Connection motifs illustrating two sources of competition. Top: Shared inhibition. Bottom: Disynaptic inhibition. **D)** Results when varying $\mathrm{CV}^{\mathrm{E\leftarrow I}}_{\mathrm{in\text{-}degree}}$, the inhibitory-to-excitatory in-degree coefficient of variation. Left: The correlation between the firing rate of an excitatory cell and its inhibitory in-degree (solid lines), or its excitatory in-degree (dashed lines). Right: Fraction of excitatory cells quiescent for hours at a time. **E)** Fraction of cells quiescent in networks where all connection probabilities are covaried from very sparsely connected, to densely connected. The dotted line is the fraction of excitatory cells with zero reciprocal connections with interneurons. The dashed line is the average fraction of overlap between cells in both the presynaptic and postsynaptic population for an interneuron.

Many principal cells project to any particular interneuron, and in turn that interneuron projects back onto most of the nearby principal cells. This "shared inhibition" motif is pervasive in cortex where excitatory cells typically out-number inhibitory cells by a

factor of at least four to one.

In such a setting, the homosynaptic rules will tend to over-inhibit most of the principal cells. To see why, picture a small excitatory-inhibitory circuit where one interneuron provides inhibition to two excitatory cells as depicted at the top of Fig 5.4C. If excitatory cell 1 has a net synaptic input that causes it to fire above the set-point, while cell 2 fires at or below it, cell 1 will increase its excitatory drive to the interneuron until it receives sufficient inhibition to reach the target rate. Because in our network the inhibitory synapses are static, as the activity of the interneuron increases, it will push the firing rate of cell 2 below the set-point. In response, it will begin to decrease its synaptic weight onto the interneuron, which in turn causes cell 1 to further increase its synaptic weight. This competition continues until the synapse from cell 2 reaches its lower bound of zero. At that point cell 1 will continue to increase its weight to satisfy its need for inhibition, while cell 2 can do nothing to oppose it.

Though this three cell circuit provides a simple intuition, in larger assemblies the picture becomes more complicated. There will be a large diversity of excitatory cells, all with different net inputs, and many sharing multiple interneurons. One might assume that it is the cell with the strongest excitatory input that will dictate the total level of inhibition within the assembly. This is only true for networks with very homogeneous inhibitory in-degree (corresponding to a low $\mathrm{CV}^{\mathrm{E}\leftarrow\mathrm{I}}_{\text{in-degree}}$), where excitatory in-degree predicts the firing rate of each excitatory cell (Fig 5.4D, left).

As inhibitory in-degree becomes increasingly heterogeneous (high $\mathrm{CV}^{\mathrm{E}\leftarrow\mathrm{I}}_{\text{in-degree}}$), the excitatory inputs become less important and inhibitory in-degree begins to predict excitatory firing rates. To see why, imagine two excitatory cells with similar excitatory input, but one with twice as many inhibitory inputs as the other. If the assembly they are part of requires more inhibition on average, as local interneuron activity increases, the rate of the cell with more inhibitory inputs will decrease more rapidly than the one with less. So despite each cell having a large number of reciprocal connections with interneurons—for the default network parametrization, fully one third of all excitatory-to-inhibitory connections are reciprocated—they have less control over the amount of inhibition they receive than those connections might imply.

In the regime of highly heterogeneous inhibitory in-degree the non- and anti-Hebbian rules begin to silence large portions of the network, comparable to that of the Hebbian rule (Fig 5.4D, right). This suggests that they too can be highly competitive, though the positive feedback from the postsynaptic term in the Hebbian rule differentiates it in the default network. It is likely that other sources of heterogeneity, such as the afferent input, would also cause the non- and anti-Hebbian rules to exhibit behaviour more in line with that of the Hebbian rule.

Disynaptic inhibition

The effect of disynaptic inhibition is illustrated by the circuit depicted at the bottom of Fig 5.4C, with excitatory cell 1 driving a single interneuron, which does not reciprocate the connection, though it does inhibit excitatory cell 2. If cell 1 is driven to fire above the set-point, it will increase all of its synapses onto interneurons whether or not the connection is reciprocated. In this case, because it receives no additional inhibition, its firing rate remains above the set-point and thus over time sends the synapse to

its upper limit. So even if cell 2 receives strong excitatory drive, it will eventually be silenced by ever-increasing inhibition (assuming the synaptic and rate bounds are sufficiently large).

The role of disynaptic inhibition becomes more apparent in simulations when connection probability is decreased, thus increasing the chances that an excitatory cell makes no reciprocal connections with any interneuron (Fig 5.4E). At connection probabilities below the default value of 0.25 the fraction of cells without any reciprocal connections becomes non-zero, and the non- and anti-Hebbian rules—which produce almost no quiescent cells with the default parameters—begin to silence part of the population, rising above one-tenth of the network at the most sparse connectivity, with $p = 0.05$.

Anterograde plasticity is robust in cortical networks

The anterograde rule is less susceptible to the competitive behaviour observed with the homosynaptic rules. Across the full range of synaptic weights the population rate is closer to the set-point, and individual cells are roughly evenly split between firing above and below it. Even in highly heterogeneous networks the fraction of silent cells remains below 0.15, compared to the homosynaptic rules which all rise to nearly 4 of 10 cells silent. What makes this rule comparatively robust?

The first reason is simply that the synapses onto interneurons do not change in response to each presynaptic cell individually. Rather, they are changed in response to the presynaptic population rate. While this is not optimal—-in the sense of minimizing error with respect to the homeostatic set-point—in our network it is sufficient to produce an equitable outcome for the population.

It is worth noting that while up until this point the anterograde rule seems non-competitive, that is not entirely the case. Whereas the competition for homosynaptic pre-IP rules is between the *individual cells*, for the anterograde rule the competition is between the *population of cells* presynaptic to each interneuron. The rule remains unresponsive to the activity of cells postsynaptic to the interneuron, unless those two populations are identical. Over the range of connection probabilities tested, the overlap between the pre- and postsynaptic populations range from as many as 40 of 100 cells belonging to both (at $p = .5$), to as few as 3 out of 100 cells at $p = 0.05$. At connection probabilities $p < 0.2$, the degree of overlap is insufficient to compensate for disynaptic inhibition and the number of silent cells increases along with the fraction of cells possessing no reciprocal connections. Despite this, the anterograde rule *is* more robust to changes in network parameters than the homosynaptic rules. It is only in the presence of significant disynaptic inhibitory connections that the competitive nature of the anterograde plasticity rule is manifest.

We have shown that for interneurons to minimize the total deviation of excitatory cells from the homeostatic set-point, their activity should be regulated by a rule responsive to the excitatory population they inhibit. In the next section, we assume that interneurons do just that, and investigate the effect of the retroaxonal rule on processing in sensory cortex.

5.2 Sensory processing

In Section 5.1 of the results, amongst other rules, we derive a novel form of homeostatic synaptic plasticity reliant upon retroaxonal signalling. That rule modifies excitatory synapses onto interneurons to minimize deviations from a homeostatic set-point in the excitatory population inhibited by each interneuron. Those results were obtained in a mostly static setting, where the only source of time-variability was uncorrelated Ornstein-Uhlenbeck noise on the afferent inputs to the network. We will now investigate what functional consequences the retroaxonal rule has for the processing of time-varying stimuli within sensory cortex.

5.2.1 What connectivity in A1 can account for a delay between the shift of excitatory and inhibitory currents?

As stated in Section 2.1 of the introduction, Harris and Mrsic-Flogel (2013) propose that co-tuning of excitatory and inhibitory currents in primary auditory cortex arises due to local feedback (recurrent) inhibition within the tonotopically ordered network. While this seems a plausible explanation for E/I co-tuning, Froemke et al. (2007) report that pairing a pure-tone auditory stimulus with release of acetylcholine from nucleus basalis (NB) can shift the preferred frequency of excitatory synaptic currents measured in the membrane of A1 principal cells of adult rats. As shown in Fig 5.5, this shift occurs within 30 minutes, but the preferred frequency of inhibitory currents can take more than 100 minutes to do the same. Restoring normal co-tuning of excitatory-inhibitory currents might take as long as 3 hours.

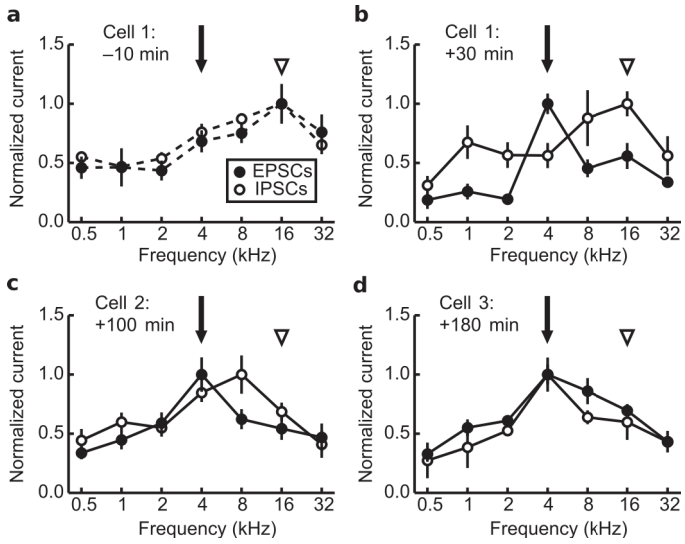

Figure 5.5: **Membrane currents recorded from a neuron in rat auditory cortex, in response to pure auditory tones.** a) Baseline response. b) Response 10 minutes after pairing 4 kHz tones with stimulation of NB. Response at c) 100 and d) 180 minutes after pairing. Black arrow indicates paired-tone frequency, white arrow indicates original preferred frequency. Reproduced with permission from Froemke et al. (2007).

We begin by asking whether this peculiar lag between changes in excitatory tuning

and inhibitory tuning is possible in a network where the co-tuning arises exclusively due to local recurrent connectivity within A1. On its face, it seems that if interneurons having similar selectivity to nearby principal cells due to local recurrent connectivity, there could be no lag on the time scale observed. Rather, we argue that in such a network, any change in excitatory tuning—reflected in the measured excitatory currents—would be immediately reflected in the inhibitory currents, on the time scale of network dynamics. We hypothesize that stimulus selectivity in interneurons must arise not from local recurrent connectivity, but rather from tuned afferent input.

To test this hypothesis we simulate networks with several combinations of connectivity: synapses from principal cells to interneurons are chosen to be either spatially broad and random, or local, and with their weights weak or strong. Similarly, afferent connections to interneurons are either spatially random, or tonotopically local, and weak or strong. The entire network is driven by a set of afferent input patterns, representing pure-tone auditory stimuli of different frequencies. We impose tonotopically ordered receptive fields upon the principal cells, with each having a slightly different characteristic frequency, covering the entire frequency spectrum. Following Froemke et al. (2007), we measure excitatory and inhibitory currents onto a single principal cell, and then to simulate the reported changed in excitatory current tuning we shift the characteristic frequency of a subpopulation of principal cells towards a new frequency as depicted in the upper left of Fig 5.6, a shift intended to simulate the change resulting from pairing that tone with electrical stimulation of nucleus basalis. We then immediately repeat the measurement of currents onto the same cell following the shift.

The objective here is to observe the response of the network to the shifted tuning of excitatory cells, within the first 30 minutes following the paired-tone protocol, prior to the observed shift in inhibitory currents (which presumably occur due to some plasticity acting on a different time scale than that which causes the excitatory shift). Consequently, these first results are obtained from simulating the network with no plasticity rule active, and thus any changes in inhibitory currents arise purely due to the changes of excitatory tuning within the network.

As we hypothesized, in networks with strong, tonotopically local recurrent synapses, the inhibitory currents shift immediately—in the absence of any inhibitory plasticity, and in contradiction with the experimentally observed behaviour (Fig 5.6, top left & right, bottom left & right). On the other hand, if afferent inputs are strong and tonotopically tuned, while recurrent synapses are random or weak, the preferred frequency of the inhibitory current remains unchanged (Fig 5.6, middle row, centre and right; bottom centre), reminiscent of the experimental results.

Notably, in the network with strong tonotopic afferent and random recurrent connections (Fig 5.6, bottom centre), there is an increase of inhibition at the new preferred frequency. Though it is not large enough to shift the preferred frequency of inhibition, it does stand in contrast to the experimental results, which actually exhibit the opposite behaviour: Inhibition at the paired-tone frequency is suppressed (Fig 5.5, upper right). In reporting their results, the authors argue that this suppression is likely due to the effects of acetylcholine, which in auditory cortex can enhance excitation while also suppressing inhibition during the pairing protocol. As we do not model any aspect of the protocol aside from simply shifting the tuning of excitatory cells, it is to be expected that we do not observe a similar suppression in our results.

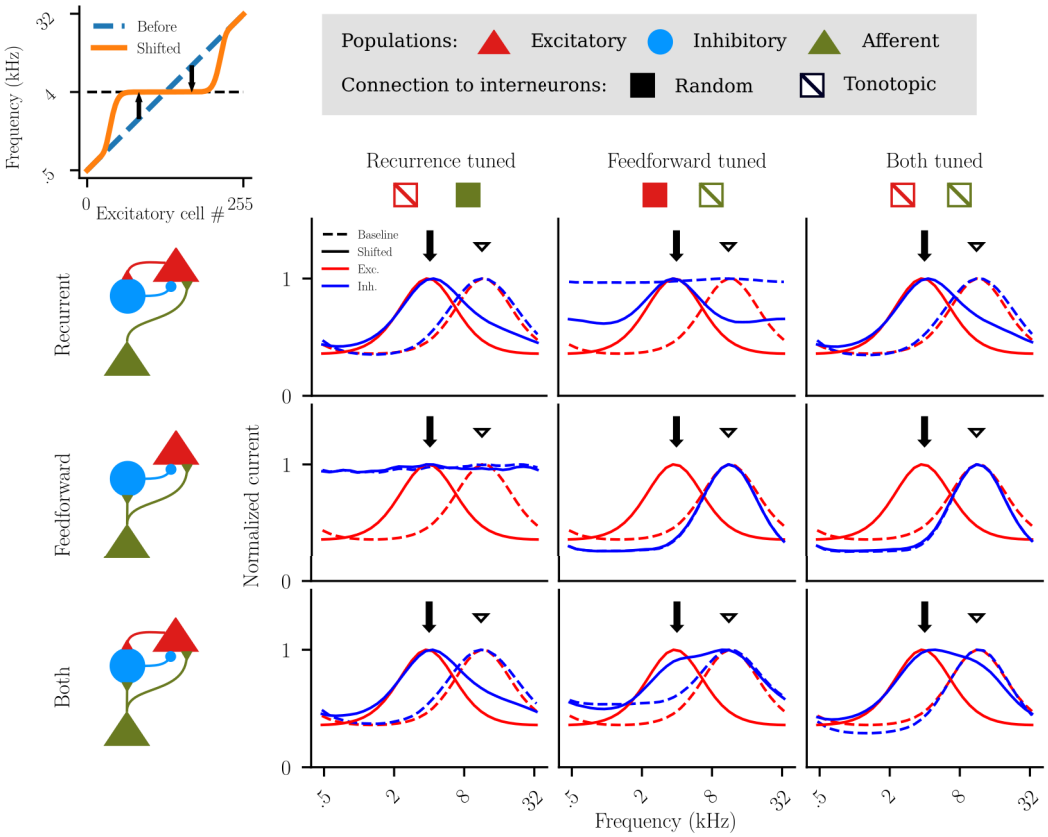

Figure 5.6: **Strong tonotopic afferent input to interneurons, with random or weak excitatory input are sufficient to reproduce auditory cortex experimental result.** Results from nine network types, with no plasticity present. Rows correspond to different strength of synapses onto interneurons. Top row: Recurrent excitatory inputs are much stronger than the afferent inputs (inhibition is feedback). Middle row: Afferent inputs are much stronger than recurrent inputs (inhibition is feedforward). Bottom row: Afferent and recurrent inputs are strong (both). Columns correspond to different connectivity patterns onto the interneurons. Left column: Recurrent excitation is local while afferent connections are random. Centre column: Afferent input is tonotopic while recurrent excitation is random. Right column: Recurrent inputs are local, and afferent inputs are tonotopic. Upper left depicts the preferred-tone frequency for excitatory cells before and after the experimental intervention.

We now proceed by exploiting our hypothesis that the network with strong afferent and recurrent (though random) input to interneurons is the operating regime of primary auditory cortex, and that these preliminary results are compatible with the observations of Froemke et al. (2007).

5.2.2 Under what conditions is inhibitory synaptic plasticity sufficient for precise co-tuning?

The previous section shows that to account for the experimentally observed delay between a shift in excitatory currents and inhibitory currents, the stimulus selectivity of interneurons in our model of A1 should be—in large part—due to stimulus-selective afferent inputs. This would be the end of the story, if E/I co-tuning were present at all stages of development. But Dorrn et al. (2010) report that in young rat pups co-tuning is initially weak, due to inhibitory currents being non-selective to frequency. According to those results, co-tuning emerges over the first month of development, eventually becoming correlated with $r = 0.71 \pm 0.05$ in adult rats.

This initially weak co-tuning suggests some form of plasticity is necessary, rather than simply being a product of anatomically structured connectivity. It has previously been shown in a feedforward setting that inhibitory synaptic plasticity (iSP), is sufficient to produce precise co-tuning (Clopath et al., 2016; Vogels et al., 2011). We now consider whether if in our recurrent network iSP alone is sufficient to account for this precise co-tuning.

We simulate iSP in the tonotopically arranged networks from before, where the connections from excitatory to inhibitory cells are strong but random, as suggested by our previous results (corresponding to the network in Fig 5.6, bottom middle). The iSP rule we use changes inhibitory synapses onto excitatory cells, such that the postsynaptic cell fires on average at a rate equal to a homeostatic set-point (note that there is no structural plasticity). In this setting, we consider two variables of interest (Fig 5.7A): In the presence of iSP, how selective does interneuron activity need to be? That is, should afferent-to-inhibitory connections be random or tonotopically ordered? And, how spatially localized are interneuron-to-principal cell connections?

To mimic recording in young rat pups, before enabling plasticity we expose the network to randomized pure-tone auditory stimuli, and record its response. In networks with both tuned afferent and local inhibitory connections, E/I co-tuning is present at the outset, whereas for the other network configurations inhibition is initially un-tuned (Fig 5.7B). Correlation of E/I currents confirms the visual inspection, with only the local-tonotopic network showing non-zero mean correlation of E/I currents (Fig 5.7C). It is important to note that this latter result is incompatible with the data from Dorrn et al. (2010).

Once iSP is enabled, we stimulate the network with pure-tone stimuli for hundreds of hours. Afterwards, in these "adult" neuronal networks, we turn off plasticity, and again measure the network response to auditory stimuli. In the networks with tonotopic afferent input to the interneurons, inhibition appears highly stimulus selective (Fig 5.7B, left column). The measured correlation between excitation and inhibition reveals a substantial increase in co-tuning within networks with either local or random inhibitory connectivity (Fig 5.7C, left column).

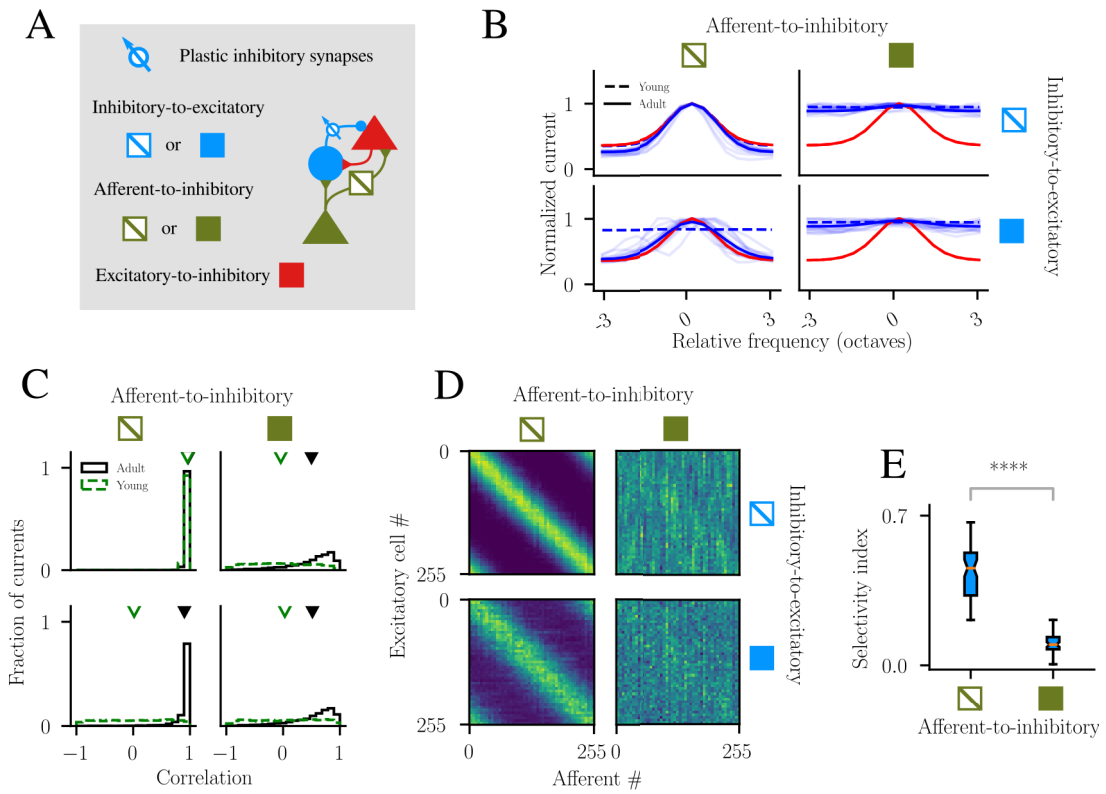

Figure 5.7: **Inhibitory plasticity creates co-tuning if interneurons are stimulus selective. A)** Networks simulated have plastic inhibitory synapses, that are either local or random, and afferent-to-inhibitory synapses that are local or random. Excitatory-to-inhibitory synapses are random, while afferent-to-excitatory are tonotopic. **B)** Membrane currents measured in young (dashed) and adult (solid) excitatory cells. Red lines are excitation, blue lines are inhibition. Dark lines are averaged over all cells, light blue lines show instances of inhibitory currents in adult cells. **C)** Correlations between excitatory and inhibitory currents. Mean values indicated with a $\triangledown$. Percentage of cells with correlated currents in Tonotopic-local (top left): Young 99.8% Adult 99.9%; Random-local (top right): Young 18% Adult 34%; Tonotopic-random (bottom left): Young 24% Adult 92%; Random-random (bottom right): Young 21% Adult 36%. Criteria is $p < 10^{-3}$ using Spearman rank-order correlation, where the two-sided p-value is for the null hypothesis that currents are uncorrelated. **D)** Effective connectivity from afferent inputs to excitatory cells, ordered according to stimulus preference. Yellow indicates a strong connection, blue indicates a weak connection. **E)** Stimulus selectivity index of interneurons. Results are the same for both local and random inhibitory connections. ***** indicates $p < 10^{-4}$ for Mann-Whitney U test.

The source of this co-tuning can be visualized by computing the "effective connectivity" from afferent inputs to excitatory cells (this is the matrix product of the afferent-to-inhibitory and inhibitory-to-excitatory connections). Because both excitatory cells and afferent inputs are ordered tonotopically according to their preferred frequency the effective connectivity matrix reveals a stimulus specific feedforward inhibitory connection from afferent to excitatory cell in the form of a diagonal band (Fig

5.7D, left column). In other words, excitatory cells with a particular preferred frequency receive feedforward inhibition from afferents with a similar preferred frequency.

In contrast, networks with random afferent-to-interneuron connections develop only weak co-tuning of currents (Fig 5.7B, right column), reflected in the comparatively lower, though non-zero correlation between currents measured in adult networks (Fig 5.7C, right column). Likewise, the effective connectivity exhibits very limited stimulus specificity in the structure of feedforward inhibition.

Why this failure to develop strong co-tuning? The answer lies in the degree of interneuron selectivity. The weak stimulus selectivity induced in interneurons by random afferent input (Fig 5.7E) provides an insufficient basis for iSP to produce precise co-tuning, in contrast to that induced by tonotopic inputs. For iSP to be effective in tuning inhibition to match excitatory currents that are strongly unimodal (tuned), the interneurons themselves must likewise be sufficiently unimodal in their responses to stimuli. This lack of unimodal response in interneurons is exactly what is quantified by a low selectivity index.

Compensating for shifted tuning with iSP

With these "adult" networks in hand, we shift the preferred stimulus frequency of a subpopulation of principal cells, as we did with the static networks (Section 5.2.1). For networks with interneurons strongly selective for stimulus (those with tonotopic afferent inputs), inhibitory current tuning is initially unchanged (Fig 5.8B, left column), which is as expected given our previous results. Once iSP re-establishes balance—at least on average—only networks with broad inhibitory connectivity can successfully shift the preferred frequency of inhibition to match excitation. This is because if inhibitory connections are too spatially localized, then some of the principal cells receive no inhibitory inputs from interneurons with a preferred frequency sufficiently similar to their new characteristic frequency.

Once again, the difference between these two cases can be visualised using the effective connectivity of feedforward inhibition (Fig 5.8E, left column). For broad inhibitory-to-excitatory connections, the previously diagonal band has been distorted, reflecting the very shift in preferred tuning imposed on the subpopulation of excitatory cells, indicating that stimulus specific feedforward inhibition has been preserved. For local inhibitory connections, the diagonal exhibits a mild distortion, but the subpopulation of excitatory cells with shifted stimulus preference no longer receives strong feedforward inhibition from afferents with similar stimulus preference. As stated earlier, this is because there simply are no anatomical connections from interneurons with the appropriate preferred stimulus frequency.

In networks with non-selective interneurons (due to their random afferent-to-inhibitory connections), inhibition is immediately biased towards the frequency of the paired-tone (Fig 5.8B, right column). This abrupt change of inhibition is reflected in the comparatively low correlation between the receptive fields of interneurons before and after the shift (Fig 5.8C). The new receptive field for more than half the interneurons has become uncorrelated with its field prior to the shift. What is causing this decorrelation? By fitting Gaussian-like curves to the interneuron receptive fields, we can show that while networks with strongly selective interneurons preserve their preferred frequency after

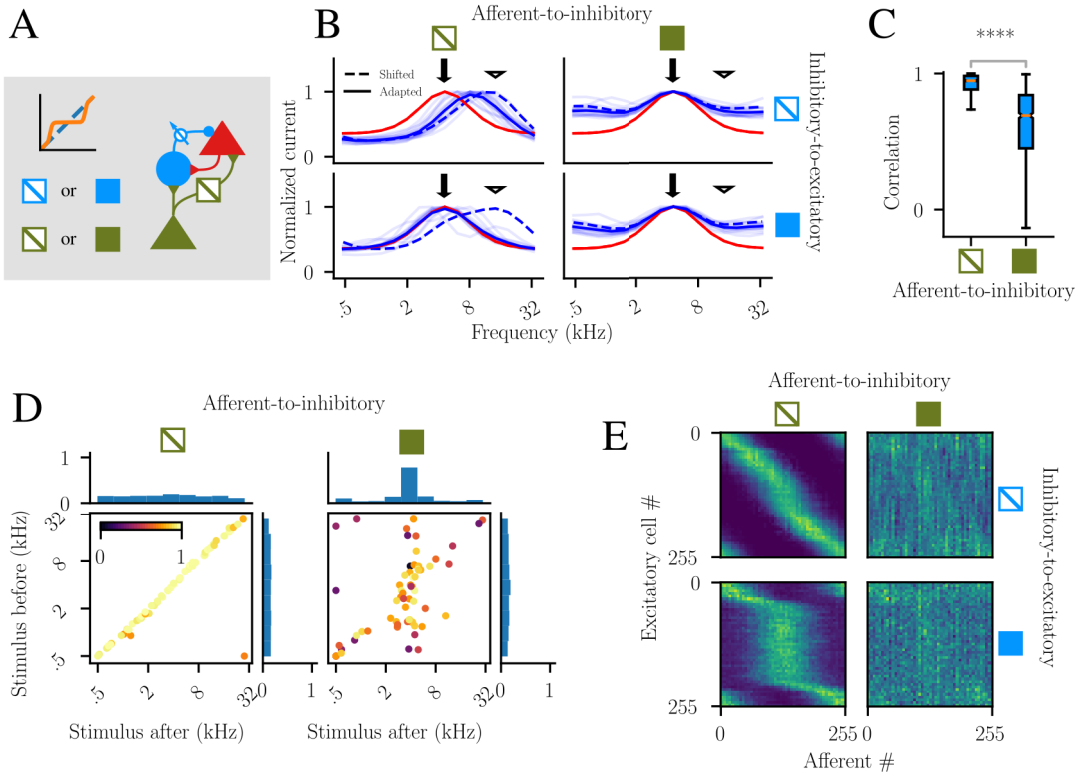

Figure 5.8: **In networks with only iSP, inhibitory connections must be spatially broad to compensate shifted tuning. A)** Same networks simulated as in 5.7A, but stimulus preference of some excitatory cells have been shifted. **B)** Membrane currents as measured in a single excitatory cell immediately following the shift in excitation (dashed), and after plasticity has adapted (solid). Dark lines are the average of 16 simulations of the same cell, light blue lines are single instances of the adapted inhibitory currents. **C)** Correlation between the receptive field of interneurons before and after the shift. Results are unaffected by inhibitory connection distance. Percentage of interneurons with new receptive fields uncorrelated with their previous field is Tonotopic: 7%, Random: 57%. Criteria is $p < 10^{-3}$ using Spearman rank-order correlation, where the two-sided p-value is for the null hypothesis that receptive fields are uncorrelated. **** indicates $p < 10^{-4}$ for Mann-Whitney U test. **D)** Preferred auditory frequency of interneurons before and after the shift, determined by fitting a Gaussian-like curve. Colour corresponds to goodness-of-fit. **E)** Effective connectivity from afferents to excitatory cells, after plasticity has adapted to the shift.

the shift (Fig 5.8D, left), networks with weakly selective interneurons exhibit a shift in preferred frequency towards the paired-tone frequency. This shift occurs because prior to the shift, excitatory current preferences are uniformly distributed over the entire auditory spectrum, and combined with random afferent and recurrent connectivity, interneurons also exhibit a uniform distribution of preferred frequencies. But because these preferences are induced by the details of random connectivity, they are weak, and so once much of the recurrent excitation produced within the network shifts to prefer the paired-tone frequency, the majority of interneurons likewise shift their preferred

stimulus to the same frequency.

So in networks with iSP as the only plasticity mechanism, our results thus far suggest that

1. Interneuron activity must be stimulus-selective (arising from strong tonotopic afferent inputs) to allow development of precise co-tuning of currents, and

2. Principal cells must initially receive inhibition from interneurons with a diverse range of preferred frequencies to account for initially weak tuning of inhibition observed in rat pups, and furthermore to permit compensatory changes of inhibition if the characteristic frequency of principal cells are significantly modified.

While these constraints do not seem impossible to satisfy, they do require a very particular pattern of anatomical and functional connectivity early in development. The breadth of inhibitory connectivity itself may be difficult to reconcile with the anatomy of auditory cortex as it is currently understood (Levy and Reyes, 2012; Li et al., 2014b), which suggests that IN to PN connections are typically not more than 200 μm (which would span about 1 octave of the tonotopic map; Zhang et al., 2001). We can loosen these constraints by introducing another form of plasticity, as we do in the following section.

5.2.3 Learning stimulus selectivity in interneurons

The results of the previous section suggest that iSP is sufficient to account for the different degrees of co-tuning in young and adult rats. But this is true only if interneuron activity is initially stimulus selective, and anatomical inhibitory connectivity spans unrealistic distances within A1. We now consider the opposite scenario, wherein interneurons are initially unselective, and inhibitory connectivity is anatomically local. Since we've established that interneurons must eventually exhibit some degree of selectivity, we need some kind of plasticity for its development. And that plasticity should operate on the afferent-to-inhibitory connections since as we have shown, interneuron selectivity cannot be imparted by recurrent excitation.

For the development of interneuron selectivity, we investigate the effects of two distinct forms of plasticity applied to the afferent-to-inhibitory connections: the retroaxonal rule studied in Section 5.1, and for comparison, the Bienenstock, Cooper, and Munro (BCM) rule. The BCM rule is a Hebbian plasticity with a sliding threshold for potentiation shown to lead to development of stimulus selectivity in the postsynaptic neuron (Bienenstock et al., 1982).

We simulate networks of four different configurations, summarized in Fig 5.9A: networks have retroaxonal or BCM plasticity on initially random afferent-to-inhibitory connections, along with spatially local inhibitory synapses that are either static or plastic. In the young networks, inhibition is un-tuned (Fig 5.9B) producing uncorrelated currents (Fig 5.9C). This is despite local inhibitory connections, because similar to the previous results for which inhibitory connections were local and afferent inputs were random, interneurons are initially unselective for stimulus (Fig 5.9D).

So far, there is little to distinguish the networks, with all producing the un-tuned inhibition observed in young rat pups. Once plasticity is enabled, and the network is

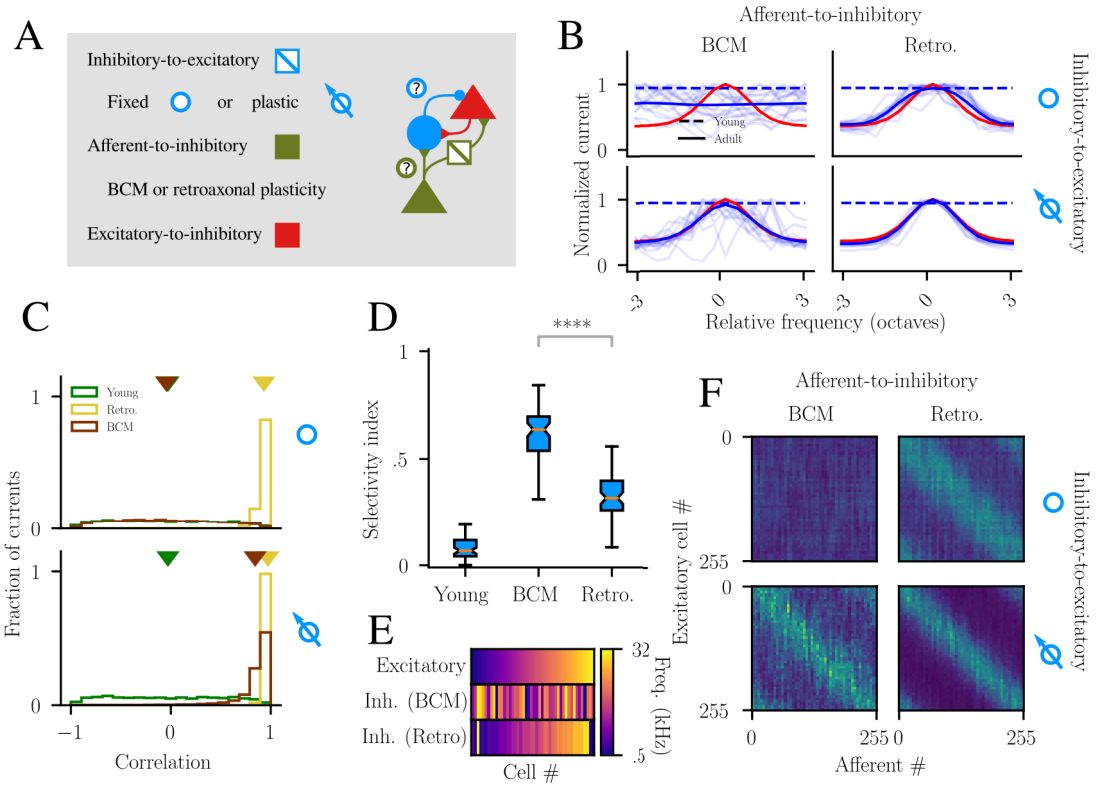

Figure 5.9: **Retroaxonal plasticity alone develops co-tuned currents, while BCM requires iSP. A)** Networks simulated have local inhibitory connections that are fixed or plastic, and afferent-to-inhibitory synapses that are random and subject to either retroaxonal or BCM plasticity. Excitatory-to-inhibitory are random. **B)** Membrane currents measured in young (dashed) and adult (solid) excitatory cells. Dark lines the average of all cells, light blue lines show instances of inhibitory currents in adult cells. **C)** Correlation between excitatory and inhibitory currents. Mean values indicated with a $\triangledown$. Percentage of cells with correlated currents in Young 19%; Top, No iSP: Retro. 99%, BCM 14%; Bottom, With iSP: Retro. 100%, BCM 87%. Criteria is $p < 10^{-3}$ using Spearman rank-order correlation, where the two-sided p-value is for the null hypothesis that currents are uncorrelated. **D)** Stimulus selectivity index of interneurons. Results are unchanged by the presence of iSP. ***** indicates $p < 10^{-4}$ for Mann-Whitney U test. **E)** Preferred auditory frequency of excitatory cells and interneurons, ordered by location along the tonotopic axis. Values shown are for a single simulation. **F)** Effective connectivity from afferent inputs to excitatory cells, ordered according to stimulus preference.

exposed to hundreds of hours of randomized pure-tone stimuli the various networks begin differentiate in their behaviour. We turn off plasticity, and measuring the adult network response to tonal stimuli reveals a lack of inhibitory tuning in the BCM network with no iSP (5.9B, upper left), and consequently no co-tuning of currents—as reflected in their lack of correlation (Fig 5.9C, top). This is despite robust stimulus selectivity in the interneurons (Fig 5.9D) and local inhibitory connections, the combination of which one might expect to yield co-tuning. The lack of stimulus specific inhibition is reflected in the effective connectivity matrix depicted in Fig 5.9F (upper left). In

contrast, the network with retroaxonal plasticity and static inhibitory synapses develop selective inhibition, co-tuned with excitation in 99% of cells. Its effective connectivity in Fig 5.9F (upper right), reflects this fact with clear preferential connections between afferent and principal cells with similar preferred stimuli. This difference can be explained by measuring the preferred frequency of interneurons. For the retroaxonal rule, interneurons tend to have a preferred frequency similar to that of nearby excitatory cells as in Fig 5.9E, with a correlation of $\rho_c = 0.74$ (angular correlation from Matusita, 1988; it is analogous to Pearson's). The BCM rule however, produces a random distribution of interneuron stimulus selectivity along the tonotopic axis, and so the preferred frequency of each interneuron is uncorrelated with nearby principal cells, with $\rho_c = 0.0$.

That the retroaxonal rule produces a tonotopic arrangement of interneuron selectivity while BCM results in random selectivity directly follows from the nature of both rules. The BCM rule does cause the interneurons to learn stimulus selectivity, but in a manner that is unaffected by the activity of the cells they inhibit, whereas the retroaxonal rule specifically strengthens synapses onto the interneuron from afferent cells that are co-active with its postsynaptic targets.

So given that with BCM, the interneuron stimulus selectivity is randomly distributed along the tonotopic axis, each principal cell will receive inhibition from interneurons with an approximately uniform distribution of preferred frequencies. The sum of these synaptic currents is the sum of the interneuron firing rate tuning curves, each scaled by their inhibitory synaptic weight, and thus flat on average. Conversely, since retroaxonal plasticity causes interneurons to prefer the same stimuli as their postsynaptic targets means that inhibitory currents are certain to be co-tuned with excitation.

The picture changes considerably when iSP is introduced to the networks. In the adult BCM network, iSP is able to exploit the stimulus selectivity of the interneurons to produce co-tuning of currents in 88% of cells (Fig 5.9C, bottom), although this means that the remaining 12% of cells still receive uncorrelated currents. Despite the the high mean correlation, a visual inspection of the inhibitory current onto single principal cells reveals a coarse co-tuning, with individual inhibitory curves appearing jagged (Fig 5.9B, bottom left, faint blue lines). Compare those with the inhibitory currents of the retroaxonal rule (bottom right), which are smooth and almost perfectly matched to excitation. This difference is apparent when looking at their effective connectivities in Fig 5.9F (bottom row). Both exhibit a diagonal band, but the relatively strong off-diagonal connections for the BCM rule indicates that feedforward inhibition is not as selective as that of the retroaxonal network with iSP.

Compensating for shifted tuning with retroaxonal plasticity

With three of the four networks capable of developing co-tuned currents, we now expose them to the same experimental protocol as before, shifting the preferred stimulus frequency of a subpopulation of principal cells (Fig 5.10A). In the BCM network without iSP, inhibition immediately becomes weakly selective for the paired-tone frequency (Fig 5.10B, upper left). This is not the result of any change in the stimulus tuning of the interneurons, which remain unaltered (Fig 5.10D, left). Rather, this is caused by the sudden increase in the number of excitatory cells cells firing at that frequency, and with those cells projecting randomly onto all interneurons, their firing rates for that stimulus

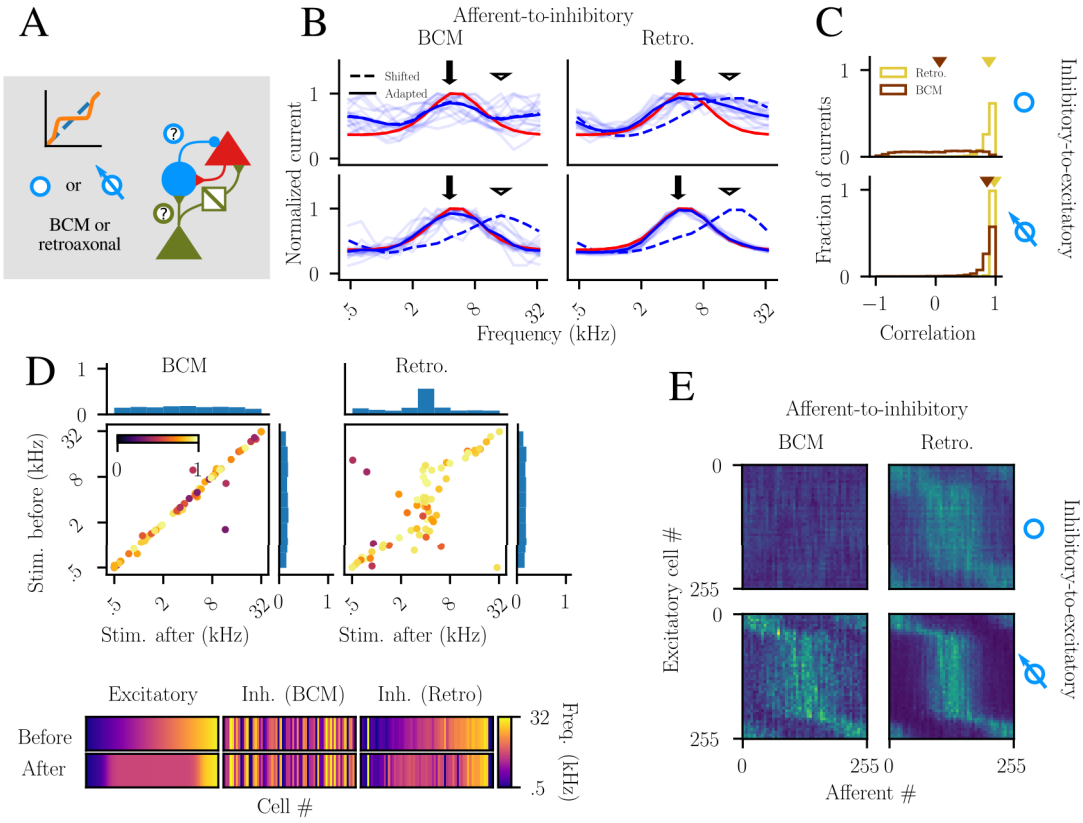

Figure 5.10: **Both retroaxonal and BCM plasticity require iSP to co-tune currents after shift. A)** Networks simulated same as in 5.9A, but stimulus preference of some excitatory cells have been shifted. **B)** Membrane currents as measured in a single excitatory cell immediately following the shift in excitation (dashed), and after plasticity has adapted (solid). Dark lines are the average of 16 simulations of the same network, light blue lines are single instances of the adapted inhibitory currents. **C)** Angular correlation between excitatory and inhibitory currents. Mean values indicated with a $\triangledown$. Percentage of cells with correlated currents in Top, No iSP: Retro. 92%, BCM 17%; Bottom, With iSP: Retro. 100%, BCM 88%. Criteria is $p < 10^{-3}$ using Spearman rank-order correlation, where the two-sided p-value is for the null hypothesis that currents are uncorrelated. **D)** Top: Preferred auditory frequency of interneurons before and after the shift, determined by fitting a Gaussian-like curve. Colour corresponds to goodness-of-fit. Results are unchanged by the presence of iSP. Bottom: Preferred auditory frequency of excitatory cells and interneurons, ordered by location along the tonotopic axis. Values shown are for a single simulation. **E)** Effective connectivity from afferents to excitatory cells, after plasticity has adapted to the shift.

rise. This manifests as an increase of inhibitory currents onto principal cells when the network is stimulated with the paired-tone, a tendency reflected in the modest increase in the percentage of cells receiving currents with any significant correlation (now 17%, up from 14% before the shift; Fig 5.10C, top).

In the retroaxonal network without iSP, inhibitory currents initially maintain the previous frequency preference. Eventually the retroaxonal rule adapts the stimulus

tuning of interneurons to match that of the cells they inhibit (Fig 5.10D, right). As a result inhibitory currents shift to prefer the paired-tone frequency, but still exhibit an excess of inhibition at the old frequency. This is because some of the interneurons contributing inhibition to shifted principal cells also innervate principal cells that retain their original preferred frequency. The attempt to satisfy postsynaptic cells with disparate tuning results in interneurons with preferred frequencies that are now less similar on average to nearby principal cells, with correlation $\rho_c = 0.53$ (down from 0.74, prior to the shift). This is because an interneuron cannot simultaneously prefer drastically different tonal frequencies. The retroaxonal rule attempts to tune those interneurons to match both types of cells, resulting in a receptive field that is responsive to a broader range of stimuli, reducing the precision of co-tuning (Fig 5.10C, top). Further evidence for this is seen in the effective connectivity, which shows a cross-coupling from a range of stimulus frequencies that do not identically match the new preferred stimulus of the excitatory cells, as revealed by the broad vertical band in the centre of the matrix.

The introduction of iSP enables networks with either the BCM or retroaxonal rule to co-tune inhibitory currents, with the former achieving a coarser precision than the latter once again (Fig 5.10C, bottom). The adaptation of both rules to the shifted tuning, and their differences can be seen in the effective connectivity in the bottom row of Fig 5.10E, which reveals a cleaner structure in the feedforward inhibition from afferents to excitatory cells for retroaxonal plasticity. The feedforward connectivity of the BCM rule has adapted to the shifted tuning, but there are some cross-connections, explaining the lower correlation in currents.

These results suggest that plasticity of afferent-to-interneuron synapses can account for the co-tuning of excitatory and inhibitory currents within the tonotopically arranged auditory cortex. Next we consider its effect on recurrent synapses within a network without such an orderly arrangement of cells.

5.2.4 Development of excitatory-inhibitory cell assemblies and interneuron selectivity in V1

In a feedforward setting both retroaxonal and BCM plasticity can account for the development of interneuron selectivity. When combined with iSP they also reproduce the experimentally observed compensation for shifts in the preferred frequency of excitatory cells in auditory cortex. We now investigate their behaviour in a recurrent setting: A network that mimics the salt-and-pepper structure of rodent primary visual cortex.

As the expression "salt-and-pepper" alludes, nearby cells in rodent V1 often have drastically different preferred orientation of moving bar stimuli. This dissimilarity of stimulus preference within the local population, in conjunction with pervasive local connectivity seems to suggest that there should be little interneuron stimulus selectivity and consequently, very weak co-tuning of excitation and inhibition. Experimental results from Runyan et al. (2010), Zariwala et al. (2011), and Fino et al. (2013) seem to defy this prediction, with them reporting orientation selectivity values in PV cells to be 0.25–0.35 on average—close to the value for principal cells, as reported by the former group. Tan et al. (2011) measured inhibitory currents to be well tuned to stimulus, with

a preferred orientation matching that of excitation. The origin of this co-tuning, in the presence of dense and apparently unspecific connectivity, may be due to the presence of excitatory-inhibitory cell assemblies. These are sets of strongly connected cells responsive to similar stimuli, and coexist alongside overlapping assemblies selective to different stimuli. Such assemblies can be detected by a correlation between the response similarity of reciprocally connected excitatory-inhibitory cells pairs and the strength of their synapses onto each other. Recently published work from Znamenskiy et al. (2018) offers data demonstrating just this sort of correlation in mouse V1.

In the following, our objective is to verify that the retroaxonal plasticity rule, when combined with iSP, can reproduce both the observed interneuron orientation selectivity, as well as co-tuning of excitatory and inhibitory synaptic currents. We will also compare its behaviour with that of BCM. The model networks we simulate are randomly connected, with static afferent-to-interneuron synapses, and plasticity of both excitatory-to-inhibitory and inhibitory-to-excitatory synapses (Fig 5.11A). The afferent inputs to principal cells remain structured, to impose orientation selectivity upon them.

Similar to our work with auditory networks, prior to enabling plasticity we measure the networks response to visual stimuli of every orientation. Subsequently plasticity is turned on and the network permitted to develop in response to hundreds of hours of randomized stimuli. Finally, we disable plasticity and measure the "adult" network response to all stimuli.

In the young networks, inhibitory currents are unselective and are not co-tuned with excitation (Fig 5.11B), with zero correlation. After development, inhibition becomes stimulus selective and co-tuned to excitation with $r > 0.9$ for both types of network (Fig 5.11C, left). The BCM rule is less precise though, a difference that is visible in the individual inhibitory currents shown in the same plot. This is reflected in the effective inhibitory connectivity in Fig 5.11C, right, which is the matrix product of the excitatory-to-inhibitory and inhibitory-to-excitatory connections and then sorted by preferred orientation. As with the auditory networks, BCM produces more cross-coupling between stimuli, but in this case it is excitatory cells providing inhibition to other excitatory cells through lateral disynaptic inhibitory connections. Nonetheless, both rules produce strong feedback inhibition amongst excitatory cells with similar stimulus preferences, which is the signature of excitatory-inhibitory cell assemblies.

In the young networks, stimulus selectivity of excitatory cells does not resemble those found in mouse V1, with values clustered around 0.55 and no highly selective or unselective cells (Fig 5.11D, top). Interneurons have a mean selectivity of less than 0.1 and little variation. In the adult network both retroaxonal and BCM plasticity produce distributions more closely resembling those observed in mouse V1 (Fig 5.11D, bottom). The retroaxonal network performs slightly better in this regard, with interneurons slightly less selective than principal cells. Both distributions are similar in shape, skewed towards higher values of selectivity, closely resembling the data of Runyan et al. (2010). The BCM rule on the other hand produces a more tightly clustered distribution of interneuron selectivities skewed towards lower values, with a higher mean than that of the principal cells.

Having shown that both rules produce the desired co-tuning of currents onto cells and the resultant stimulus selectivity, we now look for the expected synaptic markers

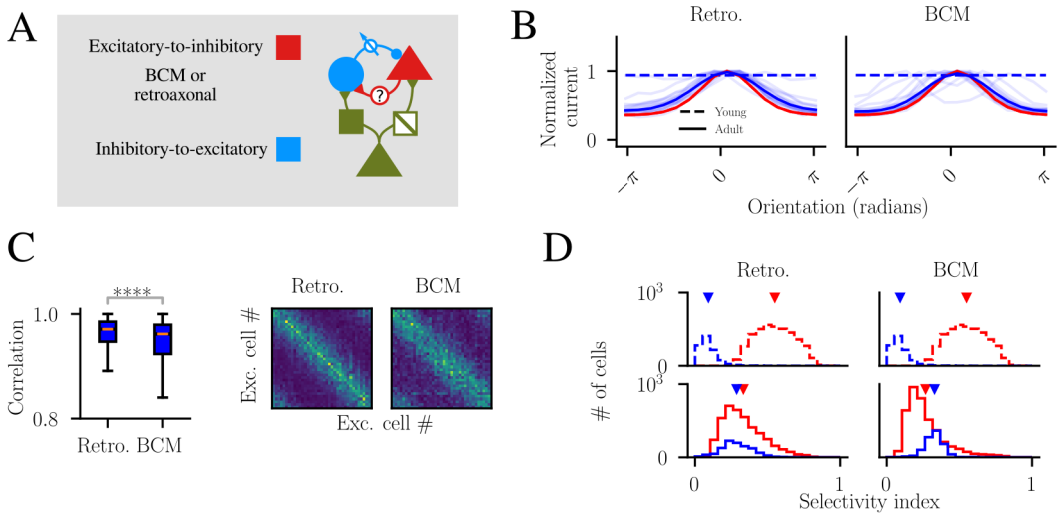

Figure 5.11: **V1 network develops co-tuning and realistic orientation selectivity with retroaxonal or BCM. A)** Networks simulated have random connectivity, inhibitory synapses with iSP and either retroaxonal or BCM plasticity on excitatory-to-inhibitory synapses. **B)** Membrane currents measured in young (dashed) and adult (solid) excitatory cells. Dark lines the average of all cells, light blue lines show instances of inhibitory currents in adult cells. **C)** Left: Correlation between excitatory and inhibitory currents. Percentage of cells with correlated currents in Retro. 98%, BCM 94%. Before plasticity, both are 20%. Criteria is $p < 10^{-3}$ using Spearman rank-order correlation, where the two-sided p-value is for the null hypothesis that currents are uncorrelated. ***** indicates $p < 10^{-4}$ for Mann-Whitney U test. Right: Effective inhibitory connectivity from excitatory cells to themselves, sorted in order of preferred orientation. **D)** Orientation selectivity in excitatory (red) and inhibitory cells (blue). Top are measured in the "young" network, bottom are the "adult" network. Mean values indicated with a $\triangledown$. Mean and standard deviation are Top: Young Exc. 0.55 ± 0.13, Inh. 0.09 ± 0.06 (Retro. and BCM are identical); Bottom left: Adult Retro. Exc. 0.33 ± 0.14, Inh. 0.29 ± 0.11; Bottom right: Adult BCM Exc. 0.27 ± 0.14. Inh. 0.33 ± 0.08.

of this behaviour. We begin by inspecting the distribution of synaptic weights in the adult network, shown in Fig 5.12A. Retroaxonal plasticity causes excitatory weights to be almost log-normally distributed, with some weights clustering near the maximum permitted value. The BCM rule produces a qualitatively different distribution, with nearly 4 in 10 synapses near the maximum value, and the remaining synapses are almost all weaker than any weight value produced by the retroaxonal rule. This is typical for BCM, as the rule induces competition between synapses. The synapse most effective in driving the cell will potentiate, while remaining synaptic weights are depressed. While the sliding threshold of BCM should limit growth of the synapse undergoing potentiation, in practice is useful to impose a reasonable upper bound. In our case, the upper bound was chosen to limit the peak firing rate of interneurons, which tend to be very high even when the target firing rate for BCM was set very low, with $\Theta = 2$ Hz.

While the two networks exhibit distinct distributions of excitatory weights, in both

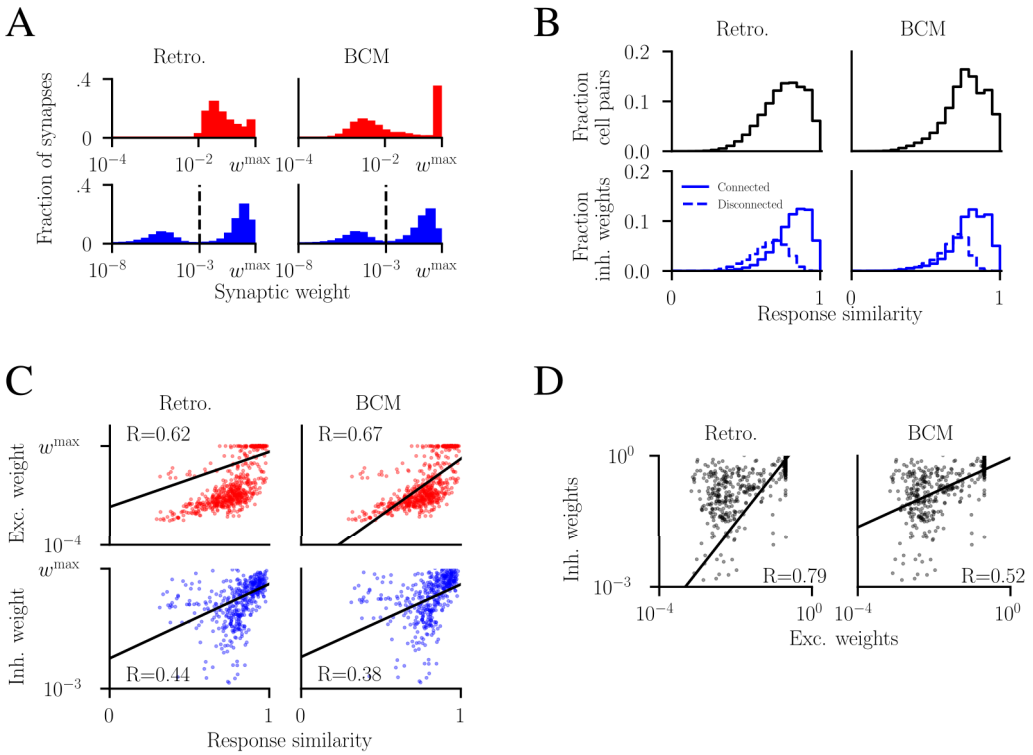

Figure 5.12: **V1 networks show correlation between response similarity and synaptic weights** A) Distribution of excitatory (top) and inhibitory weights (bottom). Dashed line is the threshold value for "disconnected" synapses. B) Response similarity between all pairs of excitatory and inhibitory cells (top). Response similarity between pre- and postsynaptic cell for connected (solid line) and disconnected inhibitory synapses (dashed line). C) Correlation between excitatory weight and response similarity of pre- and postsynaptic cell (top). Same plot for inhibitory synaptic weights (bottom). Only reciprocally connected cell pairs are included. D) Correlation between excitatory and inhibitory synaptic weight value for reciprocally connected cell pairs. Black lines in C & D depict linear least-squares regression, with two-sided p value $p \ll 10^{-4}$ for a hypothesis test whose null hypothesis is that the slope is zero.

iSP produces a bimodal distribution of inhibitory synaptic weights. We infer that this means some interneurons are "decoupling" from the postsynaptic principal cell. As such, we define synapses with weight less than 10^{-3} as "disconnected". This disconnection presumably results from the presynaptic interneuron preferring stimuli dissimilar from the preference of the postsynaptic excitatory cell. To verify this, we compute the response similarity for all excitatory-inhibitory cell pairs, and plot their distributions in Fig 5.12B, top. We now compare the response similarities for all cell-pairs sharing an inhibitory synapse, separated into two distributions: connected and disconnected (Fig 5.12B, bottom). While the two distributions are overlapping, the disconnected synapses are associated with cell pairs that exhibit lower response

similarity than those remaining connected. Now confident with our classification, the remaining analysis is performed only for connected synapses, as just defined.

Next we consider only reciprocally connected pairs of excitatory and inhibitory cells, and compare their response similarity with the strength of synapses between them. Both of the networks show strong correlation between excitatory synaptic weight and the corresponding cell-pair response similarity (Fig 5.12C). Inhibitory synapses are likewise correlated with response similarity, though the correlation is weaker. Finally, comparing the weight of excitatory and inhibitory synapses forming reciprocal connections shows a clear correlation, with the BCM rule exhibiting a lower correlation as seen in Fig 5.12D.

Ultimately, the results presented here show that retroaxonal plasticity of interneurons leads them to develop stimulus selectivity that matches postsynaptic cells. This matching of stimulus preference co-tunes inhibitory currents with excitatory currents onto principal cells. The rule accomplishes this in a model of auditory cortex without the aid of any other plasticity, something BCM cannot achieve. In a model of rodent primary visual cortex, this time with the aid of iSP, retroaxonal plasticity again develops stimulus selectivity and co-tuned currents. While BCM achieves comparable results within the V1 network, the retroaxonal rule exhibits co-tuning that is more precise, and with a more realistic distribution of stimulus selectivity in both interneuron and principal cell populations. Finally, both retroaxonal and BCM plasticity, when combined with iSP cause a V1 network model to form excitatory-inhibitory cell assemblies.

6 Discussion

The aim of this thesis is to study the homeostatic and functional properties of interneuron plasticity. This line of research was motivated by a diverse set of experimental results from others showing that some interneurons respond to chronic changes in network activity by altering their firing rates in a fashion that is plausibly homeostatic for excitatory cells. Such results leave an important question unanswered: For interneuron plasticity to be homeostatic for excitatory firing rates, what should the rule governing plasticity be responsive to? Should it be the activity of the excitatory cells driving an interneuron (the presynaptic population), or rather the activity of the cells inhibited by it (the postsynaptic population)?

To answer that question, in Section 5.1 we first use a normative approach to derive and then investigate a set of plasticity rules that regulate interneuron activity to be homeostatic for the excitatory population. The rules that emerge from this approach cause each interneuron to change its activity in response to its postsynaptic excitatory cells (post-IP). We then study a set of *ad hoc* rules that are responsive to presynaptic excitatory cells instead (pre-IP), and show that they induce competition between those cells.

We next consider the functional consequences of regulated interneuron activity for the operation of networks within sensory cortex. In Section 5.2, we show that one of the post-IP rules could account for certain experimentally reported features of inhibition in sensory cortex. In a model of auditory cortex, we account for the detailed balance of excitatory and inhibitory currents, as well as a peculiar delay between shifts of those currents in response to an experimental manipulation. Finally, in a rodent visual cortex model we provide a mechanism to account for co-tuning of excitation and inhibition despite the salt-and-pepper topography, as well as the reported stimulus selectivity reported in both classes of neuron.

6.1 Homeostatic plasticity

At the outset of this thesis in Section 5.1, we evaluate the capabilities of interneurons to act as homeostatic regulators of excitatory cell activity. We first derive a gradient-based plasticity rule for that purpose, and show that it is inherently limited in its ability to regulate the firing rates of individual excitatory cells, in contrast to plasticity that modifies inhibitory synapses onto those same cells. We then offer two biologically plausible approximations to the gradient, that are distinguished by how they detect deviations in the activity of excitatory cells postsynaptic to each interneuron. The first rule measures deviations using an unspecific diffusion-like signal, while the second rule responds only to deviations in the activity of the excitatory cells an interneuron inhibits, via retroaxonal signalling. Both rules produce results that are qualitatively similar to the gradient-based plasticity from which they derive, even when the simplifying

assumption is violated. In Section 6.1.1 we discuss the biological and theoretical implications of such forms of homeostatic plasticity.

We then investigate four additional plasticity rules, all of which modify interneuron activity in response to homeostatic deviations in the excitatory cells presynaptic to the interneuron. We show that these rules induce competition between excitatory cells. This competition arises because in contrast with the post-IP rules, synaptic changes are determined by presynaptic cells, in conjunction with two mechanisms: 1) Shared inhibition, wherein excitatory cells that each require different amounts of inhibition project to the same interneurons, and in return those INs project back onto many of the same cells. 2) Disynaptic inhibition, which occurs when an excitatory cell projects to an interneuron that does not reciprocate the connection. While these pre-IP rules control neither the firing rate of individual cells nor the population, in Section 6.1.2 we discuss a potential role for these rules in homeostasis.

6.1.1 Post-IP is an effective homeostatic controller with limitations

Through analytic and numerical techniques, we show that homeostatic control of excitatory firing rates by regulation of interneuron activity is effective though limited. The primary limiting factor is the fanout of inhibitory-to-excitatory connectivity, which makes control of individual excitatory cells impossible. This lack of individual control is not necessarily problematic, and in fact recapitulates experimental results which exhibit just that behaviour (Barnes et al., 2015; Slomowitz et al., 2015). Furthermore, the resulting set-point heterogeneity has been shown to improve performance in a working memory model (Renart et al., 2003). But regardless of any theoretical benefit to this approach, do any biological interneurons actually regulate their activity in a manner homeostatic for excitatory firing rates?

Does interneuron activity homeostatically control network firing rates?

The notion of using interneuron activity as a homeostatic regulator as we propose has some evidential support. The most direct evidence for a mechanism is the scaling of glutamatergic synapses onto interneurons following chronic changes of activity. It is important to note that some of those results were obtained in dissociated cell cultures (Doyle et al., 2010; Rutherford et al., 1998). So while they demonstrate that there is some mechanism *capable* of making such putatively homeostatic changes, there is no guarantee that it is active within the intact brain. And furthermore, the homeostatic mechanisms at play seem to be dependent on the experimental paradigm, e.g. visual deprivation by eyelid suture versus enucleation involve different homeostatic mechanisms (Maffei and Turrigiano, 2008). There are some data from intact brain tissue, including visual cortex (Keck et al., 2011) and hippocampal slices (Chang et al., 2010), which suggest that those synapses are homeostatically regulated within living animals.

Whatever the mechanism, the activity of some interneurons does respond to chronic changes of network activity in a fashion that seems to implicate it in network homeostasis. The most striking result is from Hengen et al. (2013), who report that FS interneurons exhibit a rapid drop in activity following eyelid suture, while RS cells (including PNs) maintain their pre-intervention rates for the first day. The following day, FS cell activity recovers, while RS cell rates drop, until the third day when both

populations return to their pre-intervention rates. These differing time courses suggest that FS activity is not simply dictated by RS rates, because if it were, FS rates would presumably rise and fall with the firing rate of RS cells.

While suggestive, that study reported only population activity. Later work published by Barnes et al. (2015) looked at the activity of individual units in response to monocular enucleation. They report that after 72 hours only a fraction of cells start firing again, and of those that do, the mean excitatory activity returns to its pre-intervention rate, while mean inhibitory rates do not. Although plasticity of inhibitory synapses is involved in this recovery, the fact that interneuron activity does not completely return to its prior level suggests that it might be regulated in a fashion homeostatic for the PNs. Furthermore, the interneurons that recovered were the ones that, prior to the intervention, exhibited correlated activity with the PNs whose firing rates were also later restored. This provides some support for our claim that the activity of some interneurons could be directly responsive to the firing rates of the PNs they inhibit.

Retroaxonal plasticity

One of the post-IP rules we propose is a form of plasticity that modifies the activity of interneurons in response only to the homeostatic deviations of excitatory cells directly inhibited by each interneuron. This rule requires a retroaxonal signal which is integrated centrally, and determines the activity level of the interneuron via changes to its glutamatergic synapses (or possibly the intrinsic properties, though results with those variants are not reported here). This sort of plasticity is not unprecedented, with Harris (2008) advocating for just such an retroaxonal signal playing a crucial role for regulating neuronal activity. As that author points out, retrosynaptic signals like cannabinoids, gasses, and neurotrophins can provide information about postsynaptic activity. Our model does not require that this signal includes information about the strength of individual synapses, making it less dependent on the molecular details of each connection, in contrast to Lewis and Harris (2014).

The importance of retroaxonal signals is becoming better understood, though mostly in a developmental context as yet (Ginty and Segal, 2002; Zweifel et al., 2005). The influence of retroaxonal signals upon the plasticity of the incoming synapses onto a cell has been shown in cultured hippocampal neurons (Fitzsimonds et al., 1997; Tao et al., 2000), and in the sympathetic nervous system (Sharma et al., 2010). So while retroaxonal plasticity remains exotic, the extent of its role in the brain may be currently under appreciated.

Diffusion-like plasticity

The other post-IP rule we propose relies on a diffusion-like signal to detect changes in the firing rates of excitatory cells. This signal might actually be carried by the diffusion of molecules, secreted from excitatory cells due to their activity, through the extracellular space (Syková and Nicholson, 2008). Candidate molecules include BDNF, Narp, and NO (among others gasses; Rodriguez-Grande and Konsman, 2018). Unfortunately, few of those candidates are likely to diffuse far enough at biologically relevant concentrations to be effective mediators for network-level homeostasis. For example, the Trk receptors that bind BDNF are present along the axon, and due to

their high affinity are unlikely to permit diffusion at distances much greater than between synapses (Lu, 2003). And NO is unlikely to diffuse at relevant concentrations to distances greater than 1 μm (Garthwaite, 2016).

Instead of relying on a network-scale diffusion process, it may be that detection of population activity is carried out by glial cells. Stellwagen and Malenka (2006) report that in cell cultures the cytokine TNF-α is released by glia in response to excitatory activity, and is involved with AMPA receptor accumulation (being at least permissive, if not instructive). In TNF-α knock-out mice synaptic scaling in response to visual stimulus deprivation is abolished, which suggests that it is also involved in homeostatic regulation of the intact brain (Kaneko et al., 2008). Another possibility is that the axonal arbour of the INs acts as the sensor for population activity. It is conceivable that the axon can detect activity dependent secretions of nearby PNs, even absent anatomical synapses onto those cells. If this were the case, it would effectively implement a diffusion-like signal adequate for our requirements.

Finally, many of the signal molecules we have mentioned are implicated exclusively in up-scaling, but not down-scaling of synapses, or vice versa (Fernandes and Carvalho, 2016). This suggests that our post-IP rules, if present in the brain, are likely to be implemented by more than one molecular mechanism. It is also possible that both retroaxonal and diffusion-like plasticity are active, with one responsible for down-scaling, and the other for up-scaling. Whatever the case, and however informative our results prove to be, it is important to bear in mind that these models are more phenomenological than mechanistic.

Biological time-scales of post-IP mediated homeostasis

One crucial limitation of both retroaxonal and diffusion-like plasticity is the time scale at which they would likely operate. Fast retrograde axonal transport of signalling endosomes occurs at just a few micrometers per second (Maday et al., 2014), and consequently signals would take hundreds of seconds to propagate up a typical interneuron axon. Likewise, if our proposed diffusion-like plasticity relied on glial cells to sense local population activity, it might be capable of propagating its signals an order of magnitude faster (Newman and Zahs, 1998), but this still introduces a latency on the order of tens of seconds.

Any plasticity responding to a signal with tens or hundreds of seconds of latency would necessarily have to change its plastic variable very slowly to avoid instability. This does not reduce the biological feasibility of our proposed rules, but it does mean their applicability is limited to compensating chronic changes of network activity.

Theoretical perspectives on post-IP mediated homeostasis

One of the key characteristics that distinguishes the two biological post-IP rules are their effect on population rates. The diffusion-like rule reliably brings the population rate to the homeostatic set-point irrespective of network parameters. This is in contrast with the retroaxonal rule, where the population rate is dictated by the heterogeneity of the inhibitory in-degree to excitatory cells. In Section 7.3 we derive an expression for the expected population rate, which is proportional to the variability in the sensor used to detect the excitatory firing rates. The diffusion-like rule uses a smooth sensor, and

consequently always reaches the set-point. The retroaxonal rule uses the inhibitory-to-excitatory connection matrix to propagate errors in the rate of excitatory cells to the interneurons. The variability of this sensor due to differences of in-degree, in conjunction with a negative correlation between the number of incoming inhibitory connections and the firing rate of each excitatory cell, is what produces the positive offset in population rate. As such, for the retroaxonal rule we show that the derived expression predicts the population rate as a function of the in-degree heterogeneity.

With respect to controlling the total error of excitatory firing rates, the effectiveness of the post-IP rules can be understood from the perspective of the classic back-propagation algorithm, or 'backprop' (Rumelhart et al., 1986). In that algorithm, a feedforward network of neurons learns to approximate a target mapping from inputs to outputs. It accomplishes this by measuring the error in the output neurons relative to the target, and then propagates the error backward in the network, using the matrix that defines the incoming connections to those neurons—this is the 'ideal' backprop matrix. That back-propagated error is then used to adjust the activity of neurons in the previous layer of the network, and reduce the output error.

From this description it should be clear that our post-iP rules perform something analogous, but where the backprop operates on a feedforward network that is embedded within our recurrent network. Recent work from Lillicrap et al. (2016) showed that to effectively learn—that is, reduce the error—the backprop matrix need not be the ideal one, but must merely be sufficiently similar to the ideal (as measured by the angle between them in a vector space). We show for the diffusion-like rule that the total error is proportional to the difference between the back-propagation matrix used (the diffusion matrix) and the ideal back-propagation matrix (the inhibitory-to-excitatory connection matrix). We control that difference by changing the connection probability, with the angle being inversely related to the connection probability. And the end result is that the more dense the connections from inhibitory to excitatory cells, the lower the error.

Yet another perspective on how the post-IP rules accomplish their objective comes from signal processing. As explained in detail on page 40, transforming the error signal measured across the entire excitatory population to the appropriate changes of interneuron activity is equivalent to a down-sampling operation: a reduction of N^{E} to N^{I} samples. To appropriately perform that reduction involves largely ignoring deviations in single excitatory cells that are unrelated to population-level errors. This sort of population averaging requires a diffusion process, or anti-aliasing filter, of an appropriate spatial width. By default, the diffusion-like rule uses a sensor that is responsive to the population inhibited by the average interneuron. But from a signal processing perspective it is sufficient to make interneurons responsive to a much smaller excitatory population. We demonstrate that to reach 95% of the homeostatic firing rate for population activity, it is sufficient for interneurons to respond to the population contained within about one quarter of the inhibitory connection distance.

Likewise, retroaxonal plasticity can be viewed as a form of sparse sampling. Sparse sampling is a method for reducing the negative consequences of down-sampling a signal despite not applying an appropriate anti-aliasing filter (Cook, 1986). This is achieved by distributing the power of each higher (supra-Nyquist) frequency in the original signal evenly across the spectrum, instead of erroneously representing them

at their corresponding aliased frequency in the down-sampled signal. Such stochastic sampling is believed to be responsible for the absence of perceptible aliasing that would be expected from the low density of visual receptors in the peripheral retina.

In our setting, the retroaxonal propagation matrix could be considered a stochastic sampling operator, due to the randomness present in the inhibitory-to-excitatory connections. And as a result, if there are large errors in just a few excitatory cells scattered across the network, instead of producing a large erroneous change in the activity of the interneurons, it instead produces a smaller change randomly distributed across them. The competitive behaviour of the homosynaptic pre-IP rules, discussed below, is an example of what happens when the errors in excitatory firing are translated to changes of interneuron activity without applying an appropriate down-sampling procedure, and might be thought of as a form of homeostatic aliasing.

Existing theoretical models

We are not the first to propose a model of diffusion-mediated homeostatic plasticity. Savin et al. (2009) describe a homeostatic excitatory synaptic scaling mediated by TNF-α released from glial cells. They use their model to show that in normal conditions such plasticity can homeostatically regulate network activity, but localised lesions leads to the development of seizure-like activity. This epileptiform activity is likely related to the local deviations in steady-state firing rates that we observe in the presence of spatial input heterogeneities, though the manner in which we record our data does not permit its detection. The same lab later investigated the effect of combining reward-modulated spike-timing dependent plasticity, synaptic scaling, and homeostatic diffusion-mediated intrinsic plasticity (Savin and Triesch, 2014). They report that these components in a working-memory circuit contribute to the development of task-dependent representations.

A different group studied the effect of NO mediated homeostatic intrinsic plasticity, and contrasted that with a non-diffusive variant of the same rule (Sweeney et al., 2015). They report that while both rules provide effective homeostatic control, diffusion-mediated plasticity maintained greater heterogeneity in the firing rates of cells, which enhances representation of input heterogeneity, and linearity of network response. We also observe that our diffusion-like plasticity, and in fact any post-IP rule, preserves heterogeneity in the firing rate distribution.

For retroaxonal models, Lewis and Harris (2014) consider its effect in a recurrent network, as a mechanism for slowly consolidating memories. In their system, a non-homeostatic retroaxonal plasticity acts on recurrently connected synapses, meaning that the signals are recursively propagated backwards through the network. This is in contrast to our rule, where the retroaxonal signals terminate at the interneurons. As we stated in the discussion of back-propagation, the retroaxonal and diffusion-like rules are operating on a feedforward network, embedded in our recurrent network.

6.1.2 Pre-IP plasticity

We also characterise the effectiveness of three ad hoc homosynaptic plasticity rules, and one heterosynaptic rule, all designed to be homeostatic for excitatory firing rates. Those rules attempt to achieve this by modifying excitatory synapses onto INs in response

to chronic deviations in the firing of excitatory cells that drive those synapses. In fact, an early investigation of the homosynaptic rules was what ultimately spurred all of the research presented here. The first variant investigated, the Hebbian rule, was designed by analogy to iSP. At the time, it seemed plausible that such a rule operating on the excitatory-to-inhibitory synapses should have homeostatic capabilities comparable to iSP, given that they both operate within the inhibitory feedback loop. Early computational results indicated that rather than bringing excitatory cells to fire at the homeostatic set-point, such pre-IP can actually produce total quiescence in large fractions of the excitatory population, induced by a competition between excitatory cells.

Competition and homosynaptic pre-IP

The Hebbian, non-Hebbian and anti-Hebbian rules neither bring the population rate to the homeostatic set-point, nor minimize the total error in the network. In networks where recurrent synapses are strong, almost all excitatory cells fire below the set-point, and many are driven to seemingly permanent quiescence. This is the result of competition between excitatory cells. The primary mechanism of competition is what we have called shared inhibition. This is especially relevant in the locally connected networks we investigate here, where nearby excitatory cells project to many of the same interneurons, and those interneurons are likely provide inhibition to the same population of excitatory cells. This shared inhibition, when combined with heterogeneous inputs to the excitatory cells, creates a situation in which some cells need much more inhibition than others. And because the homosynaptic rules modify synapses in response to the needs of each individual presynaptic cell, the cells whose net input is relatively more depolarising will tend to dictate the activity of all the interneurons it provides excitation to.

What predicts the steady-state firing rate of excitatory cells? In networks with homogeneous inhibitory-to-excitatory connections (where in-degree variability is low), it is the excitatory in-degree of PNs that predicts their steady-state activity. But as heterogeneity of their inhibitory input increases, that value becomes much more predictive of their final firing rate. To make sense of this, picture two excitatory cells, the first with slightly more excitatory input than the second. But imagine that the first has twice as many inhibitory inputs from local interneurons. Due to dense local connectivity, if the local population needs more inhibition, interneuron firing rates will tend to all rise together, and as a result the cell with more inhibitory inputs will have its firing rate reduced much more rapidly than the cell with fewer.

The second mechanism responsible for competition is disynaptic inhibition. It is rare in most of the networks we simulated, due to dense local connectivity, but reducing connection probability can reveal its effect. As connection probability decreases, an increasing fraction of excitatory cells project to interneurons that do not reciprocate the connection but do provide inhibition to other excitatory cells. Any cell that has no reciprocal inhibitory connections, that fires above the homeostatic set-point will increase their excitatory drive onto their postsynaptic interneurons. This in turn produces disynaptic inhibition onto other excitatory cells, which will tend to become over inhibited.

In the mammalian brain, these two mechanisms are likely to be relevant for different interneuron populations. Shared inhibition is most likely to be important for PV interneurons, on synapses which provide dense recurrent inhibition for local PNs. Disynaptic inhibition would be especially relevant for SOM interneurons, which tend to lack reciprocal connections with the PNs they inhibit (Berger et al., 2009). So if glutamatergic synapses onto either of those neurons are responsive in a homosynaptic fashion to chronic changes in the activity presynaptic cells, those interneurons would likely mediate the sort of competition we describe here.

Anterograde plasticity

We also investigate the effect of a heterosynaptic form of interneuron plasticity, which we obtain from substituting the inhibitory-to-excitatory propagation matrix of the retroaxonal rule with the excitatory-to-inhibitory matrix. We call this anterograde plasticity, and it causes each interneuron to be responsive to chronic changes in firing rate of their entire presynaptic population of excitatory cells. The anterograde rule results in population firing rates that are closer to the target rate, with excitatory cells evenly split between firing above and below the homeostatic set-point. Nevertheless, the rule is capable of producing some competition between excitatory cells, especially in sparsely connected networks where disynaptic inhibition begins to dominate. The key factor that differentiates the anterograde rule from the homosynaptic rules is that it is responsive to presynaptic population activity, rather than single presynaptic cells. So while competition is still present, the rule being responsive to populations mitigates the impact, especially in local densely connected networks.

The anterograde rule is especially interesting because Chang et al. (2010) reported a form of homeostatic plasticity acting on hippocampal PV interneurons that is highly reminiscent of it. They show that release of Narp from PNs increases with excitatory activity, and leads to its accumulation at their excitatory synapses onto PV interneurons, and induces AMPA receptor clustering, which should increase IN firing rates in a homeostatic fashion. This accumulation of Narp is dependent on the presence of perineuronal nets, which enclose the dendrite, soma, and axon initial segment of the PV interneurons (Dityatev et al., 2007). Removal of the perineuronal nets eliminates Narp accumulation, and so appears essential for the homeostatic regulation of IN activity. Furthermore, the dendrites of some PV cells in hippocampus are smooth, meaning that plasticity is not synapse-specific (Cowan et al., 1998). These are all the components required for those interneurons to exhibit the anterograde plasticity we present here, and suggests further study of its functional implications could be important for understanding the hippocampus.

Another perspective on homeostasis

The tendency for the pre-IP rules to produce sparse firing rate distributions in a network is not necessarily pathological. In fact, it might prove to be homeostatic in a sense distinct from that used throughout this text. We have assumed that plasticity that is homeostatic for excitatory cells would attempt to bring their firing rates to some common set-point. And while the post-IP rules do yield firing rate distributions that are more sparse than iSP, they might not go far enough. In fact according to Barnes

et al. (2015), about one in three excitatory cells exhibit no firing for the duration of their recording.

It is an open question both just how sparse activity in cortex is (Barth and Poulet, 2012; Shoham et al., 2006), and exactly what functional consequences this has (Buzsáki and Mizuseki, 2014). We have shown that the pre-IP rules tend to increase sparseness, which might prove to be effective in preserving an important functional aspect of cortical networks. It is worth mentioning that there are already accounts of how this sparseness might emerge due to the nonlinearity of neurons in the presence of normal inputs (Roxin et al., 2011). Regardless, it is interesting that pre-IP rules actually increase heterogeneity within cortical networks without artificially injecting randomness by varying the homeostatic set-point for each cell.

One problem with a rule that increases heterogeneity is that it can have seriously deleterious effects on E/I balance (Landau et al., 2016). We found this to be the case when we simulated the pre-IP rules in spiking networks. Network with those rules would inevitably start to oscillate, with large populations of neurons firing synchronously (this effect is precisely what led to our use of rate-based networks). So it seems likely that if pre-IP rules are actually at responsible for increasing sparseness in biological networks, they most certainly rely on the many other homeostatic mechanisms to keep from behaving pathologically.

That other homeostatic mechanisms could potentially counteract the disruptive effect of the pre-IP rules might seem a serious criticism of the conclusions we draw regarding their homeostatic properties. But this misses the crucial point of this work: We set out to determine the in-principle homeostatic capabilities of regulated excitatory-to-inhibitory synapses. Characterising the effect of a plasticity mechanism when no other compensatory mechanisms are at play enables the investigator to clearly assess the strengths and weaknesses of the mechanism. Only after completing this first step does it make sense to proceed with more complicated models. Naturally, if this line of research is continued, it is clear that evaluating pre-IP plasticity in the presence of compensatory mechanisms should yield insight into the resulting functional properties, and might even reveal some benefit for computation inside the brain.

6.2 Interneuron plasticity and sensory processing

Having found plasticity rules that homeostatically control the firing rates of excitatory cells by regulating the firing rates of interneurons, we next show how one of those rules might affect sensory processing in cortex. We begin by investigating the origin of excitatory-inhibitory membrane current co-tuning in primary auditory cortex, and use a result published by Froemke et al. (2007) to constrain connectivity within our auditory cortex model. We demonstrate that interneuron stimulus selectivity in our model of A1 must arise from the afferent inputs to them being tuned to tonal frequency, while local recurrent excitatory connections should be untuned or very weak. Next, we show that anatomical inhibitory-to-excitatory connections must be unrealistically broad if the only plasticity present is inhibitory synaptic plasticity. More plausible connectivity is feasible if the afferent synapses onto interneurons are plastic, with both retroaxonal and BCM rules producing co-tuning of membrane currents, but with a distinct prediction for the effect of either rule on interneuron selectivity within the tonotopic map.

Finally, we show the effect of those plasticity rules on networks with rodent visual cortex-like connectivity. Within networks with such a salt-and-pepper arrangement of PN orientation selectivity, inhibitory synaptic plasticity combined with plasticity of excitatory-to-inhibitory connections reproduces several notable experimental findings. If excitatory synapses on the interneurons are plastic with the retroaxonal or BCM rules, those cells develop increased stimulus selectivity. Compared to the BCM rule, the retroaxonal rule yields more biologically realistic distributions of selectivity for both the PN and IN populations. Both rules also produce correlations between excitatory and inhibitory synaptic weights for reciprocally connected PN-IN cell pairs, as well as a correlation between the strength of a synapse connecting them and the response similarity of the two cells. These latter two results match well with experimental data reported by Znamenskiy et al. (2018), and though it does not refute their hypothesis that a standard Hebbian plasticity of excitatory-to-inhibitory synapses is sufficient to reproduce those data, we do offer a novel alternative of how it might arise—with retroaxonal plasticity.

6.2.1 Auditory cortex

The results from Froemke et al. (2007) exhibiting a delay between the shift of excitatory and inhibitory currents in A1 provide a starting point for our results here. Their data suggest that while excitatory currents change their preferred frequency within 30 minutes of pairing a 4 kHz tone with nucleus basalis stimulation, inhibitory currents take as many as 3 hours to compensate by likewise shifting their preferred frequency. We show that this delay is incompatible with the hypothesised origin of co-tuning as per Harris (2008), wherein the co-tuning arises due to the local recurrent connectivity within a tonotopically arranged auditory cortex. Our simulation results indicate that if that were the case then inhibitory currents would immediately shift in response to a change in excitatory currents. Instead, inhibition must acquire its stimulus selectively from afferent inputs to account for the observed delay.

So from where do these afferent inputs originate? While the most likely candidate is the thalamus, intracortical inputs are conceivable. If the afferent inputs of our model arise from elsewhere in auditory cortex, we require that presynaptic population to not change its tuning in response to the paired-tone protocol. This is plausible if the presynaptic population are in another cortical layer that is unaffected by that same protocol. The recordings of Froemke et al. (2007) are all at depths greater than 400 μm from the pia, meaning they likely miss L2/3, which could be where our putative afferent population resides. Nonetheless, this would be a odd arrangement, and it seems more plausible that thalamus provides the afferent inputs in our model.

Thalamic input targets L4 predominantly, though not exclusively, with meaningful inputs to L2/3, L5 and L6 (Intskirveli et al., 2016). Those inputs are not only onto excitatory cells, with INs receiving significant drive from thalamus (Wall et al., 2016). Our model might best correspond to the network of L4 excitatory and PV cells. Li et al. (2014b) report that the excitatory cells of L4 are more broadly responsive to tones than those of L2/3, consistent with more precisely co-tuned currents in the deeper layer than that of the more superficial layer. This is reflected in a stimulus selectivity for frequency that is similar in both excitatory and inhibitory cells in L4, in contrast to broader tuning

in the interneurons of L2/3. This is in agreement our claim that the recordings from Froemke et al. (2007) are likely from layers deeper than L2/3.

Having established that interneuron frequency selectivity likely arises due to tuned afferent inputs, we next determine under what conditions inhibitory synaptic plasticity is sufficient by itself to account for development of co-tuning and the eventual compensatory changes of inhibitory currents. It is notable that in rat pups, inhibitory currents are not selective for frequency (Dorrn et al., 2010; although Sun et al., 2010 report contradictory results). This means that afferent inputs to interneurons and inhibitory connections to PNs cannot both simultaneously be tonotopic early on, otherwise there is an immediate co-tuning of synaptic currents. According to Barkat et al. (2011), the tonotopic map in thalamus is unchanged through the critical period, suggesting that if thalamic inputs to A1 are initially untuned, that lack of specificity resides in the afferent synapses.

In the networks that exhibit the desired untuned inhibitory currents initially, iSP is not by itself sufficient to produce significant co-tuning, in absence of at least some interneuron selectivity. This is comports well with what is known about interneuron selectivity in A1. Furthermore, if interneuron stimulus preferences are tonotopically arranged, iSP cannot compensate for the shift in excitatory currents reported by Froemke et al. (2007) unless inhibitory anatomical connections span at least two octaves of the tonotopic map, which strains the limit of what has been observed in the rodent brain (Levy and Reyes, 2012).

This extraordinary requirement is relaxed if we permit plasticity of afferent synapses onto the interneurons. Not only does this account for development of the necessary interneuron selectivity, it also eases the need for unrealistic long-distance inhibitory connections. Retroaxonal plasticity accomplishes this by causing interneurons to prefer the same stimulus frequencies as the PNs they inhibit. Consequently, when those PNs change their tuning, the interneurons eventually become responsive to the same frequencies. This is in contrast to BCM plasticity of afferent synapses, which produce interneuron selectivity that is randomised within the tonotopic map. This means that while retroaxonal plasticity can produce co-tuning by itself, BCM needs iSP for co-tuning to develop. And furthermore, when nearby PNs change their preferred frequency, the interneurons are unresponsive to that change, with BCM once again relying on iSP to compensate for the shift. This results in a clear prediction: if the preferred stimulus of interneurons is tonotopically arranged, it suggests that retroaxonal plasticity is more likely to be responsible than the basic BCM rule we test. Moore and Wehr (2013) report just that, with PV cells exhibiting tuning similar to that of the nearby population. This is not definitive proof of a retroaxonal plasticity, since it is possible to imagine a kind of plasticity for which recurrent excitation provides a teaching signal for afferent synapses, which would mimic the effect of our retroaxonal rule.

Finally, although A1 is macroscopically tonotopic, there is some evidence that there is substantial microscopic heterogeneities. Both Rothschild et al. (2010) and Bandyopadhyay et al. (2010) present evidence that within small regions of A1, the preferred frequency of individual cells are largely uncorrelated with their position, and span most of the spectrum. It is only when the entirety of A1 is considered that a correlation between neuron location and preferred frequency is measured. Such local heterogeneities would make plasticity of the sort we investigate here doubly important for co-tuning

of currents.

6.2.2 Visual cortex

The visual cortex of rodents exhibit a different topographic arrangement of cells than in A1, with PNs often exhibiting drastically different preferred orientation preferences than their neighbours. Despite this, there is some degree of co-tuning of synaptic membrane currents. As we showed in our A1-like network, even modest co-tuning requires some amount of stimulus preference in interneurons along with iSP, or some mechanism for developing strong stimulus selectivity in interneurons. We show that in V1-type networks, either retroaxonal or BCM plasticity of local excitatory-to-inhibitory connections result in interneuron selectivity. Retroaxonal plasticity reliably produces a distribution of selectivity in both INs and PNs that is more biological than the BCM rule (Runyan et al., 2010). While that result is compelling, it must be admitted that BCM is especially sensitive to parameters, and it is conceivable that some parametrisation could produce more realistic results. Nonetheless, the retroaxonal rule is extremely robust, and yields similar results almost irrespective of parameter choice. In contrast, the BCM rule required significant effort to find parameters that produced a reasonable result (an unpublished observation).

The stimulus selectivity we report would seem unlikely if local connectivity was unspecific, as speculated by some (Fino et al., 2013; Fino and Yuste, 2011). In recent years, an increasing body of evidence suggests that some PV interneurons in V1 are nearly as stimulus selective as PNs. Various groups have reported modest stimulus selectivity in PV neurons (Hofer et al., 2011; Kerlin et al., 2010; Runyan et al., 2010; Tan et al., 2011; Zariwala et al., 2011). Intriguingly, Khan et al. (2018) recently showed that learning a visual discrimination task causes PVs to develop the same degree of selectivity as pyramidal neurons. So it seems a virtual certainty that some interneurons develop strong stimulus selectivity, similar to what we have shown.

Stimulus selectivity of interneurons is only half of the story, with plasticity of inhibitory synapses being the other half. Notably, iSP seems to decouple certain PN-IN cell pairs based upon the similarity of their stimulus preference. While not the focus of this thesis, iSP proves to be sensitive to the postsynaptic cell tuning, regardless of which plasticity is operating upon the excitatory synapse onto a given interneuron. This tendency for inhibitory synapses to disconnect from PNs preferring dissimilar stimulus orientations enhances the correlation between the synaptic strength of both types and the cells that are joined by that synapse. So as one would expect, even within cortex lacking a topographic arrangement, plasticity produces a relationship between the similarity of cell pairs and the strength of the synapses joining them (Znamenskiy et al., 2018). What distinguishes the BCM and retroaxonal rule is that this is driven by different mechanisms: the BCM rule is essentially Hebbian, with a homeostatic component for the interneuron, while the retroaxonal rule is primarily homeostatic to the activity of its postsynaptic PNs. Whatever the rule at work on excitatory-to-inhibitory synapses, with the aid of iSP our V1 networks formed excitatory-inhibitory cell assemblies. With evidence for the formation of such assemblies *in vivo* having recently been found (Khan et al., 2018), it is all the more important to obtain insight into just how that might occur.

Existing theoretical models

There exist many theoretical studies of how plasticity of inhibition contributes to cortical processing. In a model of auditory cortex Clopath et al. (2016) show that excitatory and inhibitory plasticity interact to produce co-tuning of synaptic currents. They argue that iSP causes Hebbian plasticity of recurrent excitatory synapses to act similar to BCM. Although we do not consider how excitatory assemblies form, this proposed mechanism might account for it and points to potentially fruitful future research.

Formation of excitatory cell assemblies sensory cortex has been studied using a diffusion-mediated plasticity by Sweeney and Clopath (2017). They proposed a variation on BCM, where the activity of the postsynaptic cell is replaced by the activity of the local population. Such a rule causes nearby neurons to develop similar stimulus preference. In contrast to our model, their results are concerned with the development of excitatory cell assemblies, and do not address selectivity of interneurons, or their place within the assemblies.

Finally, the formation of cell assemblies has also been studied by Litwin-Kumar and Doiron (2014) as well as Zenke et al. (2015). Both of those groups relied upon iSP in conjunction with an excitatory synaptic plasticity that includes homeostatic terms. Litwin-Kumar and Doiron (2014) do not include any plasticity of interneuron activity, and do not report the selectivity within the inhibitory population. Zenke et al. (2015) include a number of plasticity mechanisms, including short-term plasticity, metaplasticity, and homeostatic plasticity of excitatory synapses. They also report the effect on afferent as well as recurrent connections, but once again do not explicitly characterise whether their model forms excitatory-inhibitory cell assemblies. Without such information being available, it is difficult to compare their results with our own.

6.3 Conclusion

In this thesis, we demonstrate that inhibitory neurons can be effective homeostatic controllers of excitatory activity. If the objective is to increase sparseness of activity, making plasticity rules responsive to the rate of individual cells presynaptic to an interneuron might suffice. But if it is to limit the total firing rate error across the excitatory population, then the plasticity rule at play should be responsive to the activity of cells inhibited by each interneuron. And finally, if plasticity does make interneurons responsive to their postsynaptic population, such a rule can also account for interneuron stimulus selectivity, the co-tuning of excitatory and inhibitory synaptic currents, and the presence of excitatory-inhibitory cell assemblies within sensory cortex.

7 Appendix

7.1 Derivation of the homeostatic gradient

We begin by deriving a rule that minimizes the total error of the time-averaged excitatory membrane potentials $\boldsymbol{h}^{\mathrm{E}}$, with respect to the homeostatic target rate ρ_0, by changing a given parameter v (either $V_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}$, $V_{ji}^{\mathrm{E}\leftarrow\mathrm{I}}$, v_i^{θ}, or v_i^{g}; see the Methods for their definitions). First we define an error function

$$\mathcal{E}(\boldsymbol{h}^{\mathrm{E}};\boldsymbol{v}^{g},\boldsymbol{v}^{\theta},\mathrm{V}^{\mathrm{I}\leftarrow\mathrm{E}},\mathrm{V}^{\mathrm{E}\leftarrow\mathrm{I}}) = \frac{1}{2}\sum_{l=1}^{N^{\mathrm{E}}}\left(\left\langle h_l^{\mathrm{E}}\right\rangle_t - \rho_0\right)^2 \tag{7.1}$$

where h_l^{E} is the membrane potential of excitatory cell l which in the steady-state is

$$h_l^{\mathrm{E}} = \sum_{m=1}^{N^{\mathrm{E}}} W_{lm}^{\mathrm{E}\leftarrow\mathrm{E}} r_m^{\mathrm{E}} - \sum_{n=1}^{N^{\mathrm{I}}} W_{ln}^{\mathrm{E}\leftarrow\mathrm{I}} r_n^{\mathrm{I}} + \mu_{\mathrm{ou}}^{\mathrm{E}}. \tag{7.2}$$

Taking the derivative of Eq 7.1 with respect to some underlying parameter v yields

$$\frac{\partial}{\partial v}\mathcal{E}(\boldsymbol{h}^{\mathrm{E}};\boldsymbol{v}^{g},\boldsymbol{v}^{\theta},\mathrm{V}^{\mathrm{I}\leftarrow\mathrm{E}},\mathrm{V}^{\mathrm{E}\leftarrow\mathrm{I}}) = \sum_{l=1}^{N^{\mathrm{E}}}\left(\left\langle h_l^{\mathrm{E}}\right\rangle_t - \rho_0\right)\frac{\partial h_l^{\mathrm{E}}}{\partial v}. \tag{7.3}$$

Thus we need the derivative of Eq 7.2, which is

$$\frac{\partial h_l^{\mathrm{E}}}{\partial v} = \sum_{m=1}^{N^{\mathrm{E}}} W_{lm}^{\mathrm{E}\leftarrow\mathrm{E}}\frac{\partial r_m^{\mathrm{E}}}{\partial h_m^{\mathrm{E}}}\frac{\partial h_m^{\mathrm{E}}}{\partial v} - \sum_{n=1}^{N^{\mathrm{I}}}\left[\frac{\partial W_{ln}^{\mathrm{E}\leftarrow\mathrm{I}}}{\partial v} r_n^{\mathrm{I}} + W_{ln}^{\mathrm{E}\leftarrow\mathrm{I}}\frac{\partial r_n^{\mathrm{I}}}{\partial v}\right]. \tag{7.4}$$

Note that the derivative of any parameter that is not a function of a plastic parameter is zero, and so those terms are excluded. Next we need

$$\frac{\partial r_n^{\mathrm{I}}}{\partial v} = \frac{\partial r_n^{\mathrm{I}}}{\partial h_n^{\mathrm{I}}}\frac{\partial h_n^{\mathrm{I}}}{\partial v} + \sum_{z=1}^{N^{\Gamma}}\frac{\partial r_n^{\mathrm{I}}}{\partial \Gamma_n^z}\frac{\partial \Gamma_n^z}{\partial v} \tag{7.5}$$

where $\Gamma_n^1 = v_n^g$ and $\Gamma_n^2 = v_n^\theta$ (so $N^\Gamma = 2$), the intrinsic parameters of the interneuron n. Eq 7.5 depends on $\frac{\partial h^{\mathrm{I}}}{\partial v}$ which is

$$\begin{aligned}\frac{\partial h_n^{\mathrm{I}}}{\partial v} &= \frac{\partial}{\partial v}\left(\sum_{m=1}^{N^{\mathrm{E}}} W_{nm}^{\mathrm{I}\leftarrow\mathrm{E}} r_m^{\mathrm{E}} - \sum_{q=1}^{N^{\mathrm{I}}} W_{nq}^{\mathrm{I}\leftarrow\mathrm{I}} r_q^{\mathrm{I}} + \mu_{\mathrm{ou}}^{\mathrm{I}}\right) \\ &= \sum_{m=1}^{N^{\mathrm{E}}}\left(\frac{\partial W_{nm}^{\mathrm{I}\leftarrow\mathrm{E}}}{\partial v} r_m^{\mathrm{E}} + W_{nm}^{\mathrm{I}\leftarrow\mathrm{E}} \frac{\partial r_m^{\mathrm{E}}}{\partial h_m^{\mathrm{E}}} \frac{\partial h_m^{\mathrm{E}}}{\partial v}\right) - \sum_{q=1}^{N^{\mathrm{I}}} W_{nq}^{\mathrm{I}\leftarrow\mathrm{I}} \frac{\partial r_q^{\mathrm{I}}}{\partial v}\end{aligned}$$

substituted into Eq 7.5

$$\frac{\partial r_n^{\mathrm{I}}}{\partial v} = \frac{\partial r_n^{\mathrm{I}}}{\partial h_n^{\mathrm{I}}}\left[\sum_{m=1}^{N^{\mathrm{E}}}\left(\frac{\partial W_{nm}^{\mathrm{I}\leftarrow\mathrm{E}}}{\partial v} r_m^{\mathrm{E}} + W_{nm}^{\mathrm{I}\leftarrow\mathrm{E}} \frac{\partial r_m^{\mathrm{E}}}{\partial h_m^{\mathrm{E}}} \frac{\partial h_m^{\mathrm{E}}}{\partial v}\right) - \sum_{q=1}^{N^{\mathrm{I}}} W_{nq}^{\mathrm{I}\leftarrow\mathrm{I}} \frac{\partial r_q^{\mathrm{I}}}{\partial v}\right] + \sum_{z=1}^{N^\Gamma} \frac{\partial r_n^{\mathrm{I}}}{\partial \Gamma_n^z} \frac{\partial \Gamma_n^z}{\partial v}$$

replace $n \to p$, and rearrange

$$\begin{aligned}\frac{\partial r_p^{\mathrm{I}}}{\partial v} + \sum_{q=1}^{N^{\mathrm{I}}}\left(\frac{\partial r_p^{\mathrm{I}}}{\partial h_p^{\mathrm{I}}} W_{pq}^{\mathrm{I}\leftarrow\mathrm{I}} \frac{\partial r_q^{\mathrm{I}}}{\partial v}\right) &= \sum_{q=1}^{N^{\mathrm{I}}}\left(\delta_{pq} + \frac{\partial r_p^{\mathrm{I}}}{\partial h_p^{\mathrm{I}}} W_{pq}^{\mathrm{I}\leftarrow\mathrm{I}}\right)\frac{\partial r_q^{\mathrm{I}}}{\partial v} \\ &= \frac{\partial r_p^{\mathrm{I}}}{\partial h_p^{\mathrm{I}}} \sum_{m=1}^{N^{\mathrm{E}}}\left(\frac{\partial W_{pm}^{\mathrm{I}\leftarrow\mathrm{E}}}{\partial v} r_m^{\mathrm{E}} + W_{pm}^{\mathrm{I}\leftarrow\mathrm{E}} \frac{\partial r_m^{\mathrm{E}}}{\partial h_m^{\mathrm{E}}} \frac{\partial h_m^{\mathrm{E}}}{\partial v}\right) + \sum_{z=1}^{N^\Gamma} \frac{\partial r_p^{\mathrm{I}}}{\partial \Gamma_p^z} \frac{\partial \Gamma_p^z}{\partial v}.\end{aligned} \tag{7.6}$$

Now define the matrix $\mathcal{M}$ as:

$$\mathcal{M}_{pq} = \delta_{pq} + \frac{\partial r_p^{\mathrm{I}}}{\partial h_p^{\mathrm{I}}} W_{pq}^{\mathrm{I}\leftarrow\mathrm{I}}$$

and multiply both sides of Eq 7.6 by its matrix inverse to solve for $\partial r_n^{\mathrm{I}}/\partial v$

$$\begin{aligned}\sum_{q=1}^{N^{\mathrm{I}}}\sum_{p=1}^{N^{\mathrm{I}}}\left[\mathcal{M}^{-1}\right]_{np}\left[\mathcal{M}\right]_{pq}\frac{\partial r_q^{\mathrm{I}}}{\partial v} &= \sum_{q=1}^{N^{\mathrm{I}}}\delta_{nq}\frac{\partial r_q^{\mathrm{I}}}{\partial v} = \frac{\partial r_n^{\mathrm{I}}}{\partial v} \\ &= \sum_{p=1}^{N^{\mathrm{I}}}\left[\mathcal{M}^{-1}\right]_{np}\left[\frac{\partial r_p^{\mathrm{I}}}{\partial h_p^{\mathrm{I}}}\sum_{m=1}^{N^{\mathrm{E}}}\left(\frac{\partial W_{pm}^{\mathrm{I}\leftarrow\mathrm{E}}}{\partial v} r_m^{\mathrm{E}} + W_{pm}^{\mathrm{I}\leftarrow\mathrm{E}}\frac{\partial r_m^{\mathrm{E}}}{\partial h_m^{\mathrm{E}}}\frac{\partial h_m^{\mathrm{E}}}{\partial v}\right)\right. \\ &\quad \left. + \sum_{z=1}^{N^\Gamma}\frac{\partial r_p^{\mathrm{I}}}{\partial \Gamma_p^z}\frac{\partial \Gamma_p^z}{\partial v}\right].\end{aligned}$$

Thus,

$$\frac{\partial r_n^{\mathrm{I}}}{\partial v} = \sum_{p=1}^{N^{\mathrm{I}}}\left[\mathcal{M}^{-1}\right]_{np}\left[\frac{\partial r_p^{\mathrm{I}}}{\partial h_p^{\mathrm{I}}}\sum_{m=1}^{N^{\mathrm{E}}}\left(\frac{\partial W_{pm}^{\mathrm{I}\leftarrow\mathrm{E}}}{\partial v} r_m^{\mathrm{E}} + W_{pm}^{\mathrm{I}\leftarrow\mathrm{E}}\frac{\partial r_m^{\mathrm{E}}}{\partial h_m^{\mathrm{E}}}\frac{\partial h_m^{\mathrm{E}}}{\partial v}\right) + \sum_{z=1}^{N^\Gamma}\frac{\partial r_p^{\mathrm{I}}}{\partial \Gamma_p^z}\frac{\partial \Gamma_p^z}{\partial v}\right]. \tag{7.7}$$

Substitute Eq 7.7 into Eq 7.4

$$\frac{\partial h_l^{\mathrm{E}}}{\partial v} = \sum_{m=1}^{N^{\mathrm{E}}} W_{lm}^{\mathrm{E}\leftarrow\mathrm{E}} \frac{\partial r_m^{\mathrm{E}}}{\partial h_m^{\mathrm{E}}} \frac{\partial h_m^{\mathrm{E}}}{\partial v} - \sum_{n=1}^{N^{\mathrm{I}}} \Bigg\{ \frac{\partial W_{ln}^{\mathrm{E}\leftarrow\mathrm{I}}}{\partial v} r_n^{\mathrm{I}} - \sum_{p=1}^{N^{\mathrm{I}}} W_{ln}^{\mathrm{E}\leftarrow\mathrm{I}} [\mathcal{M}^{-1}]_{np} \left[\frac{\partial r_p^{\mathrm{I}}}{\partial h_p^{\mathrm{I}}} \sum_{q=1}^{N^{\mathrm{E}}} \left(\frac{\partial W_{pq}^{\mathrm{I}\leftarrow\mathrm{E}}}{\partial v} r_q^{\mathrm{E}} + W_{pq}^{\mathrm{I}\leftarrow\mathrm{E}} \frac{\partial r_q^{\mathrm{E}}}{\partial h_q^{\mathrm{E}}} \frac{\partial h_q^{\mathrm{E}}}{\partial v} \right) + \sum_{z=1}^{N^{\Gamma}} \frac{\partial r_p^{\mathrm{I}}}{\partial \Gamma_p^z} \frac{\partial \Gamma_p^z}{\partial v} \right] \Bigg\},$$

gather terms in $\frac{\partial \boldsymbol{h}^{\mathrm{E}}}{\partial v}$ on the LHS, and then factor it out:

$$\begin{aligned}
&\frac{\partial h_l^{\mathrm{E}}}{\partial v} + \sum_{m=1}^{N^{\mathrm{E}}} \left(-W_{lm}^{\mathrm{E}\leftarrow\mathrm{E}} \frac{\partial r_m^{\mathrm{E}}}{\partial h_m^{\mathrm{E}}} \frac{\partial h_m^{\mathrm{E}}}{\partial v} + \sum_{n=1}^{N^{\mathrm{I}}} \sum_{p=1}^{N^{\mathrm{I}}} W_{ln}^{\mathrm{E}\leftarrow\mathrm{I}} [\mathcal{M}^{-1}]_{np} \frac{\partial r_p^{\mathrm{I}}}{\partial h_p^{\mathrm{I}}} W_{pm}^{\mathrm{I}\leftarrow\mathrm{E}} \frac{\partial r_m^{\mathrm{E}}}{\partial h_m^{\mathrm{E}}} \frac{\partial h_m^{\mathrm{E}}}{\partial v} \right) \\
&= \sum_{m=1}^{N^{\mathrm{E}}} \left(\delta_{lm} - W_{lm}^{\mathrm{E}\leftarrow\mathrm{E}} \frac{\partial r_m^{\mathrm{E}}}{\partial h_m^{\mathrm{E}}} + \sum_{n=1}^{N^{\mathrm{I}}} \sum_{p=1}^{N^{\mathrm{I}}} W_{ln}^{\mathrm{E}\leftarrow\mathrm{I}} [\mathcal{M}^{-1}]_{np} \frac{\partial r_p^{\mathrm{I}}}{\partial h_p^{\mathrm{I}}} W_{pm}^{\mathrm{I}\leftarrow\mathrm{E}} \frac{\partial r_m^{\mathrm{E}}}{\partial h_m^{\mathrm{E}}} \right) \frac{\partial h_m^{\mathrm{E}}}{\partial v} \\
&= -\sum_{n=1}^{N^{\mathrm{I}}} \left\{ \frac{\partial W_{ln}^{\mathrm{E}\leftarrow\mathrm{I}}}{\partial v} r_n^{\mathrm{I}} + \sum_{p=1}^{N^{\mathrm{I}}} W_{ln}^{\mathrm{E}\leftarrow\mathrm{I}} [\mathcal{M}^{-1}]_{np} \left(\sum_{q=1}^{N^{\mathrm{E}}} \frac{\partial r_p^{\mathrm{I}}}{\partial h_p^{\mathrm{I}}} \frac{\partial W_{pq}^{\mathrm{I}\leftarrow\mathrm{E}}}{\partial v} r_q^{\mathrm{E}} + \sum_{z=1}^{N^{\Gamma}} \frac{\partial r_p^{\mathrm{I}}}{\partial \Gamma_p^z} \frac{\partial \Gamma_p^z}{\partial v} \right) \right\}.
\end{aligned}$$

Multiplying both sides by the matrix

$$\mathcal{W} = \mathbb{I} - W^{\mathrm{E}\leftarrow\mathrm{E}} \mathrm{diag}\left(\frac{\partial \boldsymbol{r}^{\mathrm{E}}}{\partial \boldsymbol{h}^{\mathrm{E}}} \right) + W^{\mathrm{E}\leftarrow\mathrm{I}} \mathcal{M}^{-1} \mathrm{diag}\left(\frac{\partial \boldsymbol{r}^{\mathrm{I}}}{\partial \boldsymbol{h}^{\mathrm{I}}} \right) W^{\mathrm{I}\leftarrow\mathrm{E}} \mathrm{diag}\left(\frac{\partial \boldsymbol{r}^{\mathrm{E}}}{\partial \boldsymbol{h}^{\mathrm{E}}} \right)$$

yields

$$\frac{\partial h_k^{\mathrm{E}}}{\partial v} = -\sum_{l=1}^{N^{\mathrm{E}}} [\mathcal{W}^{-1}]_{kl} \sum_{n=1}^{N^{\mathrm{I}}} \left[\frac{\partial W_{ln}^{\mathrm{E}\leftarrow\mathrm{I}}}{\partial v} r_n^{\mathrm{I}} + \sum_{p=1}^{N^{\mathrm{I}}} W_{ln}^{\mathrm{E}\leftarrow\mathrm{I}} [\mathcal{M}^{-1}]_{np} \left(\sum_{q=1}^{N^{\mathrm{E}}} \frac{\partial r_p^{\mathrm{I}}}{\partial h_p^{\mathrm{I}}} \frac{\partial W_{pq}^{\mathrm{I}\leftarrow\mathrm{E}}}{\partial v} r_q^{\mathrm{E}} + \sum_{z=1}^{N^{\Gamma}} \frac{\partial r_p^{\mathrm{I}}}{\partial \Gamma_p^z} \frac{\partial \Gamma_p^z}{\partial v} \right) \right]$$

Substituting this into Eq 7.3 produces the full gradient, with respect to all potentially plastic parameters $v = \{V_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}, V_{ji}^{\mathrm{E}\leftarrow\mathrm{I}}, v_i^{\theta}, v_i^{g}\}$:

$$\begin{aligned}
\frac{\partial \mathcal{E}}{\partial v} = -\sum_{k=1}^{N^{\mathrm{E}}} \sum_{l=1}^{N^{\mathrm{E}}} \left(\left\langle h_k^{\mathrm{E}} \right\rangle_t - \rho_0 \right) \left[\mathcal{W}^{-1} \right]_{kl} \sum_{n=1}^{N^{\mathrm{I}}} \Bigg[&\frac{\partial W_{ln}^{\mathrm{E}\leftarrow\mathrm{I}}}{\partial v} r_n^{\mathrm{I}} \\
&+ W_{ln}^{\mathrm{E}\leftarrow\mathrm{I}} \sum_{p=1}^{N^{\mathrm{I}}} \left[\mathcal{M}^{-1} \right]_{np} \left(\sum_{q=1}^{N^{\mathrm{E}}} \frac{\partial r_p^{\mathrm{I}}}{\partial h_p^{\mathrm{I}}} \frac{\partial W_{pq}^{\mathrm{I}\leftarrow\mathrm{E}}}{\partial v} r_q^{\mathrm{E}} + \sum_{z=1}^{N^{\Gamma}} \frac{\partial r_p^{\mathrm{I}}}{\partial \Gamma_z} \frac{\partial \Gamma_z}{\partial v} \right) \Bigg]
\end{aligned} \tag{7.8}$$

To obtain gradients with respect to a particular type of parameter (a type of synapse, or some intrinsic property), we simply substitute the chosen parameter into Eq 7.8. So if

we set v to be the excitatory-to-inhibitory synaptic parameter $V_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}$, it reduces to

$$\begin{aligned}\frac{\partial \mathcal{E}}{\partial V_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}} &= -\sum_{kl}\left(\left\langle h_k^{\mathrm{E}}\right\rangle_t - \rho_0\right)\left[\mathcal{W}^{-1}\right]_{kl}\sum_{npq} W_{ln}^{\mathrm{E}\leftarrow\mathrm{I}}\left[\mathcal{M}^{-1}\right]_{np}\delta_{pi}\frac{\partial r_i^{\mathrm{I}}}{\partial h_i^{\mathrm{I}}}\frac{\partial W_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}}{\partial V_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}} r_j^{\mathrm{E}}\delta_{jq} \\ &= -\left[\left(\left\langle \boldsymbol{h}^{\mathrm{E}}\right\rangle_t - \rho_0\right)^{\mathsf{T}}\mathcal{W}^{-1}W^{\mathrm{E}\leftarrow\mathrm{I}}\mathcal{M}^{-1}\right]_i \frac{\partial r_i^{\mathrm{I}}}{\partial h_i^{\mathrm{I}}}\frac{\partial W_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}}{\partial V_{ij}^{\mathrm{I}\leftarrow\mathrm{E}}} r_j^{\mathrm{E}},\end{aligned} \tag{7.9}$$

whereas setting v to be the inhibitory-to-excitatory synaptic parameter $V_{ji}^{\mathrm{E}\leftarrow\mathrm{I}}$, results in

$$\begin{aligned}\frac{\partial \mathcal{E}}{\partial V_{ji}^{\mathrm{E}\leftarrow\mathrm{I}}} &= -\sum_{kl}\left(\left\langle h_k^{\mathrm{E}}\right\rangle_t - \rho_0\right)\left[\mathcal{W}^{-1}\right]_{kl}\delta_{lj}\sum_{n}\frac{\partial W_{ji}^{\mathrm{E}\leftarrow\mathrm{I}}}{\partial V_{ji}^{\mathrm{E}\leftarrow\mathrm{I}}} r_i^{\mathrm{I}}\delta_{in} \\ &= -\left[\left(\left\langle \boldsymbol{h}^{\mathrm{E}}\right\rangle_t - \rho_0\right)^{\mathsf{T}}\mathcal{W}^{-1}\right]_j \frac{\partial W_{ji}^{\mathrm{E}\leftarrow\mathrm{I}}}{\partial V_{ji}^{\mathrm{E}\leftarrow\mathrm{I}}} r_i^{\mathrm{I}}.\end{aligned} \tag{7.10}$$

To obtain the gradient for the interneuron firing threshold θ_i, substitute the underlying plastic parameter v_i^{θ}, which yields

$$\begin{aligned}\frac{\partial \mathcal{E}}{\partial v_i^{\theta}} &= -\sum_{kl}\left(\left\langle h_k^{\mathrm{E}}\right\rangle_t - \rho_0\right)\left[\mathcal{W}^{-1}\right]_{kl}\sum_{np} W_{ln}^{\mathrm{E}\leftarrow\mathrm{I}}\left[\mathcal{M}^{-1}\right]_{np}\delta_{pi}\frac{\partial r_i^{\mathrm{I}}}{\partial \theta_i}\frac{\partial \theta_i}{\partial v_i^{\theta}} \\ &= -\left[\left(\left\langle \boldsymbol{h}^{\mathrm{E}}\right\rangle_t - \rho_0\right)^{\mathsf{T}}\mathcal{W}^{-1}W^{\mathrm{E}\leftarrow\mathrm{I}}\mathcal{M}^{-1}\right]_i \frac{\partial r_i^{\mathrm{I}}}{\partial \theta_i}.\end{aligned} \tag{7.11}$$

Finally, for the gradient of the interneuron gain g_i, we substitute v_i^{g} for v

$$\begin{aligned}\frac{\partial \mathcal{E}}{\partial v_i^{g}} &= -\sum_{kl}\left(\left\langle h_k^{\mathrm{E}}\right\rangle_t - \rho_0\right)\left[\mathcal{W}^{-1}\right]_{kl}\sum_{np} W_{ln}^{\mathrm{E}\leftarrow\mathrm{I}}\left[\mathcal{M}^{-1}\right]_{np}\delta_{pi}\frac{\partial r_i^{\mathrm{I}}}{\partial g_i}\frac{\partial g_i}{\partial v_i^{g}} \\ &= -\left[\left(\left\langle \boldsymbol{h}^{\mathrm{E}}\right\rangle_t - \rho_0\right)^{\mathsf{T}}\mathcal{W}^{-1}W^{\mathrm{E}\leftarrow\mathrm{I}}\mathcal{M}^{-1}\right]_i \frac{\partial r_i^{\mathrm{I}}}{\partial g_i}\frac{\partial g_i}{\partial v_i^{g}}.\end{aligned} \tag{7.12}$$

7.2 Approximating the gradient rules

Within the gradient-based rules derived in the previous section, the $\mathcal{W}^{-1}$ and $\mathcal{M}^{-1}$ terms make their implementation in a biological system difficult to conceive. To resolve this challenge, we begin by noting that

$$\mathcal{W}^{-1} \equiv \left(\mathbb{I} - \hat{\mathcal{W}}\right)^{-1} = \sum_{k=0}^{\infty} \hat{\mathcal{W}}^k,$$

which holds if $||\hat{\mathcal{W}}|| < 1$, or equivalently that $\mathbb{I} - \hat{\mathcal{W}}$ is invertible. Naturally, this relation holds for $\mathcal{M}^{-1}$ as well. Since both those matrices are contained in Eqs 7.9, 7.11 & 7.12, we substitute the equivalent sums into the relevant sub-expression as in

$$\begin{aligned}
\mathcal{W}^{-1} W^{\mathrm{E}\leftarrow\mathrm{I}} \mathcal{M}^{-1} &= \left(\sum_{k=0}^{\infty} \hat{\mathcal{W}}^k\right) W^{\mathrm{E}\leftarrow\mathrm{I}} \left(\sum_{k=0}^{\infty} \hat{\mathcal{M}}^k\right) \\
&= W^{\mathrm{E}\leftarrow\mathrm{I}} + \hat{\mathcal{W}} W^{\mathrm{E}\leftarrow\mathrm{I}} + W^{\mathrm{E}\leftarrow\mathrm{I}} \hat{\mathcal{M}} + \left(\sum_{k=1}^{\infty} \hat{\mathcal{W}}^k\right) W^{\mathrm{E}\leftarrow\mathrm{I}} \left(\sum_{k=1}^{\infty} \hat{\mathcal{M}}^k\right) \\
&\approx W^{\mathrm{E}\leftarrow\mathrm{I}}.
\end{aligned}$$

The approximate equivalence in the last line follows from an assumption that synapses are sufficiently weak that the effect of the lowest-order term is much stronger than that of the remaining terms, which can thus be ignored. This approximation can be substituted into Eqs 7.9, 7.11 & 7.12 to yield physically realizable plasticity rules with increased biological plausibility. The same logic is employed to recover the inhibitory synaptic plasticity rule of Vogels et al. (2011) from Eq 7.10, by retaining only the zeroth-order term in the expansion of $\mathcal{W}^{-1}$.

7.3 Effect of propagation kernel choice on expected population rate

In a network with retroaxonal plasticity, the population firing rate can systematically deviate from the target rate ρ_0, whereas networks with diffusion-like plasticity do not. To understand why, we consider how inhibitory firing rates are expected to change given the excitatory firing rates $\boldsymbol{r}^{\mathrm{E}}$ and a particular propagation kernel K. The inhibitory rates have a fixed point $r^{\mathrm{I}*}$ when

$$\sum_j \left(r_j^{\mathrm{E}} - \rho_0\right) K_{ji} \stackrel{!}{=} 0. \tag{7.13}$$

We average this, with respect to all interneurons i

$$0 = \frac{1}{N^{\mathrm{I}}} \sum_i \sum_j \left(r_j^{\mathrm{E}} - \rho_0\right) K_{ji} = \frac{1}{N^{\mathrm{I}}} \sum_j \left(r_j^{\mathrm{E}} - \rho_0\right) \sum_i K_{ji}. \tag{7.14}$$

Where for diffusion-like plasticity $K = \left\langle \Delta^{\mathrm{E}\leftarrow\mathrm{I}} \right\rangle$, so $\sum_i K_{ji} = p^{\mathrm{E}\leftarrow\mathrm{I}} N^{\mathrm{I}}$ is constant, and $\left\langle \boldsymbol{r}^{\mathrm{E}} \right\rangle = \rho_0$, thus the population rate is the target rate.

In contrast, for retroaxonal plasticity $\boldsymbol{K} = \Delta^{\mathrm{E}\leftarrow\mathrm{I}}$, which can be factored into two components $\bar{\boldsymbol{K}} = \left\langle \Delta^{\mathrm{E}\leftarrow\mathrm{I}} \right\rangle$, the ensemble average of the adjacency matrix, and $\boldsymbol{\delta K}$ which is a zero-mean noise term.

$$\begin{aligned}
0 &= \frac{1}{N^{\mathrm{I}}} \sum_j \left(r_j^{\mathrm{E}} - \rho_0\right) \sum_i \left(\bar{K}_{ji} + \delta K_{ji}\right) && (7.15)\\
&= \frac{1}{N^{\mathrm{I}}} \sum_j \left(p^{\mathrm{E}\leftarrow\mathrm{I}} N^{\mathrm{I}} \left(r_j^{\mathrm{E}} - \rho_0\right) + \sum_i \left(r_j^{\mathrm{E}} - \rho_0\right) \delta K_{ji} \right) && (7.16)\\
&= p^{\mathrm{E}\leftarrow\mathrm{I}} N^{\mathrm{E}} \left(\left\langle \boldsymbol{r}^{\mathrm{E}} \right\rangle - \rho_0 \right) + \frac{1}{N^{\mathrm{I}}} \sum_{ji} r_j^{\mathrm{E}} \delta K_{ji} && (7.17)
\end{aligned}$$

with the ρ_0 in the second term eliminated by summing over the zero-mean matrix.

If $\mathrm{corr}\left(r_j^{\mathrm{E}},\, \sum_i \delta K_{ji}\right) < 0$, i.e. excitatory cells with low (high) inhibitory in-degree typically fire at higher (lower) rates, the second term in the last equation will be negative, and thus

$$\left\langle \boldsymbol{r}^{\mathrm{E}} \right\rangle = \rho_0 - \frac{1}{p^{\mathrm{E}\leftarrow\mathrm{I}} N^{\mathrm{E}} N^{\mathrm{I}}} \sum_{ji} r_j^{\mathrm{E}} \delta K_{ji} > \rho_0. \tag{7.18}$$

Hence, the expected population rate for the retroaxonal propagation kernel is greater than the homeostatic set-point, due to a shift by a constant term dependent on the connectivity statistics.

Bibliography

Agnati, L., Zoli, M., Strömberg, I., and Fuxe, K. (1995). Intercellular communication in the brain: wiring versus volume transmission. *Neuroscience*, 69(3):711–726.

Antoniou, A. (1993). *Digital filters*. McGraw Hill.

Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nature reviews neuroscience*, 7(5):358.

Bandyopadhyay, S., Shamma, S. A., and Kanold, P. O. (2010). Dichotomy of functional organization in the mouse auditory cortex. *Nature neuroscience*, 13(3):361.

Barkat, T. R., Polley, D. B., and Hensch, T. K. (2011). A critical period for auditory thalamocortical connectivity. *Nature neuroscience*, 14(9):1189.

Barnes, S. J., Sammons, R. P., Jacobsen, R. I., Mackie, J., Keller, G. B., and Keck, T. (2015). Subnetwork-specific homeostatic plasticity in mouse visual cortex in vivo. *Neuron*, 86(5):1290–1303.

Barth, A. L. and Poulet, J. F. (2012). Experimental evidence for sparse firing in the neocortex. *Trends in neurosciences*, 35(6):345–355.

Bartley, A. F., Huang, Z. J., Huber, K. M., and Gibson, J. R. (2008). Differential activity-dependent, homeostatic plasticity of two neocortical inhibitory circuits. *Journal of Neurophysiology*, 100(4):1983–1994.

Berger, T. K., Perin, R., Silberberg, G., and Markram, H. (2009). Frequency-dependent disynaptic inhibition in the pyramidal network: a ubiquitous pathway in the developing rat neocortex. *The Journal of physiology*, 587(22):5411–5425.

Bienenstock, E. L., Cooper, L. N., and 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.

Branco, T., Staras, K., Darcy, K. J., and Goda, Y. (2008). Local dendritic activity sets release probability at hippocampal synapses. *Neuron*, 59(3):475–485.

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of computational neuroscience*, 8(3):183–208.

Buzsáki, G. and Mizuseki, K. (2014). The log-dynamic brain: how skewed distributions affect network operations. *Nature Reviews Neuroscience*, 15(4):264.

Cajal, S. R. y. (1899). *Comparative study of the sensory areas of the human cortex*. Clark University.

Cajal, S. R. y. (1923). *Recuerdos de mi vida: Obra ilustrada con numerosos fotograbados. 1. P. Mi infancia y juventud.(2. P. Historia de mi labor científica.).* Imprenta de Juan Pueyo.

Cannon, W. B. (1932). The wisdom of the body.

Chance, F. S., Abbott, L. F., and Reyes, A. D. (2002). Gain modulation from background synaptic input. *Neuron*, 35(4):773–782.

Chang, M. C., Park, J. M., Pelkey, K. A., Grabenstatter, H. L., Xu, D., Linden, D. J., Sutula, T. P., McBain, C. J., and Worley, P. F. (2010). Narp regulates homeostatic scaling of excitatory synapses on parvalbumin-expressing interneurons. *Nature neuroscience*, 13(9):1090–1097.

Clopath, C., Vogels, T. P., Froemke, R. C., and Sprekeler, H. (2016). Receptive field formation by interacting excitatory and inhibitory synaptic plasticity. *bioRxiv*, page 066589.

Cook, R. L. (1986). Stochastic sampling in computer graphics. *ACM Transactions on Graphics (TOG)*, 5(1):51–72.

Cowan, A., Stricker, C., Reece, L., and Redman, S. (1998). Long-term plasticity at excitatory synapses on aspinous interneurons in area ca1 lacks synaptic specificity. *Journal of Neurophysiology*, 79(1):13–20.

Creutzfeldt, O. D. (1977). Generality of the functional structure of the neocortex. *Naturwissenschaften*, 64(10):507–517.

Dale, H. (1935). Pharmacology and nerve-endings.

Desai, N. S., Rutherford, L. C., and Turrigiano, G. G. (1999a). Bdnf regulates the intrinsic excitability of cortical neurons. *Learning & Memory*, 6(3):284–291.

Desai, N. S., Rutherford, L. C., and Turrigiano, G. G. (1999b). Plasticity in the intrinsic excitability of cortical pyramidal neurons. *Nature neuroscience*, 2(6):515.

Dityatev, A., Brückner, G., Dityateva, G., Grosche, J., Kleene, R., and Schachner, M. (2007). Activity-dependent formation and functions of chondroitin sulfate-rich extracellular matrix of perineuronal nets. *Developmental neurobiology*, 67(5):570–588.

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(7300):932.

Douglas, R. J., Koch, C., Mahowald, M., Martin, K., and Suarez, H. H. (1995). Recurrent excitation in neocortical circuits. *Science*, 269(5226):981–985.

Douglas, R. J. and Martin, K. A. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.*, 27:419–451.

Doyle, S., Pyndiah, S., De Gois, S., and Erickson, J. D. (2010). Excitation-transcription coupling via calcium/calmodulin-dependent protein kinase/erk1/2 signaling mediates the coordinate induction of vglut2 and narp triggered by a prolonged increase in glutamatergic synaptic activity. *Journal of Biological Chemistry*, 285(19):14366–14376.

Dudek, F. E. and Sutula, T. P. (2007). Epileptogenesis in the dentate gyrus: a critical perspective. *Progress in brain research*, 163:755–773.

Fernandes, D. and Carvalho, A. L. (2016). Mechanisms of homeostatic plasticity in the excitatory synapse. *Journal of neurochemistry*, 139(6):973–996.

Fino, E., Deniau, J.-M., and Venance, L. (2008). Cell-specific spike-timing-dependent plasticity in gabaergic and cholinergic interneurons in corticostriatal rat brain slices. *The Journal of physiology*, 586(1):265–282.

Fino, E., Packer, A. M., and Yuste, R. (2013). The logic of inhibitory connectivity in the neocortex. *The Neuroscientist*, 19(3):228–237.

Fino, E. and Yuste, R. (2011). Dense inhibitory connectivity in neocortex. *Neuron*, 69(6):1188–1203.

Fitzsimonds, R. M., Song, H.-j., and Poo, M.-m. (1997). Propagation of activity-dependent synaptic depression in simple neural networks. *Nature*, 388(6641):439.

Froemke, R. C., Merzenich, M. M., and Schreiner, C. E. (2007). A synaptic memory trace for cortical receptive field plasticity. *Nature*, 450(7168):425.

Garthwaite, J. (2016). From synaptically localized to volume transmission by nitric oxide. *The Journal of physiology*, 594(1):9–18.

Ginty, D. D. and Segal, R. A. (2002). Retrograde neurotrophin signaling: Trk-ing along the axon. *Current opinion in neurobiology*, 12(3):268–274.

Graham, J. and Gerard, R. (1946). Membrane potentials and excitation of impaled single muscle fibers. *Journal of cellular and comparative physiology*, 28(1):99–117.

Haider, B., Duque, A., Hasenstaub, A. R., and McCormick, D. A. (2006). Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *Journal of Neuroscience*, 26(17):4535–4545.

Hardingham, N., Dachtler, J., and Fox, K. (2013). The role of nitric oxide in pre-synaptic plasticity and homeostasis. *Frontiers in cellular neuroscience*, 7.

Harris, K. D. (2008). Stability of the fittest: organizing learning through retroaxonal signals. *Trends in neurosciences*, 31(3):130–136.

Harris, K. D. and Mrsic-Flogel, T. D. (2013). Cortical connectivity and sensory coding. *Nature*, 503(7474):51.

Hartman, K. N., Pal, S. K., Burrone, J., and Murthy, V. N. (2006). Activity-dependent regulation of inhibitory synaptic transmission in hippocampal neurons. *Nature neuroscience*, 9(5):642.

Hengen, K., Lambo, M., Van Hooser, S., Katz, D., and Turrigiano, G. (2013). Firing rate homeostasis in visual cortex of freely behaving rodents. *Neuron*, 80(2):335 – 342.

Hennequin, G., Vogels, T., and Gerstner, W. (2012). Nonnormal amplification in random balanced neuronal networks. *Physical Review E*, 86:011909(1–12).

Hofer, S. B., Ko, H., Pichler, B., Vogelstein, J., Ros, H., Zeng, H., Lein, E., Lesica, N. A., and Mrsic-Flogel, T. D. (2011). Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex. *Nature neuroscience*, 14(8):1045–1052.

Horikawa, J., Ito, S., Hosokawa, Y., HOMMA, T., and MURATA, K. (1988). Tonotopic representation in the rat auditory cortex. *Proceedings of the Japan Academy, Series B*, 64(8):260–263.

Huang, Z. J., Kirkwood, A., Pizzorusso, T., Porciatti, V., Morales, B., Bear, M. F., Maffei, L., and Tonegawa, S. (1999). Bdnf regulates the maturation of inhibition and the critical period of plasticity in mouse visual cortex. *Cell*, 98(6):739–755.

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3):574–591.

Intskirveli, I., Joshi, A., Vizcarra-Chacón, B. J., and Metherate, R. (2016). Spectral breadth and laminar distribution of thalamocortical inputs to a1. *Journal of neurophysiology*, 115(4):2083–2094.

Isaacson, J. and Scanziani, M. (2011). How inhibition shapes cortical activity. *Neuron*, 72(2):231 – 243.

Joglekar, M. R., Mejias, J. F., Yang, G. R., and Wang, X.-J. (2017). Inter-areal balanced amplification enhances signal propagation in a large-scale circuit model of the primate cortex. *bioRxiv*, page 186007.

Kameyama, K., Sohya, K., Ebina, T., Fukuda, A., Yanagawa, Y., and Tsumoto, T. (2010). Difference in binocularity and ocular dominance plasticity between gabaergic and excitatory cortical neurons. *Journal of Neuroscience*, 30(4):1551–1559.

Kaneko, M., Stellwagen, D., Malenka, R. C., and Stryker, M. P. (2008). Tumor necrosis factor-α mediates one component of competitive, experience-dependent plasticity in developing visual cortex. *Neuron*, 58(5):673–680.

Keck, T., Keller, G. B., Jacobsen, R. I., Eysel, U. T., Bonhoeffer, T., and Hübener, M. (2013). Synaptic scaling and homeostatic plasticity in the mouse visual cortex in vivo. *Neuron*, 80(2):327–334.

Keck, T., Scheuss, V., Jacobsen, R. I., Wierenga, C. J., Eysel, U. T., Bonhoeffer, T., and Hübener, M. (2011). Loss of sensory input causes rapid structural changes of inhibitory neurons in adult mouse visual cortex. *Neuron*, 71(5):869–882.

Kepecs, A. and Fishell, G. (2014). Interneuron cell types are fit to function. *Nature*, 505(7483):318–326.

Kerlin, A. M., Andermann, M. L., Berezovskii, V. K., and Reid, R. C. (2010). Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron*, 67(5):858–871.

Khan, A. G., Poort, J., Chadwick, A., Blot, A., Sahani, M., Mrsic-Flogel, T. D., and Hofer, S. B. (2018). Distinct learning-induced changes in stimulus selectivity and interactions of gabaergic interneuron classes in visual cortex. *Nature neuroscience*, page 1.

Kim, J. and Alger, B. E. (2010). Reduction in endocannabinoid tone is a homeostatic mechanism for specific inhibitory synapses. *Nature neuroscience*, 13(5):592.

Kirkwood, A., Rioult, M. G., and Bear, M. F. (1996). Experience-dependent modification of synaptic plasticity in visual cortex. *Nature*, 381(6582):526.

Kubota, Y. (2014). Untangling {GABAergic} wiring in the cortical microcircuit. *Current Opinion in Neurobiology*, 26(0):7 – 14. SI: Inhibition: Synapses, Neurons and Circuits.

Kuczewski, N., Porcher, C., and Gaiarsa, J.-L. (2010). Activity-dependent dendritic secretion of brain-derived neurotrophic factor modulates synaptic plasticity. *European Journal of Neuroscience*, 32(8):1239–1244.

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(7468):543–546.

Kullmann, D. and Lamsa, K. (2007). Long-term synaptic plasticity in hippocampal interneurons. *Nature Reviews Neuroscience*, 8:687–699.

Kullmann, D. M. and Lamsa, K. P. (2011). {LTP} and {LTD} in cortical {GABAergic} interneurons: Emerging rules and roles. *Neuropharmacology*, 60(5):712 – 719. Synaptic Plasticity & Interneurons.

Kullmann, D. M., Moreau, A. W., Bakiri, Y., and Nicholson, E. (2012). Plasticity of inhibition. *Neuron*, 75(6):951–62.

Landau, I., Egger, R., Dercksen, V., Oberlaender, M., and Sompolinsky, H. (2016). The impact of structural heterogeneity on excitation-inhibition balance in cortical networks.

Lánskỳ, P. (1984). On approximations of stein's neuronal model. *Journal of theoretical biology*, 107(4):631–647.

Larkum, M. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends in neurosciences*, 36(3):141–151.

Le Roux, N., Cabezas, C., Boehm, U. L., and Poncer, J. C. (2013). Input-specific learning rules at excitatory synapses onto hippocampal parvalbumin-expressing interneurons. *The Journal of physiology*.

Levy, R. B. and Reyes, A. D. (2012). Spatial profile of excitatory and inhibitory synaptic connectivity in mouse primary auditory cortex. *Journal of Neuroscience*, 32(16):5609–5619.

Lewis, S. N. and Harris, K. D. (2014). The neural marketplace: I. general formalism and linear theory. *bioRxiv*, page 013185.

Li, L.-y., Ji, X.-y., Liang, F., Li, Y.-t., Xiao, Z., Tao, H. W., and Zhang, L. I. (2014a). A feedforward inhibitory circuit mediates lateral refinement of sensory representation in upper layer 2/3 of mouse primary auditory cortex. *Journal of Neuroscience*, 34(41):13670–13683.

Li, L.-y., Xiong, X. R., Ibrahim, L. A., Yuan, W., Tao, H. W., and Zhang, L. I. (2014b). Differential receptive field properties of parvalbumin and somatostatin inhibitory neurons in mouse auditory cortex. *Cerebral cortex*, 25(7):1782–1791.

Lillicrap, T. P., Cownden, D., Tweed, D. B., and Akerman, C. J. (2016). Random synaptic feedback weights support error backpropagation for deep learning. *Nature communications*, 7.

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

Lorente de Nó, R. (1949). Cerebral cortex: architecture, intracortical connections, motor projections. *Physiology of the nervous system*, pages 288–330.

Lu, B. (2003). Bdnf and activity-dependent synaptic modulation. *Learning & memory*, 10(2):86–98.

Lu, J.-t., Li, C.-y., Zhao, J.-P., Poo, M.-m., and Zhang, X.-h. (2007). Spike-timing-dependent plasticity of neocortical excitatory synapses on inhibitory interneurons depends on target cell type. *The Journal of Neuroscience*, 27(36):9711–9720.

Maday, S., Twelvetrees, A. E., Moughamian, A. J., and Holzbaur, E. L. (2014). Axonal transport: cargo-specific mechanisms of motility and regulation. *Neuron*, 84(2):292–309.

Maffei, A., Nataraj, K., Nelson, S. B., and Turrigiano, G. G. (2006). Potentiation of cortical inhibition by visual deprivation. *Nature*, 443(7107):81–84.

Maffei, A., Nelson, S. B., and Turrigiano, G. G. (2004). Selective reconfiguration of layer 4 visual cortical circuitry by visual deprivation. *Nat Neurosci*, 7(12):1353–1359.

Maffei, A. and Turrigiano, G. (2008). The age of plasticity: developmental regulation of synaptic plasticity in neocortical microcircuits. *Progress in brain research*, 169:211–223.

Mahanty, N. K. and Sah, P. (1998). Calcium-permeable ampa receptors mediate long-term potentiation in interneurons in the amygdala. *Nature*, 394(6694):683.

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., and Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nature Reviews Neuroscience*, 5(10):793.

Matusita, K. (1988). *Statistical Theory and Data Analysis II*. North-Holland.

McCormick, D. A., Connors, B. W., Lighthall, J. W., and Prince, D. A. (1985). Comparative electrophysiology of pyramidal and sparsely spiny stellate neurons of the neocortex. *Journal of neurophysiology*, 54(4):782–806.

Moore, A. K. and Wehr, M. (2013). Parvalbumin-expressing inhibitory interneurons in auditory cortex are well-tuned for frequency. *Journal of Neuroscience*, 33(34):13713–13723.

Murphy, B. K. and Miller, K. D. (2009). Balanced amplification: a new mechanism of selective amplification of neural activity patterns. *Neuron*, 61(4):635–48.

Newman, E. A. and Zahs, K. R. (1998). Modulation of neuronal activity by glial cells in the retina. *Journal of Neuroscience*, 18(11):4022–4028.

Nissen, W., Szabo, A., Somogyi, J., Somogyi, P., and Lamsa, K. P. (2010). Cell type-specific long-term plasticity at glutamatergic synapses onto hippocampal interneurons expressing either parvalbumin or cb1 cannabinoid receptor. *Journal of Neuroscience*, 30(4):1337–1347.

O'Leary, T., van Rossum, M. C., and Wyllie, D. J. (2010). Homeostasis of intrinsic excitability in hippocampal neurones: dynamics and mechanism of the response to chronic depolarization. *The Journal of physiology*, 588(1):157–170.

Parra, P., Gulyas, A. I., and Miles, R. (1998). How many subtypes of inhibitory cells in the hippocampus? *Neuron*, 20(5):983–993.

Peng, Y.-R., Zeng, S.-Y., Song, H.-L., Li, M.-Y., Yamada, M. K., and Yu, X. (2010). Post-synaptic spiking homeostatically induces cell-autonomous regulation of inhibitory inputs via retrograde signaling. *Journal of Neuroscience*, 30(48):16220–16231.

Peters, A., Paley, S., and Webster, H. d. F. (1976). The fine structure of the nervous system (philadelphia: Saunders).

Pi, H.-J., Hangya, B., Kvitsiani, D., Sanders, J. I., Huang, Z. J., and Kepecs, A. (2013). Cortical interneurons that specialize in disinhibitory control. *Nature*, 503(7477):521–524.

Pouille, F., Marin-Burgin, A., Adesnik, H., Atallah, B. V., and Scanziani, M. (2009). Input normalization by global feedforward inhibition expands cortical dynamic range. *Nature neuroscience*, 12(12):1577.

Pouille, F. and Scanziani, M. (2001). Enforcement of temporal fidelity in pyramidal cells by somatic feed-forward inhibition. *Science*, 293(5532):1159–1163.

Renart, A., Song, P., and Wang, X.-J. (2003). Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron*, 38(3):473–485.

Renshaw, B., Forbes, A., and Morison, B. (1940). Activity of isocortex and hippocampus: electrical studies with micro-electrodes. *Journal of neurophysiology*, 3(1):74–105.

Rodriguez-Grande, B. and Konsman, J.-P. (2018). Gas diffusion in the cns. *Journal of neuroscience research*.

Rothschild, G., Nelken, I., and Mizrahi, A. (2010). Functional organization and population dynamics in the mouse primary auditory cortex. *Nature neuroscience*, 13(3):353.

Roxin, A., Brunel, N., Hansel, D., Mongillo, G., and van Vreeswijk, C. (2011). On the distribution of firing rates in networks of cortical neurons. *Journal of Neuroscience*, 31(45):16217–16226.

Rudy, B., Fishell, G., Lee, S., and Hjerling-Leffler, J. (2011). Three groups of interneurons account for nearly 100% of neocortical gabaergic neurons. *Developmental neurobiology*, 71(1):45–61.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088):533.

Runyan, C. A., Schummers, J., Van Wart, A., Kuhlman, S. J., Wilson, N. R., Huang, Z. J., and Sur, M. (2010). Response features of parvalbumin-expressing interneurons suggest precise roles for subtypes of inhibition in visual cortex. *Neuron*, 67(5):847–857.

Rutherford, L. C., DeWan, A., Lauer, H. M., and Turrigiano, G. G. (1997). Brain-derived neurotrophic factor mediates the activity-dependent regulation of inhibition in neocortical cultures. *The Journal of Neuroscience*, 17(12):4527–4535.

Rutherford, L. C., Nelson, S. B., and Turrigiano, G. G. (1998). Bdnf has opposite effects on the quantal amplitude of pyramidal neuron and interneuron excitatory synapses. *Neuron*, 21(3):521–530.

Sarihi, A., Jiang, B., Komaki, A., Sohya, K., Yanagawa, Y., and Tsumoto, T. (2008). Metabotropic glutamate receptor type 5-dependent long-term potentiation of excitatory synapses on fast-spiking gabaergic neurons in mouse visual cortex. *Journal of Neuroscience*, 28(5):1224–1235.

Savin, C. and Triesch, J. (2014). Emergence of task-dependent representations in working memory circuits. *Frontiers in Computational Neuroscience*, 8(57).

Savin, C., Triesch, J., and Meyer-Hermann, M. (2009). Epileptogenesis due to glia-mediated synaptic scaling. *Journal of The Royal Society Interface*, 6(37):655–668.

Sharma, N., Deppmann, C. D., Harrington, A. W., Hillaire, C. S., Chen, Z.-Y., Lee, F. S., and Ginty, D. D. (2010). Long-distance control of synapse assembly by target-derived ngf. *Neuron*, 67(3):422–434.

Shoham, S., O'Connor, D. H., and Segev, R. (2006). How silent is the brain: is there a "dark matter" problem in neuroscience? *Journal of Comparative Physiology A*, 192(8):777–784.

Silberberg, G. and Markram, H. (2007). Disynaptic inhibition between neocortical pyramidal cells mediated by martinotti cells. *Neuron*, 53(5):735–746.

Slomowitz, E., Styr, B., Vertkin, I., Milshtein-Parush, H., Nelken, I., Slutsky, M., Slutsky, I., and Nelson, S. B. (2015). Interplay between population firing stability and single neuron dynamics in hippocampal networks. *eLife*, 4.

Somogyi, P., Freund, T., and Cowey, A. (1982). The axo-axonic interneuron in the cerebral cortex of the rat, cat and monkey. *Neuroscience*, 7(11):2577–2607.

Stellwagen, D. and Malenka, R. C. (2006). Synaptic scaling mediated by glial tnf-α. *Nature*, 440(7087):1054.

Steriade, M., McCormick, D. A., and Sejnowski, T. J. (1993). Thalamocortical oscillations in the sleeping and aroused brain. *Science*, 262(5134):679–685.

Sun, Y. J., Wu, G. K., Liu, B.-h., Li, P., Zhou, M., Xiao, Z., Tao, H. W., and Zhang, L. I. (2010). Fine-tuning of pre-balanced excitation and inhibition during auditory cortical development. *Nature*, 465(7300):927.

Sweeney, Y. and Clopath, C. (2017). Emergent spatial synaptic structure from diffusive plasticity. *European Journal of Neuroscience*, 45(8):1057–1067.

Sweeney, Y., Hellgren Kotaleski, J., and Hennig, M. H. (2015). A diffusive homeostatic signal maintains neural heterogeneity and responsiveness in cortical networks. *PLoS Comput Biol*, 11(7):e1004389.

Syková, E. and Nicholson, C. (2008). Diffusion in brain extracellular space. *Physiological reviews*, 88(4):1277–1340.

Takesian, A. E., Kotak, V. C., Sharma, N., and Sanes, D. H. (2013). Hearing loss differentially affects thalamic drive to two cortical interneuron subtypes. *Journal of neurophysiology*, 110(4):999–1008.

Tan, A. Y., Brown, B. D., Scholl, B., Mohanty, D., and Priebe, N. J. (2011). Orientation selectivity of synaptic input to neurons in mouse and cat primary visual cortex. *Journal of Neuroscience*, 31(34):12339–12350.

Tao, H., Zhang, L. I., Bi, G.-q., and Poo, M.-m. (2000). Selective presynaptic propagation of long-term potentiation in defined neural networks. *Journal of Neuroscience*, 20(9):3233–3243.

Tetzlaff, T., Helias, M., Einevoll, G. T., and Diesmann, M. (2012). Decorrelation of neural-network activity by inhibitory feedback. *PLoS Comput Biol*, 8(8):e1002596.

Turrigiano, G. (2012). Homeostatic synaptic plasticity: Local and global mechanisms for stabilizing neuronal function. *Cold Spring Harbor Perspectives in Biology*, 4(1).

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(6670):892.

van Vreeswijk, C. and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726.

Vogels, T. P. and Abbott, L. F. (2009). Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nature neuroscience*, 12(4):483–91.

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 (New York, N.Y.)*, 334(6062):1569–73.

Wall, N. R., De La Parra, M., Sorokin, J. M., Taniguchi, H., Huang, Z. J., and Callaway, E. M. (2016). Brain-wide maps of synaptic input to cortical interneurons. *Journal of Neuroscience*, 36(14):4000–4009.

Wehr, M. and Zador, A. M. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature*, 426(6965):442.

Wenner, P. (2011). Mechanisms of gabaergic homeostatic plasticity. *Neural plasticity*, 2011.

Yu, L. M. and Goda, Y. (2009). Dendritic signalling and homeostatic adaptation. *Current opinion in neurobiology*, 19(3):327–335.

Zariwala, H. A., Madisen, L., Ahrens, K. F., Bernard, A., Lein, E. S., Jones, A. R., and Zeng, H. (2011). Visual tuning properties of genetically identified layer 2/3 neuronal types in the primary visual cortex of cre-transgenic mice. *Frontiers in systems neuroscience*, 4:162.

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

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

Zenke, F., Hennequin, G., and Gerstner, W. (2013). Synaptic plasticity in neural networks needs homeostasis with a fast rate detector. *PLoS Comput Biol*, 9(11):e1003330.

Zhang, L. I., Bao, S., and Merzenich, M. M. (2001). Persistent and specific influences of early acoustic environments on primary auditory cortex. *Nature neuroscience*, 4(11):1123.

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

Zweifel, L. S., Kuruvilla, R., and Ginty, D. D. (2005). Functions and mechanisms of retrograde neurotrophin signalling. *Nature Reviews Neuroscience*, 6(8):615.

List of Figures

List of Tables

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 24.06.2018 Owen John Mackwood