Technical University of Munich

TUM School of Life Sciences

Experience-dependent development of functional cortical circuits

Leonidas Maximilian Alban Richter

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

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:

Prof. Dr. Harald Luksch

Prüfende der Dissertation:

1. Prof. Dr. Julijana Gjorgjieva
2. Prof. Dr. Ruben Portugues
3. Prof. Dr. Raoul-Martin Memmesheimer

Die Dissertation wurde am 09.02.2022 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 20.06.2022 angenommen.

Prior publication of parts of this thesis

Some of the results, figures, and tables in this thesis (Chapters 3 and 4) are part of an article entitled *Interneuron subtypes enable independent modulation of excitatory and inhibitory firing rates after sensory deprivation* which I wrote together with my supervisor Julijana Gjorgjieva. The article has been uploaded to the preprint server *bioRxiv* [1]. After submission of this Thesis, a revised version of this paper was published in *Proceedings of the National Academy of Sciences of the United States of America* under the title *A circuit mechanism for independent modulation of excitatory and inhibitory firing rates after sensory deprivation* [2].

The results presented in Section 3.5 are part of an article entitled *Sensory experience inversely regulates feedforward and feedback excitation-inhibition ratio in rodent visual cortex* published in *eLife* and written with Nathaniel J. Miska, Brian A. Cary, Julijana Gjorgjieva, and Gina G. Turrigiano as part of a collaboration with the Turrigiano Lab at Brandeis University [3]. The summary of the modeling insights into the development of cortical circuits in Chapter 2 was partially based on a literature review entitled *Understanding neural circuit development through theory and models* published in *Current Opinions in Neurobiology* which I wrote together with my supervisor Julijana Gjorgjieva [4]. All methods, results, figures, and tables from these articles which are part of this thesis were my contribution to the articles unless specifically mentioned otherwise.

Abstract

Networks of neurons in the brain are shaped by the sensory stimuli that an animal experiences. The changes induced by such experiences refine the connectivity and activity in these neural circuits and serve their adaptation to the structure of the environment. Such adaptation is especially important during development when fundamental aspects of the neural circuits that underlie behavior are formed. During so-called *critical periods* sensory circuits are exceptionally plastic. A perturbation of the sensory experience during such periods can lead to irreversible defects in sensory perception throughout life.

Experimental studies have used sensory deprivations, e.g., monocular deprivation, the closure of one eye for times ranging from hours to weeks, to understand the mechanisms shaping sensory circuits during critical periods. Even two days of monocular deprivation induce remarkable plasticity in both feedforward and recurrent connectivity in the primary visual cortex. Measurements of the activity in the cortex *in vivo* have shown that firing rates of cortical neurons strongly decrease and subsequently recover with a specific time course during ongoing monocular deprivation. Crucially, the timing of suppression and recovery is different in excitatory and inhibitory neurons.

How do these two experimental observations – the reorganization of synaptic connectivity and the regulation of activity – relate to each other? I have addressed this problem using a network model of the cortical circuit that allows measuring the activity for arbitrary synaptic changes associated with monocular deprivation. Through systematic scans of the biologically constrained parameter space of synaptic changes, I investigated how experience-dependent plasticity can regulate the activity in highly recurrent model networks of sensory cortices.

I found that the operating regime of the network, that is, whether or not dynamics are stabilized by recurrent inhibition, has a major impact on how facilitation and suppression of activity distribute in these parameter spaces. In an analytical model of the average rates of the neural populations, I prove that the mechanism underlying the different responses is the paradoxical effect only present in inhibition stabilized networks. This effect is defined by a counterintuitive response of interneurons to external input. Paradoxical is the property that additional excitation injected into an interneuron population leads to a decrease of its activity. In a network with one stabilizing inhibitory population, this effect prevents the time course of suppression and recovery from being different between the excitatory and inhibitory neurons. However, inhibitory interneurons in the cortex are not a single, homogeneous class, but they are distinguished into multiple subtypes with distinct connectivity and activity patterns. The consensus emerged that most of all cortical interneurons can

be classified into one of three subtypes. Interestingly, the interaction of multiple subtypes of interneurons can give rise to circuits in which some subtypes show the paradoxical effect while others do not show it. I studied the overall ability of different network structures with a single and with two subtypes of interneurons to produce the temporally distinct modulation of excitatory and inhibitory firing rates and analytically characterized the conditions on the network structure that enable this.

These results generalize earlier work on the paradoxical effect in inhibition stabilized networks to multidimensional perturbations of connections. Furthermore, these perturbations also target recurrent connections, whose effects have not yet been established with respect to the paradoxical effect.

I furthermore investigated how spike-timing-dependent plasticity shapes the structure of networks with multiple interneuron subtypes. Through ongoing stimulation, strongly connected groups of neurons with similar response properties, so-called Hebbian assemblies, emerge in such networks. Experiments found that such assemblies emerge during development and that their formation is dependent on sensory experience. I included two streams of inputs, representing sensory signals and feedback arising from other brain areas. The competition between the two input streams strongly disturbs the development of Hebbian assemblies, while their cooperation aids it.

Overall, these results point at an intricate interplay of the operating regime, the multiplicity of interneuron subtypes, developmental plasticity, and the structure of the signals impinging onto sensory circuits, and partially uncover the roles of this interplay for the development of functional networks in the cortex.

Zusammenfassung

Netzwerke von Neuronen im Gehirn werden durch die sensorischen Stimuli, die ein Tier erfährt geformt. Die Änderungen durch diese Erfahrungen verfeinern die Struktur der Aktivität und Konnektivität in neuronalen Schaltkreisen so, dass sie an die Struktur der Umwelt angepasst werden. Besonders wichtig sind diese Anpassungen während der Entwicklung des Gehirns, wenn fundamentale Aspekte der neuronalen Schaltkreise entstehen. In sogenannten *kritischen Perioden* sind Netzwerke in sensorischen Kortizes ausnehmend plastisch. Wenn die sensorische Wahrnehmung während solcher Perioden gestört wird, kann es zu irreversiblen Defekten führen, die ein Leben lang anhalten.

Experimentelle Studien haben sensorische Deprivation, z. B. monokulare Deprivation, das Verschließen eines Auges von einigen Stunden bis hin zu Wochen, benutzt, um die Mechanismen, die sensorische Schaltkreise während kritischer Perioden formen, zu untersuchen. Schon zwei Tage monokularer Deprivation führen zu erheblicher Plastizität sowohl in vorwärtsgerichteten als auch in rekurrenten Synapsen im primären visuellen Kortex. Messungen der Aktivität im Kortex *in vivo* haben gezeigt, dass die Feuerraten kortikaler Neurone sich mit einem spezifischen Zeitverlauf während fortbestehender monokularer Deprivation stark verringern und danach auf ihr ursprüngliches Niveau erholen. Insbesondere zeigt sich, dass der Zeitverlauf in exzitatorischen und in inhibitorischen Neuronen unterschiedlich ist.

Wie können wir diese zwei Arten von experimentellen Beobachtungen – die Reorganisation der synaptischen Konnektivität und die Regulation der Aktivität – miteinander verbinden?

Ich habe mich mit diesem Problem mit Hilfe eines Netzwerkmodells des kortikalen Schaltkreises befasst. Das Modell erlaubte mir die Simulation der Aktivität für beliebige synaptische Änderungen, die mit monokularer Deprivation zusammenhängen. Durch systematisches Abtasten von Parameterbereichen, die durch biologische Kriterien umgrenzt sind, habe ich untersucht, wie erfahrungsabhängige Plastizität die Aktivität in rekurrenten Netzwerkmodellen von sensorischen Kortices reguliert.

Ich fand heraus, dass das Operationsregime des Netzwerks, das heißt, ob die Dynamik des Netzwerks rekurrente Inhibition zur Stabilität benötigt oder nicht, zentral bestimmt, wie Verstärkung und Unterdrückung der Aktivität in diesen Parameterbreichen verteilt sind. In einem analytischen Modell der mittleren Feuerraten der neuronalen Populationen konnte ich zeigen, dass der Mechanismus, der den verschiedenen Antworteigenschaften in den verschiedenen Operationsregimen zugrundeliegt, der *paradoxe Effekt* ist, der nur in inhibitions-stabilisierten Netzwerken auftritt. Dieser Effekt ist durch eine kontraintuitive Antwort der Interneurone auf ein externes Eingangssignal definiert. Paradox ist die Eigenschaft, dass zusätzliche Exzi-

tation, die in eine Population von Interneuronen injiziert wird, zu einer Verringerung der Aktivität dieser Population führt. In einem Netzwerk mit einer einzigen stabilisierenden inhibitorischen Population verhindert dieser Effekt, dass der Zeitablauf der Unterdrückung und Wiederherstellung der Feuerraten sich zwischen exzitatorischen und inhibitorischen Neuronen unterscheidet.

Allerdings sind inhibitorische Interneurone im Kortex nicht eine einzige homogene Klasse, sondern es gibt mehrere verschiedene Suptypen von Interneuronen mit unterschiedlicher Konnektivität und Aktivität. Als Konsens hat sich eine Einteilung in drei Subtypen etabliert. Interessanterweise kann die Wechselwirkung zwischen mehreren Subtypen von Interneuronen zu Schaltkreisen führen, in denen manche Subtypen den paradoxen Effekt zeigen und andere nicht. Ich habe die Gesamtfähigkeit, den unterschiedlichen Zeitablauf der Modulation der Feuerraten in exzitatorischen und inhibitorischen Neuronen herzustellen, in verschiedenen Netzwerkstrukturen mit einer einzigen und mit zwei Subtypen von Interneuronen studiert und analytisch die Bedingungen, die eine unterschiedliche Modulation erlauben, charakterisiert.

Diese Ergebnisse verallgemeinern bisherige Erkenntnisse zum paradoxen Effekt in inhibitions-stabilisierten Netzwerken hin zu mehrdimensionalen Modifikationen der Verbindungen im Netzwerk. Insbesondere habe ich auch Modifikationen von rekurrenten Verbindungen untersucht, deren Rolle im Zusammenhang mit dem paradoxen Effekt bisher unbekannt ist.

Ich habe auch untersucht, wie synaptische Plastizität, die vom Zeitpunkt einzelner Aktionspotenziale abhängt (*spike-timing-dependent plasticity*), die Struktur von Netzwerken mit mehreren Subtypen von Interneuronen beeinflusst und formt. Durch anhaltende spezifische Stimulation können sich Gruppen von stark verbundenen Neuronen mit ähnlichen Antworteigenschaften in einem Netzwerk bilden, sogenannte *Hebbian Assemblies* (Hebb'sche Verbünde). In Experimenten fand man heraus, dass sich solche Verbünde während der Entwicklung des Gehirns bilden und dass ihre Entstehung von sensorischer Erfahrung abhängig ist. Ich habe hierbei zwei verschiedene Eingangssignale, die sensorische Signale und Rückkopplungen aus höheren Hirnarealen repräsentieren, berücksichtigt. Konkurrenz zwischen diesen beiden Eingangssignalen stört die Entwicklung von Hebb'schen Verbünden, wohingegen ihre Kooperation diese Entwicklung unterstützt.

Zusammengefasst zeigen diese Ergebnisse ein komplexes Zusammenspiel des Operationsregimes, der Vielfalt von Interneuronen, der Plastizität in der Hirnentwicklung und der Struktur der Signale, die auf Schaltkreise in sensorischen Kortizes während der Entwicklung einwirken. Diese Arbeit trägt dazu bei, die Rolle dieses Zusammenspiels in der Entwicklung funktioneller Schaltkreise im Kortex besser zu verstehen.

Contents

List of Figures

List of Tables

Acronyms

ISN	Inhibition Stabilized Network
LGN	Lateral Geniculate Nucleus in the thalamus
MD	Monocular Deprivation
PV	Parvalbumin-Positive Interneurons
S1	Primary Somatosensory Cortex
STDP	Spike-Timing-Dependent Plasticity
SST	Somatostatin-Positive Interneurons
V1	Primary Visual Cortex
V1m (V1b)	Monocular (Binocular) Portion of V1
VIP	Vasoactive-Intestinal-Peptide-Positive Interneurons
WD	Whisker Deprivation

1 Introduction

τὸ γὰρ αὐτὸ νοεῖν ἐστίν τε καὶ εἶναι
because the same are thinking and being
Parmenides

φύσις κρύπτεσθαι φιλεῖ
nature loves to hide itself
Heraclitus

How do physical systems, as us, acquire knowledge and understanding about the physical structure of the world and themselves? Starting from this problem, central questions of epistemology become questions about the brain, as the part of our body that is first and foremost responsible for this acquisition of knowledge. The brain is involved in this inquiry in a double sense of being the target of the scientific investigation and the physical structure that primarily undertakes it.

Computations in neural circuits, networks of interconnected neurons, are performed through the dynamics of the activities of the neurons in the network [5, 6]. Circuits involved in sensory processing must transform input signals from the senses into dynamical signals that trigger adequate behavioral responses adapted to the situation. To execute computations properly, that is, appropriate for the ecological demands of the animal, the network structure underlying these dynamics, most prominently the synaptic connectivity is shaped during an extended period of development. In mammalian cortical circuits, this development is a multistage process shaped by a manifold of mechanisms, which are invoked either in parallel or in specific sequences [7, 8]. Circuits for sensory processing in the cortex are fundamentally shaped by experience, particularly during later stages of development [9, 10]. During so-called *critical periods*, these developing cortical circuits are particularly susceptible to sensory experience of the environment [11]. The experience-dependence of development is a fascinating gateway into investigating how physical systems can acquire knowledge about their environment. The signals that shall be adequately processed are the same signals that shape the structure of the processing circuits in the first place.

Since the seminal work of Wiesel and Hubel, perturbations of an animal's sensory inputs by *sensory deprivation* have been used to experimentally investigate the mechanisms underlying the high susceptibility of circuits to experience [12, 13]. Sensory deprivation includes a varied set of manipulations that disrupt the sensory experience of an animal during development, either by manipulating the sensory organ or the environment. In vision, for example, possible ways of sensory deprivation

are the sustained (in rodents, usually from one day up to ten days), closure of one eye, or both eyes (monocular and binocular deprivation, respectively). Alternatively, young animals are also reared in the dark, removing most of the visual experience even though their eyes are intact.

In my research, I have addressed how the circuit changes induced by sensory deprivation shape the activity in network models of sensory cortices. I have used networks of excitatory and inhibitory model neurons whose recurrently generated activity is a good model of the activity of cortical neurons [14, 15] (see also [16]). I investigated how such circuits respond to sensory deprivation-induced changes and, in a second step, how they are shaped by activity-dependent plasticity to understand the mechanisms that shape circuits in sensory cortices during development in an experience-dependent manner.

1.1 Understanding neural circuit development through theory and models

Development is a universal feature of multicellular organisms. Cell division and differentiation underly growth and the emergence of many different tissues. During the developmental increase in volume and differentiation into areas in the brain, we find an extensive developmental adaptation of the circuitry performing specific computations.

Circuits in the adult brain are of enormous complexity, and we barely grasp how they perform specific computations. Investigating these circuits in their developmental history, at first sight, seems to further burden us with multiple stages of this exceptionally complex system. So why should one study developing neural circuits? We may find that the developmental perspective can reduce the complexity we need to consider. In their developmental history, young animals and their nervous systems acquire specific capabilities and functions during specific developmental phases (Fig 1.1). Many of these skills are acquired sequentially. For example, the topographic order of inputs to the visual cortex emerges before eye-opening, followed by the development of receptive fields and the selectivity of visual cortical neurons for bars with a specific orientation during a period just before and after eye-opening [7, 17]. In parallel to the emergence of this orientation selectivity, neurons initially only driven by the contralateral eye start to be also innervated by the ipsilateral eye (binocularity). However, the matching of the orientation selectivity of the inputs from the two eyes only emerges after this binocularity of inputs has developed. Only then fully binocular vision is established.

When we find a capability of the animal, or a specific computational property of the corresponding neural circuit, to emerge at some point in time, we can hypothesize that the best candidate circuit elements that implement these functions are the elements most substantially changing at that time. Therefore, through studying development, we may constrain our search for the physiological elements that implement a computation. Related to this is the idea that understanding

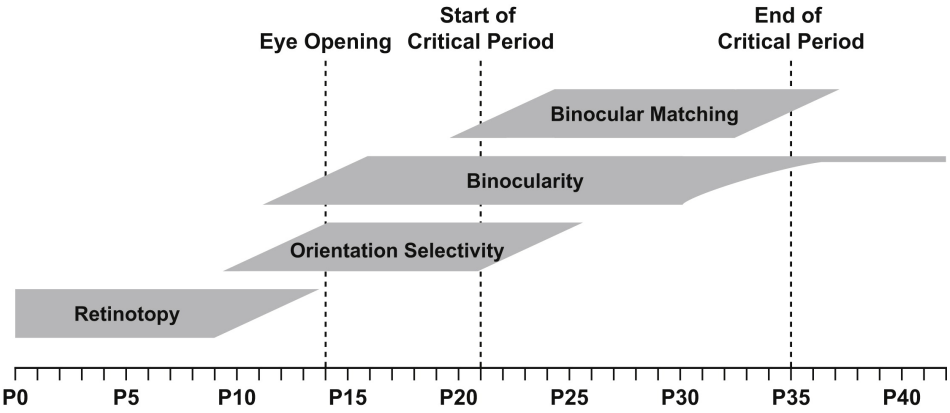

Figure 1.1: Development of visual processing in rodents. Timeline of events in the development of visual processing in rodents. Before eye-opening (around postnatal day 14), development is driven by genetic factors and spontaneously generated activity. One week after eye-opening, the critical period of experience-dependent development starts. Figure from [7].

the function of a neural system is best approached by understanding the learning and plasticity mechanisms underlying its function instead of studying a specific circuit with all its biological detail. The high variability of the circuits in different individuals, even within an individual over time, can make it seem impossible to understand how the structure of the circuit underlies its function [18]. Crucially, differences between circuits in specific parameters may be compensatory responses to changes in other parameters induced by homeostatic plasticity mechanisms, such that the combination of multiple differences in the circuitry produces functionally similar circuits [19]. By understanding the mechanisms that produce functionally similar circuits with a different microscopic structure, we may be able to understand the specific structure of the variability and the overarching principle that abstracts from the microscopic variability. As an engineering example, we may also think of machine learning approaches. The engineer often primarily controls the network's learning mechanisms and global properties, but not the detailed connectivity that emerges to solve the task.

Theoretical models are a valuable tool to investigate mechanisms in neural circuits in a way that abstracts from the biological variability to understand the principle behind a computation in a neural circuit. The mathematical formulation of these models forces us to be precise and to set up models that are complete and self-consistent [16]. A further benefit of mathematical models is the fast turnover of hypotheses. At its best, a computational model can generate, and potentially confirm or reject, hypotheses at a pace inaccessible to experimental investigation. The goal of modeling is to gain insight into a process and extract a mechanism hidden within the system's complexity. Such a model only includes relevant variables of the process investigated instead of being a one-to-one mapping of a specific circuit. The challenge is to include the necessary biological detail without importing confounding factors.

Whether a circuit element is a relevant mechanism or a confounding factor depends on the question at hand, that is, in models investigating different circuit levels, different circuit elements become the relevant mechanistic or causal components. Theoretical models on multiple levels have generated many insights into the development of neural circuits [4]. Using a detailed single neuron model with elaborate morphology, my colleague Jan Kirchner studied the development of the organization of synapses on individual dendrites and dendritic branches, finding specific conditions for the emergence of clusters of synapses with similar stimulus preferences [20]. In a network of biophysically realistic model neurons, yet abstracting from the complex morphology, it was shown that the development of the ion channel composition of single neurons changes the mode of information transmission in a network. This mode changes from a reliable and selective transmission of large and slow fluctuations of an incoming current to selective transmission of fast fluctuations [21]. On a yet higher level of abstraction, the physiology of neurons may be reduced to just a rate input-output relation. On this level, it is possible to study the development of large and complex populations. In classical work with such a model, Ken Miller investigated the emergence of different domains in the visual cortex that receive inputs just from one of the two eyes and the specific distribution of these domains [22].

In my research, I studied the activity emerging in large networks of simplified spiking neurons. The model abstracts from the physiology of the neurons, such as elaborate dendrites and interacting ion channels, and reduces each neuron to a single variable, the membrane potential. However, the spiking behavior is maintained, meaning that the neurons exchange signals through distinctly timed pulses. The spike is a causal element of model networks of spiking neurons, even if we describe the networks on the level of rates [23]. With such simplified neurons, the essential variables of the system are the parameters of the connectivity, which determine the activity structure and the dynamics of the network. Through my modeling, I investigated how changes of the connectivity investigated during sensory deprivation in previous work shape the activity in networks and how these changes relate to activity changes characterized *in vivo*.

1.2 Thesis overview

In Chapter 2, I present the relevant biological and theoretical background of my research from the neuroscientific literature. My work rests on experimental research investigating the connectivity in cortical circuits and how it is shaped during development, specifically how it is influenced by early sensory experience. As the crucial paradigm, I review relevant results from sensory deprivation experiments. I then also discuss previous modeling work which investigated sensory deprivation using multiple network architectures. Finally, I introduce the basics of the type of networks I used to further our understanding of the organization processes during experience-dependent development.

In Chapter 3, I investigate how experimentally observed synaptic changes induced by two days of *monocular deprivation* (MD) affect the activity in network models composed of excitatory and inhibitory neurons. Two days of MD is a comparatively brief period. It is not long enough to observe complete homeostasis of activity to the changed input after eyelid closure, yet long enough to observe profound effects on the activity and connectivity of the cortical circuit. I am specifically interested in how the observed synaptic changes relate to the experimentally observed regulation of the cortical activity during the first two days of MD. I use parameter scans of the potential synaptic changes and extract the whole landscape of firing rate changes. For these scans, I define ranges of parameters associated with MD-induced synaptic changes. I then vary multiple parameters in parallel on a grid spanning the thus defined parameter spaces. Thereby, I extract firing rates for all potential synaptic changes within these ranges. My research rests on simulations of large-scale networks composed of spiking neurons and on mathematical analysis of simplified dynamical models representing the average firing rates of the excitatory and inhibitory populations. This research shows a strong dependence of the possible regulation of firing rates on the coupling strengths in the network and specifically on the stabilizing role of inhibition in the circuit. It will become apparent that a network of just one excitatory and one inhibitory population is problematic if we want to explain how synaptic changes in the cortical circuit lead to the regulation of firing rates observed in experiments *in vivo*.

In Chapter 4, I thus expand the model to include an additional subtype of interneurons as a direct extension of the model investigated in Chapter 3. Cortical circuits are composed of multiple interneuron subtypes, all of which have specific connectivity and specific functional roles. The main classes of interneurons are comprised of three subtypes which are major contributors to local cortical processing [24, 25]. My modeling shows that the inclusion of two subtypes of interneurons is a viable solution to explain how experimentally observed synaptic changes during brief MD underlie the regulation of firing rates in the animal. I find that the additional subtype of interneurons directly operates on the mechanism that hinders the explanation of the firing rate regulation with just one subtype of interneurons, providing a mechanistic understanding of the necessary elements to model the firing rates *in vivo*.

In Chapter 5, I investigate how biologically inspired plasticity mechanisms shape such a circuit that includes two subtypes of interneurons. For this inquiry, I use a common framework for learning and development of neural networks with functionally specific connectivity, namely the training and emergence of *Hebbian assemblies*. Such assemblies are groups of neurons that consistently coactivate and strongly connect with each other, proposed by Hebb as the fundamental units of neural computation [26]. The strong connectivity among neurons that coactivate emerges during development in an experience-dependent manner [27,28]. A prevalent proposal is that they emerge from Hebbian forms of plasticity (recently reviewed in [29]), colloquially referred to as *fire together–wire together* (after Carla Shatz). In models with just one subtype of interneurons, this training has been successfully

investigated in the literature [30–32]. I extend the model to include biologically inspired plasticity mechanisms and diversity of interneuron subtypes in large-scale recurrent spiking networks. I investigate the emergence of Hebbian assemblies and their consequences for the circuit dynamics under multiple training paradigms.

Finally, I discuss the results in their broader perspective, scrutinize the limitations of my approaches, and propose future research directions in Chapter 6.

2 Biological and theoretical foundations

Experience-dependent development of cortical circuits plays a significant role in the refinement of connectivity and processing capabilities of cortical circuits. These refinements are driven by activity-dependent plasticity of synaptic connections. In this section, I first introduce the central elements of the visual processing machinery in mammalian brains (Sec. 2.1). Before visual signals arrive in the primary visual cortex (V1), they go through multiple stages of neural processing (Sec. 2.1.1). Within V1, neurons receiving these signals form complex circuits, which, in conjunction with the preprocessing in the previous stages, shape their functional response properties, e.g., the selectivity of single cells to respond to the presence of specific features in the visual scene (Sec. 2.1.2). These functional properties and the circuits underlying them are not fully established at birth but form during an extended period of postnatal development that includes multiple stages (Fig. 1.1). At the earliest stages, before visual experience, the development is driven by intrinsic, genetic factors and spontaneously generated activity. After these early stages, some of the response properties of cells found in the adult are established. Thus, to understand the context and relevance of the circuit development during the later, experience-dependent stages, which were the focus of my research, I will introduce the structures that already have formed through the early stages of development (Sec. 2.1.3).

In Sec. 2.2, I introduce the mechanisms acting during experience-dependent stages of development, most notably in so-called *critical periods*, during which cortical circuits are particularly susceptible to perturbations of their sensory environment. To study the mechanisms and the functional role of these periods of high susceptibility to the sensory experience of the animal, a rich strand of research used different paradigms of *sensory deprivation*. I give a brief overview of these paradigms and some of the insights they produced. My research was primarily concerned with the effects of monocular deprivation (MD), specifically, its effects on the circuitry in the monocular portion of the visual cortex (V1m), which is exclusively driven by one of the eyes. I outline the experimental foundations underlying my research in more detail in Sec. 2.3.

The study of sensory deprivation has also attracted attention from theory and modeling. I introduce important insights from the literature in this field in Sec. 2.4. In Sec. 2.5 I introduce general aspects of the models which I used to study the effects of sensory deprivation. I studied large-scale spiking networks with excitation and inhibition that balance each other dynamically, as well as a population-rate description of the average firing rates in these networks. Specifically, for the spiking networks the notion of the *balanced state* and its dynamical consequences is central (Sec. 2.5.1). I have used simplified rate models to analyze how the different neuron

types interact on a population level. I introduce the notion and analytical description of *inhibitory stabilization* of the excitatory activity in such population-rates models composed of a single excitatory and inhibitory population (Sec. 2.5.2), and how this inhibitory stabilization behaves with multiple subtypes of interneurons (Sec. 2.5.3). Finally, I give a brief outlook for the relations between inhibitory stabilization and deprivation-induced plasticity in the developing cortex (Sec. 2.5.4).

2.1 Structure and development of the mammalian visual system

2.1.1 The visual pathway

Many areas of the mammalian brain are involved in the processing of visual information. These different areas have rich interconnectivity of feedforward, feedback, and local recurrent connections [33]. The principal stream of visual information flow from the outside world to the neocortex can be reduced to a canonical structure of feedforward signal transmission. Starting at the retina, where visual signals enter the nervous system, this stream goes through the lateral geniculate nucleus (LGN) in the thalamus to V1 (Fig. 2.1).

At its entrance into the visual system, active optical components transform and select visual information, for example, by choosing the depth and localization of optical focus through a muscle-controlled flexible lens and eye movements. This initial transformation of, and selection from, all the visual signals in the surroundings shapes the statistics of the visual inputs the brain receives. These statistics shape cortical circuits, which are especially plastic during development [4]. However, to give rise to properly functioning circuits, the plasticity and input statistics must accord with one another [34], otherwise dysfunctional circuit structure and responses can result [35].

The retina is the first neural stage of visual processing. It comprises a diverse set of different, specialized cell types embedded in an elaborate circuitry [36]. The output of the retina emanates from *retinal ganglion cells*. Retinal ganglion cells have particular spatiotemporal response properties, such as selective firing in response presence of light (ON-cells) or its absence (OFF-cells). Such selectivity often pairs with a temporal component, where some cell types have transient responses that signal changes of the luminance conditions in their field of view, or cells with sustained responses that signal overall brightness [37].

The retinal output is transmitted through the optic nerve to the thalamus, specifically the *lateral geniculate nucleus* (LGN). Before the thalamus, the nerves emanating from either eye cross over to the respective opposite brain hemisphere (the contralateral hemisphere) in the *optic chiasm*. In rodents, most of the fibers from an eye cross to the contralateral hemisphere (Fig. 2.1, left).

Excitatory cells in the LGN, the *relay cells*, transmit the visual signal further to V1. The area V1 is a part of the mammalian neocortex, which is, in turn,

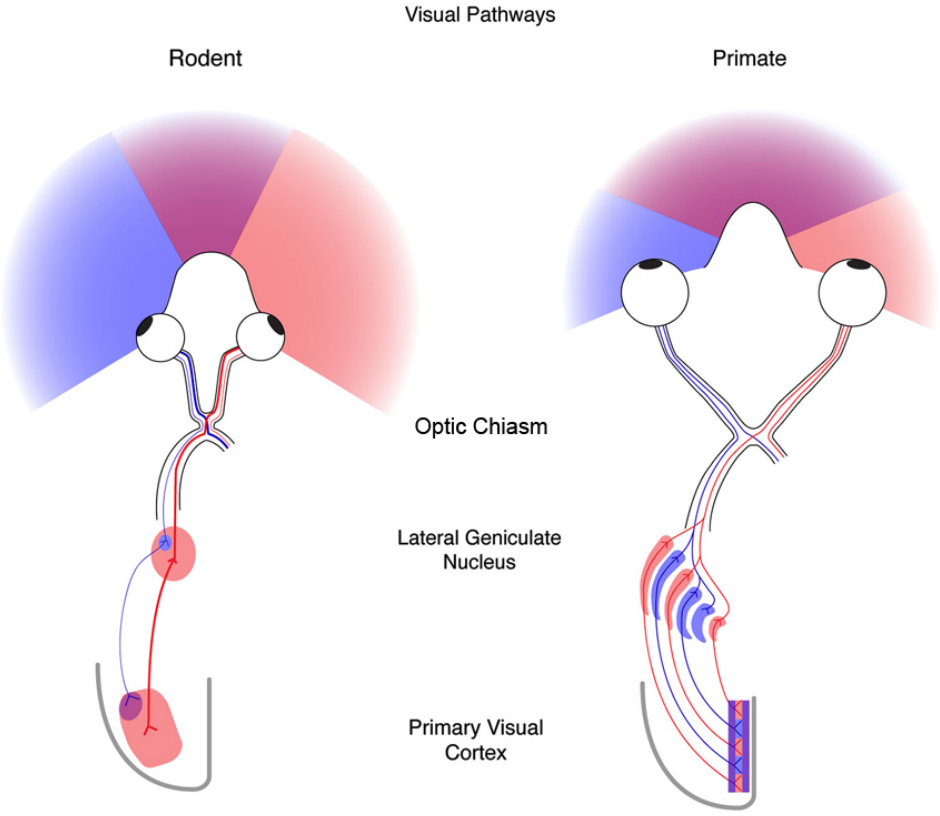

Figure 2.1: The mammalian visual pathway. Structure and comparison between rodents (left) and primates (right). Regions of monocular vision of the left (blue) and right eye (red), and the region of binocular vision (purple). Pathway in one hemisphere of the ipsilateral eye (blue), and the contralateral eye (red). Modified from [39].

a part of the cerebral cortex. A six-layered arrangement of neural cell bodies is characteristic for the neocortex. As is custom, I will simply use "cortex" to reference the six-layered neocortex[1]. The LGN inputs arriving in V1 target the middle layer 4 as the primary input layer and, to a lesser extent, the deep layer 6 [38]. These projections are excitatory but synapse onto both excitatory and inhibitory cells in their target location. From the primary input layers 4 and 6, the visual signal is sent to the upper layer 2/3 and then to higher-order cortices (from layer 4) or subcortical targets (from the deep layers 5 and 6). The latter also includes feedback projections involved in recursive thalamo-cortical/cortico-thalamic connectivity.

Due to the crossover in the optic chiasm, the largest fraction of the cells in one hemisphere of the LGN in rodents, as well as V1, primarily receive signals from the contralateral eye (Fig. 2.1, red). A minority of the fibers from an eye remain on

[1]This terminology is not unique, since one should distinguish, first, the cerebral cortex from the cerebellar cortex, and, second, within the cerebral cortex the six-layered neocortex from the three-layered paleocortex and the four-layered archicortex.

the same side of the brain (seen from LGN or V1, this is then the ipsilateral eye, Fig. 2.1, blue). While the fibers from the different eyes target separate portions of the LGN, their signals partially mix in V1. Contralateral fibers target cells in all of V1, whereas the fibers from the ipsilateral eye only target cells in a small portion of V1. The region in V1 where signals from both sides arrive is the *binocular* primary visual cortex (V1b), while the larger part that only receives direct signals from the contralateral eye comprises the monocular primary visual cortex (V1m). Single cells in V1b receive inputs from both eyes, where usually one of the two is dominant. The quantitative aspects of the binocular vs. monocular portions of V1 depend on the species. While in rodents, we find a large monocular fraction, in primates and carnivores, most (if not all) of the V1 receives inputs from both eyes. However, in these species, so-called *ocular dominance columns* form. These are domains in which cells are primarily responsive to one of the eyes. The distribution of these domains has a regular structure [22] and is highly sensitive to the experience of the animal, especially to monocular deprivation, the sustained closure of one eye [9] (for more details see Sections 2.2 and 2.3).

Ocular dominance is one among many tuning properties of neurons in V1, expressing a selectivity of a cell to respond to signals to either one of the eyes. Visual cortical neurons have a localized receptive field describing the part of the visual field from which a response of the cells can be elicited, but furthermore also defines the tuning to more specific characteristics of visual stimuli [40]. Hubel and Wiesel found that cells in the cat V1 strongly responded to edges of some orientation or another [40]. Since then, the response to such edges has been investigated extensively, using drifting gratings as one of the most widespread patterns used for stimulation. These are circular patches of sinusoidally modulated brightness, building a pattern of parallel edges. From the properties of these drifting gratings, basic tuning properties of neurons in V1 can be analyzed, like orientation selectivity, but furthermore direction selectivity (for moving drifting gratings), spatial-frequency tuning, i.e., the selectivity for the speed of the luminance modulation in the drifting grating, and size tuning, or surround suppression, the selectivity for the total radius of the drifting grating with some preferred size. These tuning properties are shaped by the synaptic connections in feedforward and recurrent pathways, and are modulated by long-range horizontal inputs and higher-order feedback.

Dependent on the species, tuning properties such as the orientation tuning and the ocular dominance have an orderly distribution within the cortical sheet, called *maps* [41]. Probably the most basic and best-preserved of maps (across species) is the retinotopic map. The positions of the localized receptive fields determine the selectivity of neurons to be driven only from specific locations in the visual field. The neighborhood-relations between these receptive fields align with the neighborhood-relations of the cells: cells that are nearby to each other in the cortex also receive visual inputs from nearby positions. In primates and carnivores, but not in rodents, also orientation selectivity of V1 cells is arranged on an ordered map, where cells in neighboring positions have similar orientation preferences (except around so-called pinwheels [41]). The distribution of single cells' selectivities influences the statistics

of the inputs cortical cells receive from their closest neighbors, with which they form highly recurrent local circuits.

2.1.2 Circuits in the primary visual cortex

The cortical sheet is densely packed with neurons, the largest fraction of which are excitatory. Pyramidal cells are the most common morphological type of excitatory neurons in the cortex, especially in V1 [42]. The excitatory pyramidal neurons within one layer highly interconnect through synaptic contacts with thousands of cells [43]. They receive synapses from both local excitatory and local interneurons at their basal and apical dendrites [24, 44].

The large amount of connections suggests a statistical view, that is, to assume that connections are determined by chance and to focus on the study of expectation values and variability in ensembles of random networks [45]. However, local circuits in the cortex are not purely random networks in which the pairs of neurons connecting are chosen by chance. Rather, certain motifs of connections, especially bidirectional connections between pairs of cells, where a synapse from either of the neurons targets the other, are much more likely found in cortex than is expected by chance [46]. This over-representation of bidirectional connections suggests that neurons in the cortex form *Hebbian assemblies* [47]. Hebb proposed that neurons with correlated activity could be more strongly connected than uncorrelated neurons, with groups of such correlated neurons ultimately forming highly interconnected groups that serve as the primary computational units in the brain [26]. The emergence of such groups could be driven by plasticity of connections in the network that increases synaptic weights based on correlations, referred to as *Hebbian plasticity*. Using interconnected groups of neurons as the computational units can be seen as a counterproposal to the prevalent view of feedforward processing [48]. In the feedforward view, the function of cells is primarily determined through their tuning properties, the receptive field, essentially seeing them as feature-detectors that progressively become selective for ever more complex features along the processing stream, ultimately giving rise to complex computations such as the categorization of complex scenes [49]. The assembly view proposes that the selectivity of single neurons is flexible and context-dependent, shaped by the recurrent interactions within an assembly and between the assemblies [48]. Although this account emphasizes the recurrent processes, it does not contradict the fact that cortical neurons have certain feature-detector properties expressed through their tuning. Beyond the purely statistical over-representation of bidirectional connections, it was indeed found that neurons in V1, which have highly correlated receptive fields, do connect with higher weight and higher probability than neurons whose receptive fields are uncorrelated [50, 51].

Even though layer 4 is the primary input layer for visual signals from the LGN, recurrent connections from excitatory neurons within cortex account for the major fraction of the excitation which an excitatory neuron in layer 4 receives [52]. Such strong recurrent excitation, likely to be present in many parts of the cortex, constitutes a strong positive feedback loop that calls for negative feedback to prevent

runaway activity. A highly diverse set of inhibitory interneurons in the cortex appears to serve this function.

The most common classification of cortical interneurons into a few non-overlapping classes that include most of the interneurons present in a circuit is based on molecular markers. Rudy and colleagues could show that three markers can cover almost all of the cortical interneurons [25]. These markers are parvalbumin (PV), somatostatin (SST), and 5HT3aR, a serotonin-receptor. The latter class is frequently further distinguished into vasoactive-intestinal peptide (VIP) positive interneurons and into non-VIP interneurons [24].

All of the interneuron classes in this general scheme are heterogeneous, for example, concerning the morphology of the cells or their electrophysiological properties [53], but also show similarities that indicate a functional significance of the categorization through molecular markers. For example, multiple morphologically defined interneuron types that express PV, like basket cells and chandelier cells, are classified as fast-spiking (FS) type based on their electrophysiological properties [24]. The fast-spiking property has two sides. First, the spike in itself is indeed faster than, for example, in an excitatory neuron electrophysiologically identified as regular spiking. In extracellular recordings, this difference of the properties of single spikes can be used to classify neurons and, for example, distinguish putative excitatory from putative PV neurons [54]. The second aspect of the fast-spiking property is that these cells can maintain a high firing rate without adaptation [24].

Somatostatin-positive (SST) interneurons are morphologically divided into the Martinotti- and non-Martinotti-cells [24]. The Martinotti-cells are found in layers 2/3, 5, and 6 and have a very distinct morphology with axonal processes targeting the most upper layers (primarily layer 1) in the visual cortex from cells with somata as deep as layer 5 or 6 [55]. In contrast, the non-Martinotti-cells mostly reside in layer 4 and have the axonal projections largely restricted to this layer. These two cell types are also electrophysiologically distinct from each other. While the Martinotti-cells show heterogeneous spiking behaviors, e.g., bursting or regular spiking, the non-Martinotti-cells show primarily fast-spiking behavior, although with some adaptation of the spiking frequency (called quasi-FS).

However, also some remarkable similarities of neurons within either of these classes appear [24]. It is robust, for example, that the parvalbumin-positive (PV) interneurons synapse onto the soma and the axon-initial segment of an excitatory cell, meaning that they directly control the output of this cell. SST-interneurons, on the other hand, preferentially target the dendritic tree of an excitatory cell, thus controlling the input to the excitatory neurons.

Pfeffer and colleagues published an experimental analysis of the connectivity of the interneuron subtypes defined by these different markers and uncovered its general patterns [56]. PV-interneurons strongly inhibit other PV-interneurons, along with excitatory neurons, but they do not inhibit SST- or VIP-interneurons (Fig. 2.2). They densely connect to nearby excitatory neurons, likely without specificity of the selectivity of the excitatory neurons [57]. This makes the PV-interneurons well suited to dynamically stabilize excitation, also suggested by the fact that the

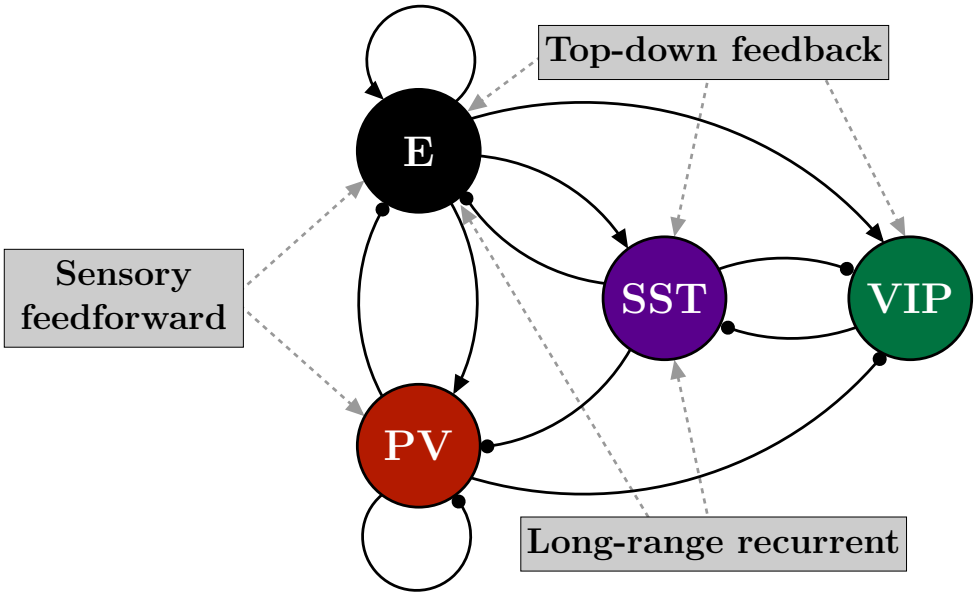

Figure 2.2: Connectivity of interneuron classes in the primary visual cortex. Excitatory cells (E, black), and three major classes of interneurons identified by molecular markers: PV, SST and VIP. Presence of connections determined following [56]. Inputs to the neuron classes can broadly be classified into feedforward sensory inputs to E and PV, e.g. from LGN to layer 4 [38], into top-down feedback inputs from higher order cortical centers to E, SST and VIP [61], e.g. locomotion related inputs to VIP [62], and into long-range inputs from within the same cortical area to E and SST [59].

PV-interneurons' output strength onto an excitatory neuron is correlated with the excitation this cell receives [58]. SST-interneurons do not inhibit other SST-interneurons, but they do inhibit all other types of interneurons, PV and VIP, as well as excitatory neurons. Thus, SST-interneurons seem to distribute inhibition throughout the circuit, which seems consistent with their putative role in surround suppression of excitatory and PV responses [59, 60]. VIP-interneurons appear highly specialized in inhibition of SST-interneurons.

PV- and SST-interneurons play different functional roles. Optogenetic activation of either of the interneuron subtypes suppresses the activity of excitatory neurons. However, while SST-activation leaves orientation tuning of the excitatory neurons unchanged, activation of PV-interneurons leads to more sharply tuned responses [63]. Underlying this sharpening is a specific form of the modulation of tuning curves. Subtraction from activity, homogeneous for all activity values, leads to a sharpening of the tuning (through an iceberg effect), whereas division that scales with the level of activity leaves the overall tuning sharpness unaffected. The specific form of this modulation by different interneuron subtypes, specifically PV and SST, elicited considerable debate [63–65], where the consensus emerged that it is highly dependent

on the dynamical state and the details of the experimental setup, e.g., the depth of anesthesia and the laser intensity used in optogenetic stimulation [66].

The different interneuron subtypes receive their inputs from different sources. PV-interneurons receive strong feedforward drive, for PV-interneurons in layer 4 this drive comes from the LGN [38]. VIP-interneurons are driven specifically by higher-order feedback. The opposite preference of PV- and VIP interneurons to be driven by bottom-up vs. top-down inputs was recently used to determine the hierarchy of cortical areas in sensory processing [67]. Through their inhibition of SST-interneurons, VIP-interneurons govern contextual disinhibition during locomotion [62]. SST-interneurons receive negligible inputs from the feedforward pathway, but also receive higher-order inputs [61, 68], and inputs from higher-order thalamic nuclei [69].

Although some insights have been gained, research is still ongoing on how the different properties of these interneuron subtypes, e.g., their connectivity and input sources, can give rise to the functional roles they play. Modeling can specifically aid in uncovering how these different aspects interact in highly recurrent networks.

2.1.3 Development of visual circuits, activity and processing

In the earliest embryonic development, many interesting processes take place, like the formation of the neural tube and its differentiation into different regions or the emergence of axons and dendrites from neural progenitor cells [70]. However, I will focus on the refinement of visual circuits after neurons are established, have their somata in place, and have differentiated axons and dendrites that already connect neurons into networks. The refinement of the circuits appears shortly before and, most prominently, after birth.

The eyes of a rodent remain closed until roughly two weeks after birth. During this time, circuit refinement takes place based on molecular gradients (such as the Eph/*ephrin* system), and spontaneously generated activity [17, 71]. In the initial stages, molecular gradients that signal the topographic axes of the cortical sheet set up the retinotopic maps in the LGN and V1 by guiding axons to their target locations. In parallel, spontaneously generated activity in the retina also influences the formation of these maps, as well as receptive field refinement of the single neurons [8]. Retinal spontaneous activity is highly structured in so-called retinal waves whose statistical and generating mechanisms progress through multiple stages that start before birth and last until after eye-opening. The input generated by these waves propagates throughout the visual system, and likely influences connectivity refinement along the whole visual pathway [72]. On the level of the cortex, both spontaneous, intrinsically generated activity that spreads over large areas, as well as localized events that are driven by the retinal input, emerge [73]. The spontaneously generated activity of large bursts that synchronizes cells over large areas in the cortex gradually vanishes and is indistinguishable from mature activity by the time of eye-opening around postnatal day 14 (P14) [74].

Driven by this activity, global features such as the retinotopic map, as well as single neurons' receptive fields, are established at eye-opening when experience starts to become a significant factor in the further development of the circuitry. Even though the neurons already have an established receptive field, the connectivity between cells is not specific for cells with highly correlated receptive fields at eye-opening [27]. During the first week of visual experience of the mouse (the third postnatal week, P14-P21), however, neurons with correlated receptive fields become more likely connected than neurons with uncorrelated receptive fields. The probability of bidirectional connections increases during this period, making the first week of sensory experience a possible period for the emergence of the putative Hebbian assemblies in V1 (see above). Visual experience, however, only contributes a part to this emergence of selective connectivity, as it also emerges partially when the mice are kept in the dark during this time [28]. However, some significant differences of this connectivity between animals raised in the dark and animals with normal visual experience can be identified: (a) the preferential connectivity between cells with similar receptive fields is less strongly expressed in animals raised in the dark, and (b) the reduction of connections between cells with dissimilar receptive fields is entirely prevented by dark rearing. Together these two factors lead to less selective connectivity overall.

In parallel with this refinement of the recurrent circuitry, orientation tuning and ocular dominance further refine [7]. However, especially the binocularity of inputs and the binocular matching, that is, that inputs from the two eyes impinging on one neuron have similar receptive fields, further refines in a period of high experience-dependent plasticity that has been termed the *critical period* (CP, also the canonical critical period) [11] (for binocular matching, see specifically [75]).

2.2 The canonical critical period and sensory deprivation

Investigations of the effects of sensory deprivation on the cortical circuit began with Thorsten Hubel and David Wiesel in the early 1960s when they studied the effects of long-term lid-closure in kittens [12], and have since then been undertaken in a range of species and with different forms of deprivation (Table 2.1). Hubel and Wiesel investigated in these experiments, how the distribution inputs to cortical cells from either of the two eyes (their ocular dominance) changes if one of the eyes is closed for extended periods (lasting from a few days to months), called *monocular deprivation* (MD). Their research uncovered drastic shifts of this distribution in favor of the eye that was still functional. Interestingly, these investigations of the development of specific tuning properties of neurons in V1 were undertaken at the same time as when these tuning properties were discovered in the first place [9, 40]. An important discovery beyond the changes of cell responses induced by deprivation was the discovery of the critical period [13].

MD can be induced by multiple means, e.g., through closure of one eyelid either by sewing or gluing, through inactivation of retinal activity in one eye by intra-ocular

injection of TTX (monocular inactivation), or one eye can be surgically removed (monocular enucleation). A disadvantage of the latter is that it is not reversible, and thus, does not allow to investigate if recovery would be possible when the inputs are restored. A direct extension of these monocular paradigms is to perform either of them binocularly. Another highly effective paradigm to study the development of visual cortical circuits is to keep animals in darkness for an extended period (see [10] for comparison with MD), or even rearing animals in the dark starting before the onset of vision (dark exposure, dark rearing). Even experimental procedures as closely related as MD through lid suture or MD through retinal inactivation have different effects on the developing brain. These differences are driven by differences in the spatial and temporal correlations on the spike trains in the LGN after MD of different methods [76]. I refer to MD primarily to mean monocular lid-suture unless mentioned otherwise. The comparison of experimental results across multiple monocular deprivation paradigms can be advantageous, but the procedures must be carefully distinguished because of subtle differences in the plasticity they induce (Table 2.1).

On the single-cell level, extended MD leads to a loss of responsiveness of the cortical cells to the deprived eye [10,12,77]. This loss is not apparent in the retinal or thalamic circuitry but only emerges at the level of the cortex, where it also involves thalamocortical synapses [9]. This underlies the overall shift in the ocular dominance distribution within the cortex in favor of the open eye. In animals with an ocular dominance map (primates and carnivores), MD prevents the normal development of these maps. After normal development, the domains driven by either of the eyes have similar sizes and distribute over the cortical sheet in stripes. After MD, the domains of the deprived eye are markedly reduced as small islands surrounded by domains selective to the open eye [9] (see also [22, 78]).

Rats and mice do not have an ocular dominance map, but the ocular dominance of the cells also shifts towards the open eye. This shift towards stronger dominance of the open eye is a two-stage process: within only two days of MD, the responsiveness to the deprived eye decreases, while only later-on, after five days of MD and more, a potentiation of responsiveness to the spared (open) eye is observed [77]. Binocular deprivation on this timescale has little effect [77]. Likely, MD has stronger effects because it induces competition between a deprived and a spared (non-deprived) input stream. However, longer binocular deprivation also leads to increased responsiveness of the cells [81].

Brief MD leads to depression of feedforward synapses onto excitatory neurons in both V1m and V1b [81] (for more details on V1m see Sec. 2.3). However, the responsiveness of fast-spiking interneurons shifts to favor the closed eye during early MD, which could contribute to the early decrease of responsiveness of excitatory neurons to the closed eye [79]. With longer MD, also the preference of the fast-spiking cells shifts towards the open eye.

The depression of feedforward synapses can be either caused by decreased activity arriving from the closed eye or by a decorrelation of these inputs. Linden and colleagues found that neither the amount of activity nor its temporal structure in

Table 2.1: Experimental studies of deprivation-induced plasticity. CP: critical period, DR: dark rearing, MS: monocular lid suture, MI: monocular inactivation (e.g., through injection of TTX into one eye).

Ref.	Age, Species	Type	Effects
[12]	P8 onwards, cat	MS	Drastic ocular dominance shifts by MS (1-4 months), loss of responsiveness to the deprived eye in V1
[13]	P28-90, cat	MS	Discovery of a CP for deprivation effects in kitten V1 between P28 and P56
[10]	P17-45, rat	DR & MS	Maturation of visual response properties in V1, all perturbed by dark rearing, ocular dominance shift by MS
[76]	P28 (CP), mouse	MS & MI	Neither MS nor MI change mean rates in LGN, MI increases the fraction of bursts in LGN firing, MS decreases cross-correlations while MI increases them
[77]	P28-33 (CP), mouse	MS & MI	Different ocular dominance shifts for MI and MS: MI produces potentiation of open-eye responses, MS produces depression of closed-eye responses
[78]	P27-33 (CP), cat	MS	Fast changes of ocular dominance within the first day of MS occur in upper layers 2/3 but not in input layer 4
[79]	P24-32 (CP), mouse	MS	PV shift their preference initially towards the closed eye, but later favor the open eye
[80]	>P60 (post-CP), mouse	MS	Upon repeated MS and reopening, single cells return to their preferred ocular dominance
[81]	P21-P28 (CP), rat	MS	Potentiation of responsiveness to both eyes after six days of MD requires synaptic scaling and involves plasticity of intrinsic excitability
[82]	P18-21 (CP), rat	MS	Two days of MS induce potentiation of recurrent inhibition in V1m, the plasticity mechanism underlying is identified
[83]	P15-23 (pre-CP & CP), rat	None	Plasticity mechanism of inhibitory synapses in V1m switches sign from depression to potentiation at the onset of CP
[84]	P15-26 (pre-CP & CP), mouse	MS	Potentiation of recurrent inhibition after MS during the CP is expressed in multiple synaptic properties
[85]	P18-28 (pre-CP & CP), rat	MS	Onset of CP is accompanied by switch of sign of inhibitory plasticity also in V1b

LGN changes upon eyelid closure [76]. Instead, the correlations between LGN relay cells decrease. The change of correlations rather than average activity levels has important implications for the nature of the plasticity mechanisms underlying the competitive plasticity in V1b (see below, Sec. 2.4). In contrast to the closure of the eye, inactivation through injection of TTX into the eye leads to more bursting activity and increases correlations of LGN neurons, while mean activity is also unaffected.

The most dramatic effects of MD occur when it is applied during the CP. In rodents, this developmental phase starts around the beginning of the third postnatal week (P21) and begins to decline around the end of the fifth postnatal week (P33-35) [10]. During the critical period, the shift of the ocular dominance due to MD can lead to functional blindness on the deprived eye that is irreversible (amblyopia). This can also appear in children when the eyes are misaligned during development (strabismus) or when a natural condition effectively leads to monocular deprivation like a congenital cataract [86]. In the latter case, amblyopia can only be prevented by surgery early in life, suggesting a CP in humans. MD in adult animals can also shift the ocular dominance of single cells, but this shift is reversible after re-opening of the deprived eye [80].

The induction and termination of the CP are driven by the maturation of PV-interneurons [11]. Enhancing the action of the inhibitory neurotransmitter GABA, e.g., by application of benzodiazepines, leads to an earlier onset of critical period plasticity. On the other hand, dark rearing from birth delays the maturation of GABA function and the onset of the critical period.

Neural circuits are highly plastic throughout development and remain plastic into adulthood. The CP is not only a period of heightened plasticity but is also governed by distinct plasticity mechanisms. Lefort and colleagues found a switch in the sign of plasticity induced by the same electrophysiological protocol at the onset of the CP [83]. The protocol, originally described in [82], combines a subthreshold depolarization of a postsynaptic excitatory neuron with spikes from a presynaptic inhibitory fast-spiking cell. They found that this protocol induces long-term depression before the onset of the CP (animals aged P15-P17) while it induces long-term potentiation after the onset of the CP (animals aged P21-P23) [83, 84]. A further change related to the onset of the CP they found was that plasticity in the reciprocal synapses, from excitatory onto inhibitory neurons, cannot be induced by the same protocol before the onset of the CP, but that the synapses are potentiated through it after the onset of the CP. Also, in V1b, the plasticity of inhibitory synapses switches sign with the onset of the critical period for ocular-dominance plasticity [85].

2.3 Monocular deprivation and the monocular primary visual cortex

The disruption of vision in one eye during development has strong, long-lasting effects on binocular processing in the adult. It seems natural to study the effects

of monocular deprivation in V1b, in which inputs from both eyes arrive and are integrated. However, there are multiple arguments for studying the effects of MD in V1m.

A quantitative reason is that the major fraction of V1 in rodents, such as rats and mice as two of the essential mammalian animal models, is monocular.

Secondly, closing the contralateral eye disrupts the one primary source of sensory signal in V1m. Studying the effects of this closure is an experimentally accessible path to study how brain areas are possibly "re-purposed" when a brain area can no longer serve its original function (or degenerate in the long term). For example, when rats are deprived of vision from birth by enucleation, cells in the visual areas become responsive to somatosensory stimuli [87]. MD during the critical period may not be sufficient to induce such drastic shifts but could, for example, lead to an expansion of inputs from the ipsilateral eye also into the monocular region initially driven by the contralateral eye.

Thirdly, the plasticity in V1b is driven mainly by competition between the inputs from the two eyes. The closure of one eye leads to a form of winner-take-all dynamics in V1b in favor of the open eye. Such competition makes the experimental dissection of the synaptic basis of the changes in responsiveness to the closed vs. the open eye (fast loss in responsiveness to the closed eye vs. slower the gain in responsiveness to the open eye [77]) harder and can confound interpretation of the results. Furthermore, strong winner-take-all competition is a very distinct form of plasticity, whereas the plasticity of a single input after a change of its statistical properties represents the more general case. Finally, as an experimental benefit, the focus on V1m also offers the opportunity to have the control case of the experiment within the same animal by measuring activity or plasticity in both hemispheres. The hemisphere contralateral to the closed eye is the experimental condition, and the ipsilateral hemisphere is the control case. Given the high variability of many biological characteristics between individuals, the direct comparison within one animal can reduce this source of uncertainty.

Despite the absence of a winner-take-all competition, MD also induces strong plasticity in V1m. It also triggers a sequence of activity changes in the cortex that progress from an early drop of activity within the first two days of MD, to a later recovery that homeostatically regulates activity to its baseline level within six to eight days, with distinct timescales for excitatory and inhibitory activity. How the experimentally measured plasticity gives rise to the experimentally measured regulation of activity is not yet understood. This relation was a primary focus of my research.

2.3.1 Activity in V1m during prolonged MD

Hengen and colleagues studied how the activity in V1m develops during multiple days of MD in freely behaving animals *in vivo* [54, 89]. These experiments lasted from days P24-P32 which lie within the CP (Fig. 2.3). During the first three days of measurement, the activity in the baseline state was monitored, and only then

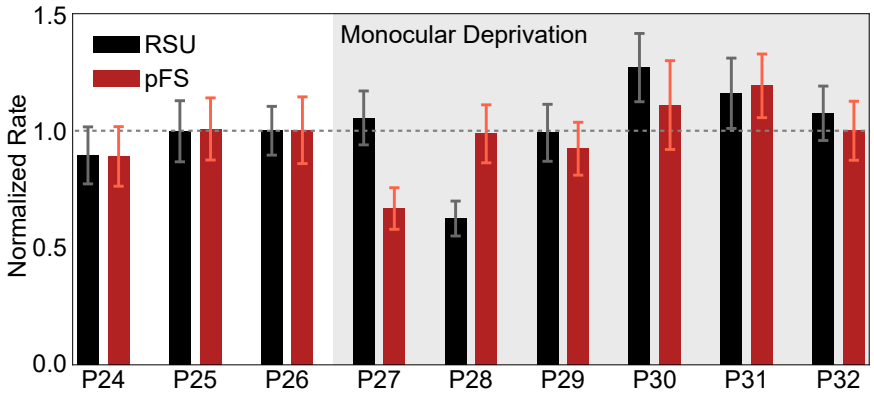

Figure 2.3: *In vivo* activity during monocular deprivation in the monocular portion of the primary visual cortex (V1m). Normalized firing rates of regular spiking units (RSU, black, presumably excitatory neurons), and putative fast-spiking cells (pFS, red, presumably PV-interneurons). Bars show mean ± standard error of the mean (SEM). Activity from *in vivo* recordings lasting nine days, between postnatal day (P)24 and P32. Drop and recovery following MD starting at the evening of P26 (gray shaded area). Figure redrawn from [54], data extracted using *Webplotdigitizer* [88].

MD was applied (on the evening of P26). Because the recordings are extracellular, neuron types can only be identified through the properties of their individual spikes. The authors were able to distinguish regular-spiking units (RSU), which have the spiking characteristics of excitatory cells, and putative fast-spiking cells (pFS), that likely represent PV-interneurons. In the following, I will assume that the electrophysiologically identified cell types are excitatory neurons and PV-interneurons.

Hengen and colleagues observed a reduction of activity in V1 during the first two days of MD (Fig. 2.3). A homeostatic recovery of the activity follows this reduction to the baseline level (before MD), which lasts until roughly five days after MD. At this point, the average activity is indistinguishable from the activity before MD. Interestingly, it is not only the overall activity that returns to its baseline level, but it is indeed single excitatory neurons that return to their individual baseline firing rates before MD [89]. Single excitatory cells in V1 have an individual firing rate set-point that is fixed on a timescale of multiple days.

The general pattern of a drop and subsequent recovery of the firing rates is true for both the excitatory and the fast-spiking inhibitory cells. However, the initial decrease and subsequent recovery of firing rates happen on distinct timescales for the two cell types. The average firing rate of excitatory cells is still around their baseline level on the first day of MD, but it is markedly reduced to around 60% of this baseline level on the second day of MD (Fig. 2.3, black bars). Recovery is already visible on the third day, but there might be a slight overshoot of the firing rates until the rates finally settled on the sixth day. In inhibitory interneurons, the dynamics of this firing rate regulation of a drop and recovery is faster than in excitatory neurons. The firing rate of this population drops to around 60% of their

baseline activity already on the first day of MD and recovers on the second day (Fig. 2.3, red bars).

Pairwise correlations in the network decrease during the early period of MD, similarly to the firing rates [90]. Unlike the firing rates, which remain constant in the control hemisphere (ipsilateral), correlations undergo a developmental increase during the MD experiment (P27-P31) in the control hemisphere. In the deprived hemisphere (contralateral), correlations decrease relative to their baseline levels before the induction of MD during the first two days of MD and then partially recover back to baseline by the fifth day of MD. The recovery of the correlations does not only apply to the overall correlations but actually recovers the structure of the connectivity matrix. A similar recovery of the correlation structure was observed after re-opening of the eyes in adult animals [80].

Different from the firing rates and correlations, Ma and colleagues found that the statistics of avalanches of spikes change immediately after closure of the eye but recover after two days of MD to their initial state [91]. The baseline state avalanches express statistics suggesting that cortical firing is in a critical state that maximizes information capacity and transmission. The recovery of this critical state suggests that the criticality of the spiking is also a homeostatic set-point.

The fact that the drop of firing rates, as well as of correlations, takes multiple days to unfold suggests that a lack of input does not generate these activity changes, as is also apparent from the preserved average activity in LGN just after MD [76]. More likely, the firing rate and correlation changes are driven by a reorganization of synaptic connections driven by activity-dependent plasticity mechanisms.

2.3.2 Synaptic plasticity in V1m after brief MD

Unfortunately, it is technically not feasible to monitor the development of synaptic strengths during MD continuously *in vivo*. Electrophysiological investigation of the strengths *in vitro* after 1-2 days of MD uncovered that the first two days induce profound changes of the synaptic connections, both within V1m and in feedforward, thalamocortical connections from LGN to neurons in layer 4 of V1m (Fig. 2.4).

The reduction of activity during the first two days of MD could be driven by either depression of the feedforward synapses that provide the external drive to the network or by recurrent changes, either depression of recurrent excitation or potentiation of inhibition within the recurrent circuit. As it turns out, specific synapses in both feedforward and recurrent pathways undergo plastic changes during the first two days of MD. Firstly, such brief MD leaves the strength of recurrent connections between excitatory neurons within layer 4 in V1m unaffected, thus excluding the reduction of recurrent excitation as one of the possible mechanisms [82,92]. Recurrent connections between fast-spiking basket cells (the most common type of PV-interneurons with the highest density in layers 4 and 5 [24]) and excitatory, regular-spiking neurons potentiated during brief MD [3,82,84] (Fig. 2.4A, B). This suggests the potentiation of recurrent inhibition to be a major component of the reduction of activity and responsiveness during the early phases during MD (discussed in [3]).

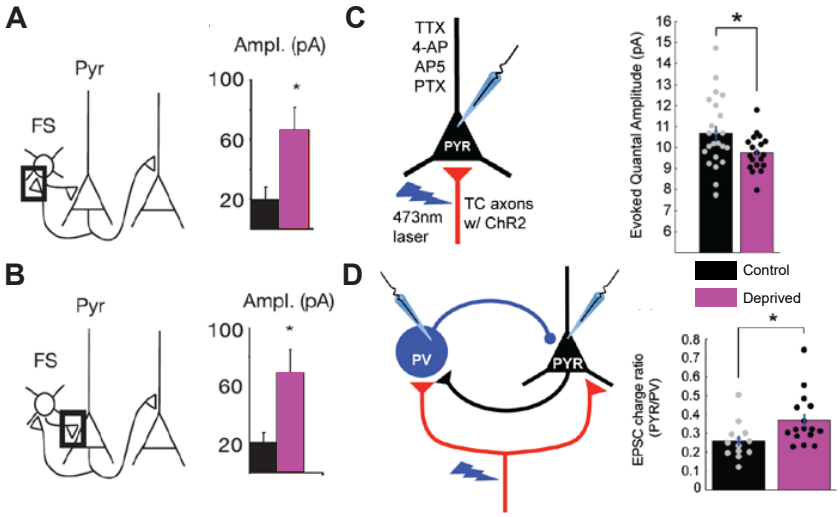

Figure 2.4: Synaptic changes induced by two days of MD in rodent V1m. **(A)** Paired-patch recordings of excitatory pyramidal cells and fast-spiking (FS), putative PV-interneurons demonstrated potentiation of recurrent excitatory synapses onto PV-interneurons in V1m as well as **(B)** in the reciprocal, inhibitory synapse from fast-spiking to pyramidal neurons. Strength was quantified using the peak-amplitude of the postsynaptic current. **(C)** Measurement of excitatory inputs through optogenetic stimulation of thalamocortical fibers together with patch-clamp recordings of excitatory pyramidal cells (Pyr) in V1m (left) demonstrated depression of feedforward synapses (right, quantified using quantal amplitudes). **(D)** Measurement of the ratio of total charge arriving in excitatory pyramidal cells and in nearby PV-interneurons in V1m by paired-patch recordings. This ratio increased (right), indicating suppression of feedforward synapses onto PV-interneurons. Panels A and B modified from [82], panels C and D modified from [3] (experiments performed by Nathaniel Miska).

The potentiation of recurrent inhibition happens in both connections from excitatory pyramidal cells (Pyr) to fast-spiking, putative PV-interneurons (Fig. 2.4A), and in the reciprocal synapse type from fast-spiking to excitatory neurons (Fig. 2.4B) [82]. The authors furthermore investigated the biophysical mechanism of the long-term potentiation in synapses from PV-interneurons to excitatory neurons. They found a novel mechanism triggered by a pairing of presynaptic spiking in a PV-interneuron with a subthreshold depolarization of the postsynaptic excitatory neuron. This pairing induced potentiation of the inhibitory synapses onto the excitatory neurons and in the reciprocal synapses. The statistical characteristics of the synaptic changes (specifically, the coefficient of variation of postsynaptically induced currents and the paired-pulse ratio quantifying short-term plasticity) were the same when plasticity was induced using the *in vitro* protocol as to when it was induced through MD, suggesting that this plasticity mechanism indeed underlies the potentiation of the synapses induced by MD. As a further confirmation that the MD-driven potentiation was based on this mechanism, the authors showed that MD occluded the induction of synaptic plasticity using their protocol. Maffei and colleagues furthermore excluded

several factors to contribute to the changes of activity and responsiveness induced by brief MD in V1m, for example, the connection probability between excitatory neurons, and changes in synapses from PV to PV interneurons, which were both found to be unaffected by brief MD.

As an additional assessment of the activity in V1m after two days of MD, Maffei and colleagues also measured the spontaneous activity in slices and found a reduction of spontaneous activity of the excitatory cells. Indeed, the changes in spontaneous activity *in vitro* can only give an indirect indication of the activity changes *in vivo*, but the results align well with the suppression of activity found in excitatory cells *in vivo* after two days of MD. Interestingly, in a comparable set of experiments that were performed clearly before the CP (P15-P17), brief MD increased the spontaneous activity of the excitatory cells [93]. Increase of activity before the CP vs. decrease of activity in the CP well aligns with the reversal of the plasticity in the synapses from PV to excitatory neurons [83, 84].

Feedforward synapses from LGN to V1m also undergo plasticity during the first two days of MD. Thalamocortical synapses impinging onto excitatory neurons depress (Fig. 2.4C) [3, 92], whereas the feedforward E/I-ratio that arrives at an excitatory cell in layer 4 of V1m increases during the same period (Fig. 2.4D). This increase indicates that the feedforward synapses onto the strongly feedforward-driven PV-interneurons in layer 4 also depressed, and that this depression was even more substantial than the depression of feedforward synapses onto excitatory neurons [3]. A similar pattern of depression of feedforward inputs was also found in the synapses within V1b going from layer 4 to layer 2/3 [94].

The depression of feedforward synapses found in [3] suggests that the feedforward pathway also contributes to the reduction of the responsiveness after MD. However, as I demonstrate in Chapter 3, a network model suggested that the shift of the E/I-ratio in favor of excitation increases the excitability of the circuit. The interaction of the recurrently generated dynamics of the network needs to be taken into account when evaluating the role of the different synaptic changes. With my research, I investigated in general how these changes in the feedforward and the recurrent pathways can give rise to the activity changes *in vivo*. Specifically, since the synaptic changes were only measured after two days of MD, while the activity is continuously regulated during these first days, there is a gap of timescales that modeling can bridge. In the model, I can apply the synaptic changes continuously and evaluate their effects on the activity, mapping out all potential modulations of the rates these changes can generate.

2.3.3 Synaptic plasticity in the primary somatosensory cortex after brief whisker deprivation

Another important sensory deprivation paradigm, which targets the somatosensory system, is whisker-deprivation (WD). During WD, a subset of the whiskers on one cheek or both are cut short or completely removed through plucking. Plasticity induced by MD and WD shows some general similarities, but on the short timescale

of the first 1-2 days, the two paradigms invoke different mechanisms [95]. Similar to brief MD, brief WD induces disinhibition in the primary somatosensory cortex (S1) [96]. However, in contrast to the depression of feedforward inputs from the LGN to L4 in V1m after brief MD, this disinhibition is driven by a decrease of the intrinsic excitability of PV-interneurons [97]. The WD studies investigated the feedforward pathway from L2/3 to L4 within S1, but note that the feedforward depression after MD is similar from L2/3 to L4 as it is from LGN to L4 [94]. Longer periods of WD lead to plasticity of both the feedforward and recurrent synapses that is comparable to the effects of brief MD. Specifically, feedforward inputs from L2/3 to L4 onto excitatory synapses as well as onto fast-spiking interneurons depress after this longer deprivation [98, 99]. Recurrent synapses from fast-spiking onto excitatory neurons potentiate. In short-term WD experiments, a tendency of these synapses to potentiate was identified, but it did not reach statistical significance [97]. The long-term *in vivo* activity following WD has not been measured yet. My modeling allowed me to investigate the consequences for the activity of the different forms of plasticity invoked. Specifically, it is not clear if the decrease of the intrinsic excitability of PV-interneurons has the same effects on activity in the circuit as the depression of feedforward input synapses.

2.4 Previous modeling of monocular deprivation

The experimental investigation of MD-induced plasticity in synapses and neuron responsiveness has produced many important insights about the experience-dependent refinement of cortical circuits. Several modeling studies investigated the different aspects of the MD-induced circuit changes, generating mechanistic or phenomenological insights into the regulation processes taking place in development (Table 2.2).

In one of the early models, Miller and colleagues investigated a phenomenological model of the formation of ocular dominance domains found in carnivores and primates [22]. This early study of the formation of ocular dominance domains was validated, in part, by predicting how the distribution of the domains changes with prolonged MD during development. Specifically, the authors tested how the domains of cells responding primarily to the open eye expand, while the domains of cells responding to the closed eye shrink. In this model, the domains and their properties emerged from spontaneous pattern formation on a two-dimensional grid of neurons. Toyoizumi and Miller investigated the conditions of the equalization of ocular-dominance domains in cat V1 from an initially biased state [100]. Before the critical period, the domains driven by the contralateral eye are bigger than those driven by the ipsilateral eye. During the critical period, the sizes of these domains equalize. Their model could explain the equalization during the CP by the maturation of intracortical inhibition interacting with a Hebbian plasticity rule. As it was shown in earlier experimental work, the maturation of inhibition induces the CP in mice. They could, furthermore, show that the equalization of domains responding to the ipsi- or contralateral eye is prevented by binocular deprivation and,

most importantly, that MD causes a shift towards the open eye, despite the drive from inhibition towards equalization that they found for the normal development.

In follow-up work, Toyoizumi and colleagues investigated further investigated the role of the maturation of inhibition in the induction of CP-plasticity [101]. Their investigation suggested that maturation of inhibition suppresses spontaneous activity more than visually evoked responses. Through this, learning in the circuit becomes more dependent on visual experience during the CP and is less dependent on internal signals. Interestingly, this result proposes that the CP could be induced purely by the maturation of inhibition with no need to change the plasticity mechanisms. In this proposal, it is not the plasticity mechanisms that change at the onset of the CP, but primarily the outcome of experience-dependent plasticity using the same mechanisms. Note, however, that changes in the plasticity mechanisms were found at the onset of the CP. Nevertheless, this work proposes an important and exciting role of recurrent inhibition in MD-induced plasticity. The maturation of inhibition intensifies competition between inputs from different eyes.

Bono and Clopath further unfolded the role of inhibition in the induction of CP-plasticity [102]. They found that, beyond the maturation of GABAergic inhibition in the circuit, the plasticity of excitatory synapses onto inhibitory neurons also plays a vital role. Blocking of this plasticity could suppress ocular dominance plasticity in their model. Additionally to the onset of CP through the maturation of inhibition, they could thus explain the termination of CP-plasticity as an effect of a potential developmental blocking or fading of plasticity in excitatory onto inhibitory synapses. This result corresponds with the experimental results that show that plasticity of excitatory synapses onto inhibitory neurons, both in feedforward and recurrent pathways, is a crucial component of MD-induced plasticity in both the binocular and monocular V1. Furthermore, it is interesting that the specific induction protocol for potentiation of recurrent synapses from excitatory to inhibitory neurons, which is related to MD-induced plasticity, becomes functional only at the onset of the CP [83].

Blais, Shouval, and Cooper aimed at a quantitative comparison of heterosynaptic and homosynaptic plasticity mechanisms through the investigation of the effects of MD on a single cell's ocular dominance [103]. In heterosynaptic plasticity, the weight changes in one synapse influence the weights of other synapses on the same cell [105], for example, through competition for receptors [106]. An implementation of this is to enforce a normalization of the overall excitatory weight that can impinge on a single excitatory cell, which induces strong competition between single synapses on a single neuron. This type of competition can be considered a spatial competition of simultaneously active synapses. For homosynaptic plasticity, Blais and colleagues investigated the *Bienenstock-Cooper-Munro* (BCM)-rule [107]. Following the BCM-rule, postsynaptic activity below a certain threshold leads to depression of synapses (if the presynaptic neuron is active, else the synapse weights do not change), whereas high postsynaptic activity leads to potentiation. Most importantly, regarding homeostasis in the BCM-rule, the threshold between depression and potentiation is a dynamic function of the preceding activity in the postsynaptic neurons. Through

Table 2.2: Modeling studies of deprivation-induced plasticity, the types of models used, and some key insights

Ref.	Model type	Description
[22]	Rate-based recurrent network, plasticity in feedforward synapses	A model for the emergence ocular dominance domains and the influence of deprivation on the domain distributions
[100]	Rate-based recurrent network, plasticity in feedforward synapses	Maturation of inhibition induces a driving force towards equal size of ocular dominance domains, MD counteracts this equalization
[101]	Rate-based feedforward network	Maturation of inhibition can induce a critical period of plasticity without direct changes to the plasticity mechanisms
[90]	Spiking recurrent network with biophysical plasticity mechanisms	Distinct roles for synaptic scaling and plasticity of intrinsic excitability for the recovery of correlation structure and firing rates during ongoing MD
[91]	Probabilistic recurrent network, plasticity in recurrent synapses	Homeostatic regulation of criticality of spike avalanches during ongoing MD
[102]	Rate-based recurrent network	An alternative model for the onset of critical period plasticity to [101] and, additionally, for its termination
[103]	Rate-based feedforward network	Comparing different forms of plasticity, the authors find the BCM-rule to be the best model for ocular dominance plasticity during MD
[104]	Rate-based feedforward network	Furthers the case for the BCM-rule as model for MD-induced feedforward plasticity based on new measurements of the correlations in LGN in [76] after different forms of deprivation (suture vs. inactivation)

this dynamical adjustment of the threshold between depression and potentiation, the plasticity rule itself is plastic (*metaplasticity*). In the BCM-rule, the competition between the two inputs is in the temporal domain, whereby the overall level of activity influences how presynaptic activity changes synapses due to the dynamical depression/potentiation threshold. If some highly active synapses keep this threshold high, they can potentiate while forcing other, less active synapses to depress.

The investigation in [103] led to quantitative predictions about the role of noise and could explain the differences between plasticity induced by monocular or binocular deprivation. In MD, the open eye keeps the threshold between potentiation and depression high, thus forcing depression of the synapses from the closed eye. In binocular deprivation, the threshold decreases, and synapses from the closed eyes do not change as much as in MD. The BCM-rule has also been used to explain the difference of plasticity induced by monocular deprivation and monocular inactivation [104] (see also supplementary discussion to [76]). Using the BCM-rule, Blais and colleagues could indeed reproduce that monocular deprivation induces a more substantial shift in ocular dominance than monocular inactivation does [77,108], and that this can be explained by the different effects on correlations found in [76] as they interact with the plasticity rule. In contrast, a heterosynaptic rule predicted the opposite, making the BCM-rule the better explanatory model of the plasticity in these two deprivation paradigms.

Two recent studies have investigated the long-term regulation and, ultimately, homeostasis of multiple aspects of network activity and its underlying plasticity in V1m during MD. Work from my colleague Wu Yue and others scrutinized the role of multiple forms of plasticity in the homeostatic recovery of firing rates of excitatory neurons after around 6 days of MD in large-scale networks of spiking neurons [90]. Extending the data analyses performed by Hengen and colleagues [54], Wu and colleagues found that not only rates of single cells are homeostatically regulated after 6 days of MD, but also the pairwise correlations between cells [90]. Crucially, individual pairwise correlations recover such that subnetworks of highly correlated neurons are preserved (similar to what has been found in [80] in adult animals). The recovery of individual pairwise correlations is also reminiscent of the recovery of individual excitatory firing rates [89]. The model predicted that different homeostatic plasticity mechanisms cause the homeostasis of rates and the homeostasis of correlations. The two crucial homeostatic mechanisms were plasticity of the cell-intrinsic excitability [109] and synaptic scaling [110]. To study firing rates and correlation structure, Wu and colleagues first imprinted Hebbian assemblies in the network with multiple interacting forms of plasticity [90]. They drove neurons in the assemblies by correlated spike trains. Subsequently, MD was modeled by (a) removing the correlations between input spike trains and (b) depression of feedforward inputs. These two changes triggered a homeostatic response in the recurrent connectivity, where different homeostatic mechanisms played different roles. Wu and colleagues found that the overall homeostasis of the activity after prolonged MD is based on the interplay of these multiple forms of homeostatic plasticity. The long-term recovery of excitatory firing rates required plasticity of the intrinsic

excitability of excitatory neurons, realized through activity-dependent plasticity of the threshold for action potential initiation. The recovery of the correlation structure furthermore required synaptic scaling.

In related work, Ma and colleagues investigated the short-term regulation and homeostasis of firing rates and, as their primary focus, of the criticality of cortical dynamics [91]. Criticality is a hypothesized property of cortical networks that would lead to powerful computational capabilities of the networks [111]. In a critical state, a network is close to a phase transition where the size-distribution of so-called *avalanches* of spikes, bouts of spikes that engage a set of neurons in a cascade, follow a power-law [91]. Their data analysis showed that criticality of cortical spiking is perturbed immediately after lid suture, different from the rates and correlations that drop only after 1-2 days. Using a phenomenological model of the cortical circuit, they could find parameter regimes in which they could reproduce the regulation of firing rates *in vivo*. The model also predicted the instantaneous deviation from criticality and its recovery over the first two days of MD.

Except for the work by Ma and colleagues [91], none of the previous work investigated the specific activity regulation during the first two days of MD in V1m. Furthermore, also this work did not investigate how the MD-induced synaptic changes during the first days of MD directly relate to the activity changes *in vivo*. Ma and colleagues could draw conclusions about plasticity mechanisms that seem necessary for the specific temporal activity regulation. However, they could only generate this regulation with tightly constrained plasticity parameters and did not relate it to the actual synaptic changes in the thalamocortical circuits.

My research has focused on connecting these two factors, the specific *in vivo* changes of synapses and how they give rise to the regulation of activity. Furthermore, a goal was to understand the conditions that can generate the observed regulation of firing rates as robust as possible in terms of the specific synaptic changes. Therefore, my approach is based on investigating the dynamics of highly recurrent networks undergoing perturbations in feedforward and recurrent synapses based on the synaptic changes observed *in vivo* (Fig. 2.4). The basis of this work comes from our understanding of the dynamics of highly recurrent networks with excitatory and inhibitory neural populations.

2.5 Dynamics of recurrent networks

Two important dynamical aspects of highly recurrent networks of excitatory and inhibitory neurons that are relevant to my modeling approach are the *balanced state* [112,113] and *inhibitory stabilization* [114]. Balance determines the microscopic statistical structure of spike-trains in recurrent networks. Inhibitory stabilization has important implications for the dynamics of rate responses. In my work, I investigated the responses to MD-induced plasticity in spiking-network models that are in an approximately balanced state and analytically described the rate responses of these networks and their relation to inhibitory stabilization. In this chapter I will

thus introduce these two concepts, the balanced state (Sec. 2.5.1) and inhibitory stabilization (Sec. 2.5.2).

2.5.1 Spiking-networks of excitatory and inhibitory neurons and the balanced state

Spiking of cortical neurons is irregular, with statistics approximately described by a Poisson-process [115, 116]. In a Poisson process with a constant rate, the occurrence of an event is equally likely at any time point. Thus, the Poisson process is the least predictable, most irregular, stochastic point process. Concerning the relation between different neurons, the firing of cortical neurons appears highly asynchronous, with low correlations of the spike-trains of pairs of neurons [117, 118]. How can such asynchronous irregular firing be produced in a recurrently coupled network? In the initial investigation of this problem, it seemed that such an asynchronous irregular state could not be self-consistent since a neuron driven by excitatory asynchronous irregular spike-trains fires more regularly than its inputs [115]. Since the irregular firing of presynaptic cells would produce more regular firing of the postsynaptic cells that feedback onto the circuit, this state seems impossible as a stable and steady solution of the activity. The problem is that, even though the inputs are irregular, the cell's firing is primarily driven by the integration of the presynaptic excitation, yielding a drift towards the threshold with only slightly varying integration times.

Contrary to this, a neuron with a subthreshold input current whose firing is primarily driven by fluctuations of its membrane potential fires irregularly [119]. Van Vreeswijk and Sompolinsky gave the seminal analytical explanation of how such a *fluctuation-driven regime* and, consequently, the high variability of the spiking of cortical neurons can emerge self-consistently in highly recurrent networks [112, 113]. They proposed that the cells receive large excitatory and inhibitory currents that cancel on average, balancing each other (Fig. 2.5A). Nevertheless, the total current, though balanced on average, will fluctuate in the network. When a neuron is driven by such a current, its spiking is determined by random fluctuations of the membrane potential that cross the spiking threshold. This gives rise to irregular spike times of the single neuron (Fig. 2.5B).

The balanced state and fluctuation-driven regime also have an interesting information-theoretic component, summarized by Abbott in [16]. Consider a single neuron that receives, on average, a large number N of inputs (that number including the number of synapses and expected number of presynaptic spikes). Since N is the number of expected inputs, there is $\mathcal{O}(N)$ uninformative component, and only deviations from this expectation transmit information (the fluctuations are what distinguishes different events). Under the assumption of independence of the different inputs, the variances sum up, so σ^2 is $\mathcal{O}(N)$ and, thus, the standard deviation is $\mathcal{O}(\sqrt{N})$. Calculating then the fraction of the informative component (fluctuations) of the mean signal (expected value) is $\mathcal{O}(1/\sqrt{N})$. Therefore, for many independent inputs, the informative fraction of the signal a neuron receives is small (e.g., 1% for $N = 10,000$, which is on the order of inputs a cortical cell receives).

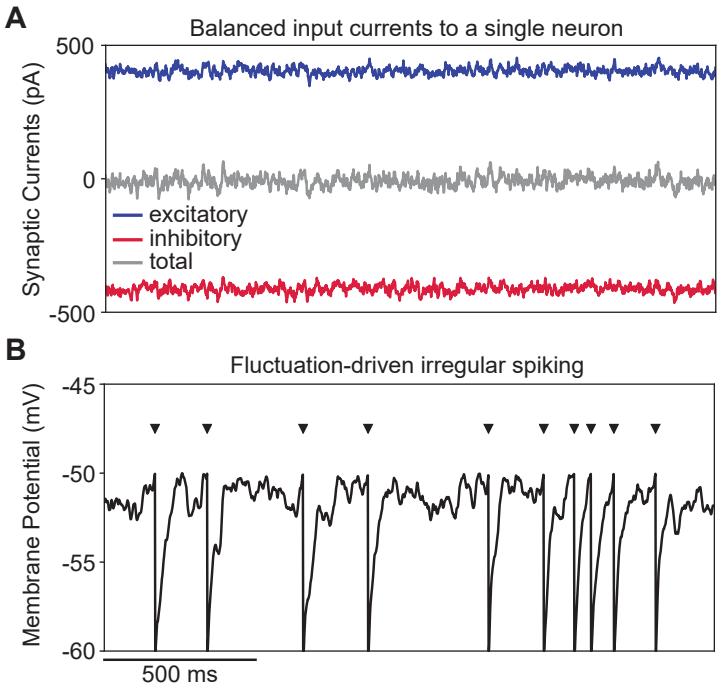

Figure 2.5: The balanced state and the fluctuation-driven regime in a simulated balanced network. **(A)** In the balanced state, large excitatory (blue) and inhibitory (red) currents cancel on average and leave only a small residual, but fluctuating, total current (gray). **(B)** Spike-times (black triangles) of a neuron subject to such a current are determined by fluctuations of the membrane potential (black curve) and not a constant drift of the mean membrane potential – called the *fluctuation-driven regime*. (A) and (B) measured in the same neuron and during the same time window.

The inefficiency of this organization from a coding perspective makes it likely that biological circuits have somehow solved this problem. Two potential solutions are to either abrogate the requirement for independence and correlate inputs (which increases the informative fluctuating component) or to reduce the number of inputs to reduce the uninformative mean component. A third way, which is the solution the balanced state uses, is to reduce the uninformative mean component by having two large mean components, excitatory and inhibitory, that cancel each other to only leave a small uninformative mean component with a comparatively large informative, fluctuating component.

A central property of the balanced state is that the input-output rate relation is strictly linear on the population level when the balance is perfect (also termed *tight balance* [120]). In simulations of spiking networks, this strict linearity is not accurate. However, the rate response are phenomenologically well approximated as threshold linear, a fact I made use of in the analytical investigation of the mechanism generating the rate responses of the spiking network. Brunel first described the conditions to generate the asynchronous irregular spiking in recurrent networks

of leaky integrate-and-fire neurons, along with conditions to find more regular or more synchronous firing [14]. He furthermore derived the self-consistent mean-field firing rate in a spiking network of leaky integrate-and-fire (LIF) neurons. Many other studies have since then extended this work, finding both new insights into the dynamics of networks of this type [121–124] and into their computational capabilities that biological neural networks could harness, such as generating sharp, contrast-invariant orientation selectivity [125, 126]. The asynchronous irregular spiking could even, in itself, represent a highly efficient form of computation, as proposed by Denève and Machens [127].

2.5.2 Inhibitory stabilization

Recurrent coupling between excitatory neurons comprises positive feedback that potentially destabilizes the circuit dynamics. The positive feedback that recurrently coupled excitatory neurons exert on each other does not destabilize the circuits for arbitrary small values of excitation. Thus, recurrent circuits without inhibition can be dynamically stable due to sufficiently small, yet non-zero, recurrent excitation.

Due to the reset of the membrane potential after a spike, neurons generate an effective negative feedback current [128]. Since this post-spike reset appears for every spike, it comprises a dynamically adjusting negative feedback that can counteract positive feedback up to a certain level. In rate models of the average population activity, this comprises rate-dependent negative feedback. A simple approach to describe the dynamics is to use a threshold-linear input-output relation, e.g., if the networks modeled are close to a balanced state (see above). Describing the rates of the different populations by a time-dependent vector $\boldsymbol{r}(t)$, where each entry represents the activity of one population[2], the external inputs to each population by a vector $\boldsymbol{x}(t)$, and the recurrent couplings among the neural populations by a matrix $\boldsymbol{W}$, the dynamics of the network in the linear regime is:

$$\tau \dot{\boldsymbol{r}}(t) = -\boldsymbol{r}(t) + \boldsymbol{W}\boldsymbol{r}(t) + \boldsymbol{x}(t), \tag{2.1}$$

where τ is the intrinsic time constant of the neural populations[3]. This model is equivalent to a mean-field description of the average firing rates in a balanced network of perfect integrate-and-fire neurons, where the constants and couplings between neurons can be derived from the network connectivity and single neuron properties (e.g., [129]). The decay term of the firing rate without external inputs also implies that the resting state of the network is zero. Thus, such networks do not generate spontaneous activity that is independent of an external drive.

This description can be used for networks with arbitrary many populations of neurons, either excitatory or inhibitory. The boundary of stability is most easily

[2] I use as convention boldface letters for vectors and matrices, and regular letters for scalar variables.

[3] These timescales can also be different for the different populations. Then the scalar parameter τ is replaced by a matrix $\boldsymbol{T}$ with the different time-constants as entries on the diagonal and zero elsewhere. I omitted this detail for simplicity.

accessible considering a single excitatory population coupling onto itself, described by the scalar equation

$$\tau \dot{r}(t) = -r(t) + wr(t) + x(t) = (w-1)r(t) + x(t). \tag{2.2}$$

This system is intrinsically stable for $w < 1$, and unstable if the recurrent excitation exceeds one ($w > 1$). In this latter case, other sources of negative feedback are necessary to stabilize the dynamics.

The connectivity of a general network with one excitatory and one inhibitory population can be described by the matrix:

$$\boldsymbol{W} = \begin{pmatrix} w_{EE} & -w_{EI} \\ w_{IE} & -w_{II} \end{pmatrix}, \tag{2.3}$$

where the parameters w_{AB} describe the strength of the connection from population B to population A (Fig. 2.6A). Note that I use all parameters w_{AB} as positive numbers, while the excitatory or inhibitory action on the postsynaptic population is given explicitly through the sign of the matrix entry.

To investigate the stability, one can study the eigenvalues of the Jacobian $\mathcal{J}$, which for the dynamics in Eq. 2.1 is:

$$\mathcal{J} = \boldsymbol{W} - \mathbb{I}, \tag{2.4}$$

where $\mathbb{I}$ is the identity matrix of the appropriate dimension (2×2 in this case). With this Jacobian, we can investigate stability using the real parts of the eigenvalues of the connectivity matrix $\boldsymbol{W}$. Stability requires that these are smaller than one, equivalent to the requirement that the eigenvalues of the Jacobian have all negative real parts: If λ is an eigenvalue of $\boldsymbol{W}$, then $\lambda - 1$ is an eigenvalue of $\mathcal{J}$. Tsodyks and colleagues analyzed the role of inhibition for stability in this general circuit [114]. The details of the stability analysis are shown in Appendix A.1. Similar to the one-dimensional example above (Eq. 2.2), stability of the network requires negative feedback from the inhibitory population when $w_{EE} > 1$. When recurrent inhibition exceeds this bound, sufficiently strong inhibition is necessary to ensure stability. Note that all the connections are involved in this stability, the loop from excitation to inhibition (w_{IE}) as well as the inhibitory feedback onto excitation ($-w_{EI}$) and even the disinhibitory connections from the inhibitory population onto itself (w_{II}).

Such a network with unstable recurrent excitation stabilized by recurrently coupled inhibition is called an *inhibition stabilized network* (ISN). When excitation is weak enough to be stable without recurrent inhibition, I call the network non-ISN. The two cases are also denoted as *operating regimes*, using as terminology also ISN-regime and non-ISN-regime to indicate networks that are either inhibition stabilized or stable without feedback inhibition.

Crucially, the ISN and non-ISN have distinct dynamical properties that only depend on recurrent excitation. Consider two networks with $w_{EE} > 1$ (ISN) and $w_{EE} < 1$ (non-ISN). Stability permitted, such two networks have qualitative

differences in their dynamics even when all other parameters remain the same. Most prominently, ISN's express the so-called *paradoxical effect*. When an external, additional current is injected into the inhibitory population, the naive expectation would be that the firing rate of the inhibitory neurons increases, which will lead to a decrease of the excitatory firing. This is indeed what happens in a non-ISN. However, in an ISN, an additionally injected current to the inhibitory population suppresses not only the firing of the excitatory neurons but also that of inhibitory neurons. One can derive the effect from the steady-state rates in a model describing the average rates of the populations.

The steady-state rate $\boldsymbol{r}_{ss}$ of the dynamical system in Eq. 2.1 is (assuming constant input $\boldsymbol{x}$):

$$\boldsymbol{r}_{ss} = \left(\mathbb{I} - \boldsymbol{W}\right)^{-1}\boldsymbol{x}. \tag{2.5}$$

Inverting the matrix $(\mathbb{I} - \boldsymbol{W})$, the solution of the steady-state for arbitrary input $\boldsymbol{x} = (x_E, x_I)^T$ is:

$$\begin{pmatrix} r_{E,ss} \\ r_{I,ss} \end{pmatrix} = \frac{1}{\det(\mathbb{I} - \boldsymbol{W})} \begin{pmatrix} (1 + w_{II}) & -w_{EI} \\ w_{IE} & (1 - w_{EE}) \end{pmatrix} \begin{pmatrix} x_E \\ x_I \end{pmatrix}, \tag{2.6}$$

where

$$\det(\mathbb{I} - \boldsymbol{W}) = (1 - w_{EE})(1 + w_{II}) + w_{IE}w_{EI} \tag{2.7}$$

is positive when the network is stable (stability requires this determinant to be positive as it is $\det(-\mathcal{J})$; for a derivation from the eigenvalues see Appendix A.1). Due to linearity, the change of the steady-state firing rate ($\Delta\boldsymbol{r}_{ss}$) upon additional stimulation of the inhibitory population with a current δ can be calculated as the steady-state rate for the input vector:

$$\Delta\boldsymbol{x} = \begin{pmatrix} x_E \\ x_I + \delta \end{pmatrix} - \begin{pmatrix} x_E \\ x_I \end{pmatrix} = \begin{pmatrix} 0 \\ \delta \end{pmatrix}, \tag{2.8}$$

where δ is a small external current that can be either positive or negative. The steady-state for this input vector $\Delta\boldsymbol{r}_{ss}$ is the deviation from a baseline firing rate before injection of the current. The elements of $\Delta\boldsymbol{r}_{ss}$ can thus be either positive or negative. The steady-state rate-change for the input $\Delta\boldsymbol{x}$ is

$$\begin{pmatrix} \Delta r_{E,ss} \\ \Delta r_{I,ss} \end{pmatrix} = \frac{1}{\det(\mathbb{I} - \boldsymbol{W})} \begin{pmatrix} -w_{EI}\delta \\ (1 - w_{EE})\delta \end{pmatrix}. \tag{2.9}$$

Thus, the rate of the excitatory population changes opposite to the externally injected current. To show this response numerically, we can simulate a network of an excitatory and an inhibitory population and inject a step current to the inhibitory population (Fig. 2.6A). This simulation confirms that the excitatory rate decreases when the current injected to inhibition is positive (Fig. 2.6B). The change of the steady-state of the inhibitory population depends on the sign of $(1 - w_{EE})$ in Eq. 2.9 and, therefore, the operating regime. In the non-ISN ($w_{EE} < 1$), the steady-state of the inhibitory changes in the same direction as the external current: positive input

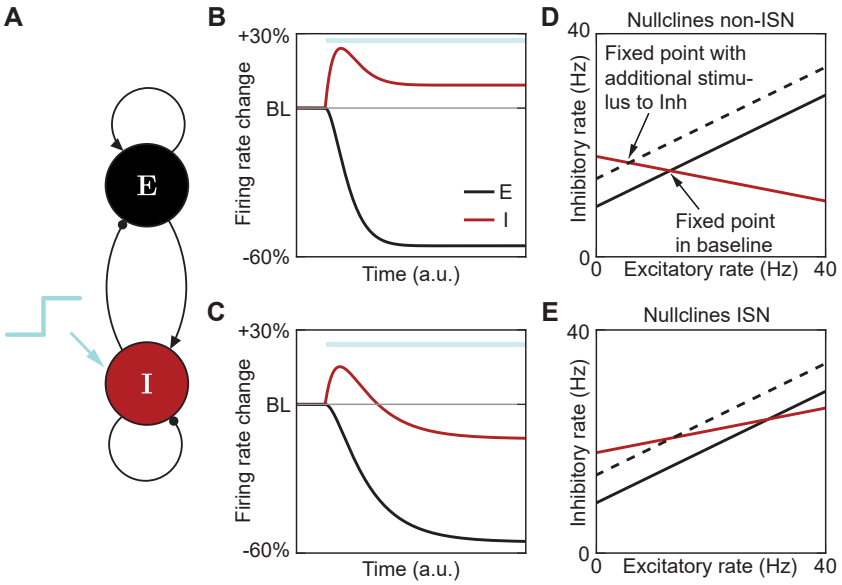

Figure 2.6: Dynamical explanation of the paradoxical effect. **(A)** General network of excitatory (E) and inhibitory (I) populations. The ISN-regime can be extracted through the paradoxical effect in response to an input to I. **(B)** Response to a step current injected to I in the non-ISN (increased input to I in the time window indicated by blue line). **(C)** Same as (B) for the ISN. **(D)** Nullclines of the non-ISN in baseline state before additional input to inhibition for E (full black line) and I (full red). The intersection of the lines determines the steady-state of the dynamics. Additional stimulus to I shifts the nullcline of E to the left (dashed black) and leads to a new steady-state. **(E)** Same as (D) for the ISN. The operating regimes differ in the slope of the nullcline of I.

leads to an increase of the inhibitory firing rate (Fig. 2.6B), negative input would lead to a decrease. In the ISN, where $w_{EE} > 1$, the responses are in the opposite direction to the applied current, the same direction as the excitatory population (Fig. 2.6C).

The dynamical mechanism underlying the paradoxical effect can be made visible from an analysis of the nullclines of the system. Starting from Eq. 2.1, one can derive nullclines for the excitatory and inhibitory populations as the lines on which the time-derivative of the respective rate is zero:

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} w_{EE} - 1 & -w_{EI} \\ w_{IE} & -w_{II} - 1 \end{pmatrix} \begin{pmatrix} r_E \\ r_I \end{pmatrix} + \begin{pmatrix} x_E \\ x_I \end{pmatrix} \tag{2.10}$$

from which one finds each of the rates of the population as a function of the respective other rate and the inputs:

$$\begin{aligned} r_E^0 &= \frac{1}{w_{IE}} \Big((w_{II} + 1) \Big) r_I - x_I \\ r_I^0 &= \frac{1}{w_{EI}} \Big((w_{EE} - 1) \Big) r_E + x_E. \end{aligned} \tag{2.11}$$

Plotting these lines in the (r_E, r_I)-plane, we find an explanation of the paradoxical effect (Fig. 2.6D, E). Increasing the stimulus to inhibition shifts the nullcline of the excitatory population (r_E^0, Eq. 2.11) to the left, independent of the network parameters, and thus, of the regime (Fig. 2.6D, E).

The direction of the inhibitory rate change is then determined from the slope of the inhibitory nullcline. This slope has a different sign dependent on the operating regime: in the non-ISN ($w_{EE} < 1$), the inhibitory nullcline has a negative slope (Fig. 2.6D), while in the ISN ($w_{EE} > 1$) it has a positive slope (Fig. 2.6E). In the non-ISN, the inhibitory rate thus increases when increasing x_I (Fig. 2.6B), whereas it decreases in the ISN, together with the excitatory rate (Fig. 2.6C).

There are two ways to interpret this behavior of the ISN: (a) the inhibitory population changes in the direction opposite to its additional input, or (b) the inhibitory population changes in the same direction as the excitatory population. In a network with a single subtype of interneurons, these two interpretations are equivalent. In a network with multiple subtypes of interneurons, it turns out that the latter is more robust for different network structures, as I will show next.

2.5.3 The paradoxical effect with multiple subtypes of interneurons

Litwin-Kumar and colleagues generalized the notion of the paradoxical effect to networks with multiple interneuron subtypes, inspired by the circuitry in V1 (Fig. 2.7A) [130]. In such networks, external perturbation of one of the subtypes or combinations does not necessarily decrease the rate of any or of all of the interneuron populations, nor the excitatory population. For example, for such a circuit based on the V1-circuitry, some subtypes may increase their rate while other decrease it, seemingly paradoxical, in response to an additional input to all the interneuron subtypes even in the non-ISN (Fig. 2.7B). Furthermore, these directions of the rate changes do not allow to determine the regime, as they are the same in the ISN as in the non-ISN in the given example (Fig. 2.7C). However, it is still possible to distinguish the non-ISN and ISN regimes by evaluating the total inhibitory current the excitatory population receives and comparing this to the change of the excitatory current that the excitatory population receives. In the non-ISN, the excitatory current and the inhibitory current the excitatory population receives change in opposite directions (Fig. 2.7D), while in the ISN, they change in the same direction (Fig. 2.7E). This is very similar to the behavior of the rates in networks with a single subtype of interneurons (Fig. 2.6C). Note that with a single excitatory population, the excitatory current is directly proportional to the excitatory rate. Thus, the excitatory rate and the inhibitory current change in the same directions in the ISN.

Litwin-Kumar and colleagues gave a straightforward analytical explanation of this phenomenon [130]. Denoting the elements of the connectivity matrix from the different interneuron populations X onto the excitatory population E as w_{EX}, without specifying the number and biological interpretation of the interneuron subtypes, the change of the current from the interneuron subtypes onto the excitatory population is the sum of the inhibitory rate changes weighted with the appropriate

connection strengths onto the excitatory population:

$$\Delta I_{E\leftarrow I} = \sum_X w_{EX} \Delta r_X^{ss}, \tag{2.12}$$

where Δr_X^{ss} are the changes of the steady state rates of the different inhibitory populations.

Without generally solving the steady-state rates, Litwin-Kumar and colleagues calculated how this current depends on the operating regime. If we have M interneuron subtypes, one can derive the rate changes $\Delta \boldsymbol{r}_{ss}$ for a stimulation of the inhibitory subtypes

$$\Delta \boldsymbol{x} = (0, \delta_1, \ldots, \delta_M)^T \tag{2.13}$$

from the steady-state

$$0 = \big(\boldsymbol{W} - \mathbb{I}\big)\Delta \boldsymbol{r}_{ss} + \Delta \boldsymbol{x} \tag{2.14}$$

and thus:

$$\Delta \boldsymbol{x} = \big(\mathbb{I} - \boldsymbol{W}\big)\Delta \boldsymbol{r}_{ss}. \tag{2.15}$$

Multiplying both sides of this equation with $\boldsymbol{e}_1^T = (1, 0, 0 \ldots)$ to extract the first component of the changes in the excitatory population, its rate and its inhibitory inputs, one finds:

$$(w_{EE} - 1)\Delta r_{E,ss} = \sum_X w_{EX} \Delta r_X^{ss} = \Delta I_{E\leftarrow I}. \tag{2.16}$$

This equation expresses the general paradoxical effect with multiple subtypes of interneurons: change of the excitatory rate ($\Delta r_{E,ss}$) and the inhibitory current into the excitatory population ($\Delta I_{E\leftarrow I}$) change in the same direction in the ISN, where $(w_{EE} - 1) > 0$. In the non-ISN, this rate and the current change in opposite directions.

Interestingly, since the operating regime only constrains the weighted sum of the inhibitory rate changes, there is some potential for different interneuron populations to respond "non-paradoxically" even in the ISN, or "paradoxically" in the non-ISN, as shown in Fig. 2.7 for a biologically motivated network.

The discovery of the paradoxical effect was driven by the study of oscillations in the hippocampus [114]. Tsodyks and colleagues found that excitatory and inhibitory neurons were in phase during theta-oscillations, which they modeled as emerging from rhythmic inhibition fed to inhibitory neurons. Without knowing about the paradoxical effect, the authors expected that the phase shift would be 180 degrees, where excitatory activity is highest when inhibitory activity is lowest. That the excitatory and inhibitory populations oscillate with the same phase could be explained in an ISN. Further indirect evidence came from the study of surround suppression in V1 in cats [131]. Surround suppression is the effect that the increase in the size of a high-contrast stimulus facilitates excitatory activity up to a certain point, but further increase in the size of the stimulus decreases the activity. A priori, one may expect that the suppression of excitatory responses

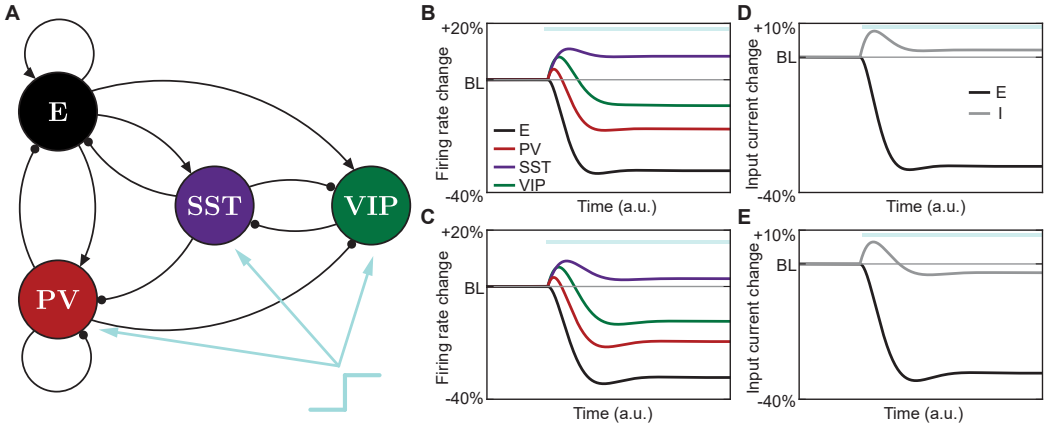

Figure 2.7: The paradoxical effect with multiple interneuron subtypes. (A) Schematic of the V1-circuit model with multiple interneuron subtypes. In the example, I inject a step current to all interneuron subtypes at the same time. **(B)** Rate response to injection of an additional current to all inhibitory subtypes in the non-ISN. Rates of interneurons may increase or decrease. Connection strengths as in [130], colors same as in (A). **(C)** Same as (B) in the ISN. Directions of the rate changes of any subtype do not allow to infer the regime. **(D)** Change of the total excitatory (E, black) and inhibitory (I, gray) currents received by the excitatory population upon stimulation of all interneurons subtypes (in the time window indicated by blue line) in the non-ISN. **(E)** Same as (D) in the ISN. The total inhibitory current received by excitatory neurons expresses the paradoxical effect [130].

through larger stimuli emerges from specific recruitment of inhibition. Without considering the paradoxical effect, this recruitment should lead to an activation of the suppressing interneurons. However, Ozeki and colleagues found that PV-interneurons are themselves surround suppressed, which could be explained through the paradoxical effect. Later work uncovered that surround-suppression of PV does not necessarily depend on a paradoxical effect in PV-interneurons, but could also be governed by spatial summation of SST-interneurons, whereby SST-interneurons, which are not surround suppressed themselves, would suppress both excitatory and PV neurons [59, 60]. Adesnik observed the paradoxical effect in mice related to surround-suppression by optogenetically driving SST [132]. This study found that optogenetically suppressing SST-interneurons led to an increase of the overall inhibitory current consistent with the general notion of the paradoxical effect in ISNs with multiple interneuron subtypes [130].

A problem for detecting paradoxical effects is that many parameters of the experimental setup can corrupt their observation, for example, the depth of anesthesia or the intensity and size of the stimulating laser in optogenetic experiments (discussed in [133]). A crucial factor for finding paradoxical effects is also the recruitment of inhibitory cells during optogenetic activation [134]. Nevertheless, recent work directly confirmed the presence of the paradoxical effect in PV-interneurons in the adult rodent V1 by direct optogenetic stimulation of the PV-interneurons [135].

The dependence on the recruitment of inhibitory interneurons was confirmed in this study, as only transgenic mice with high expression levels of the light-sensitive channel in PV-interneurons showed robust paradoxical effects.

Another study providing direct evidence of paradoxical effects by Mahrach and colleagues also added a further theoretical aspect to the relation between paradoxical effects and inhibitory stabilization with multiple subtypes of interneurons. Litwin-Kumar and colleagues showed that the absence of the paradoxical effect in the rate response of any of the interneuron subtypes does not prove that a circuit operates as a non-ISN [130]. Mahrach and colleagues showed that neither does the presence of the paradoxical effect in the rate response in an interneuron subtype imply that the circuit operates as an ISN [136]. Paradoxical effects could be present in some interneuron subtypes also in a non-ISN. However, the generalized notion of the paradoxical effect based on the total inhibitory current remains valid and is consistent with this finding.

2.5.4 Inhibitory stabilization in the developing visual cortex

The question of inhibitory stabilization in cortical circuits touches upon the broader question of the role of feedforward processing vs. the role of recurrence in cortical computations [133]. High excitatory recurrent coupling is useful for multiple computations such as pattern completion [29] and amplification of sensory signals [137]. This computational benefit comes at the cost of intrinsic dynamical instability that requires recurrently coupled inhibition. Dynamical stabilization through inhibition is a fast mechanism to stabilize the dynamics of networks. Despite the confirmation of the paradoxical effect in adult animals, it is currently unknown if it is present in the developing cortex nor if the developing cortical circuits operate in an inhibition stabilized mode. Through my study of the activity regulation through MD-induced plasticity, I found distinct roles of inhibitory stabilization (Chapter 3) and the interactions between multiple interneuron subtypes (Chapter 4). MD induces changes in multiple synaptic pathways, both feedforward and recurrent. My investigations of MD-induced plasticity thus also led to the investigation of new aspects of the paradoxical effect and inhibitory stabilization: (a) I study multidimensional perturbations of the circuit that go beyond external inputs to one or multiple subtypes of interneurons, and (b) I study perturbations of the recurrent circuitry that shape activity in the circuit in very different ways from external inputs. Investigating the developmental processes during sensory deprivation thus sheds light on the role of recurrent interactions and inhibitory stabilization during experience-dependent stages of development. At the same time, the biological processes uncovered by sensory deprivation stimulate new insights into general properties of inhibition stabilized networks.

3 Effects of deprivation-induced synaptic changes in network models of the visual cortex

Remark: Some of the results, figures, and tables in this Chapter are part of an article entitled *Interneuron subtypes enable independent modulation of excitatory and inhibitory firing rates after sensory deprivation* which has been written together with my supervisor Julijana Gjorgjieva. The article has been uploaded to the preprint server *bioRxiv* [1] and is currently under review for publication in the journal *Proceedings of the National Academy of Sciences of the United States of America*. The results presented in Section 3.5 are part of an article published in *eLife* with Nathaniel J. Miska, Brian A. Cary, Julijana Gjorgjieva, and Gina G. Turrigiano [3]. All methods, results, figures, and tables from these articles which are part of this Chapter were my contribution to the articles unless specifically mentioned otherwise.

3.1 Overview

Two days of monocular deprivation (MD) induce strong plasticity in feedforward, and recurrent pathways that include excitatory and fast-spiking, putative PV-interneurons in V1m [3, 82, 84, 93]. At the same time, activity in these two cell classes is regulated in a specific, temporally distinct fashion: the firing rates of inhibitory interneurons are suppressed after one day of MD and recover on the second day, whereas the firing rates of excitatory neurons stay at their baseline levels on the first day and are suppressed only after two days of MD. They recover back to their baseline after six days [54, 89]. However, since there is a gap of timescales between the direct measurement of synaptic weights (snapshot after two days) and the activity regulation (continuously), we do not understand how the continuous plasticity during the first two days of MD regulates activity. In order to gain a mechanistic understanding of this process, I first used a spiking network model of cortical circuits. I systematically studied the effects of all the different changes of synaptic strengths in the circuit established experimentally in MD experiments. To develop a thorough mathematical understanding of the mechanisms, I furthermore studied a population-rate description of the network dynamics that serves as an effective model of the simulated spiking circuit while allowing analytical treatment that generalizes beyond specific properties of the circuitry.

I investigated a broad range of potential MD-induced changes in the spiking network, studying the synaptic changes independently and interacting with each other (Sec. 3.2). I characterized the whole landscapes of potential dynamical responses (Sec. 3.3). Studying these landscapes, I found a fundamental difference in how ISNs and non-ISNs respond to MD-induced changes, with important implications for the possible types of activity regulation emerging in these networks. The difference between the two regimes, ISN and non-ISN, is the recurrent coupling among excitatory cells. Thus, to investigate the dichotomy in the responses of the different network types more broadly, I studied a whole range of network setups that span both ISN and non-ISN regimes with varying degrees of excitatory coupling in both of the regimes. I introduced multiple measures to quantify the global structure of the responses (Sec. 3.4). The measures enabled a detailed analysis of how MD-induced synaptic changes allow a temporally distinct firing rate regulation in excitatory and inhibitory neurons, with inhibitory changing before excitatory firing rates.

I also investigated a version of the spiking network that was tuned to reproduce firing rates in the model system of our experimental collaborators at the Turrigiano Lab (Sec. 3.5). For this network, I studied the implications of experimentally more tightly constrained synaptic changes (see also [3]). I tested and compared two experimental hypotheses regarding the specific roles of the synaptic changes in the feedforward and recurrent synaptic pathways. My modeling contributed to strengthening the view that recurrent inhibitory pathways play an essential role in controlling activity, namely the early depression of activity during brief MD. In contrast, the synaptic changes in the feedforward pathway have a facilitatory effect on the activity that contradicts the activity seen *in vivo*. To demonstrate that this is not a particular property of a specific parameter setup of the network, I used a Monte-Carlo approach that covers wide ranges of network parameters and is only constrained to have suitable firing rates.

A critical feature of the response in the ISN, which I found in both the general study of the synaptic changes (3.3 and 3.4) and in the Monte-Carlo approach with experimentally constrained synaptic changes (Sec. 3.5) is a tight coordination between excitatory and inhibitory firing rates in response to MD-induced synaptic changes. I, therefore, developed a rate model that heuristically describes the mean rates in the spiking network on the level of the populations (Sec. 3.6). This model is a two-dimensional (semi-)linear dynamical system that can be treated analytically. Through this analysis, I was able to prove that the differences of the response to MD-induced plasticity between the spiking network in the ISN and non-ISN regimes are general properties of these operating regimes (Sec. 3.7). The tight coordination of firing rates in the ISN is primarily due to the presence of the paradoxical effect.

3.2 Modeling approach

Remark: Some of the results of this section (Fig. 3.1D, Table 3.1 and Eqs. 3.1-3.8) have been published before in [1]. The figure panel 3.1D has been modified for clarity, notation in the equations and table has been adapted for consistency.

3.2.1 General network structure

In layer 4 of V1m, feedforward synapses onto excitatory as well as onto fast-spiking, putative PV, interneurons undergo substantial plasticity during the first two days of MD [3]. Furthermore, also the recurrent synapses connecting these two neuron classes are substantially affected [3, 82, 93]. Based on this current knowledge about the synaptic pathways affected by brief MD, the most salient choice for a model is a simplified network of interacting excitatory neurons and inhibitory PV-interneurons. In my model, cells of these two classes are randomly and sparsely connected. Sparsity means, that the total number of presynaptic neurons that connect to a neuron is small compared to the total number of neurons on the circuit, equivalent to a small connection probability. Networks with similar structure have been studied extensively in neuroscience [14, 112, 121, 123, 124, and many others] and have given rise to mechanistic models for multiple computations in cortical circuits, such as the contrast invariance of orientation selectivity in V1 [125, 126] or the stimulus-dependent reduction of variability observed in multiple cortical areas [138, 139].

The synaptic weights within the recurrent circuit are described by two global parameters: First J, that sets a shared scale of all synapses, excitatory or inhibitory (called the *coupling scale*); and second g_{rc}, that sets the relative strengths of inhibitory projections to guarantee overall stability (Fig. 3.1A and Table 3.1). The connection strengths within and between the excitatory (E) and PV (P) cells are specific for the pre- and postsynaptic partner in a given connection, schematically described through a matrix $\boldsymbol{W}_{\text{rec}}$ of synaptic weights W^{AB} from a cell of type B to a cell of type A:

$$\boldsymbol{W}_{\text{rec}} = \begin{pmatrix} W^{EE} & W^{EP} \\ W^{PE} & W^{PP} \end{pmatrix} = \begin{pmatrix} J & g_{rc}J \\ J & g_{rc}J \end{pmatrix}. \tag{3.1}$$

The weights in this schematic describe the peak of the postsynaptic conductance induced in a postsynaptic neuron by a spike in a connected presynaptic neuron. The network has a connection probability of 10% between neurons where this postsynaptic conductance is induced. The remaining 90% of potential connections between cells have weight zero. The different neuron classes are identical in their intrinsic properties, and all connection probabilities are the same for different connections types. I simulated networks with a total of 5000 neurons, 4000 of which are excitatory (80% of all neurons), and 1000 are inhibitory, representing PV-interneurons. Note that in the matrix Eq. 3.1, inhibitory connections have a positive sign just like excitatory ones because they represent peak-conductances for specific

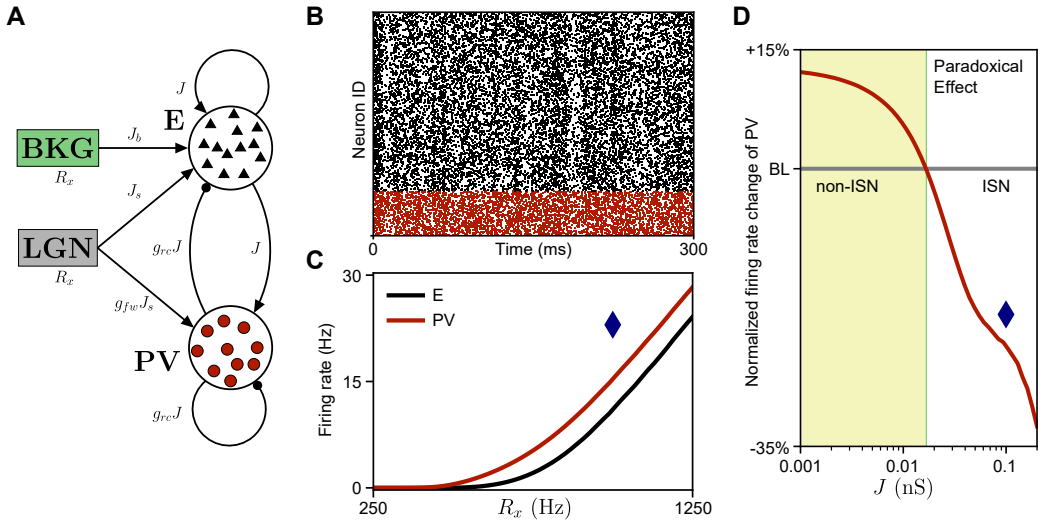

Figure 3.1: Network structure, activity and the paradoxical effect in the spiking network. (A) Schematic of the connectivity and parameters in the network (default values in Table 3.1). **(B)** Spike raster of activity in the network with coupling scale $J = 0.1$ nS, showing asynchronous irregular state. **(C)** Firing rates as a function of R_x, the firing rate of both input sources to the network (lateral geniculate nucleus, LGN, and background, BKG). The blue diamond shows the input rates used in (B). **(D)** Emergence of the paradoxical effect in PV-interneurons, and thus transition to ISN, dependent on the coupling scale J. Additional input to PV-interneurons is implemented through increasing the LGN-input weights into PV by 5%. The blue diamond shows the value of J used in (B). Panel D modified from [1].

ions. The actual effect of a presynaptic spike on the postsynaptic neuron, either excitatory or inhibitory, depends on the distance of the membrane potential to the reversal potential of the respective ion channel (sodium for excitatory and chloride for inhibitory synapses, Table 3.1) at the time the spike arrives. The increased conductance leads to a current of the respective ions through the membrane. The different ion-currents then have the excitatory (depolarizing, towards the spike threshold) or inhibitory (hyperpolarizing) effects on the postsynaptic neurons.

With sufficiently strong inhibition, activity in such a network settles in a stable state of irregular firing of single neurons and asynchronous between different neurons. This state has thus been termed the *asynchronous irregular* (AI) state and is considered a good approximation for the activity in cortical circuits [14, 113]. The network I implemented also shows such AI spiking activity over wide ranges of parameter choices (example in Fig. 3.1B).

As external input, excitatory spike-trains with Poisson statistics from two external sources are injected. One of these sources represents the lateral geniculate nucleus (LGN) in the thalamus, the major source of feedforward input to layer 4 (L4) in V1 and, therefore, of sensory input to V1 overall. Input from this source is sent to both excitatory and PV neurons with cell-type-specific weights. Experiments have found

that the LGN-input to PV-interneurons is at least twice as strong as LGN-input to excitatory neurons in both visual and auditory cortices [38]. I incorporated such a strong feedforward drive to PV-interneurons into the model. Because the absolute amplitude is uncertain, I choose a common parameter J_s for the synaptic weight of feedforward input to both neuron types and implement the relatively stronger drive to PV with a multiplicative factor g_{fw} (Fig. 3.1A). The second source of external input pools together multiple sources of potential feedback inputs, e.g., from the cortical circuit surrounding the local patch implemented in the model or feedback from higher-order cortical areas. This input is denoted as background (BKG) and is only injected into excitatory neurons. The synaptic weight of BKG-inputs is denoted as J_b. The synaptic weights of the external drives can be schematically described by

$$\boldsymbol{W}_{\text{ffw}} = \begin{pmatrix} W^E_{LGN} + W^E_{BKG} \\ W^P_{LGN} + W^P_{BKG} \end{pmatrix} = \begin{pmatrix} J_s + J_b \\ g_{fw} J_s \end{pmatrix} \tag{3.2}$$

for the postsynaptic conductances evoked by externally injected spikes. Spike trains into each neuron are statistically independent and have the same rate $R_x = 1000$ Hz regardless of their source (LGN or BKG, Table 3.1). Similar to what is predicted from the balanced network [113], the input-output relation in this network is approximately threshold-linear when the firing rates are in a reasonable range (Fig. 3.1C). Together with the asynchronous irregular spiking statistics (Fig. 3.1B), this network can thus be considered a biophysically grounded implementation of the *balanced network* [113] (but see [140]).

Single neurons are modeled as leaky integrate-and-fire (LIF) neurons. The membrane potential $V_i^A(t)$ of a neuron i in population $A \in \{E, P\}$ follows the dynamics:

$$C_m \frac{dV_i^A(t)}{dt} = g_L \cdot \left(E_L - V_i^A(t)\right) + I_i^A(t), \tag{3.3}$$

where C_m is the capacitance of the cell membrane, g_L is the conductance and E_L the reversal potential of the leak current (Tab. 3.1). The first term on the right hand side in Eq. 3.3 describes the leak current driving the neuron to the resting potential E_L when no input arrives in the cell. The total input current to this neuron, $I_i^A(t)$ is composed of two factors: an external input current $I^A_{i,\text{ext}}(t)$ and the currents from the cells in the populations $B \in \{E, P\}$, which arise from recurrent interactions within the network, $I^{AB}_{i,\text{rec}}(t)$:

$$I_i^A(t) = I^A_{i,\text{ext}}(t) + \sum_B I^{AB}_{i,\text{rec}}(t). \tag{3.4}$$

To describe these currents, I denote the general sparse recurrent connectivity matrix as $\boldsymbol{C}_{\text{rec}}$ whose entries $C^{AB}_{ij,\text{rec}}$ are (1) non-zero only in case the neurons are connected in the sparse network and (2) take values dependent on the pre- and postsynaptic partner according to the schematic in Eq. 3.1 in case they are non-zero. Every neuron receives a single external spike train with an input weight $C^A_{i,\text{ffw}}$ that depends on the cell class A the neuron belongs to, following the schematic in Eq. 3.2,

Table 3.1: Default parameters of the single neurons and networks, and parameters for modeling the synaptic changes during monocular deprivation. EPSG/IPSG: excitatory/inhibitory postsynaptic conductance; LGN: lateral geniculate nucleus; BKG: background; modified from [1].

Symbol	Value/ Range	Description
C_m	200 pF	Membrane capacitance
g_L	10 nS	Leak conductance
E_L	-70 mV	Leak reversal potential
V_θ	-50 mV	Threshold potential
V_r	-58 mV	Post-spike reset potential
E_E	0 mV	Excitatory reversal potential
E_P	-85 mV	Inhibitory reversal potential from PV interneurons
τ_{syn}	5 ms	Synaptic conductance time constant
N_E	4000	Number of excitatory neurons
N_P	1000	Number of PV-interneurons
p	0.1	Connection probability between two cells
J	0.1 nS	EPSG amplitude $E \to E$ and $E \to$PV in the ISN
	0.01 nS	EPSG amplitude $E \to E$ and $E \to$PV in the non-ISN
	$[0.001, 0.2]$ nS	Range of EPSG amplitudes $E \to E$ and $E \to$PV for the systematic quantification across network operating regimes
g_{rc}	8	Multiplicative factor for IPSG amplitude PV$\to E$ and PV$\to$ PV
J_s, J_b	0.5 nS	EPSG amplitude LGN, BKG $\to E$ and scale for EPSG LGN $\to$ PV
g_{fw}	2	Multiplicative factor for strong feedforward EPSG LGN $\to$ PV
R_x	1000 Hz	Rate of feedforward input spike trains LGN $\to E$ and LGN $\to$ PV and of background input spike trains BKG $\to E$
δ_E	$[0.5, 1.0]$	Depression of feedforward synapses onto excitatory neurons
δ_P	$[0.5, 1.0]$	Depression of feedforward synapses onto PV-interneurons
ζ_{PE}	$[1.0, 1.5]$	Potentiation of recurrent synapses from excitatory neurons to PV-interneurons
ζ_{EP}	$[1.0, 1.5]$	Potentiation of recurrent synapses from PV-interneurons to excitatory neurons
ρ_{EI}	$[1.0, 1.5]$	Increase of feedforward E/I-ratio

and takes the same value for all neurons i belonging to that population. Further denoting the spike train of neuron j in population B by $Y_j^B(t)$ and the feedforward input spike train to neuron i by $X_i(t)$, we can explicitly write the two components of the total input current as:

$$\begin{aligned} I_{i,\mathrm{rec}}^{AB}(t) &= \sum_j C_{ij,\mathrm{rec}}^{AB}(\varepsilon * Y_j^B)(t)\big[E_B - V(t)\big] \\ I_{i,\mathrm{ext}}^{A}(t) &= C_{i,\mathrm{ffw}}^{A}(\varepsilon * X_i)(t)\big[E_E - V(t)\big], \end{aligned} \tag{3.5}$$

where $\varepsilon(t)$ is an exponentially decaying synaptic kernel, such that a presynaptic spike causes an instantaneous jump of the post-synaptic conductance, followed by a decay back to zero. The operation $*$ denotes convolution:

$$(\varepsilon * x)(t) = \int_{-\infty}^{\infty} \varepsilon\big(\tilde{t}\big) x\big(t - \tilde{t}\big) d\tilde{t}, \tag{3.6}$$

where x is either of the X_i or the Y_j^B.

The network with this connectivity can switch between the ISN and the non-ISN regimes dependent on the coupling scale J, which is central to my investigation of the effects of MD-induced plasticity. As a defining property of the ISN, I investigated the paradoxical effect dependent on J. I simulated the network in an initial setup to measure the baseline rates. Then, I simulated the network with additional input to PV-interneurons and measured the relative rate change of the PV-interneurons. Doing this over a range of coupling scales J, I extracted the curve of PV firing rate changes in response to additional external input as a function of the coupling scale J (Fig. 3.1D). For small values of J, the PV-interneurons increased their rate in response to this additional stimulus, showing the network is a non-ISN. For high values of J, the PV-interneurons decreased their rate paradoxically in response to this additional stimulus, showing the network is an ISN.

3.2.2 Implementation of MD-induced plasticity

I implemented the plastic changes following brief MD by introducing four parameters which I varied systematically (Fig. 3.2). After two days of MD, feedforward synapses from LGN to L4 in V1m onto both excitatory and PV neurons are depressed [3] (similarly, also feedforward synapses from L4 to L2/3 within visual cortex in binocular V1 are depressed [94]). To implement this depression, I changed the synapses from external input onto excitatory neurons (J_s) through multiplication with a factor $\delta_E < 1$, such that the weight after MD-induction was $\delta_E J_s$. Similarly, depression of feedforward inputs to PV-interneurons (original weight $g_{fw} J_s$) was implemented through multiplication with a factor $\delta_P < 1$, such that the synaptic weight of external feedforward input after MD-induction was $\delta_P g_{fw} J_s$. Within the recurrent circuit in layer 4 of V1m, two days of MD lead to potentiation of synapses from excitatory onto PV neurons and the reciprocal synapse type from PV onto excitatory neurons [82]. I implemented this potentiation in the network through a multiplication of the

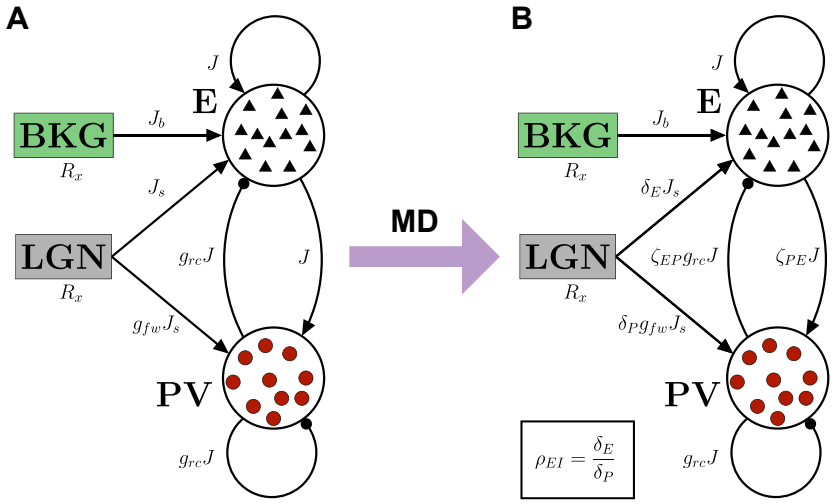

Figure 3.2: Connection strengths in the spiking network before and after MD **(A)** Connectivity in the baseline network before application of MD. **(B)** Connectivity in the network after application of MD.

respective synapses: synapses from excitatory onto PV neurons were multiplied with a factor $\zeta_{PE} > 1$, such that the initial weight J changed to $\zeta_{PE}J$ after MD induction. The reciprocal synapse from PV onto excitatory neurons was multiplied by a factor $\zeta_{EP} > 1$, such that the initial weight $g_{rc}J$ changed to $\zeta_{EP}g_{rc}J$. Taken together, plasticity after brief MD changed the recurrent and feedforward weights (Eqs. 3.1 and 3.2) to be:

$$\boldsymbol{W}^{bl}_{\text{rec}} = \begin{pmatrix} J & g_{rc}J \\ J & g_{rc}J \end{pmatrix} \quad \Longrightarrow \quad \boldsymbol{W}^{md}_{\text{rec}} = \begin{pmatrix} J & \zeta_{EP}g_{rc}J \\ \zeta_{PE}J & g_{rc}J \end{pmatrix} \tag{3.7}$$

and

$$\boldsymbol{W}^{bl}_{\text{ffw}} = \begin{pmatrix} J_s + J_b \\ g_{fw}J_s \end{pmatrix} \quad \Longrightarrow \quad \boldsymbol{W}^{md}_{\text{ffw}} = \begin{pmatrix} \delta_E J_s + J_b \\ \delta_P g_{fw} J_s \end{pmatrix}, \tag{3.8}$$

where *bl* denotes the baseline state before any MD-related synaptic changes (Fig. 3.2A) and *md* denotes the state after MD-induction (Fig. 3.2B). Note the difference in how MD affects only a part of the input synapses to excitatory neurons from the LGN, while BKG input-synapses are unaffected. In contrast to this, all of the input synapses to PV-interneurons are affected by MD since they are exclusively driven by LGN input.

3.3 Landscape of network responses to MD-induced plasticity

Remark: The results of this section (Figs. 3.3 and 3.4 and Eqs. 3.9 and 3.10) have been published before with identical content in [1]. The figures have been modified for consistency, notation in the equations has been adapted for consistency.

To understand the dynamical effects of the synaptic changes, as well as their size and interactions, I split the four types of synaptic changes into three different classes: (1) the depression of feedforward inputs onto excitatory and PV-interneurons (δ_E, δ_P); (2) the potentiation of recurrent connections between excitatory and PV-interneurons (ζ_{EP}, ζ_{PE}); and (3) where I combined feedforward and recurrent changes in a reduced manner to investigate their interaction. Feedforward depression was reduced to an upward shift in the feedforward E/I-ratio, similar to what has been measured by Miska and colleagues in synapses from LGN in the thalamus to L4 in V1m [3], and in feedforward synapses from L4 to L2/3 within the visual cortex [94]. I defined this direct feedforward E/I-ratio as:

$$\rho_{EI} = \frac{\delta_E}{\delta_P}. \tag{3.9}$$

Thus, it combines both parameters for feedforward plasticity. In order to reduce this to a one-dimensional measure, I implemented the change in ρ_{EI} through varying δ_P for fixed δ_E. This increase of the feedforward E/I-ratio was paired with the potentiation of the recurrent excitatory drive onto PV-interneurons (ζ_{PE}), also measured in [3], which effectively yields a decrease of the recurrent E/I-ratio, giving the parameter pair (ζ_{PE}, ρ_{EI}) for interacting feedforward and recurrent plasticity. For each of these three pairs of either feedforward, recurrent, or interacting feedforward and recurrent synaptic changes, I simulated networks with combinations of potentiation or depression values on a two-dimensional 21×21 grid. The resulting simulated firing rates after the induction of MD in the circuit were normalized by the baseline firing rate before any of the synaptic changes to investigate the *firing rate fold changes* caused by MD-induced synaptic plasticity.

For each pair (x, y) of synaptic changes in the two-dimensional parameter spaces, the matrix of fold-changes of the rates of each population $A \in \{E, P\}$ relative to baseline is:

$$\Psi^A(x, y) = \frac{R^A_{md}(x, y)}{R^A_{bl}}, \tag{3.10}$$

where the variables x and y depend on the type of plasticity studied (feedforward: $(x, y) = (\delta_E, \delta_P)$, recurrent: $(x, y) = (\zeta_{EP}, \zeta_{PE})$, interacting feedforward and recurrent: $(x, y) = (\zeta_{PE}, \rho_{EI})$). This matrix yields a systematic view of how the different synaptic changes quantitatively influence activity as they interact with each other. I plotted the resulting landscape of firing rate fold changes in the respective parameter spaces as a heatmap. I studied these emerging landscapes for the different cell

types (excitatory and PV) and in different operating regimes of the network (ISN or non-ISN).

3.3.1 Response in the inhibition stabilized network

I implemented the network in the ISN-regime by setting a sufficiently large coupling scale $J = 0.1$ nS (Fig. 3.3A, and Table 3.1). Stabilizing inhibitory connections scale accordingly, with a relative scale $g_{rc} = 8$ that sets the network in a stable activity state with dominant inhibition. MD-induced plasticity leads to distinct structures of the response landscape in feedforward (Fig. 3.3B), recurrent (Fig. 3.3C) and interacting feedforward and recurrent synaptic changes (Fig. 3.3D). All three classes of synaptic changes lead to substantial modulation of firing rates in the ISN, making them a suitable substrate for the significant activity variation observed during brief-MD *in vivo*. Second, regardless of this strong modulation of activity, all the firing rates stay finite. Thus, the MD-induced synaptic changes do not push the spiking network into an intrinsically unstable regime.

After depression of feedforward synapses (Fig. 3.3B), we find both facilitation (red) and suppression (blue) of activity of both excitatory and PV neurons relative to baseline. A linear relationship between the depression of input to excitatory neurons (δ_E) and the depression of input to PV-interneurons (δ_P) distinguishes the regions of suppression and facilitation. This boundary crosses the point in the parameter space where no synaptic changes have been induced, corresponding to the baseline state (upper right corners in Fig. 3.3B). Remarkably, an extended facilitatory parameter space exists in this MD-induced feedforward plasticity, especially for PV-interneurons, considering that feedforward inputs only depress. How can this depression of feedforward inputs lead to an increase in activity? The heatmaps show that facilitation of activity rests on the depression of feedforward inputs arriving at PV-interneurons (δ_P, Fig. 3.3B): for both cell types, decreasing δ_P leads to facilitation of activity. In excitatory neurons, this is at first intuitive: depression of external inputs into the local inhibitory interneurons releases the local excitatory population of inhibition and thus leads to higher excitatory firing. This intuition would naively require that the firing rate of the PV-interneurons decreases. Contrary to this intuition, however, the activity of PV-interneurons themselves increases as their external drive decreases. Here the fact that the network is an ISN comes into play. The facilitation of both excitatory and PV firing rates through a decrease of the drive to PV-interneurons is a form of the paradoxical effect [114]. Decreasing the input to PV-interneurons facilitates excitatory neurons dynamically first through a depression of the PV-firing. In the ISN, this facilitation of excitatory firing rates is so strong that it overcomes the decrease of the input to PV-interneurons, such that their firing increases, which leads to a final state of increased firing rates of both populations.

Thus, there is a tight connection between the paradoxical effect and the response of PV-interneurons to MD-induced feedforward plasticity. This connection explains how depression of feedforward inputs to the PV-interneurons can yield facilitation of

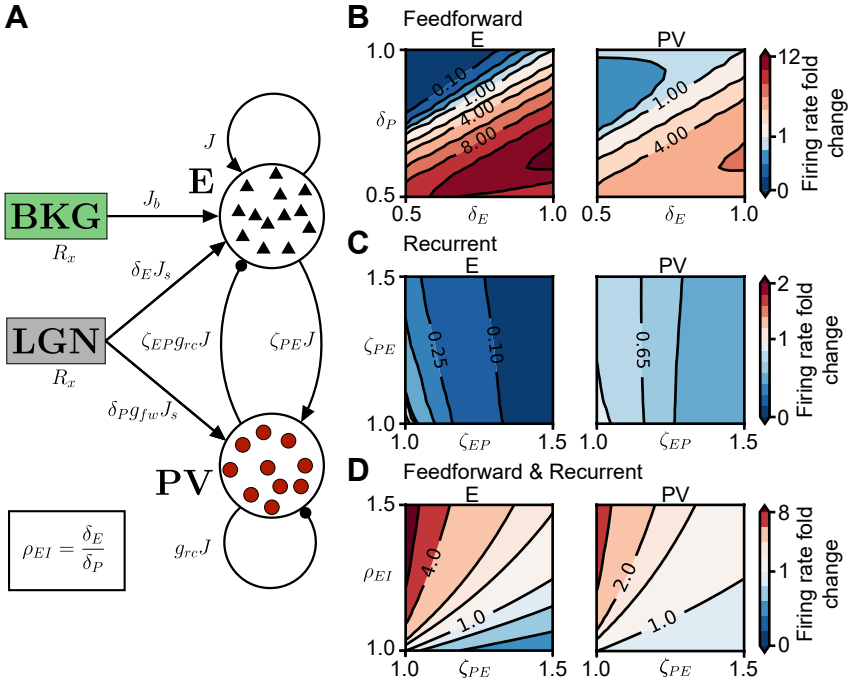

Figure 3.3: Response to synaptic changes induced by brief MD in a spiking network in ISN-regime. (A) Network schematic showing synaptic connections among neurons with J denoting the overall coupling scale, g_{rc} the dominance of recurrent inhibition and g_{fw} the dominance of feedforward inhibition. The parameters to model MD-induced synaptic plasticity are: depression of feedforward drive to excitatory neurons (δ_E) and to PV-interneurons (δ_P), and potentiation of recurrent excitation to PV-interneurons (ζ_{PE}) and of recurrent inhibition from PV to excitatory neurons (ζ_{EP}). Values of network parameters and ranges for synaptic changes due to MD are provided in Table 3.1. **(B)** Network firing rate in the (δ_E, δ_P) plane as fold-change of baseline firing rate (top right corner where $\delta_E = \delta_P = 1$) for excitatory neurons (left panel) and PV-interneurons (right panel). **(C)** Network firing rate in the (ζ_{EP}, ζ_{PE}) plane as fold-change of baseline firing rate (bottom left corner where $\zeta_{EP} = \zeta_{PE} = 1$). **(D)** Network firing rate in the (ζ_{PE}, ρ_{EI}) plane as fold-change of baseline firing rate (bottom left corner where $\zeta_{PE} = \rho_{EI} = 1$). Here $\rho_{EI} = \delta_E/\delta_P$ is the feedforward E/I ratio. Modified from [1].

their firing rate. Comparing the two heatmaps for excitatory and PV neurons, one can also see that the distribution of facilitatory and suppressive parameter spaces in the two populations are very similar (Fig. 3.3B). Specifically, when excitatory neurons increase their rate, so do PV-interneurons and vice versa. Along the δ_P-axis, this co-modulation stems directly from the paradoxical effect. However, the tightly aligned co-modulation of excitatory and PV firing rates in the whole plane is a generalization of the paradoxical effect.

The potentiation of recurrent connections between excitatory and PV neurons has a purely suppressive effect in all of the parameter range and for both cell types (Fig. 3.3C). Any combination of potentiation of synapses from excitatory

neurons to PV-interneurons (ζ_{PE}) and of the reverse synapse from PV-interneurons to excitatory neurons (ζ_{EP}) suppresses activity below the baseline activity (lower left corners Fig. 3.3C). For the excitatory population, it is clear that potentiation of both synapse types increases the inhibition arriving at the excitatory cells. This increase of inhibition happens either directly, through an increase of the inhibitory weights impinging on the excitatory cell (ζ_{EP}), or indirectly, through an increase of the excitatory drive of the local inhibition (ζ_{PE}). However, just as in the feedforward case, the intuition for suppression of excitatory activity through the latter pathway (ζ_{PE}) would rest on an increase of the firing rate of PV-interneurons. But also the PV-interneurons are suppressed throughout the range of recurrent potentiation in either of the parameters, most importantly also through potentiation of their recurrent excitatory drive (ζ_{PE}). This has not been studied in the previous literature. My investigation in the following sections showed that this suppression of activity and, consequently, the aligned co-modulation of excitatory and PV firing rates for potentiation of recurrent synapses, is based on the same mechanism as the paradoxical suppression of interneurons due to feedforward depression.

Upon careful inspection, however, there is a slight discrepancy between the qualitative behaviors of excitatory and PV firing rates: whereas in the excitatory neurons both parameters have a purely suppressive effect, such that the contours of equal firing rates lean to the left (Fig. 3.3C, left), the contour for larger values of ζ_{EP} and ζ_{PE} leans to the right for PV-interneurons (Fig. 3.3C, right). Thus, from the suppressed state of PV firing rates there is a slight facilitation at high parameter values. This is so because excitatory neurons approach rectification in this range, and is consistent with the paradoxical effect. Suppression through the paradoxical effect in PV-interneurons is brought about by a withdrawal of recurrent excitation through suppression of the excitatory firing rates. However, close to rectification excitatory firing rates can only be suppressed very little, such that the additional input, in this case through increased weight, can overcome the effect of the suppression of the excitatory firing rate. This was studied before for external feedforward input [135]. Nonetheless, the local facilitatory effect is not strong enough to bring about facilitation of PV firing rates above the baseline level. Thus, the firing rates of PV-interneurons are suppressed in all of the plane of recurrent potentiation, just as the firing rates of excitatory neurons.

For interacting feedforward and recurrent changes, I increased ρ_{EI} by decreasing δ_P while keeping δ_E fixed, yielding facilitation of both excitatory and PV rates (Fig. 3.3D). For the recurrent pathway, I made use of the potentiation of synapses from excitatory to PV-interneurons (ζ_{PE}) to decrease the recurrent E/I-ratio. This potentiation leads to a depression of activity in the network (Fig. 3.3C). These interacting feedforward and recurrent changes of the E/I-ratio show an antagonistic role of the synaptic changes in the different pathways, with a non-linear boundary distinguishing the parameter spaces of suppression and facilitation (Fig. 3.3D).

An apparent feature of all the types of synaptic changes is that the firing rates of excitatory and of PV-interneurons are closely co-modulated: when excitatory

rates increase or decrease, the same happens with PV rates. This tight coordination appears to depend on dynamical processes related to the paradoxical effect.

This phenomenon is deeply problematic when we consider the behavior of the two cell classes *in vivo*. The experiments by Hengen and colleagues explicitly point to a temporally-distinct firing-rate dynamics of the two populations over the time course of MD. On the first day of MD, PV firing rates drop to a level substantially below their baseline while excitatory firing rates remain at their baseline (Fig. 2.3, see [54]). On the second day of MD, PV firing rates recover back to baseline, being up-regulated from their depressed state, while excitatory firing rates become suppressed substantially below their baseline.

Qualitatively, it is not clear how these modulations can be achieved in the ISN, especially the transition from the first to the second day of MD requires modulation of excitatory and PV firing rates in opposite directions.

Quantitatively, we shall take into account that the boundary between facilitation and suppression of firing rates in the feedforward and interacting feedforward and recurrent cases is very close in excitatory and PV neurons, but not exactly the same in the example case I have shown (Fig 3.3). Some modulation of the firing rates in different directions would be possible, but only in tightly constrained parts of the parameter spaces. The regulation of firing rates could potentially be achieved, but it would require a coordinated fine-tuning of all the synaptic changes. It is possible that such fine-tuning emerges from orchestrated plasticity mechanisms in the cortex. However, a more promising direction would be if the network itself allowed an independent modulation of excitatory and PV firing rates more generically in large portions of the parameter spaces. Given the connection between the alignment of responses of the two populations with the paradoxical effect in the ISN, a potential network structure lending itself to generate such independent modulation of firing rates is a non-ISN that does not show the paradoxical effect.

3.3.2 Response in the non-inhibition stabilized network

To study the response in the non-ISN, I chose a small $J = 0.01$ nS (compared to $J = 0.1$ nS for the ISN). In this non-ISN, indeed, a very different picture of the responses to all the different types of MD-induced synaptic changes emerges (Fig. 3.4). While responses of each cell type in each type of plasticity change to some degree, the response structure changes more substantially in PV. Specifically, the firing rate changes of excitatory and PV neurons become quite different from each other in the non-ISN.

Excitatory neurons have areas in the parameter space in which depression or facilitation can be induced, and the two regions are separated by a linear boundary (Fig. 3.4B, left). Qualitatively this is similar to the ISN. However, this linear boundary shifts substantially while also the strength of the firing rate modulations is decreased.

The response of PV-interneurons, on the contrary, changes to purely depressive (Fig. 3.4B, right). In all of the parameter range, the firing rate of PV-interneurons

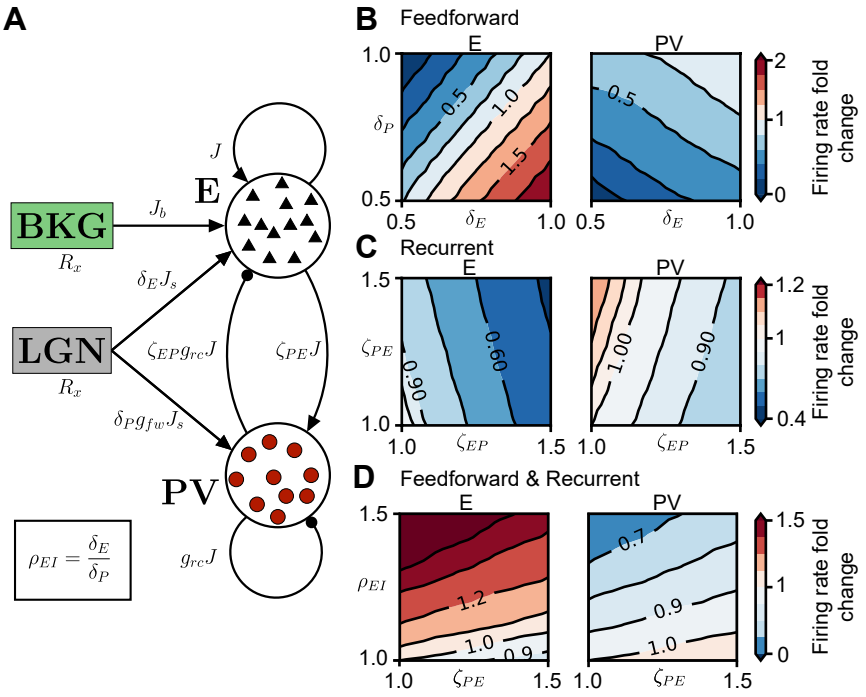

Figure 3.4: Response to synaptic changes induced by brief MD in a spiking network in non-ISN regime. (A) Network schematic showing couplings among cells and parameters to model MD-induced synaptic changes (see Fig. 3.3). In the non-ISN, J has a smaller value than in the ISN (shown responses for $J = 0.01$ nS) **(B)** Network firing rate in the (δ_E, δ_P) plane as fold-change of baseline firing rate for excitatory neurons (left panel) and PV-interneurons (right panel). In the non-ISN there is no facilitation of activity of PV. **(C)** Network firing rate in the (ζ_{EP}, ζ_{PE}) plane as fold-change of baseline firing rate. Facilitation of activity of PV-interneurons is present in the non-ISN. **(D)** Combined feedforward and recurrent plasticity: changing the recurrent E/I-ratio through the parameter ζ_{PE}, changing the feedforward E/I-ratio as $\rho_{EI} = \delta_E/\delta_P$. Network firing rate as fold-change of baseline firing rate in (ζ_{PE}, ρ_{EI}) plane. In contrast to the ISN (see Fig. 3.3), here in the non-ISN, the rate changes are opposite over large fractions of the plane. Modified from [1].

depresses below their baseline. Crucially, I found that also depression of feedforward inputs to PV ($\delta_P < 1$) suppresses PV firing rates. This behavior is opposite to the firing rate changes predicted from the paradoxical effect and conforms with the absence of the paradoxical effect in the non-ISN. Comparing the responses of excitatory and PV neurons in the non-ISN, we find large parameter spaces that enable independent or opposite directions of firing rate modulation for the two cell classes. A natural way to interpret the emerging picture in the non-ISN would be to produce a state of suppressed PV firing rates combined with excitatory firing rates that remain at baseline on the first day of MD (MD1) by making use of feedforward depression.

In response to recurrent plasticity, excitatory neurons show suppression of firing rates in all of the parameter space (Fig. 3.4C, left), similar to the response in the ISN (Fig. 3.3C). For PV-interneurons, a facilitatory area in the parameter space emerges in the non-ISN that was not present in the ISN (Fig. 3.4C, right). The combination of this facilitatory parameter regime for PV firing rates together with the purely suppressive response of excitatory neurons is again a possible parameter regimen that enables the independent modulation of firing rates in the two cell classes. Extending the idea to produce the state at MD1 using the feedforward depression, it would be possible to recover PV firing rates through recurrent plasticity (in the facilitatory area) that necessarily also leads to suppression of excitatory firing rates, producing a state reminiscent of the second day of MD.

The results for interacting feedforward and recurrent plasticity in the non-ISN complete this picture of independent or opposite modulation of excitatory and PV firing rates during brief MD (Fig. 3.4D). In this space of increased feedforward E/I-ratio (ρ_{EI}) and decreased recurrent E/I-ratio (ζ_{PE}), firing rates of the two cell classes behave exactly opposite. Wherever the network is instantiated in the parameter space, it can produce either facilitation of excitatory firing rates paired with suppression of PV firing rates (in the largest fraction of the space), or suppression of excitatory firing rates paired with facilitation of PV firing rates.

In conclusion, I found that a network of randomly and sparsely connected excitatory and PV neurons can be a suitable candidate to explain the independent modulation of firing rates *in vivo*, granted it is a non-ISN. In general, there seems to be a much larger parameter space that enables independent modulation of excitatory and PV firing rates in the non-ISN. In order to validate this interpretation and the role of the ISN/non-ISN distinction, I used a more general account that quantifies the response structure in the two regimes from a broader perspective.

3.4 Quantification of network response structure to MD-induced plasticity across operating regimes

Remark: Some of the results of this section (Fig. 3.5 in Sec. 3.4.1) have been published before with identical content in [1].

The results in the ISN and non-ISN make a role of the operating regime for the tight coordination of firing rate changes in excitatory and PV neurons very likely. However, the two networks studied so far are specific cases for coupling scales $J = 0.1$ nS (ISN) and $J = 0.01$ (non-ISN). The distinction between non-ISN and ISN is an important dichotomy. However, it is imposed on a continuum of network structures, ranging from a very weakly coupled non-ISN to a non-ISN or an ISN close to the transition between the two regimes, up to a very strongly coupled ISN. A priori, it is conceivable that the overall coupling scale J shapes the network responses to synaptic changes induced by MD beyond the simple distinction into

two regimes. To investigate the general role of the coupling scale, I varied J in more detail on a large range, spanning weakly coupled to strongly coupled networks, covering both non-ISN and ISN regimes. For all these values of J, I generated the full landscapes of responses for all three types of synaptic changes (feedforward, recurrent and interacting feedforward and recurrent, as in Fig. 3.4B-C). To quantify the firing rate changes globally, I introduced multiple scalar measures that describe the overall distribution of firing rate changes in the two neural populations and their relationship.

The first set of measures quantifies the distribution of facilitation and suppression in the parameter spaces by measuring the area of parameter space in which activity is facilitated (Sec. 3.4.1). I calculated this normalized to the total range of parameters in the simulations, such that we can extract the relative area of facilitation and suppression at once. Additionally, I quantified the relation of the respective distributions of facilitatory and suppressive parameter spaces in excitatory and PV neurons by calculating the normalized area in which both populations are modulated in the same direction (either facilitation or suppression). This quantification of the fractional area of co-modulation describes how tightly coordinated the responses of the two populations are in either direction (facilitatory or suppressive) and thus how finely tuned synaptic changes must be to allow the independent modulation of firing rates in excitatory and PV neurons measured *in vivo*. These quantifications show that the coordination of the excitatory and PV firing rate changes is indeed a general property of the ISN, whereas the non-ISN generally allows for more independent firing rate changes.

A caveat with these measures is that they do not capture the size of the effects, i.e., the amplitude of firing rate changes, but only their direction. This reduction is useful to give intuitive measures of the general co-modulation of the two populations but at the cost of leaving out important information on the strength of modulations. As a second set of measures, therefore, I quantified properties of the gradient of responses in the respective parameter spaces, for each neural population separately and for their relation (Sec. 3.4.2). The absolute direction of the gradient throughout the parameter space is of limited interest, giving only local relations of the firing rate changes. However, the length of the gradient measures how strongly firing rates are modulated overall in the respective plane. I quantified the average lengths of the gradients of firing rate changes for either neural population in all three types of MD-induced plasticity. This quantification uncovers a problem of the non-ISN: even though it permits independent modulation of the excitatory and inhibitory firing rates, the size of the firing rate modulations is very small for the small coupling scales in the non-ISN. Thus, the non-ISN calls for very large MD-induced synaptic changes in order to lead to sufficient suppression of PV on the first day of MD, and, likewise, their recovery with sufficient simultaneous suppression of excitatory rates on the second day of MD. Potentially, the large synaptic changes called for in the non-ISN are beyond what is biologically realistic.

As a third measure derived from the gradient of responses, I also calculated the angle between the gradients of the two populations (Sec. 3.4.2). This relative angle

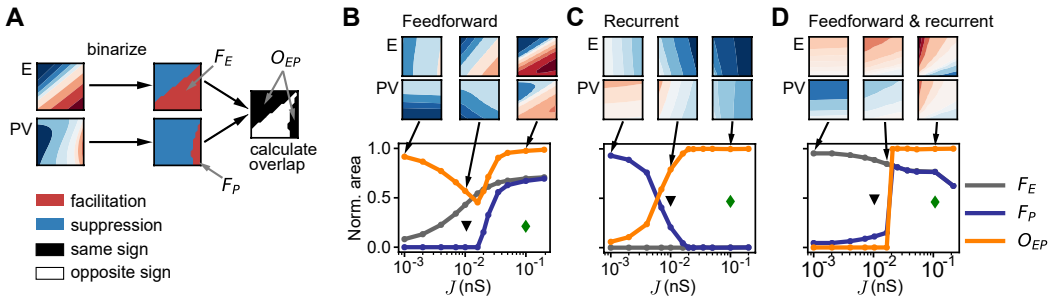

Figure 3.5: Parameter space of facilitation and suppression as a function of the coupling scale J. (A) Schematic to determine the overlap of facilitating and suppressive response areas of excitatory neurons and PV-interneurons. The thresholded plane of responses (as in Figs. 3.3 and 3.4) for both neuron types is used to compute the total facilitating area and quantify how closely excitatory and PV firing rates follow each other through the overlap of facilitation and suppression. **(B)** Network with feedforward depression only (Figs. 3.3B and 3.4B). Fractional area of facilitation for excitatory neurons (F_E, gray), PV-interneurons (F_P, blue) and the overlap between excitatory and PV response areas (O_{EP}, orange) as a function of the overall coupling scale J. Green diamond shows the value of J used in Fig. 3.3 (ISN), black triangle shows the value of J used in Fig. 3.4 (non-ISN). **(C)** Same as (B) for recurrent potentiation only (Figs. 3.3C and 3.4C). **(D)** Same as (B) for interacting feedforward depression and recurrent potentiation (Figs. 3.3D and 3.4D). Modified from [1].

is another measure that quantifies the coordination between the two populations, backing up the understanding gained from studying the overlap of the fractional area of co-modulation.

3.4.1 Parameter space of co-modulation of activity during MD

To extract the distribution of facilitation and suppression, I thresholded the plane of firing rate fold changes of each population ($\Psi^A(x, y)$ for $A \in \{E, P\}$, Eq. 3.10) at the boundary between facilitation and suppression (Fig. 3.5A). Firing-rate fold changes above one ($\Psi^A(x, y) > 1$), corresponding to facilitation of firing rates, are assigned the value $+1$ (red region in the binarized plane in Fig. 3.5A). Normalized firing rates below one ($\Psi^A(x, y) < 1$), corresponding to suppression of activity, are assigned the value -1 (blue regions in the binarized plane in Fig. 3.5A). The fractional area of facilitation is then calculated by summing all positive entries in this thresholded plane and dividing it by the total number of entries. I did this for excitatory neurons and PV-interneurons independently for the three different types of MD-induced plasticity: feedforward (Fig. 3.5B), recurrent (Fig. 3.5C), and interacting feedforward and recurrent plasticity (Fig. 3.5D). The black triangle in Fig. 3.5B-D marks the value of J used in the non-ISN (Fig. 3.4), the green diamond in Fig. 3.5B-D marks the value of J used in the ISN (Fig. 3.3).

The fractional area of facilitation for excitatory neurons as a function of the coupling scale J is shown in gray (F_E). In response to feedforward plasticity, I found

a continuous increase of F_E, followed by saturation (Fig. 3.5B). For very small values of J, the change of the feedforward input dominates the response. Since this change is depressive, the primary direction of the modulation of firing rates is suppression. Nonetheless, even for the smallest value of J studied, feedforward plasticity can induce facilitation in specific regions of the parameter space in excitatory neurons due to the recurrent interaction with PV-interneurons. The fractional area of facilitation for excitatory neurons will only reach zero when the network is completely decoupled ($J = 0$ nS, not shown due to logarithmic scale in J).

For large values of J, the fractional area of facilitation does not reach one, corresponding to pure facilitation, but saturates at a value $F_E < 1$. This hints at a fixed boundary between facilitation and suppression even for $J \to \infty$. Intuitively, it is clear that the depression of direct excitatory inputs into excitatory neurons ($\delta_E < 1$) on its own must suppress activity to some degree. This depression can only be counteracted by depression of excitatory inputs to inhibitory neurons ($\delta_P < 1$), given that the two populations interact recurrently to a sufficient level. Note that this facilitation of excitatory responses through a depression of inputs to inhibition is independent of the operating regime of the network (non-ISN or ISN) and only requires recurrent interactions to be present at all.

The fractional area of facilitation of PV-interneurons is shown in dark blue (F_P). In response to feedforward plasticity, I found this area to be constantly zero until a certain threshold in J was reached (Fig. 3.5B). Above this specific coupling scale F_P steeply increases until it also saturates at the same value as F_E. Crucially, the value of J where the fractional area of facilitation first becomes non-zero coincides with the value of J at which the network model transitions into the ISN regime and first expresses the paradoxical effect (Fig. 3.1D). From the paradoxical effect, we can also derive the intuition underlying this emergence of a facilitatory area in PV responses to MD-induced feedforward plasticity (see previous section).

I calculated the area of co-modulation of excitatory and PV firing rates as the fractional overlap between facilitatory and suppressive areas in the two cell classes. Starting from the thresholded, binarized planes for both cell classes, in which all entries have values $+1$ (facilitation) and -1 (suppression), I multiplied each entry in the excitatory response plane $\Psi^E(x, y)$ with the corresponding entry in the PV response plane $\Psi^P(x, y)$ (element-wise product). When the firing rates of the two cell classes are modulated in the same direction, either both being facilitated ($+1$ and $+1$) or both being suppressed (-1 and -1), the product of the two entries has the value $+1$ (black region in the combined plane in Fig. 3.5A). If the firing rates of the two cell classes are modulated in opposite directions ($+1$ and -1), the product has the values -1 (white region in the combined plane in Fig. 3.5A). I calculated the fractional area of co-modulation, or overlap between facilitatory and suppressive regions, in the two cell classes from this combined plane of co-modulation. I do this by summing up all positive entries and dividing this sum by the total number of entries in the plane. This fractional area of co-modulation as a function of the coupling scale J is shown in orange (O_{EP}).

In response to purely feedforward plasticity, this fractional area of co-modulation of the two cell classes changes non-monotonically, driven by the two different processes of the fractional area of facilitation in the two different cell classes (Fig. 3.5B). For very small values of J, the overlap is very close to one, which is a consequence of the very small areas of facilitation in both cell classes and, consequently, a common depression of the firing rates in both cell classes. Nonetheless, because of the non-zero fractional area of facilitation in excitatory neurons, the overlap is only in a part of the planes and further decreases as J increases. At the value of J for which the network transitions from non-ISN to ISN, the overlap reaches a minimum and increases again for higher values. This increase is driven by the emergence of a facilitatory region in the PV response. The discontinuity in the overall behavior of O_{EP} follows from the sudden emergence of $F_P > 0$. Together with the increase of F_P, the fractional area of co-modulation steeply increases until it saturates at $O_{EP} = 1$. For this value of the fractional area of co-modulation, the two cell classes respond in the same direction in the whole parameter space of MD-induced feedforward depression, making an independent modulation of the two firing rates as is observed *in vivo* impossible.

In response to purely recurrent plasticity, the behavior of the fractional area of facilitation in excitatory and PV neurons, and the overlap between facilitatory and suppressive areas between the two cell classes, is dominated by a single process (Fig. 3.5C). This process is the monotonic decay of facilitatory area in the response of PV-interneurons (F_P in Fig. 3.5C). For very small J, the PV firing rate almost exclusively facilitates above baseline. As J increases, this area decays until it reaches zero. Importantly, the value of J at which the fractional area of facilitation reaches zero is again the coupling scale at which the network transitions from non-ISN to ISN, and at which a facilitatory area for PV firing rates first emerges for feedforward plasticity. Thus, the presence of facilitation in response to recurrent plasticity in PV firing rates ($F_P > 0$) depends on the network being a non-ISN. The purely suppressive effect of recurrent plasticity on PV firing rates in the ISN ($F_P = 0$) is governed by a mechanism related to the paradoxical effect. In feedforward and recurrent plasticity studied separately, the condition for the presence of a facilitatory area is opposite.

The fractional area of facilitation for excitatory neurons constantly stays at zero over the whole range of J in response to recurrent plasticity (F_E in Fig. 3.5C). In both operating regimes of the network, excitatory firing rates exclusively depress below baseline in response to recurrent plasticity. This fortifies the conclusion from the two planes for the ISN and ISN (Figs. 3.3 and 3.4). From the combination of the two behaviors of the facilitatory area in excitatory firing rates, which constantly stays at zero, and of the facilitatory area in PV-interneurons, which monotonically decays to zero, the total area of co-modulation monotonically increases to one and stays there for all higher coupling scales. Thus we find in the ISN that excitatory and PV firing rates can exclusively be modulated in the same direction through recurrent plasticity, while in the non-ISN, an independent modulation of the firing rates is possible.

In response to interacting feedforward and recurrent plasticity, the fractional areas of facilitation in excitatory and PV neurons and the fractional area of the co-modulation of the two cell classes all show a pronounced dichotomy between the non-ISN and ISN (Fig. 3.5D). The area of co-modulation (O_{EP}) is constantly at zero for all values of the coupling scale J below the transition to ISN, indicating opposite directions of firing rate modulations of excitatory and PV neurons in the non-ISN over the whole parameter space. As the network transitions to ISN, O_{EP} jumps to one and remains there for all higher values of J, indicating fully aligned co-modulation of the firing rates in the two cell classes.

This jump emerges because, for both parameters varied in the space of interacting feedforward and recurrent plasticity, the response of the PV-interneurons reverses its direction as the network transitions from non-ISN to ISN. In the non-ISN, the increase of the feedforward E/I-ratio suppresses PV firing rates while the decrease of the recurrent E/I-ratio meditated by ζ_{PE} facilitates PV firing rates (Fig. 3.4D). In the ISN, the increase of the feedforward E/I-ratio facilitates PV firing rates while the decrease of the recurrent E/I-ratio meditated by ζ_{PE} suppresses PV firing rates (Fig. 3.3D). The combination of the two reversals of the responses in the two parameters studied in the interacting feedforward and recurrent plasticity generates a sudden in the response of PV firing rates.

In summary, this quantification shows that the tight coordination of excitatory and PV firing rates is a property of the ISN with large coupling scales. The ISN generally permits only narrowly constrained parameter ranges for independent modulation of the firing rates in the two cell classes, which even shrinks to zero the larger J is. In the non-ISN with small J, there are generally larger areas in the parameter spaces that enable an independent modulation of excitatory and PV firing rates.

3.4.2 Response gradients of the MD-induced firing rate changes

The quantification in the previous section only takes the direction of firing rate modulations into account. For a complete picture, it is necessary to also investigate the strengths of the firing rate modulations. The different coupling scales in the different operating regimes are expected to influence not only the distribution of the direction of firing rate modulations but also the size of the modulations. In order to globally quantify the size of firing rate modulations in the different types of MD-induced plasticity, I used the average length of the gradient in the different response planes for each of the cell classes.

I calculated the gradient of the firing rate fold changes numerically from the firing rate fold changes $\Psi^A(x, y)$. The two components of the gradient for the population $A \in \{E, P\}$ at each point (x, y) in the respective parameter space are:

$$\nabla\Psi^A(x,y) = \begin{pmatrix} \dfrac{\Psi^A(x+\Delta,y) - \Psi^A(x,y)}{\Delta} \\ \\ \dfrac{\Psi^A(x,y+\Delta) - \Psi^A(x,y)}{\Delta} \end{pmatrix}, \tag{3.11}$$

where Δ is the spacing of the grid underlying the numerically generated plane of firing rate fold changes $\Psi^A(x, y)$. To calculate the average lengths of the gradients, I first calculated the lengths of the gradients for each neuron type at each point in the respective parameter space

$$l_A(x, y) = \left|\nabla\Psi^A(x, y)\right|, \tag{3.12}$$

where $A \in \{E, P\}$ denoted the respective population. I then averaged over all positions (x, y) in the respective parameter space:

$$L_A = \left\langle l_A(x, y)\right\rangle_{(x,y)}. \tag{3.13}$$

In Fig. 3.6 the averaged lengths of the gradients were plotted over the range of coupling scales J for the excitatory (L_E, green) and PV neurons (L_P, light blue).

For pure feedforward plasticity, the average length of the gradient monotonically increases for both cell classes (Fig. 3.6B). Crucially, for both cell classes it is very small for small coupling scales J. Essentially, this is true throughout the range of J that corresponds to the non-ISN (black triangle in Fig. 3.6B marks $J = 0.01$ nS used in Fig. 3.4). The lengths then steeply increase for intermediate to large J that span the ISN regime (green diamond in Fig. 3.6B marks $J = 0.1$ nS used in Fig. 3.3).

For pure recurrent plasticity, the behavior of the gradient lengths is non-monotonic and rather complex with respect to the operating regimes (L_E and L_P, Fig. 3.6C). Generally, even though both gradient lengths start at a value close to zero for a non-ISN network with a very small coupling scale ($J = 0.001$ nS), they are of comparable size in the non-ISN and the ISN with intermediate values of J (black triangle and green diamond, Fig. 3.6C). For interacting feedforward and recurrent plasticity, the curves for the gradient lengths largely resemble those for pure feedforward plasticity, with very little firing rate fold changes in the non-ISN and large firing rate fold changes in the ISN (L_E and L_P, Fig. 3.6D).

A further measure of the global response structure is the angle between the gradients for excitatory and PV firing rate changes caused by MD-induced plasticity (Φ_{EP}, Fig. 3.6A). From the gradients for the two populations at each position in the parameter space ($\nabla\Psi^E(x, y)$ and $\nabla\Psi^P(x, y)$), I calculated the angle between those gradients at each point as:

$$\phi_{EP}(x, y) = \cos^{-1}\left(\nabla\Psi^E(x, y) \cdot \nabla\Psi^P(x, y)\right) \tag{3.14}$$

which I then averaged over all positions (x, y) in the respective parameter space:

$$\Phi_{EP} = \left\langle \phi_{EP}(x, y)\right\rangle_{(x,y)}. \tag{3.15}$$

This angle Φ_{EP} measures how tightly coordinated the firing rate modulations are in the two cell classes, similarly to the area of co-modulation in the previous section (O_{EP}, Fig. 3.5). The conclusions from this quantification are quite similar to the conclusions from the areas of co-modulation.

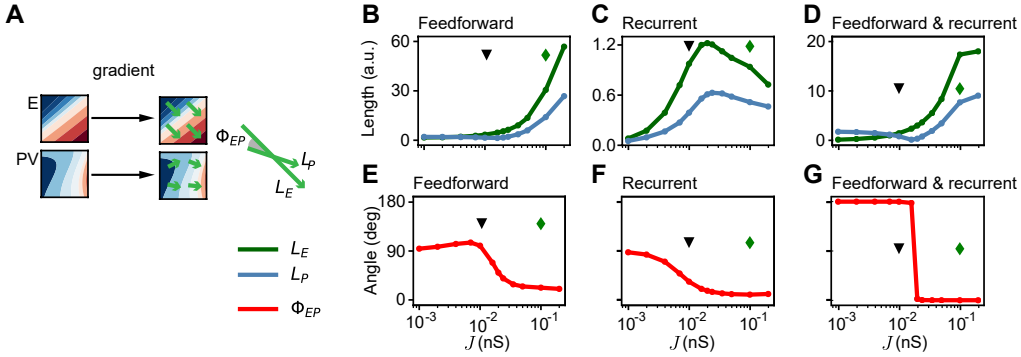

Figure 3.6: Responsiveness of excitatory and inhibitory firing rates to MD-induced plasticity as a function of the coupling scale J. **(A)** Schematic for the extraction of average gradient lengths and angle between gradients for excitatory and PV responses. We use the length of the gradient as a proxy for the magnitude of responsiveness of each population to MD-induced plasticity. The angle between the gradients denotes the amount of co-modulation of excitatory and inhibitory firing rates (0° if the firing rates follow each other as in the case of high J for feedforward and recurrent plasticity, 90° if the firing rates are independent of each other as in the case of low J for feedforward and recurrent plasticity, and 180° if the firing rates are opposite from each other as in the case of low J for interacting feedforward and recurrent plasticity). **(B)** Network with feedforward depression only. Average gradient length for excitatory responses (L_E, green) and for PV responses (L_P, blue) as a function of J. The green diamond shows the value of J used in Fig. 3.3 (ISN), the black triangle shows the value of J used in Fig. 3.4 (non-ISN). **(C)** Same as (B) for recurrent potentiation only. **(D)** Same as (B) for interacting feedforward depression and recurrent potentiation. **(E)** Network with feedforward depression only. Angle between the gradients of excitatory and PV responses (Φ_{EP}) as a function of J. **(F)** Same as (E) for recurrent potentiation only. **(G)** Same as (E) for interacting feedforward depression and recurrent potentiation.

For pure feedforward plasticity, the angle between the two gradients is around 90° throughout the non-ISN (Φ_{EP}, Fig. 3.6E). Thus, for the non-ISN regime, the firing rate modulations are on average orthogonal, independent of the area of co-modulation of the two firing rates. Interestingly, even though I found a large area of co-modulation (O_{EP}) for very small J (close to zero) in the previous section, the angle between the gradients shows that firing rates can be modulated independently within the suppressive regime of firing rate fold changes. The transition in J from non-ISN to ISN is marked by a decrease of the angle from 90° to 0°. Thus, the relative independence in the non-ISN is replaced by very tight coordination of firing rates in the ISN throughout the parameter space. Regardless of the direction in which we move in the parameter space of feedforward depression, excitatory and PV firing rates are modulated in the same fashion, independent of whether they are suppressed or facilitated. The same holds for purely recurrent plasticity, just with a more smooth change of the angle between the two regimes (Fig. 3.6F).

For interacting feedforward and recurrent plasticity, the dichotomy between the two regimes is quite pronounced (Fig. 3.6G), similar to what I found for the area

of co-modulation of the firing rates (Fig. 3.5D). In contrast to the other types of MD-induced plasticity, the angle between the gradients here is 180° throughout the non-ISN, indicating not independent but opposite rate modulations (see also Fig. 3.4D). As the network transitions to the ISN regime, the angle steeply changes to 0° and remains there for all higher J, again showing a tight alignment of the rate modulations throughout the ISN regime (see also Fig. 3.3D).

3.4.3 Summary

The investigation and quantification of the structure of the firing rate fold changes over J confirmed a dichotomy between the non-ISN and the ISN as the major mechanism underlying either independence or tight coordination in the two cell classes. It seems that generally, the non-ISN is more suited to generate independent rate modulations as seen *in vivo*. In contrast, the ISN would predict that the firing rates of the two populations vary together over the course of MD. This is not only true for the two selected cases shown in Figs. 3.3 and 3.4, but also over wide ranges of the coupling scale J (Figs. 3.5 and 3.6).

A problem in the non-ISN, however, is that the lengths of the gradients are very small, especially in feedforward plasticity (Fig. 3.5B). This casts doubt on whether feedforward plasticity in the non-ISN can modulate firing rates strongly enough to explain the firing rates modulations *in vivo* with realistic sizes of synaptic weight changes. The lengths of the gradients in the ISN leave no doubt that the effects are strong enough to suppress rates to 60% of their baseline and recover them with only modest synaptic weight changes.

However, beyond changing the coupling scale J, maybe changing other parameters in the network could enable independent modulation of firing rates, also in the ISN? I investigated how robust this coordination in the ISN is if I vary not only J but also multiple other network parameters that do not directly influence the operating regime but nonetheless change the network responses.

3.5 Robustness against network parameter variations with experimentally constrained synaptic changes

Remark: The results of this section (simulation data in Fig. 3.7B and Table 3.2) have been published before in [3]. The simulation data used in this article have been replotted here for increased clarity and notation in the table has been adapted for consistency. The simulation data and methods were my contributions to the article.

Apart from the coupling scale J, all network parameters have been fixed so far. How robust is the tight coordination between excitatory and PV firing rates in the ISN when we allow more parameters to change? I addressed this question using a Monte-Carlo approach in which multiple parameters are varied. I randomly sampled over large ranges of the relative scale of inhibition g_{rc} and of the feedforward and

background input rates R_{LGN} and R_{BKG} (Table 3.2, note that previously I have used the same rate from the two input sources, while here I also take differing input rates into account). Furthermore, also the coupling scale J is chosen randomly from a large range. However, the boundaries of this range are such that the network is an ISN in all cases.

I generated a large set of networks upon which I then imposed MD-induced synaptic changes. However, simulating the firing rates on fine-grained grids spanning the different parameter spaces is not feasible for this whole set of networks. Instead, I used specific values for MD-induced synaptic changes that match results from our collaborators in the Turrigiano Lab [3]. The loss of generality this entails at the same time enabled more direct biological predictions for the relative roles of feedforward and recurrent plasticity.

I found that (1) overall, the tight coordination of firing rates in the ISN holds, also when varying multiple parameters simultaneously. However, (2) with experimentally restricted values of MD-induced synaptic changes, firing rates closely resembling the state after two days of MD can be generated. Nonetheless, the tight coordination of excitatory and PV firing rates still prohibits the full sequence of firing rate changes on the first and second day of MD.

Methods

The network connectivity, number of neurons and their dynamics are the same as in previous sections (see Section 3.2, specifically Equations 3.3-3.5, parameters provided in Table 3.2). For the publication, we chose a slightly different setup of inputs for a specific example that achieves experimentally constrained firing rates of excitatory (5 Hz) and PV (10 Hz) as proposed by our collaborators from the Turrigiano Lab [3]. Specifically, feedforward inputs from the LGN and BKG inputs where injected to both excitatory and PV neurons (Fig. 3.7A), while the relative strength of LGN input to PV was reduced compared to the previous network ($g_{fw} = 1.25$, Table 3.2). I varied the overall coupling scale J, but combined this with variation of the dominance of recurrent inhibition g_{rc}, the rate of the feedforward input R_{LGN} and the rate of the background input R_{BKG}. Values for these parameters were sampled with a uniform distribution in ranges $J \in \{0.1, 0.4\}$ (thus, all networks are ISN), $g_{rc} \in \{7, 10\}$, $R_{LGN} \in \{500, 1500\}$ and $R_{BKG} \in \{500, 1500\}$.

In total, I generated $100,000$ networks with parameters in these ranges, each of which was run for 9s to estimate average firing rates. From the networks, I sub-selected those which have firing rates around 5 Hz for excitatory and 10 Hz for PV neurons, allowing for $\pm 10\%$ of deviation (Fig. 3.7B, left). This constraint is fulfilled by $\approx 2,500$ networks.

On all of these networks, I applied feedforward and recurrent plasticity with specific values that our collaborators measured in rat V1m layer 4 after two days of MD. In this circuit, the feedforward E/I-ratio was found to change approximately by a factor ≈ 1.5, which I chose for ρ_{EI}. The feedforward input synapses to excitatory neurons were found to depress to 90% of their values before MD (thus $\delta_E = 0.9$,

Fig. 3.7A, blue). From the shift of the feedforward E/I-ratio together with the depression of input synapses to excitatory neurons, I inferred the depression of feedforward input synapses from LGN to PV to be $\delta_P = 0.6$ (Fig. 3.7A, blue), using the relation $\rho_{EI} = \delta_E/\delta_P$. I first studied the consequences this feedforward plasticity has on its own for the firing rates in all the sub-selected networks and plotted the distribution (Fig. 3.7B, middle). In a second step, I also implemented the change of the recurrent E/I-ratio that was experimentally measured in rat V1m layer 4 in the same set of experiments. Our collaborators found the recurrent E/I-ratio to decrease by a factor of ≈ 0.73. I implemented the decrease of the recurrent E/I-ratio through $\zeta_{PE} = 1.5$ (Fig. 3.7A, orange). I studied the consequences on the firing rates when combining feedforward and recurrent plasticity with these specific values for the firing rates in all the sub-selected networks and plotted the distribution (Fig. 3.7B, right).

The resulting distributions of firing rates are visualized as box plots, shown for both populations in all three conditions (baseline, feedforward only, and interacting feedforward and recurrent). The representation of the data is based on quantiles of the underlying distribution. The central red line shows the median of the data, surrounded by a box extending to the first quartiles above and below this median (interquartile range, IQR). The whiskers extending from the box show ranges ± 1.5IQR beyond these quartiles.

The tight coordination of excitatory and PV firing rates is very robust and random sampling confirms the specific roles of the feedforward and recurrent E/I-ratios

After the subselection for suitable firing rates from all randomly generated networks, excitatory firing rates scatter around 5 Hz (Fig. 3.7B, left. Target rate: gray dashed line), and PV firing rates around 10 Hz (Fig. 3.7B, left. Target rate: gray dotted line). Increasing the feedforward E/I-ratio to the value measured in [3] ($\rho_{EI} = 1.5$ with $\delta_E = 0.9$ and $\delta_P = 0.6$) leads to strong facilitation of both the excitatory and PV firing rates (Fig. 3.7B, middle). This confirms my finding of the strong facilitatory effect of the increase of the feedforward E/I-ratio in the ISN (Fig. 3.3) and shows that this is a robust feature over large ranges of randomly chosen parameters.

When I applied recurrent plasticity that decreases the recurrent E/I-ratio ($\zeta_{PE} = 1.5$) in addition to this shift of the feedforward E/I-ratio, both excitatory and PV firing rates decreased strongly relative to the case with only feedforward plasticity (Fig. 3.7B, right). Interestingly, the suppression of excitatory firing rates is more expressed than the suppression of PV firing rates, such that excitatory firing rates seem to be suppressed even relative to the initial baseline, while PV firing rates are more modestly suppressed. This setup is reminiscent of the firing rates measured *in vivo* on the second day of MD, showing PV firing rates, at least on the verge of recovery. Thus, the combination of the feedforward and recurrent synaptic changes measured in [3] on the second day of MD generates firing rates reminiscent of the ones *in vivo* on the second day of MD.

Table 3.2: Parameters of the single neurons and networks, parameter ranges for Monte-Carlo simulation, and parameters for modeling the specific measured synaptic changes during monocular deprivation in rat V1m layer 4 [3]. EPSG/IPSG: excitatory/inhibitory postsynaptic conductance; LGN: lateral geniculate nucleus; BKG: background. Modified from [3].

Symbol	Value/ Range	Description
C_m	200 pF	Membrane capacitance.
g_L	10 nS	Leak conductance.
E_L	-70 mV	Leak reversal potential.
V_θ	-50 mV	Threshold potential.
V_r	-58 mV	Post-spike reset potential.
E_E	0 mV	Excitatory reversal potential.
E_P	-85 mV	Inhibitory reversal potential from PV (P) interneurons.
τ_{syn}	5 ms	Synaptic conductance time constant.
N_E	4000	Number of excitatory neurons.
N_P	1000	Number of PV-interneurons.
p	0.1	Connection probability between two cells.
J	$[0.1, 0.4]$ nS	EPSG amplitude $E \to E$ and $E \to$PV.
g_{rc}	$[7, 10]$	Multiplicative factor for IPSG amplitude PV$\to E$ and PV$\to$ PV.
J_s, J_b	0.5 nS	EPSG amplitude $LGN, BKG \to E$, $BKG \to$ PV and scale for EPSG $LGN \to$ PV.
g_{fw}	1.25	Multiplicative factor for strong feedforward EPSG $LGN \to$ PV.
R_{LGN}	$[500, 1500]$ Hz	Range for the rate of feedforward input spike trains $LGN \to E$ and $LGN \to$ PV.
R_{BKG}	$[500, 1500]$ Hz	Range for the rate of background input spike trains $BKG \to E$ and $BKG \to$ PV.
ρ_{EI}	1.5	Feedforward E/I-ratio after MD $\delta_E = 0.9$ and $\delta_P = 0.6$
ζ_{PE}	1.5	Potentiation of recurrent synapses from excitatory neurons to PV-interneurons, underlying decrease of reccurent E/I ratio.

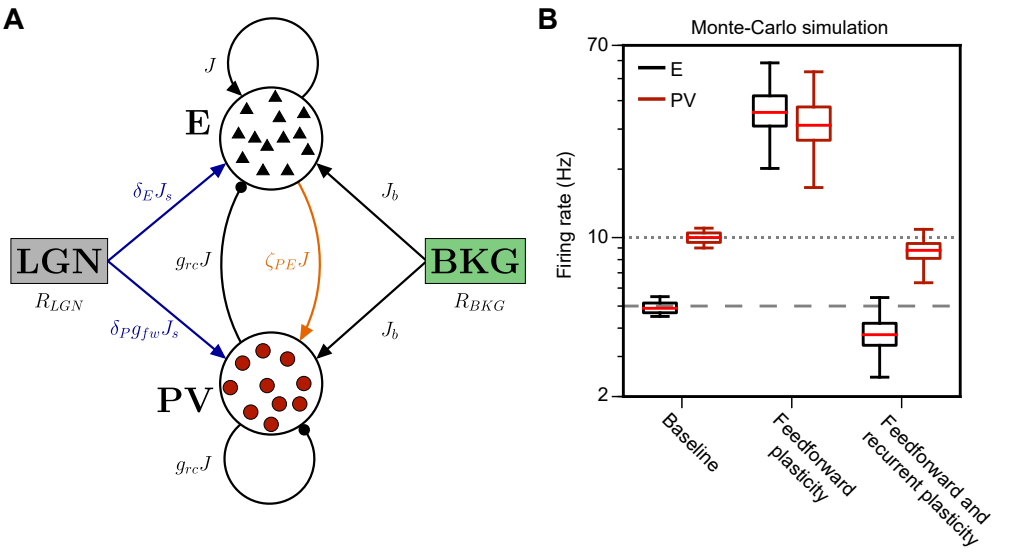

Figure 3.7: Tight coordination between excitatory and PV firing rates is robust over wide parameter ranges. (A) Network schematic showing synaptic connections among neurons with J denoting the overall coupling scale, g_{rc} the dominance of recurrent inhibition and g_{fw} the dominance of feedforward inhibition. The parameters to model MD-induced synaptic plasticity are: depression of feedforward drive to excitatory neurons (δ_E) and to PV-interneurons (δ_P), and potentiation of recurrent excitation to PV-interneurons (ζ_{PE}). Ranges of network parameters and values for synaptic changes due to MD are provided in Table 3.2. **(B)** Monte-Carlo simulation with variation of J, g_{rc} and input rates (R_{LGN}, R_{BKG}) over wide ranges. Box-plots (see Methods) for the distribution of baseline rates (left), rates after feedforward plasticity (middle, blue in A), and after interacting feedforward and recurrent plasticity (right, blue and orange in A). Data in (B) have been published in [3] and replotted here.

However, I am also interested in the mechanism that can bring about the temporally distinct modulation throughout MD, where this firing rate setup of suppressed excitatory firing rates and recovered PV firing rates is preceded by a state of suppressed PV firing rates while excitatory firing rates are homeostatically kept at their initial baseline. When we look a the global picture of facilitation and suppression of firing rates in the two populations, we still see the tight coordination of excitatory and PV firing rates. After feedforward plasticity, the firing rates of both populations strongly increase together (Fig. 3.7B, middle). The subsequent application of recurrent plasticity strongly suppresses the firing rates of both populations together (Fig. 3.7B, right). Thus, the tight coordination of excitatory and PV firing rates is a robust feature of the ISN over large ranges of parameter combinations.

With this approach, I could infer a robust role of plasticity in the feedforward and the recurrent synaptic pathways. As I have shown before, an increase of the feedforward E/I-ratio strongly facilitates rates in both excitatory and PV populations in the ISN, while a decrease of the recurrent E/I-ratio strongly suppresses their firing rates (Fig. 3.3D). From a purely experimental point of view, especially the

contribution of the change of the feedforward E/I-ratio is not trivial to determine. While the increase of the feedforward E/I-ratio makes facilitation of firing rates plausible, the fact that this change of the feedforward E/I-ratio comes about through depression of the feedforward synapses (and thus, through a depression of the total input) could also make a suppression of firing rates plausible [141, 142]. With my Monte-Carlo approach, I have shown that the shift of the feedforward E/I-ratio exclusively leads to facilitation of firing rates over large ranges of randomly chosen network parameters, granted the network is an ISN. Recurrent plasticity, on the other hand, has a purely suppressive role, which is also robust. The combination of the biologically measured feedforward and recurrent plasticity can give rise to a firing rate state resembling the *in vivo* setting after two days of MD. The major homeostatic mechanism, the one that keeps firing rates in check, is the decrease of the recurrent E/I-ratio (see also Discussion in [3]).

3.6 Mathematical analysis of the MD-induced synaptic changes in a rate description of the balanced network

Remark: Some of the results of this section (Fig. 3.10) and parts of the analysis (Eqs. 3.31 and 3.32) have been published before in [1]. The notation has been adapted for consistency, and the analysis has been extended.

In order to theoretically analyze the mechanism that underlies the tight coordination of excitatory and PV-firing rates in the ISN, I studied an analytically tractable model that phenomenologically describes the rates of the different neural populations. I will first introduce the model structure and analytically derive the network parameters that lead to an ISN. This condition is identical to the condition of the emergence of the paradoxical effect (Sec. 3.6.1). I will then introduce the expressions that describe the effects of MD-induced plasticity on the steady-state rate of the network (Sec. 3.6.2). I analyzed the dependence of MD-induced firing rate changes on the operating regime of the network (Sec. 3.7). Specifically, I derived the conditions to find facilitation of PV firing rates in feedforward and recurrent plasticity.

The emergence of a facilitatory area in response to feedforward plasticity was one of the most salient qualitative changes of the firing rate fold changes of PV-interneurons that happens in the spiking network at the value of the coupling scale where the paradoxical effect emerges (Fig. 3.5B). Similarly, in response to recurrent plasticity, the facilitatory area vanished for the PV firing rate changes at the same value of the coupling scale (Fig. 3.5C). I derived that, as the numerical evidence suggested, the PV firing rates cannot be facilitated above baseline in response to feedforward plasticity if the network is a non-ISN. Likewise, the PV firing rate cannot be facilitated above baseline in response to recurrent plasticity if the network is an ISN. I also proved that excitatory rates always show a facilitatory region in

response to feedforward plasticity, with a specific linear boundary that approaches a fixed limit as the coupling scale goes to infinity. Finally, I also proved that recurrent plasticity can never lead to facilitation of excitatory rates. I derived the exact conditions on the parameters for MD-induced synaptic changes to facilitate (or suppress) firing rates in response to either type of plasticity (feedforward or recurrent) in both the excitatory and PV populations.

3.6.1 Network structure, stability, and the paradoxical effect

A general population-rate model describes the firing rates of multiple interacting neural populations by the dynamical equation:

$$\tau \dot{\boldsymbol{r}} = -\boldsymbol{r} + \boldsymbol{f}\big(\boldsymbol{W}\boldsymbol{r} + \boldsymbol{s}\big), \tag{3.16}$$

where $\boldsymbol{r}$ is the vector of different population rates, $\boldsymbol{W}$ is the connectivity matrix of the network, and $\boldsymbol{s}$ is the vector of external inputs. The function $\boldsymbol{f}$ is a vector-valued function that describes the input-output relation of the different populations. Based on the approximate linear-threshold shape of the input-output relation of the spiking network I have studied (Fig. 3.1C), I used a rectified linear function for $\boldsymbol{f}$. Taking the vector $\boldsymbol{x}$ as argument, the different elements f_i of the function $\boldsymbol{f}$ are

$$f_i(\boldsymbol{x}) = \Gamma_i \big[\boldsymbol{x}\big]_+ = \begin{cases} 0 & \text{for} \quad x_i < 0 \\ \Gamma_i x_i & \text{for} \quad x_i \geq 0, \end{cases} \tag{3.17}$$

where Γ_i are the respective slopes of the linear part of the function, which is the input-output gain of each neural population. I used the same gain Γ for both populations in the network. The rectified-linear population-rate model then follows the dynamics

$$\tau \dot{\boldsymbol{r}} = -\boldsymbol{r} + \Gamma \big[\boldsymbol{W}\boldsymbol{r} + \boldsymbol{s}\big]_+. \tag{3.18}$$

Without loss of generality, I absorbed the scalar factor Γ into $\boldsymbol{W}$ and $\boldsymbol{s}$ by rescaling them. Assuming that the external input drives the network enough so that it is in the linear regime of the input-output function, the dynamics of the population-rate model takes the form

$$\tau \dot{\boldsymbol{r}} = -\boldsymbol{r} + \boldsymbol{W}\boldsymbol{r} + \boldsymbol{s}. \tag{3.19}$$

In analogy to the spiking network (Equation 3.2), the network is driven by LGN- and BKG-inputs (Fig. 3.8A). In the rate model, we can write input weights and rates together in the stimulus vector:

$$\boldsymbol{s} = \begin{pmatrix} w_s r_s + w_b r_b \\ g_{fw} w_s r_s \end{pmatrix}, \tag{3.20}$$

where w_s and r_s are the input weight and rate for inputs from the LGN, and w_b and r_b are the respective weight and rate from the BKG. The variable g_{fw} describes the relatively stronger LGN-inputs into the PV-population. As in the spiking model, I

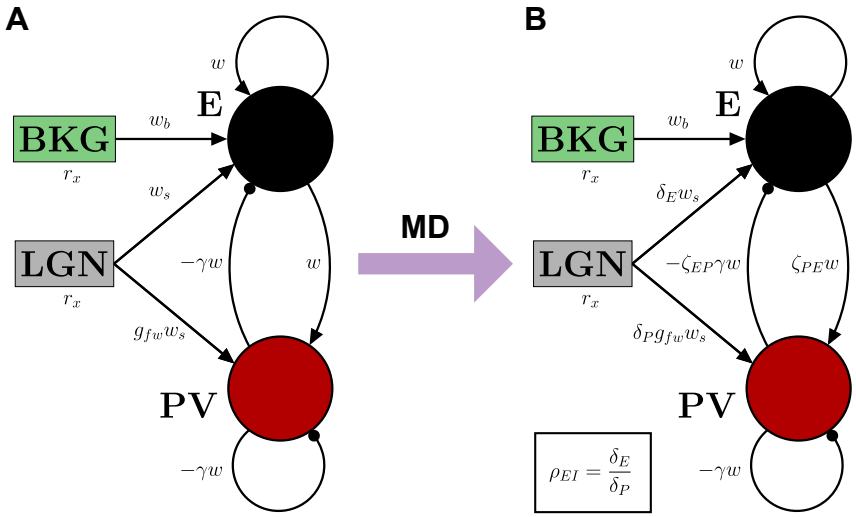

Figure 3.8: Connectivity in the population-rate model with a single subtype of interneurons before and after MD. (A) Connectivity in the baseline network before application of MD. **(B)** Connectivity in the network after application of MD.

used $w_s = w_b$ and $r_s = r_b$. Furthermore, the input weights can be absorbed into the input rates. Since for both populations the input weights and rates are the same, apart from g_{fw}, I use a single variable r_x to describe the external inputs. The rescaled stimulus vector is:

$$\boldsymbol{s} = \begin{pmatrix} 2r_x \\ g_{fw} r_x \end{pmatrix}. \tag{3.21}$$

The recurrent connectivity is also structured in analogy to the spiking network (Fig. 3.8A, spiking see Eq. 3.1). I used w for the coupling scale and γ for the relative scale of inhibition from the PV-population. These parameters represent the averaged effect of a presynaptic population on a postsynaptic population, in contrast to the parameters in the spiking case that represent the weight of a single synapse. Nonetheless, with a grain of salt, we can say the parameters J and w, as well as g_{rc} and γ, are proportional to each other[1]. The connectivity matrix of the rate model is:

$$\mathbf{W} = \begin{pmatrix} W_{EE} & W_{EP} \\ W_{PE} & W_{PP} \end{pmatrix} = \begin{pmatrix} w & -\gamma w \\ w & -\gamma w \end{pmatrix}, \tag{3.22}$$

where both $w > 0$ and $\gamma > 0$, and negative signs are explicitly written in the matrix to denote inhibitory action. For a general analysis, I first derived the conditions for stability of the dynamics and the analytical expression of the steady-state.

[1]The proportionality also needs to take the size of the respective presynaptic populations into account. Furthermore, one needs to consider the activity-dependent effective weight in the spiking network to derive the actual proportionality, which is not generally possible with conductance based-synapses as I have used them.

Stability conditions

The Jacobian of the system in Eq. 3.19 is $\mathcal{J} = \mathbf{W} - \mathbb{I}$. If λ_W is an eigenvalue of the connectivity matrix $\mathbf{W}$, then $\lambda_W - 1$ is an eigenvalue of the Jacobian $\mathbf{W} - \mathbb{I}$. Therefore, to determine the conditions for dynamical stability in the system, which requires that all eigenvalues of the Jacobian have real parts smaller than zero, I investigated the conditions for all eigenvalues of $\mathbf{W}$ to have real parts smaller than one. The eigenvalues of $\boldsymbol{W}$ are:

$$\begin{aligned} \lambda_{1/2} &= -\frac{w(\gamma-1)}{2} \pm \sqrt{\frac{\big(w(\gamma-1)\big)^2}{4}} \\ \lambda_1 &= 0 \\ \lambda_2 &= -w(\gamma-1). \end{aligned} \tag{3.23}$$

The first eigenvalue λ_1 is thus equal to 0 independent of the network parameters, and both eigenvalues are real. From λ_2 I derived as the only condition for the two parameters γ and w that needs to be fulfilled for stability:

$$-w(\gamma-1) = w(1-\gamma) < 1 \quad \Longleftrightarrow \quad \gamma > 1 - \frac{1}{w}. \tag{3.24}$$

For $w < 1$ this condition holds for any $\gamma \geq 0$ since then $(1 - 1/w) < 0$. Note that $\gamma > 0$ simply means inhibitory projections from the second population because I write signs explicit in the connectivity matrix Eq. 3.22. That γ can be zero for all $w < 1$ while the network is still stable means that in this range of w stability does not require any inhibition. Thus, a network with $w < 1$ is a non-ISN, defined by this property. Stability would even allow the second population to add further excitation as long as the condition in Eq. 3.24 is fulfilled.

As soon as $w > 1$, stability requires $\gamma > 0$. Thus, the network needs non-zero inhibition for dynamical stability, i.e., it is an ISN, as it was originally termed by Tsodyks and colleagues [114]. The amount of inhibition necessary for dynamical stability is determined by the coupling scale w. As a limiting case, for $w \gg 1$ we can state the general lower bound $\gamma \geq 1$ since then $(1 - 1/w) \to 1$. This means that for $\gamma \geq 1$, the network is stable for arbitrary coupling scale w. This is also directly visible from $\lambda_2 \leq 0$ in Eq. 3.23 while the condition for stability is only the weaker condition $\lambda_2 < 1$.

Steady state firing rate

The dynamical system described by Equation 3.19 is in a steady state if:

$$0 = -\boldsymbol{r} + \boldsymbol{W}\boldsymbol{r} + \boldsymbol{s}. \tag{3.25}$$

This equation can be solved to give the analytical expression of the steady state firing-rate that only depends on the connectivity and the external input:

$$\boldsymbol{r}_{ss} = \big(\mathbb{I} - \boldsymbol{W}\big)^{-1}\boldsymbol{s}. \tag{3.26}$$

Solving this for my specific connectivity matrix yields

$$\begin{pmatrix} r_{ss}^E \\ r_{ss}^P \end{pmatrix} = \frac{1}{\eta} \begin{pmatrix} 1+\gamma w & -\gamma w \\ w & 1-w \end{pmatrix} \begin{pmatrix} 2r_x \\ g_{fw} r_x \end{pmatrix} \tag{3.27}$$

with

$$\eta = \det(\mathbb{I} - \boldsymbol{W}) = 1 - w + \gamma w. \tag{3.28}$$

Stability of the network requires that this determinant is positive, as $\eta = \det(-\mathcal{J})$. Most easily, one can show that $\eta > 0$ by substituting γ with its lower bound for stability, derived from the requirement that all eigenvalues have positive real parts (Eq. 3.24):

$$\eta > 1 - w + \Big(1 - \frac{1}{w}\Big) w = 0. \tag{3.29}$$

The sign of η was especially important when I studied the direction of steady-state rate changes. This sign is relevant for both the paradoxical effect (see below) and for distinguishing facilitatory or suppressive responses to MD-induced plasticity (Sec. 3.7).

Derivation of the paradoxical effect in the ISN

I calculate the changes of the steady state firing rates ($\Delta \boldsymbol{r}_{ss}$) induced in the two neural populations through an additional external input ξ into PV as:

$$\begin{pmatrix} \Delta r_{ss}^E \\ \Delta r_{ss}^P \end{pmatrix} = \frac{1}{\eta} \begin{pmatrix} 1+\gamma w & -\gamma w \\ w & 1-w \end{pmatrix} \left[\begin{pmatrix} 2r_x \\ g_{fw} r_x \end{pmatrix} - \begin{pmatrix} 2r_x \\ g_{fw} r_x + \xi \end{pmatrix} \right]. \tag{3.30}$$

The stimulus baselines subtract from each other, thus, the firing rate changes of the two populations are:

$$\begin{pmatrix} \Delta r_{ss}^E \\ \Delta r_{ss}^P \end{pmatrix} = \frac{1}{\eta} \begin{pmatrix} 1+\gamma w & -\gamma w \\ w & 1-w \end{pmatrix} \begin{pmatrix} 0 \\ \xi \end{pmatrix} = \frac{1}{\eta} \begin{pmatrix} -\gamma w \xi \\ (1-w)\xi \end{pmatrix}. \tag{3.31}$$

Since $\eta > 0$, I reduce the firing rate changes and just keep their direction:

$$\begin{aligned} \Delta r_{ss}^E &\propto -\gamma w \xi \\ \Delta r_{ss}^P &\propto (1-w)\xi. \end{aligned} \tag{3.32}$$

For the excitatory population, the sign of the activity change for any stable parameter set is negative. Thus, the rate change of excitatory neurons is always in the opposite direction to the current ξ injected into PV: injecting additional current to PV ($\xi > 0$) leads to suppression of the excitatory firing rate, while withdrawing current ($\xi < 0$) leads to facilitation of the excitatory firing rate. As an example, I measured the firing rate changes of the excitatory population when injecting an additional positive current to the PV population. For arbitrary w, this is negative (Fig. 3.9A, black). These results are a direct application of the results from [114] to my specific parametrization of the ISN.

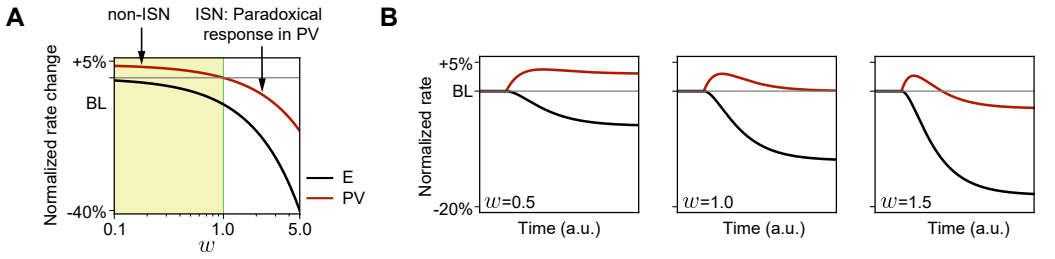

Figure 3.9: Paradoxical effect in PV-interneurons depends on the coupling scale w in the population-rate model. (A) Relative change of steady-state rate of PV-interneurons induced by current injection to PV-interneurons as a function of the coupling scale w. The rate of each population is normalized to its baseline firing rate before additional current injection (BL). This is very similar to the behavior of spiking model when changing J (Fig. 3.1D). **(B)** Dynamics of the linear population-rate model following the onset of a step current to PV-interneurons for $w = 0.5$ (left), $w = 1.0$ (middle) and $w = 1.5$ (right). Colors are the same as in (A).

For the PV-population, the direction of the activity change depends on w. When $w < 1$ rates of PV-interneurons change in the same direction as the injected current (Fig. 3.9A, dark red). This is the originally expected behavior: when an extra current is injected to the PV-population, its firing rate goes up, which, consequently, suppresses excitatory firing rates. If $w > 1$, however, the response of the PV-population shows the paradoxical effect: injection of an additional current into PV leads to suppression of their own activity (Fig. 3.9A, dark red). Equivalently, withdrawal of external current leads to facilitation of activity. I have shown before that the condition $w > 1$ is the exact condition for stability to require non-zero inhibition in my parametrization (Eq. 3.24). Thus, also in this specific parametrization, the conditions for an ISN and for the expression of the paradoxical effect are identical [114].

In the ISN, the activity changes in the PV-population are in the same direction as the rate changes in excitatory neurons, which is opposite to the direction of the external input current. In the example in Fig. 3.9B, the dynamics leading to this paradoxical suppression after injection of an additional current shows the dynamical mechanism underlying this. For all values of the coupling scale w, the PV population initially responds with an increase of its activity. When $w < 1$, the firing rate of the PV population then settles at steady-state activity higher than the activity before current injection (Fig. 3.9B, left). When $w > 1$ the PV activity also initially increases in direct response to the input current but ultimately settles in a steady state with decreased activity (Fig. 3.9B, right). This decreased steady state is an effect of the strongly suppressed recurrent excitation. At a critical value of $w = 1$, the additional external input current and the withdrawal of the recurrent excitation balance each other so that the PV population goes to its initial steady-state after the initial transient rate increase (Fig. 3.9B, middle).

3.6.2 Analytic description of MD-induced steady-state activity changes

The synaptic changes induced by brief MD substantially affect both the inputs to the network and its recurrent connectivity. It is first necessary to analyze if and how these changes of the network affect stability. I will then derive the general changes of the steady-state firing rates that follow from this MD-induced plasticity.

MD-induced changes of the network structure and stability

Given the network was stable before MD-induced synaptic changes, which we shall require for a biologically reasonable model, does MD-induced plasticity potentially destabilize the circuit? I have used dimensionless parameters for MD-induced plasticity in the spiking network, which are multiplied with the synaptic weights in the baseline setup. We can therefore use these same parameters to describe MD-induced plasticity in the rate model (Fig. 3.8B). Feedforward inputs from the LGN to both excitatory and PV neurons depress ($\delta_E < 1$, $\delta_P < 1$). This depression affects the stimulus vector (Eq. 3.21 in baseline), specifically the LGN-inputs:

$$\mathbf{s}^{md} = \begin{pmatrix} (\delta_E + 1) r_x \\ \delta_P g_{fw} r_x \end{pmatrix}. \tag{3.33}$$

Recurrent connections from excitatory to PV and the reciprocal synapses from PV to excitatory neurons potentiate ($\zeta_{PE} > 1$, $\zeta_{EP} > 1$). The connectivity matrix Eq. 3.22 with these changes of recurrent synaptic strengths becomes:

$$\mathbf{W}^{md} = \begin{pmatrix} w & -\zeta_{EP}\gamma w \\ \zeta_{PE} w & -\gamma w \end{pmatrix}. \tag{3.34}$$

Feedforward changes do not affect stability in the linear model since the Jacobian is only influenced by the recurrent connectivity. Calculating the eigenvalues of the connectivity matrix with MD-induced synaptic changes $\mathbf{W}^{md}$ I find:

$$\lambda_{1,2}^{md} = -\frac{w(\gamma - 1)}{2} \pm \sqrt{\frac{\big(w(\gamma - 1)\big)^2}{4} - (\zeta_{EP}\zeta_{PE} - 1)\gamma w^2}. \tag{3.35}$$

In comparison to the eigenvalues of the connectivity matrix before induction of MD (Equation 3.23), recurrent potentiation thus leads to a subtractive term $(\zeta_{EP}\zeta_{PE} - 1)\gamma w^2$ within the root. This term is always positive for potentiation of the recurrent synapses. Two different effects on the eigenvalues need to be considered: either the term in the root becomes smaller but remains positive (if $(\zeta_{EP}\zeta_{PE} - 1)\gamma w^2 < (w(\gamma - 1))^2/4$), in which case the eigenvalues after MD induction are two real numbers just as in the case without MD-induced plasticity. Alternatively, the term in the root becomes negative (if $(\zeta_{EP}\zeta_{PE} - 1)\gamma w^2 > (w(\gamma - 1))^2/4$), in which case the connectivity matrix has two conjugate eigenvalues with identical real parts. In both cases, the largest real part of the two eigenvalues can be derived from the

eigenvalue λ_1^{md}:

$$\lambda_1^{md} = -\frac{w(\gamma-1)}{2} + \sqrt{\frac{\big(w(\gamma-1)\big)^2}{4} - (\zeta_{EP}\zeta_{PE}-1)\gamma w^2}. \tag{3.36}$$

In the case that this eigenvalue is real (positive term in the root), one can derive the stability condition

$$\begin{aligned} \lambda_1^{md} = \max\Big(\Re\big(\lambda_{1,2}^{md}\big)\Big) &< 1 \\ \implies \frac{w(\gamma-1)}{2} + 1 &> \sqrt{\frac{\big(w(\gamma-1)\big)^2}{4} - (\zeta_{EP}\zeta_{PE}-1)\gamma w^2} \\ \implies w(\gamma-1) + 1 &> -(\zeta_{EP}\zeta_{PE}-1)\gamma w^2. \end{aligned} \tag{3.37}$$

Stability of the baseline setup before MD-induction requires that the left hand side is positive (inverting the sign in Equation 3.24 yields $w(\gamma-1) > -1$), while recurrent potentiation renders the right hand side negative. Thus, this condition is satisfied for arbitrary amounts of recurrent potentiation[2] of recurrent connections between excitatory and PV-interneurons, provided the network was stable before MD-induction.

In the case that the term in the root becomes negative and the eigenvalues are complex conjugate to each other, both eigenvalues have real part $-w(\gamma-1)/2$. Stability then requires

$$-\frac{w(\gamma-1)}{2} < 1. \tag{3.38}$$

This condition is independent of the values for recurrent potentiation. It is, in fact, weaker than the condition for stability in the baseline setup, which is $-w(\gamma-1) < 1$ (Eq. 3.24). Thus, also in the case of complex-conjugate eigenvalues, the network after MD-induction is stable if the network was stable in the baseline setup.

In conclusion, for arbitrary values of recurrent potentiation, leading either to two complex-conjugate or two real eigenvalues, dynamical stability is maintained when the baseline setup was stable. The system is even stabilized to some degree as the conditions on stability are becoming weaker than in the network without recurrent plasticity. This finding is also interesting for the biological role of MD-induced plasticity. The analysis suggests that the plasticity mechanisms invoked in the recurrent circuit are stabilizing the dynamics.

[2]Within the bounds imposed by the condition that both eigenvalues are real.

General steady-state firing rate fold changes following from MD-induced synaptic changes

Including MD-changes in both feedforward and recurrent synapses, we can calculate the steady state rate after MD from Eq. 3.26:

$$\begin{pmatrix} r_{ss}^{E,md} \\ r_{ss}^{P,md} \end{pmatrix} = \frac{1}{\eta^{md}} \begin{pmatrix} 1+\gamma w & -\zeta_{EP}\gamma w \\ \zeta_{PE} w & 1-w \end{pmatrix} \begin{pmatrix} (\delta_E+1) r_x \\ \delta_P g_{fw} r_x \end{pmatrix}, \tag{3.39}$$

where η^{md} is the determinant of $\mathbb{I} - \mathbf{W}^{md}$:

$$\eta^{md} = 1 - w + \gamma w + (\zeta_{EP}\zeta_{PE} - 1)\gamma w^2. \tag{3.40}$$

For $\zeta_{EP} = \zeta_{PE} = 1$, this determinant is the same as baseline setup (η, Eq. 3.28). To better distinguish the two determinants, I will call the original determinant from the baseline setup η^{bl} instead of η.

The determinant of the negative of the Jacobian in the baseline setup, η^{bl}, is positive when the network is stable[3]. Recurrent plasticity leads to the addition of a positive term to η^{bl}. Thus, η^{md} is also always positive:

$$\eta^{md} = \eta^{bl} + (\zeta_{EP}\zeta_{PE} - 1)\gamma w^2. \tag{3.41}$$

From the matrix equation (Eq. 3.39), we find the rates of the two populations to be:

$$\begin{aligned} r_{ss}^{E,md} &= \frac{1}{\eta^{md}} \Big((1+\gamma w)(\delta_E+1) r_x - \zeta_{EP}\gamma w \delta_P g_{fw} r_x \Big) \\ r_{ss}^{P,md} &= \frac{1}{\eta^{md}} \Big(\zeta_{PE} w (\delta_E+1) r_x + (1-w)\delta_P g_{fw} r_x \Big). \end{aligned} \tag{3.42}$$

I was most interested in the firing rate changes induced by MD-plasticity relative to the baseline firing rate, similar to the experimental studies. To better distinguish the rates in baseline and after MD, I used index *bl* to denote the rates in the baseline state before any MD-induced synaptic changes (Eq. 3.27):

$$\begin{aligned} r_{ss}^{E,bl} &= \frac{1}{\eta^{bl}} \Big(2(1+\gamma w) r_x - \gamma w g_{fw} r_x \Big) \\ r_{ss}^{P,bl} &= \frac{1}{\eta^{bl}} \Big(2 w r_x + (1-w) g_{fw} r_x \Big). \end{aligned} \tag{3.43}$$

The firing rate fold change for a population $A \in \{E, P\}$ for MD-induced plasticity was calculated the same way as in the spiking network:

$$\Psi^A = \frac{r_{ss}^{A,md}}{r_{ss}^{A,bl}}. \tag{3.44}$$

[3] Also derived from the requirement that the largest real part of the eigenvalues is negative (Eq. 3.29).

Numerically solving these firing rate fold changes in the different parameter spaces for feedforward, recurrent and interacting feedforward and recurrent plasticity reproduced the results of the spiking network very well (Fig. 3.10A-C). This confirms that the population-rate model with a linear-threshold input-output function is a good approximation to study the mechanisms underlying the rate modulations in the spiking networks. Analytically, however, it is easier to study the difference of steady-state firing rates rather than the ratio. The most important property is the induction of facilitation or suppression of firing rates above or below baseline. The relation linking the divisive and the subtractive case, for the case of facilitation of firing rates, is:

$$\Psi^A = \frac{r_{ss}^{A,md}}{r_{ss}^{A,bl}} > 1 \quad \Longleftrightarrow \quad \Xi^A = r_{ss}^{A,md} - r_{ss}^{A,bl} > 0. \tag{3.45}$$

To analytically understand the mechanism underlying the tight alignment of firing rate changes between excitatory and PV neurons in feedforward and recurrent changes in the ISN, as well as the independent modulation in the non-ISN, I studied the conditions for MD-induced synaptic changes to induce facilitation or suppression in the two populations, and how they interrelate with the coupling parameters of the underlying network, especially w as the parameter defining the operating regime. The boundary is $w = 1$ between non-ISN ($w < 1$) and ISN ($w > 1$). In general, the explicit condition for induction of facilitation of firing rates in the excitatory population, including all MD-induced synaptic changes ($r_{ss}^{E,md} - r_{ss}^{E,bl} > 0$) is:

$$\begin{aligned} \Xi^E \;=\; & \frac{1}{\eta^{md}}\Big((1+\gamma w)(\delta_E+1)r_x - \zeta_{EP}\gamma w \delta_P g_{fw} r_x\Big) \\ & -\frac{1}{\eta^{bl}}\Big(2(1+\gamma w)r_x - \gamma w g_{fw} r_x\Big) \end{aligned} \tag{3.46}$$

and for the PV-population ($r_{ss}^{P,md} - r_{ss}^{P,bl} > 0$):

$$\begin{aligned} \Xi^P \;=\; & \frac{1}{\eta^{md}}\Big(\zeta_{PE} w(\delta_E+1)r_x + (1-w)\delta_P g_{fw} r_x\Big) \\ & -\frac{1}{\eta^{bl}}\Big(2w r_x + (1-w) g_{fw} r_x\Big). \end{aligned} \tag{3.47}$$

I have shown that both $\eta^{bl} > 0$ and $\eta^{md} > 0$ for stable systems. This is important to know so that we can multiply by these determinants without switching signs or, equivalently, the direction of the inequality. To more clearly and intuitively understand the role each parameter plays, it is useful to again study the condition for facilitation separately for feedforward ((δ_E, δ_P)-plane) and recurrent ((ζ_{EP}, ζ_{PE})-plane) plasticity. The combined feedforward and recurrent changes are a direct application of these results.

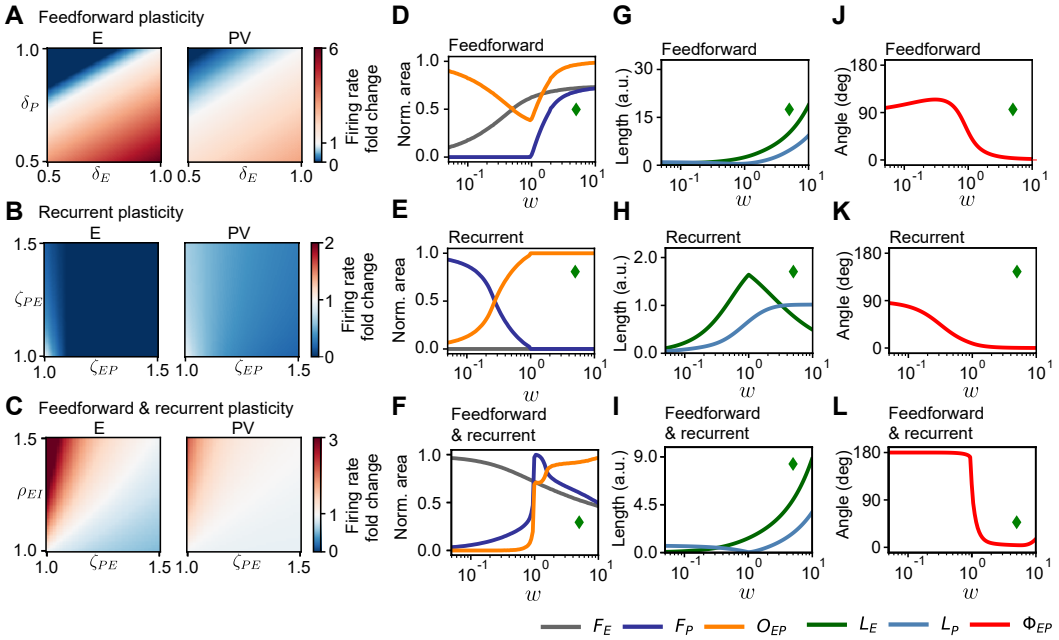

Figure 3.10: Linear model of the network with a single subtype of interneurons (PV). (A) Network firing rate for an ISN in the (δ_E, δ_P) plane (feedforward depression) as fold-change of baseline firing rate for excitatory (left) and PV populations (right). **(B)** Same as (A) for recurrent potentiation ((ζ_{EP}, ζ_{PE}) plane). **(C)** Same as (A) for interacting feedforward and recurrent plasticity: decrease of the recurrent E/I-ratio through ζ_{PE}, increase of the feedforward E/I-ratio as $\rho_{EI} = \delta_E/\delta_P$. **(D)** Network with feedforward depression only. Fractional area of facilitation for excitatory neurons (F_E, gray), PV-interneurons (F_P, blue) and the overlap between excitatory and PV response areas (O_{EP}, orange) as a function of overall coupling scale w. **(E)** Same as (D) for recurrent potentiation only. **(F)** Same as (D) for interacting feedforward and recurrent plasticity. **(G)** Network with feedforward depression only. Average length of the gradient for excitatory neurons (L_E, green) and PV-interneurons (L_P, light blue) as a function of overall coupling scale w. **(H)** Same as (G) for recurrent potentiation only. **(I)** Same as (G) for interacting feedforward and recurrent plasticity. **(J)** Network with feedforward depression only. Average angle between the gradients of excitatory PV neurons (Φ_{EP}, red) as a function of overall coupling scale w. **(K)** Same as (J) for recurrent potentiation only. **(L)** Same as (J) for interacting feedforward and recurrent plasticity. Note that w in the linear population model is the equivalent to the coupling scale J in the spiking network. Green diamond in (D-L) shows the value of the coupling scale ($w = 5$) used in (A-C). Panels (A-F) modified from [1].

3.7 Mathematical analysis of the role of the operating regime

Remark: Parts of the results of this section have been published before in [1]. The notation has been adapted for consistency, and the analysis has been extended.

As in the spiking case (Section 3.3), I split up the different types of plasticity, focusing on the feedforward depression (index ffw) and recurrent potentiation (index rec). Numerical solution of the linear model shows very good qualitative agreement of the rate changes produced in the linear rate model with those in the spiking network, beyond the response plane in the ISN (Fig. 3.10A-C) also in the curves for the different quantifications I have introduced in Sec. 3.4, for the facilitatory areas and overlap (Fig. 3.10D-F), the gradient lengths (Fig. 3.10G-I, compare to Fig. 3.5B-D) and the angle between the gradients (Fig. 3.10J-L, compare to Fig. 3.6).

I have discussed the planes and these curves in detail for the spiking network. However, it is worthwhile to point out that the most salient changes that relate to the alignment between excitatory and PV firing rate changes happen in the linear rate model at exactly $w = 1$, as predicted from the connection between the paradoxical effect and these changes in the spiking case.

3.7.1 Conditions on facilitation for feedforward plasticity

Excitatory population

The response to pure feedforward plasticity in the excitatory population is:

$$\begin{aligned}\Xi^E_{\text{ffw}} &= \frac{1}{\eta^{bl}}\Big((1+\gamma w)(\delta_E+1)r_x - \gamma w \delta_P g_{fw} r_x\Big) \\ &\quad -\frac{1}{\eta^{bl}}\Big(2(1+\gamma w)r_x - \gamma w g_{fw} r_x\Big) > 0. \end{aligned} \tag{3.48}$$

This can be simplified to the condition:

$$\Big((1+\gamma w)(\delta_E-1) - \gamma w(\delta_P-1)g_{fw}\Big) > 0. \tag{3.49}$$

From this, I derived where in the parameter space of MD-induced feedforward plasticity, excitatory rates are facilitated. I expressed this as a condition on the δ_P-value necessary to induce facilitation for a given δ_E, that is: the δ_P at the boundary between facilitation and suppression as a function of δ_E.

$$\delta_P < 1 - (1-\delta_E)\frac{1+\gamma w}{g_{fw}\gamma w}. \tag{3.50}$$

On the right hand side of Eq. 3.50 is a term of the form $1 - x$. Since $(1 - \delta_E) > 0$, $w > 0$ and $\gamma > 0$, the term x is positive. Thus, this condition to find facilitation of

the excitatory rate can be fulfilled within the biological constraint that $\delta_P < 1$, and facilitation is generally possible for depression of feedforward inputs, independent of the value of the coupling scale w. The condition for facilitation of excitatory rates in Eq. 3.50 imposes an upper bound on δ_P that depends on δ_E, confirming the numerical insight that for a given depression of feedforward inputs to excitatory neurons, facilitation can only be achieved if the depression of feedforward inputs to PV-interneurons is strong enough.

Replacing the inequality in Eq. 3.50 by an equality, I extracted the exact values of δ_P that determine the boundary between facilitation and suppression over the whole range of δ_E. The family of curves for the boundaries over a wide range of coupling scales w is plotted in Fig. 3.11B (left).

To extract a global picture of the behavior of this boundary between facilitation and suppression, two limiting cases in w are important: First, if $w \to 0$: then $(1+\gamma w)/g_{fw}\gamma w \to \infty$. Thus, facilitation can only happen for $\delta_E = 1$ because otherwise $(\delta_E - 1)(1+\gamma w)/g_{fw}\gamma w$ also goes to $-\infty$ and cannot be fulfilled by $\delta_P > 0$. This is the analytical expression for the fact that, for very small coupling scale w, responses of excitatory neurons almost exclusively suppress, and, in the limiting case of no coupling, exclusively suppress. Second, if $w \to \infty$, one can simplify the linear relationship between δ_E and δ_P through $(1+\gamma w)/g_{fw}\gamma w \to 1/g_{fw}$ and obtain:

$$\delta_P < \frac{1}{g_{fw}}\delta_E - \frac{1}{g_{fw}} + 1 \tag{3.51}$$

which is approached in Fig. 3.11B (left) as w increases. This limit of the boundary will always leave a part of the parameter space depressive, visible in the fact that the facilitatory area for excitatory firing rate fold changes saturates at a value smaller than one for large w (F_E, Fig. 3.10D).

PV-population

The response to pure feedforward plasticity in the PV population is:

$$\begin{aligned}\Xi^P_{\text{ffw}} &= \frac{1}{\eta^{bl}}\Big(w(\delta_E+1)r_x + (1-w)\delta_P g_{fw} r_x\Big) \\ &\quad -\frac{1}{\eta^{bl}}\Big(2wr_x + (1-w)g_{fw}r_x\Big) > 0\end{aligned} \tag{3.52}$$

from which the simplified condition

$$\Big(w(\delta_E - 1) + (1-w)g_{fw}(\delta_P - 1)\Big) > 0 \tag{3.53}$$

follows. In contrast to the excitatory population, for the PV population it is not possible for general w to derive the value of δ_P necessary for facilitation given δ_E, because the term $(1-w)$ switches sign at $w = 1$. In a first step we can derive:

$$(w-1)g_{fw}\delta_P < w(\delta_E - 1) + (w-1)g_{fw}. \tag{3.54}$$

To further simplify, we have to distinguish the two different operating regimes: If $w < 1$ (non-ISN), the factor $(w - 1)$ on the left hand side in Eq. 3.54 is negative. Division by it leads to an inversion of the direction of the inequality:

$$\delta_P > 1 + \frac{w}{(1-w)g_{fw}}(1 - \delta_E), \tag{3.55}$$

where I rearranged terms such that all negative signs are written explicitly. Given $\delta_E < 1$, we see that this condition cannot be fulfilled for $\delta_P < 1$, since we add a positive term to one on the right hand side of Eq. 3.55. Therefore, there can be no facilitation of PV firing rates induced by feedforward depression in the non-ISN, as the quantification of the facilitatory area shows for both the linear model for $w < 1$ (L_P, Fig. 3.10D) and in the spiking network (Fig. 3.5B).

If $w > 1$ (ISN) then $(w - 1) > 0$ and we can divide in Eq. 3.54 by this term without inverting the direction of the inequality (again all negative signs are written explicitly)

$$\delta_P < 1 - \frac{w}{(w-1)g_{fw}}(1 - \delta_E). \tag{3.56}$$

In this case, a positive term is subtracted from one. Thus, in the ISN, PV firing rates can be facilitated through feedforward depression. Replacing the inequality in Eq. 3.56 by an equality, I derived and plotted the boundary between facilitation and suppression (Fig. 3.11B, right). I also show ranges of δ_E and δ_P that fall outside the biological constraint of feedforward depression in order to visualize how the boundary behaves over the whole range of w. The plot shows how the boundary falls outside the biological range in the non-ISN but crosses it in the ISN.

The boundary is similar to the condition in the excitatory population (Eq. 3.50).The emergence and growth of a facilitatory area of PV-responses in feedforward plasticity is a major driver behind the tight alignment of excitatory and PV-responses in the whole feedforward parameter space (Fig. 3.5). This can be shown considering the limiting case $w \to \infty$, which yields:

$$\delta_P < \frac{1}{g_{fw}}\delta_E - \frac{1}{g_{fw}} + 1. \tag{3.57}$$

This is the same condition as in the case of the excitatory neurons. Thus, the larger w, the more tightly coordinated are the firing rate changes of the two populations. For a strongly connected ISN, the firing rates of excitatory and PV neurons are perfectly aligned.

3.7.2 Conditions on facilitation for recurrent plasticity

Excitatory population

From the steady state rate of the excitatory population in response to MD (Eq. 3.42) and their baseline steady state rate (Eq. 3.43) the condition to induce facilitation in

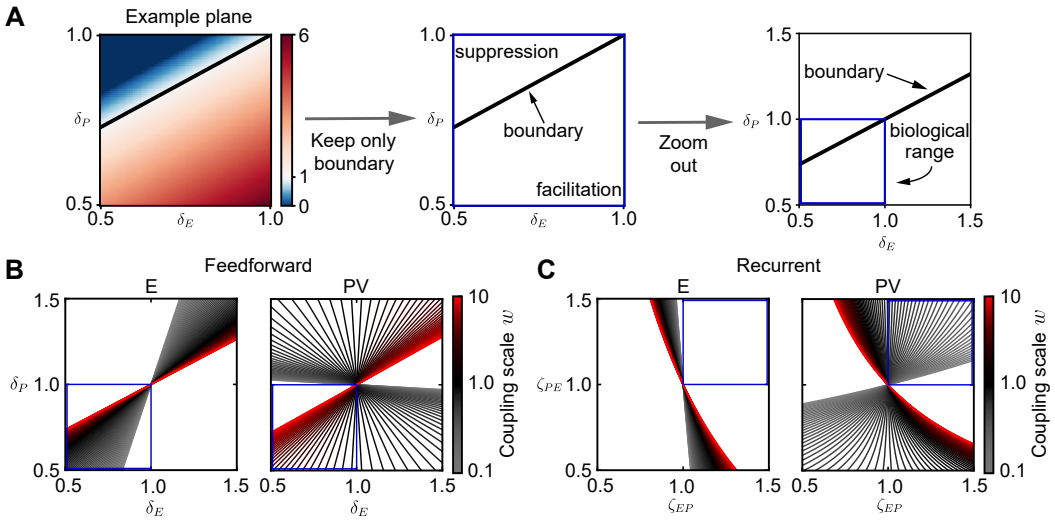

Figure 3.11: The boundary between facilitation and suppression in feedforward and recurrent plasticity. (A) Extraction of the boundary between facilitation and suppression. The boundary is defined through the parameter combinations for which rate after MD is the same as in baseline (white areas left). For further visualization, I only keep the boundary and dismiss the firing rate fold changes (middle, analytically derived boundary). I also include non-biological parameter ranges to see how the boundary behaves when we find no facilitatory area within the biological range of MD-induced plasticity. Blue box shows biological range (depression of feedforward synapses: $\delta_E \leq 1$ and $\delta_P \leq 1$, potentiation of recurrent synapses: $\zeta_{EP} \geq 1$ and $\zeta_{PE} \geq 1$). **(B)** Boundary between facilitation and suppression for feedforward plasticity in the network with a single subtype of interneurons as a function of the coupling scale w for excitatory (left, Eq. 3.50) and PV neurons (right, Eq. 3.56). **(C)** Same as (B) for recurrent plasticity, based on Eq. 3.62 for excitatory (left) and Eq. 3.80 for PV (right).

recurrent plasticity follows:

$$\begin{aligned}\Xi^E_{\text{rec}} &= \frac{r_x}{\eta^{md}}\Big(2(1+\gamma w) - \zeta_{EP}\gamma w g_{fw}\Big) - \frac{r_x}{\eta^{bl}}\Big(2(1+\gamma w) - \gamma w g_{fw}\Big) > 0 \\ &\Leftrightarrow -(\zeta_{EP}-1) - (\zeta_{PE}\zeta_{EP}-1)\frac{w}{g_{fw}\eta^{bl}}\Big(2(1+\gamma w) - \gamma w g_{fw}\Big) > 0, \end{aligned} \tag{3.58}$$

where I made use of of the relation between the determinants in the baseline state and after MD (Eq. 3.41):

$$\eta^{md} = \eta^{bl} + (\zeta_{EP}\zeta_{PE} - 1)\gamma w^2 \tag{3.59}$$

and of the fact that these determinants are always positive in stable networks. To further simplify the expression in Eq. 3.58 it is necessary to know the sign of the factor

$$A_E = \frac{w}{g_{fw}\eta^{bl}}\Big(2(1+\gamma w) - \gamma w g_{fw}\Big). \tag{3.60}$$

By definition, $w > 0$ and $g_{fw} > 1$. Thus, the term A_E is directly proportional to the baseline rate (Eq. 3.43, only missing $r_x > 0$). In order to study a biologically meaningful setup, rates shall be positive in the baseline setup before MD-induced plasticity is applied. This implies an upper bound for g_{fw}

$$g_{fw} < 2 + \frac{2}{\gamma w}. \tag{3.61}$$

Using the shortcut in Eq. 3.60, and knowing that this term is positive, I derived from Eq. 3.58 the lower bound of ζ_{PE} necessary to achieve facilitation of excitatory rates for a given ζ_{EP}

$$\zeta_{PE} > \frac{-\zeta_{EP} + A_E + 1}{\zeta_{EP} A_E}. \tag{3.62}$$

Due to the non-linear dependencies of ζ_{PE} on ζ_{EP}, this condition is not as intuitive to explain as it was in the feedforward case. However, numerically solving the boundary between facilitation and suppression (replacing the inequality in Eq. 3.62 by equality) over large ranges of w shows that there is no solution within the biological ranges of ζ_{EP} and ζ_{PE}, neither for the non-ISN nor for the ISN (Fig. 3.11C, left). This confirms the results from quantifying the area of facilitation over a large range of w spanning both operating regimes (Fig. 3.10).

Furthermore, a general proof is possible that excitatory firing rates cannot be facilitated by recurrent plasticity without using the explicit expression of ζ_{PE} in Eq. 3.62. To simplify this derivation, I first substitute in the explicit expression for the steady state rates (Eqs. 3.43 and 3.42, see also Eq. 3.58):

$$\begin{aligned} X_1 &= 2(1 + \gamma w) r_x \\ X_2 &= \gamma w g_{fw} r_x \end{aligned} \tag{3.63}$$

which are both positive by definition. With these substitutions, I can express the steady state rate in the baseline state as:

$$r_{ss}^{E,bl} = \frac{1}{\eta^{bl}} \Big(X_1 - X_2 \Big). \tag{3.64}$$

I also substitute $\zeta_{EP} > 1$ in the expression for the steady state firing rate after recurrent plasticity by:

$$\zeta_{EP} = 1 + \varepsilon_{EP} \tag{3.65}$$

such that we can write:

$$r_{ss}^{E,\mathrm{rec}} = \frac{1}{\eta^{md}} \Big(X_1 - (1 + \varepsilon_{EP}) X_2 \Big). \tag{3.66}$$

The firing rate change induced by recurrent plasticity in excitatory neurons can then be written as

$$\begin{aligned} \Xi_{\mathrm{rec}}^{E} &= r_{ss}^{E,\mathrm{rec}} - r_{ss}^{E,bl} \\ &= \frac{1}{\eta^{md}} \Big(X_1 - (1 + \varepsilon_{EP}) X_2 \Big) - \frac{1}{\eta^{bl}} \Big(X_1 - X_2 \Big) \\ &= \left(\frac{1}{\eta^{md}} - \frac{1}{\eta^{bl}} \right) (X_1 - X_2) - \frac{\varepsilon_{EP}}{\eta^{md}} X_2, \end{aligned} \tag{3.67}$$

with $(X_1 - X_2) = \eta^{bl} r_{ss}^{E,bl}$, the rate change Ξ_{rec}^{E} becomes

$$\Xi_{\text{rec}}^{E} = \left(\frac{\eta^{bl}}{\eta^{md}} - 1 \right) r_{ss}^{E,bl} - \frac{\varepsilon_{EP}}{\eta^{md}} X_2. \tag{3.68}$$

Since $\eta^{md} > \eta^{bl}$ we have $(\eta^{bl}/\eta^{md} - 1) < 0$, and with positive baseline firing rate the whole first term on the right hand side in Eq. 3.68 is negative. From this, a positive term is subtracted when recurrent synapses potentiate ($\varepsilon_{EP} > 0$). Therefore, from this expression, one can conclude that it is impossible to find facilitation in excitatory neurons after recurrent potentiation, independent of the value of $w > 0$ and thus regardless of the operating regime (ISN or non-ISN).

PV population

From the baseline steady state rates of the PV population (Eq. 3.43) and their steady state rates in response to MD (Eq. 3.42) the condition to induce facilitation in recurrent plasticity can be derived using the same relation between η^{md} and η^{bl} as for the excitatory case (Eq. 3.41):

$$\begin{aligned} \Xi_{\text{rec}}^{P} &= \frac{r_x}{\eta^{md}} \Big(2\zeta_{PE} w + (1-w) g_{fw} \Big) - \frac{r_x}{\eta^{bl}} \Big(2w + (1-w) g_{fw} \Big) > 0 \\ &\Leftrightarrow (\zeta_{PE} - 1) - (\zeta_{PE}\zeta_{EP} - 1) \frac{\gamma w}{2\eta^{bl}} \Big(2w + (1-w) g_{fw} \Big) > 0. \end{aligned} \tag{3.69}$$

I have shown numerically that the rates of the PV population can facilitate in the non-ISN in response to recurrent plasticity (Fig. 3.10B and E). In the ISN I did not find any facilitation of PV firing rates numerically. I will prove that this is indeed a general property the ISN by showing that the boundary between facilitation and suppression, specifically the ζ_{EP} necessary to achieve facilitation of firing rates for a given ζ_{PE} does not fall within the biological range.

In a first step, the general mathematical argument that facilitation of PV firing rates is possible follows the same route as the argument showing that excitatory rates cannot be facilitated (Eq. 3.68). I used the variables

$$\begin{aligned} Y_1 &= 2wr_x \\ Y_2 &= (1-w) g_{fw} r_x. \end{aligned} \tag{3.70}$$

Using the substitution for $\zeta_{PE} > 1$

$$\zeta_{PE} = 1 + \varepsilon_{PE} \tag{3.71}$$

in addition, the PV firing rates in the baseline state and after recurrent plasticity can be expressed as

$$\begin{aligned} r_{ss}^{P,bl} &= \frac{1}{\eta^{bl}} \Big(Y_1 + Y_2 \Big) \\ r_{ss}^{P,\text{rec}} &= \frac{1}{\eta^{md}} \Big((1 + \varepsilon_{PE}) Y_1 + Y_2 \Big), \end{aligned} \tag{3.72}$$

from which the expression for the rate changes Ξ^P_{rec} follows:

$$\Xi^P_{\text{rec}} = \left(\frac{\eta^{bl}}{\eta^{md}} - 1\right) r^{P,bl}_{ss} + \frac{\varepsilon_{PE}}{\eta^{md}} Y_1. \tag{3.73}$$

Similar to the excitatory case in Eq. 3.68, the first term in Eq. 3.73 is negative because $\eta^{bl} < \eta^{md}$. Contrary to the excitatory case, however, a positive term is added to this. Thus, generally facilitation is possible. A further insight is that this facilitation centrally requires $\varepsilon_{PE} > 0$ or, equivalently, $\zeta_{PE} > 1$. This is also biologically reasonable, since ζ_{PE} describes the potentiation of recurrent excitatory inputs onto the PV population. To derive the exact boundary between facilitation and suppression and, more crucially, to proof that facilitation of PV firing rates is not possible through recurrent plasticity in the ISN, I rewrote Eq. 3.69 with the abbreviation

$$A_P = \frac{\gamma w}{2\eta^{bl}} \Big(2w + (1-w) g_{fw}\Big). \tag{3.74}$$

Thus,

$$\zeta_{PE} \cdot (1 - \zeta_{EP} A_P) > 1 - A_P. \tag{3.75}$$

The variable A_P is proportional to the PV firing rate in the baseline state and should thus to be positive for a biologically reasonable setup. However, to proceed further from this equation, one needs to know the sign of $1 - \zeta_{EP} A_P$, because this will determine the direction of the inequality in Eq. 3.75 after division. With $\zeta_{EP} > 1$ as the only constraint, this is not easily determined. However, since facilitation of firing rates requires $\zeta_{PE} > 1$ (Eq. 3.73), I first studied the one-dimensional case by setting $\zeta_{EP} = 1$. Solving Eq. 3.75 then yields:

$$\zeta_{PE} \begin{cases} < 1 & \text{if: } A_P > 1 \\ > 1 & \text{if: } A_P < 1. \end{cases} \tag{3.76}$$

Thus, if $A_P > 1$, no facilitation of PV firing rates is possible with potentiation of the recurrent synapse from the excitatory to the PV population ($\zeta_{PE} > 1$). For A_P to be larger than one, and without considering the operating regime, the condition can be derived

$$(2 - g_{fw})(w - 1) > \frac{2}{\gamma w}(1 - w). \tag{3.77}$$

Since w and γ are only constrained by stability and positive firing rates, I derived from Eq. 3.77 the following: is $A_P > 1$ for arbitrary w and γ, as long as g_{fw} is below the upper bound necessary to have positive baseline firing rate in the excitatory population (Eq. 3.61)? To resolve the inequality Eq. 3.77, one needs to divide by $(w - 1)$ whose sign depends on the operating regime. In the ISN with $(w - 1) > 0$ the boundary follows

$$g_{fw} < \frac{2}{\gamma w} + 2 \tag{3.78}$$

which is exactly the same condition on g_{fw} as in Eq. 3.61. Thus $A_P > 1$ in the ISN as long as the excitatory rate in the baseline state is positive. Therefore, PV

firing rates cannot be facilitated above baseline (Eq. 3.76). Furthermore, with this proof that $A_P > 1$ in the ISN, it is possible to prove that also in the two-dimensional case, allowing ζ_{EP} to vary, no facilitation can be found. With $A_P > 1$ and $\zeta_{EP} > 1$ it certainly holds true that $\zeta_{EP} A_P > 1$. Furthermore, from $\zeta_{EP} > 1$ and $1 - \zeta_{EP} A_P < 0$ follows:

$$\zeta_{PE} < \frac{1 - A_P}{1 - \zeta_{EP} A_P} < 1. \tag{3.79}$$

Thus, the argument that recurrent plasticity cannot induce facilitation of PV firing in the ISN holds generally for the (ζ_{EP}, ζ_{PE})-plane.

For the non-ISN, one can prove that $A_P < 1$ in the range permitted for g_{fw} using the same route. The only difference to the proof given above is to start with the assumption $A_P < 1$, then the inequality in Eq. 3.77 is the opposite, but division by $(w - 1)$ does switch its direction ($w < 1$ in non-ISN). This leads to the same solution as in Eq. 3.78. It follows that $A_P < 1$ in the non-ISN and, therefore, that in this reduced one dimensional derivation based on Eq. 3.76, PV firing rates can be facilitated above baseline with $\zeta_{PE} > 1$.

However, the generalization to two dimensions is not as simple as in the ISN. With respect to Eq. 3.75, $A_P < 1$ by no means implies that $\zeta_{EP} A_P < 1$ and thus, that ζ_{PE} can induce facilitation for arbitrary values of ζ_{EP}. However, this also aligns with my numerical results. The conclusion we can draw is that facilitation is possible in the non-ISN, but it is not in the whole (ζ_{EP}, ζ_{PE})-plane. Rather, for any $\zeta_{PE} > 1$ that induces facilitation in the biological range of recurrent synaptic plasticity, there will also be a $\zeta_{EP} > 1$ that overcomes this and will ultimately lead to suppression of PV firing rates. The trade-off between these two parameters is visible in the boundary between facilitation and suppression (Fig. 3.11C, right). I derived the expression for this boundary between facilitation and suppression, assuming equality. One can solve Eq. 3.75 without considering the sign $(1 - \zeta_{EP} A_P)$ and find

$$\zeta_{PE} = \frac{1 - A_P}{(1 - \zeta_{EP} A_P)}. \tag{3.80}$$

I plotted this in Fig. 3.11C (right) as a family of curves over a range of w spanning both non-ISN and ISN. The boundary between facilitation and suppression indeed crosses the biological range of the (ζ_{EP}, ζ_{PE})-plane in the non-ISN, not if the network is an ISN.

3.8 Discussion

I studied the repertoire of activity changes in recurrent networks undergoing experimentally established synaptic changes during the first two days of MD. Generally, this repertoire seems rich enough to explain the activity changes measured during brief MD *in vivo*. However, the effects strongly depend on the overall recurrent coupling in the circuit, revealing a dichotomy between the ISN and the non-ISN in the possible activity changes following MD, specifically for the PV-interneurons.

Excitatory firing rates during the first two days of MD stay at baseline on the first day and are suppressed on the second day [54]. In the model, their rate can either be kept at baseline, be suppressed or facilitated, in different portions of the parameter space of feedforward depression (Figs. 3.3 and 3.4). In the parameter space of recurrent potentiation, their firing rate is exclusively suppressed. Interestingly, these general properties are independent of whether the network is an ISN or non-ISN. The suppression of the excitatory firing rate could underly their suppression in the animal on the second day of MD. The parameter ranges that keep excitatory rates around their baseline could explain the preserved baseline firing on the first day. Alternatively, a combination of suppression and facilitation through different synaptic changes could also model the preserved firing rates or their suppression, depending on the balance between facilitation and suppression. I found an antagonistic interaction between shifts of the E/I-ratio in the feedforward and recurrent connectivity. For a large set of networks with randomly generated connectivity parameters, the increase of the feedforward E/I-ratio robustly drives a strong increase of the network firing rate, which is counteracted by the suppression of firing rates through a decrease of the recurrent E/I-ratio (Fig. 3.7B). A possible biological interpretation is that the feedforward changes destabilize the circuit, driving the firing rates up to boost the residual, less reliable visual signals arriving in V1 after one eye is closed. The recurrent plasticity invoked in the biological circuit could be a form of stabilizing, homeostatic plasticity. I found further evidence that recurrent plasticity plays a homeostatic role through the analysis of the population rate dynamics, which shows that recurrent plasticity leads to more stable networks (Sec. 3.6.2).

In contrast to the excitatory neurons, PV firing rates are suppressed on the first day of MD, but already recover back to baseline on the second day *in vivo* [54]. In the model, the effect of MD-induced plasticity on the PV firing rates strongly depends on the operating regime. In the ISN, feedforward plasticity can induce both facilitation and suppression of PV firing rate, whereas recurrent plasticity induces only suppression (Fig. 3.3). These types of firing rate changes of PV-interneurons in the ISN are tightly coordinated to those in the excitatory neurons. In the non-ISN, feedforward plasticity acts exclusively suppressive on PV firing rates while recurrent potentiation can induce facilitation and could therefore explain their recovery on the second day (Fig. 3.4).

The non-ISN allows an independent modulation of the firing rates of the excitatory and PV neurons in large parts of the parameter space of MD-induced plasticity. In contrast, in the ISN, the tight coordination of the firing rate changes of excitatory and PV neurons prohibits an independent modulation of the firing rates for the largest parts of the parameter space of MD-induced plasticity (Fig. 3.5). Therefore, the non-ISN seems to be a suitable model for the independent modulation of firing rates *in vivo* that was shown in [54]. In the ISN, in contrast, this independent modulation requires fine-tuned synaptic changes. However, a problem in the non-ISN is that the amount of the firing rate changes is very small (Fig. 3.6). Firing-rate regulation thus requires very large synaptic modifications to achieve the same amplitude of firing rate modulations as has been found *in vivo*. A second problem of the non-ISN is that

experimental evidence suggests that cortical circuits are ISN [131,135,136]. However, none of the studies of the operating regime was conducted in the developing animal. It not clear if the operating regime of the visual cortex during the critical period matches with the experimentally investigated operating regime in the adult animal.

Taken together, both the non-ISN and the ISN come with specific problems that impede their usefulness to explain the full sequence of firing rate changes of excitatory and PV neurons. The experimental evidence that suggests the cortical circuit to be ISN raises doubts on an explanation of the independent firing rate changes in a non-ISN. Furthermore, strong coupling also has computational benefits that mammalian cortices could make use of [15, 133]. For example, strongly coupled networks with balancing inhibition can explain sharp orientation tuning, even when neurons in the network are randomly wired, that is, without a requirement to specifically tune recurrent connections to the different stimulus preferences of the single cells [125,126]. Such networks also show a very interesting form of amplification of external signals, called *balanced* or *non-normal amplification* [137]. This balanced amplification enables a network to strongly enhance a signal transiently, leading to strong amplification that can reliably signal to downstream areas on a behaviorally relevant timescale. Lastly, independent of the transient amplification, circuits with strong coupling can respond very fast to external stimulation [112, 143]. This would be a beneficial property for sensory processing in an ecological context.

My analysis has shown that the main mechanism that drives the tight coordination between excitatory and PV firing rates in the ISN is the paradoxical effect. Thus, the problem I found for the ISN, this tight coordination, may not be a property of the strong coupling per se, but of a network in which inhibitory stabilization and the paradoxical effect underly the exact same condition. This identity is generally true for networks with only a single population of interneurons [114]. However, a multiplicity of interneuron subtypes, a prominent property of cortical circuits, has been shown to enable non-paradoxical responses in certain subtypes of interneurons also in the ISN [130]. Such a network could combine the computational benefits of the strong coupling scale with that of a network in which PV-interneurons do not express the paradoxical effect. I, therefore, extended my model to include another subtype of interneurons beyond the PV.

4 The role of interneuron diversity for the effects of deprivation

Remark: Some of the results, figures, and tables in this chapter are part of an article entitled *Interneuron subtypes enable independent modulation of excitatory and inhibitory firing rates after sensory deprivation* which has been written together with my supervisor Julijana Gjorgjieva. The article has been uploaded to the preprint server *bioRxiv* [1] and is currently under review for publication in the journal *Proceedings of the National Academy of Sciences of the United States of America*. All methods, results, figures, and tables from this article which are part of this chapter were my contribution to the article unless specifically mentioned otherwise.

4.1 Overview

Circuits in the mammalian neocortex contain many subtypes of interneurons that play different functional roles [144]. The PV-interneurons I have modeled in the previous chapter are the most numerous of interneurons in primary sensory cortices. The second most numerous class are SST-interneurons [24]. Experimental work made clear that both of these interneuron types, PV and SST, play essential and distinct roles in shaping the dynamics and functional responses in cortical circuits [55, 144, 145]. For example, in orientation tuning and surround suppression in V1 [59, 60, 63, 64], in working memory [146, 147] and in the generation of cortical oscillations [148]. These interneuron subtypes also play a role in the control of plasticity during development and learning [149–152]. I therefore extended my network model to include both PV- and SST-interneurons.

A multiplicity of interneuron subtypes makes the relation between the paradoxical effect in the rates of inhibitory interneurons and the operating regime (ISN or non-ISN) more complex [130, 134] (also outlined in Sec. 2.5.3). The the total inhibitory current still shows the paradoxical effect in an ISN, but this only constrains the sum of the rate changes (Fig. 2.7). Thus, different inhibitory subpopulations can behave non-paradoxically as long as this is canceled and overcome by a paradoxical response in other subtypes, such that the total inhibitory current shows the paradoxical behavior [130]. Since it became clear from my results in the previous chapter that the paradoxical effect in PV-interneurons is a major determinant of the tight coordination of excitatory and PV neurons, the question emerged if there is a way that the paradoxical effect could be reversed in PV-interneurons even in an ISN. Crucially, the question is if such a reversal could be generated with biologically

justified connectivity, e.g., with SST-interneurons that couple onto the other cell-types similar to the experimental findings from V1. If this is the case, the follow-up question would be, whether an ISN with such a "reversed" paradoxical effect could enable the temporally independent modulation of excitatory and PV firing rates also in an ISN. In such a setup, one could combine the benefits of both the non-ISN, with its great potential to independently modulate excitatory and PV firing rates, and the ISN, with its higher responsiveness that enables sufficiently strong rate modulations with realistic synaptic changes. Therefore, I investigated the firing rate changes in an ISN that includes two subtypes of interneurons, PV and SST (Sec. 4.2). The basic questions I want to ask are: Firstly, can a biologically justified connectivity of SST-interneurons lead to a reversal of the paradoxical effect in PV-interneurons? I found that this is indeed possible. Secondly, how does such a reversal of the paradoxical effect change the response to MD-induced plasticity? I studied the landscape of potential responses to MD-induced plasticity (Sec. 4.3). In an ISN with multiple interneurons subtypes, the reversal of the paradoxical effect indeed leads to a complete inversion of the response landscape of PV-interneurons to MD-induced plasticity. This inversion of the PV response landscape in an ISN leads to opposite modulation of the excitatory and the PV firing rates, making this network a suitable candidate to model the temporally distinct firing rate modulations *in vivo*, even when it operates in the ISN regime. To analyze how distinct the firing rate changes of excitatory and PV neurons are as a function of feedback from SST-interneurons into the circuit, I did a similar analysis as for the network with a single subtype of interneurons, where I varied the coupling scale J (Sec. 3.4). For the network with two subtyoes of interneurons, I quantified how the area of co-modulation and the gradients of the firing rate modulations behave over an extensive range of SST-feedback strengths (Sec. 4.4).

I furthermore studied two alternative types of plasticity in the circuit that could be related to deprivation (Sec 4.5). First, the ISN with multiple subtypes of interneurons could explain the decrease of the recurrent E/I-ratio (measured in [3]) also through plasticity in synapses that originate from SST-interneurons. I investigated the effects of this potential plasticity and found them to be very similar to the effects of the plasticity involving only the excitatory-PV circuit in the network with strong SST-feedback (Sec. 4.5.1). Deprivation-induced plasticity has not been directly measured in synapses originating from SST-interneurons, so this is a new prediction of our model that points to a different pathway that enables an independent modulation of excitatory and PV firing rates. Second, beyond MD I investigated the effects of plasticity identified for a closely related experimental paradigm to understand experience-dependent developmental plasticity in sensory cortices, whisker deprivation (WD, Sec. 4.5.2). The deprivation in WD consists of a partial plucking of whiskers in young animals. This plucking of whiskers induces substantial changes in the neural circuitry, similar to the sustained closure of one eye in MD, but in different circuit elements. Crucially, also following WD plasticity, an increase of the feedforward E/I-ratio is found experimentally. However, the mechanism underlying this is a decrease of the intrinsic excitability of PV-

interneurons rather than a depression of direct feedforward inputs [96, 97, 99]. My modeling predicted that the effect of the decrease of PV intrinsic excitability is very similar to the depression of feedforward inputs. This hints at a possible shared principle of activity regulation involved, regardless of the underlying biophysical mechanism.

To analytically investigate the circuit that includes both PV and SST interneuron subtypes, I extended the linear population rate model (Sec. 4.6). I was able to analytically show the connection between the reversal of the paradoxical effect and the inversion of the normalized firing rates of PV-interneurons in response to MD-induced plasticity (Sec. 4.7).

Finally, I summarize the distinct abilities of the different networks with a single and with two subtypes of interneurons to produce independent or opposite firing rate changes in excitatory and PV neurons in the spiking network using a Monte-Carlo approach and directly compare the predictions of the different networks to experimental data (Sec. 4.8) and discuss the results (Sec. 4.9).

4.2 Modeling approach

Remark: Some of the results of this section (Fig. 4.1A-C, Tables 4.1 and 4.2, and Eqs. 4.1-4.4) have been published before in [1]. The figure panel has been modified for clarity, notation in the equations and table has been adapted for consistency.

The network with multiple subtypes of interneurons I investigated is a direct extension of the previous model, with the same sparse connectivity and single neuron dynamics (Equations 3.3-3.5), but with specific connectivity and coupling parameters for the SST-feedback. In addition to the $N_E = 4000$ excitatory and $N_P = 1000$ PV neurons, I added $N_S = 500$ SST-interneurons (Table 4.1). To build on the understanding of the circuit with a single subtype of interneurons, connections among excitatory and PV neurons have the same weights as before, with parameters J for the coupling scale and g_{rc} for the relative strength of inhibition from PV-interneurons (Fig. 4.1A). SST-interneurons receive recurrent inputs from the excitatory neurons with synaptic strength J. In contrast to PV-interneurons, SST-interneurons do not make inhibitory synapses among themselves, nor do they receive inhibition from PV-interneurons [56]. On the other hand, outputs from SST-interneurons are send to both excitatory and PV neurons. I included such SST-outputs to both other neuron classes (excitatory and PV) with a common SST-feedback strength K. The use of this single parameter allowed me to continuously transition between the network with a single subtype of interneurons (equivalent to $K = 0$ nS) and the network with multiple subtypes of interneurons, with adjustable levels of overall SST-feedback. With indices E for excitatory, P for PV and S for SST neurons, the matrix schematic describing the recurrent connectivity is (compare

to Eq. 3.1):

$$\boldsymbol{W}_{\text{rec}} = \begin{pmatrix} W^{EE} & W^{EP} & W^{ES} \\ W^{PE} & W^{PP} & W^{PS} \\ W^{SE} & W^{SP} & W^{SS} \end{pmatrix} = \begin{pmatrix} J & g_{rc}J & K \\ J & g_{rc}J & K \\ J & 0 & 0 \end{pmatrix}. \tag{4.1}$$

In the visual cortex, SST-interneurons receive very little feedforward input from the thalamus [38], but they receive a variety of external inputs from the surrounding cortical tissue, pooling over much larger areas than PV [59], and a manifold of inputs from higher-order cortical areas [61, 69]. These inputs provide a background (BKG) that is independent of the first order thalamic input. BKG is the only external input delivered to SST-interneurons, with weight J_b. The BKG-input is also connected to excitatory neurons (weight J_b), in addition to their feedforward inputs from LGN (weight J_s). PV-interneurons are exclusively driven by strong LGN-input with weight $g_{fw}J_s$. Schematically, the feedforward connectivity is:

$$\boldsymbol{W}_{\text{ffw}} = \begin{pmatrix} W^{E}_{LGN} + W^{E}_{BKG} \\ W^{P}_{LGN} + W^{P}_{BKG} \\ W^{S}_{LGN} + W^{S}_{BKG} \end{pmatrix} = \begin{pmatrix} J_s + J_b \\ g_{fw}J_s \\ J_b \end{pmatrix}. \tag{4.2}$$

LGN and BKG inputs are delivered as excitatory spike trains with Poisson statistics. The spike trains are statistically independent between the neurons. Activity in the network with this connectivity and inputs is asynchronous and irregular over wide ranges of SST-feedback K (example in Fig. 4.1B), as is desirable for a model of cortical circuits. As a function of SST-feedback K, firing rates of excitatory and PV neurons in the circuit are moderately suppressed but are in an appropriate range over large ranges of K (Fig. 4.1C). SST-rates are relatively high for very small K, but also reach a biologically reasonable range for modest K.

I specifically tested the paradoxical effect in PV-interneurons as a function of K in an ISN (with $J = 0.1$ nS). With multiple interneuron subtypes, one can study the paradoxical effect in multiple ways (injecting to both populations, only to PV or only to SST). Concerning my previous model and the role the paradoxical effect plays in the firing rate modulations in response to MD-induced feedforward and recurrent plasticity, I tested the rate change of PV-interneurons in response to an increase of the external input to PV. In the ISN without or with small SST-feedback, this leads to a decrease of the PV firing rate (Fig. 4.1D, yellow area). As K increases, the rate decrease of PV-interneurons becomes smaller until it becomes positive (Fig. 4.1D, green line). The larger K becomes, the more positive is the response, and the PV-interneurons react non-paradoxically, even though the network as a whole is an ISN.

Little is known about how brief MD affects synapses that involve SST-interneurons, either in the recurrent circuit or in their external inputs. I therefore investigated the effects of the same MD-induced synaptic changes as in the previous chapter

Table 4.1: Default parameters of the single neurons and connectivity in the network with two subtypes of interneurons. EPSG/IPSG: excitatory/inhibitory postsynaptic conductance; LGN: lateral geniculate nucleus; BKG: background;

Symbol	Value/ Range	Description
C_m	200 pF	Membrane capacitance.
g_L	10 nS	Leak conductance.
E_L	-70 mV	Leak reversal potential.
V_θ	-50 mV	Threshold potential.
V_r	-58 mV	Post-spike reset potential.
E_E	0 mV	Excitatory reversal potential.
E_P, E_S	-85 mV	Inhibitory reversal potential from PV (P) and SST (S) interneurons.
τ_{syn}	5 ms	Synaptic conductance time constant.
N_E	4000	Number of excitatory neurons.
N_P	1000	Number of PV-interneurons.
N_S	500	Number of SST-interneurons.
p	0.1	Connection probability between two cells.
J	0.1 nS	EPSG amplitude $E \to E$, $E \to$PV and $E \to$SST.
g_{rc}	8	Multiplicative factor for IPSG amplitude PV$\to E$ and PV$\to$ PV.
K	1.6 nS	IPSG amplitude for SST $\to E$ and SST $\to$ PV connections.
	$[0.1, 4.0]$ nS	Range of IPSG amplitudes SST$\to E$ and SST$\to$PV for the systematic quantification across levels of overall SST-feedback.
J_s, J_b	0.5 nS	EPSG amplitude $LGN, BKG \to E$, $BKG \to$ SST and scale for EPSG $LGN \to$ PV
g_{fw}	2	Multiplicative factor for strong feedforward EPSG $LGN \to$ PV.
R_x	1000 Hz	Rate of feedforward input spike trains $LGN \to E$ and $LGN \to$ PV and of background input spike trains $BKG \to E$ and $BKG \to$ SST.

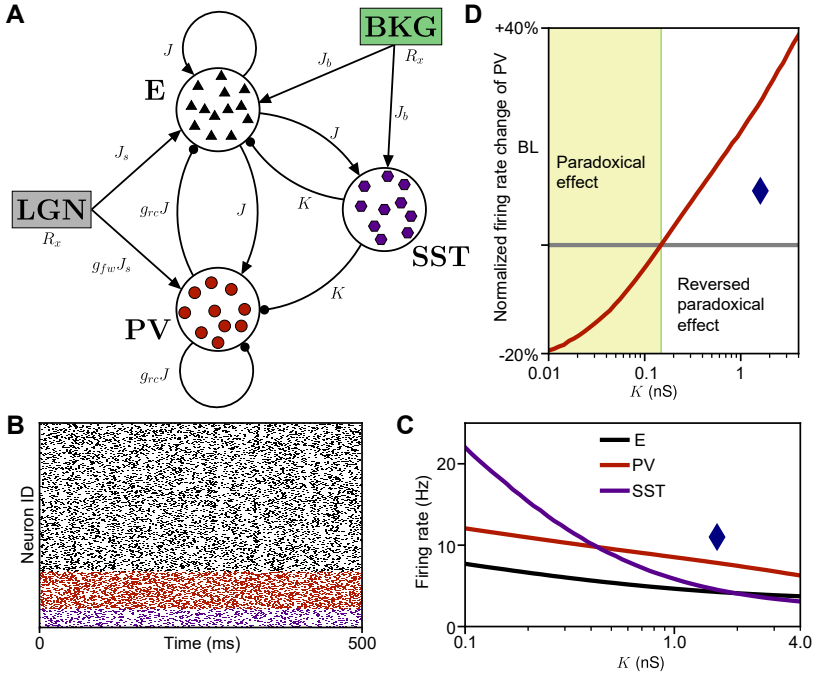

Figure 4.1: Network structure, activity and the paradoxical effect in the spiking network with two subtypes of interneurons. (A) Schematic of the connectivity and parameters in the network (default values in Table 4.1). **(B)** Spike raster of activity in the network with coupling scale $J = 0.1$ nS and SST-feedback $K = 1.6$ nS, showing asynchronous irregular state also with strong SST-feedback. **(C)** Firing rates of excitatory (black), PV (red) and SST (purple) cells as a function of SST-feedback K. The blue diamond shows the value of K used in (B). **(D)** Reversal of the paradoxical effect in PV-interneurons through SST-feedback in an ISN. Panels (A-C) modified from [1].

(Equations 3.8 and 3.7). The recurrent connectivity changes through MD as:

$$\boldsymbol{W}_{\text{rec}}^{BL} = \begin{pmatrix} J & g_{rc}J & K \\ J & g_{rc}J & K \\ J & 0 & 0 \end{pmatrix} \quad \Longrightarrow \quad \boldsymbol{W}_{\text{rec}}^{MD} = \begin{pmatrix} J & \zeta_{EP}g_{rc}J & K \\ \zeta_{PE}J & g_{rc}J & K \\ J & 0 & 0 \end{pmatrix} \tag{4.3}$$

(see also Fig. 4.2). The feedforward connectivity changes as:

$$\boldsymbol{W}_{\text{ffw}}^{BL} = \begin{pmatrix} J_s + J_b \\ g_{fw}J_s \\ J_b \end{pmatrix} \quad \Longrightarrow \quad \boldsymbol{W}_{\text{ffw}}^{MD} = \begin{pmatrix} \delta_E J_s + J_b \\ \delta_P g_{fw}J_s \\ J_b \end{pmatrix}. \tag{4.4}$$

I investigated the response landscape of firing rate fold changes for this MD-induced plasticity for K large enough for a strongly reversed paradoxical effect in PV (blue diamond in Fig. 4.1D). As a further type of plasticity of recurrent connections, I investigated a potential pathway for MD-induced plasticity involving the output-synapses of SST-interneurons. A potentiation of synapses from SST onto excitatory

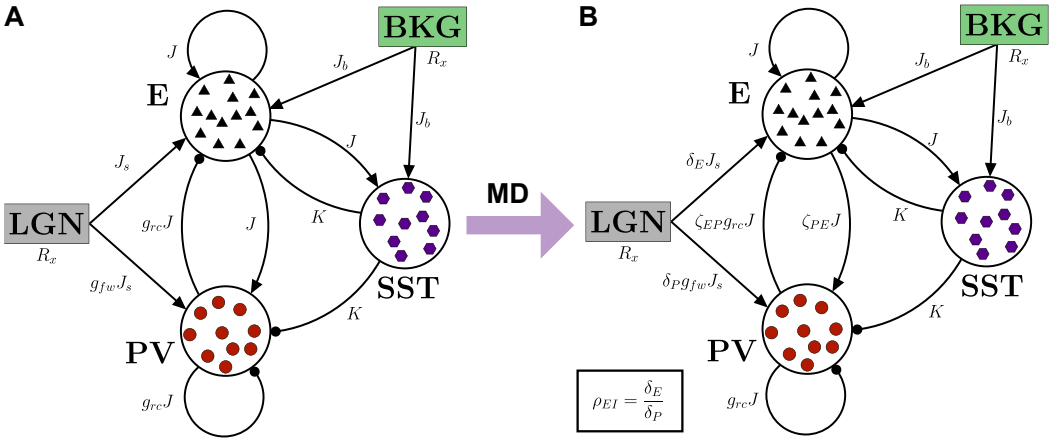

Figure 4.2: Connection strengths in the spiking network with two subtypes of interneurons before and after MD. (A) Connectivity in the baseline network before application of MD. **(B)** Connectivity in the network after application of MD.

neurons or a suppression of synapses from SST- onto PV-interneurons could also underlie the observed decrease of the recurrent E/I-ratio measured at excitatory neurons (see Discussion in [3]. The matrix describing this hypothetical plasticity is:

$$\boldsymbol{W}_{\text{rec}}^{BL} = \begin{pmatrix} J & g_{rc}J & K \\ J & g_{rc}J & K \\ J & 0 & 0 \end{pmatrix} \quad \Longrightarrow \quad \boldsymbol{W}_{\text{rec}}^{MD,hypo} = \begin{pmatrix} J & g_{rc}J & \zeta_{ES}K \\ J & g_{rc}J & \zeta_{PS}K \\ J & 0 & 0 \end{pmatrix}. \tag{4.5}$$

As for the experimentally established MD-induced synaptic changes, I investigated the effects of these plastic changes in a systematic scan over the (ζ_{PS}, ζ_{ES})-plane, from which I obtained a landscape of responses to this hypothetical MD-induced plasticity.

4.3 Landscape of network responses to MD-induced plasticity: strong SST-feedback inverts PV response

Remark: The results of this section (Fig. 4.3) have been published before with identical content in [1]. The figure has been modified for consistency.

As in the case with a single subtype of interneurons, I studied the MD-induced synaptic changes split up into three different types of plasticity: (1) depression of feedforward inputs onto excitatory and PV-interneurons ($\delta_E < 1$, $\delta_P < 1$, Table 4.2); (2) potentiation of recurrent connections between excitatory and PV-interneurons ($\zeta_{EP} > 1$, $\zeta_{PE} > 1$); and (3) interacting feedforward and recurrent plasticity ($\zeta_{PE} > 1$, $\rho_{EI} > 1$). In all of these types of plasticity, I simulated the networks on a two-dimensional grid spanning each respective parameter space, generating multiple

landscapes of firing rate fold-changes for each cell type ($\Psi^A(x, y)$, Eq. 3.10, with x and y from the respective parameter spaces and $A \in \{E, P, S\}$). I simulated all the networks with multiple interneurons subtypes using $J = 0.1$ nS, thus they are ISN, and with strong SST-feedback $K = 1.6$ nS that reverses the paradoxical effect in PV-interneurons (blue diamond in Fig. 4.1D).

For feedforward plasticity, excitatory neurons have a facilitatory and a suppressive region of firing rate modulations distinguished by a linear boundary (Fig. 4.3B, left). The position of this boundary appears to be the same as in the ISN without SST-feedback (Fig. 3.3B). Also for recurrent and interacting feedforward and recurrent plasticity, excitatory neurons respond similar to the network with a single subtype of interneurons (Fig. 4.3C, D, compare to Fig. 3.3C, D).

Interestingly, the response plane of PV-interneurons in the ISN with two subtypes of interneurons and reversed paradoxical effect in PV is not comparable to the response of PV in a non-ISN (Fig. 4.3B-D, compare to Fig. 3.4B-D). Rather, for all types of plasticity it is inverted relative to the excitatory neurons, showing facilitatory and suppressive regions in opposite parts of the feedforward parameter space as the excitatory neurons do (Fig. 4.3B-D, middle). This is also opposite to the responses of PV-interneurons in the ISN with a single subtype of interneuron (Fig. 3.3B-D). For feedforward plasticity, one can see the reversal of the paradoxical effect in the response plane as a purely suppressive action of δ_P, the depression of direct external inputs to the PV-interneurons (Fig. 4.3B, y-axis of the middle plane).

Table 4.2: Parameters for modeling the synaptic changes during monocular deprivation in layer 4 of V1, as well as for the changes during whisker deprivation in layer 2/3 of S1.

Symbol	Value/ Range	Description
δ_E	$[0.5, 1.0]$	Depression of feedforward synapses onto excitatory neurons.
δ_P	$[0.5, 1.0]$	Depression of feedforward synapses onto PV-interneurons.
ζ_{PE}	$[1.0, 1.5]$	Potentiation of recurrent synapses from excitatory neurons to PV-interneurons.
ζ_{EP}	$[1.0, 1.5]$	Potentiation of recurrent synapses from PV-interneurons to excitatory neurons.
ρ_{EI}	$[1.0, 1.5]$	Increase of feedforward E/I-ratio in V1.
ζ_{ES}	$[1.0, 1.5]$	Hypothetical potentiation of recurrent synapses from SST to excitatory neurons.
ζ_{PS}	$[0.5, 1.0]$	Hypothetical depression of recurrent synapses from SST- to PV-interneurons.
ξ_θ	$[0, 3]$ mV	Increase of firing threshold of PV-interneurons in S1.

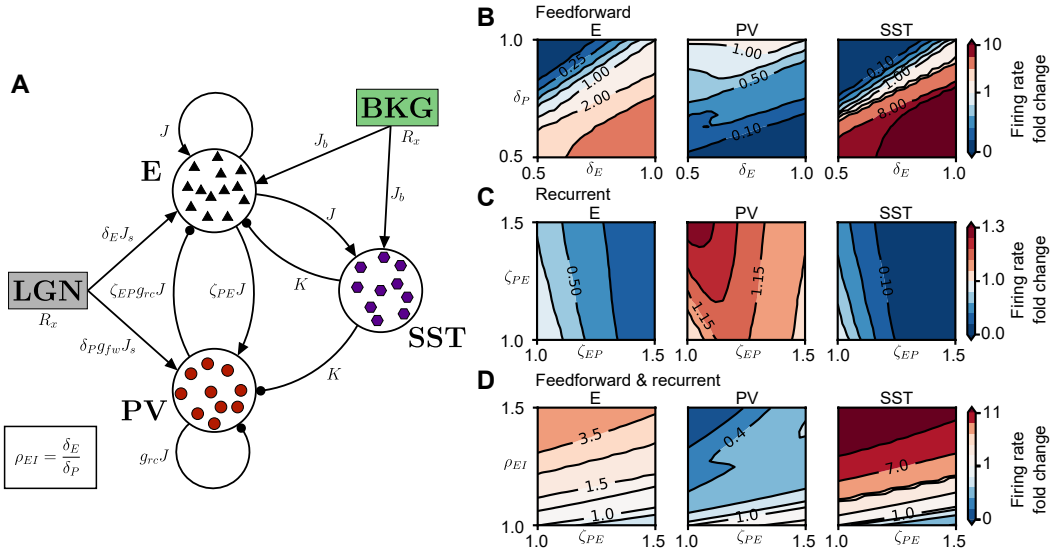

Figure 4.3: Response to synaptic changes induced by brief MD in a spiking network with multiple subtypes of interneurons in ISN regime (compare to Fig. 3.3). **(A)** Network schematic showing synaptic connections among neurons with J denoting the overall coupling scale, g_{rc} the dominance of recurrent PV-inhibition, g_{fw} the dominance of feedforward inhibition and K the strength of SST-feedback. The parameters to model MD-induced synaptic plasticity are: depression of feedforward drive to excitatory neurons (δ_E) and to PV-interneurons (δ_P), and potentiation of recurrent excitation to PV-interneurons (ζ_{PE}) and of recurrent inhibition from PV to excitatory neurons (ζ_{EP}). Values of network parameters are provided in Tables 4.1 and 4.2. **(B)** Network firing rate in the (δ_E, δ_P) plane as fold-change of baseline firing rate (top right corner where $\delta_E = \delta_P = 1$) for excitatory neurons (left panel) and PV-interneurons (right panel). **(C)** Network firing rate in the (ζ_{EP}, ζ_{PE}) plane as fold-change of baseline firing rate (bottom left corner where $\zeta_{EP} = \zeta_{PE} = 1$). **(D)** Network firing rate in the (ζ_{PE}, ρ_{EI}) plane as fold-change of baseline firing rate (bottom left corner where $\zeta_{PE} = \rho_{EI} = 1$). Here $\rho_{EI} = \delta_E/\delta_P$ is the feedforward E/I ratio. Modified from [1].

SST firing rates are modulated the same way as excitatory neurons for all types of MD-induced plasticity (Fig. 4.3B-D). This is intuitive since none of the experimentally established plastic changes I applied here directly involved synapses onto or from SST-interneurons (Eq. 4.3). Thus, the only aspect that changes during brief MD from the perspective of SST-interneurons is the firing rate of the excitatory neurons from which they receive recurrent inputs. SST-interneurons therefore closely follow the excitatory rate in either being suppressed or facilitated. This predicts that during brief MD, the SST firing rates should be modulated in the same direction as excitatory neurons, staying at baseline on the first day of MD and being suppressed on the second day MD.

Beyond this intuitive explanation of why SST firing rate fold changes resemble those in the excitatory neurons, one can also see an aspect of the paradoxical effect with multiple subtypes of interneurons at work here, though more general than it was

studied before. Previous work generalized the classical paradoxical effect to multiple interneuron subtypes with the result that the inhibitory current as a whole, i.e., the weighted sum of the rates, should show the paradoxical effect. Thus, if specific interneuron subtypes do not show a paradoxical response, others subtypes should compensate for this. In the higher dimensional responses to MD-induced plasticity, I found a similar type of compensation to happen: PV-interneurons change their firing rates opposite to the excitatory neurons and opposite to how they change in the ISN without SST-feedback. However, SST neurons, in turn, follow the excitatory neurons very tightly, reminiscent of the compensation in the paradoxical effect.

Overall, I thus found that a reversal of the paradoxical effect in PV-interneurons through feedback from SST-interneurons selectively inverted the response of PV-interneurons to MD-induced plasticity. This leads to opposite firing rate fold changes in excitatory and PV neurons and is a mechanism to explain the temporally distinct firing rate modulations of excitatory and PV-interneurons during brief MD *in vivo*. In both cases that allow such a distinct modulation, the non-ISN and the ISN with strong SST-feedback, the central property seems to be a non-paradoxical response of the PV. The portion of the parameter spaces in which excitatory and PV neurons are modulated in opposite directions appears to be even larger in the ISN with multiple interneuron subtypes than it is in the non-ISN. To better understand this, I quantified the global response structure with varying degrees of SST-feedback K, on a range that spans networks with a paradoxical and with reversed paradoxical response in PV-interneurons.

4.4 Quantification of network response structure to MD-induced plasticity across levels of SST-feedback

Remark: Some of the results of this section (Fig. 4.4 in Sec. 4.4.1) have been published before with identical content in [1].

How does the selective inversion of PV-responses happen as a function of SST-feedback K? Moreover, is the invariance of the excitatory responses that appeared with the selected plane a feature that generally holds for a wide range of K? I addressed these questions using the measures I have introduced for the global response structure of excitatory and PV-interneurons in the study of the network with just a single subtype of interneurons (Sec. 3.4).

I used the fractional area of co-modulation of excitatory and PV firing rates to quantify how much of the respective parameter spaces allow for independent modulation of the two cell classes as a function of the SST-feedback K (Sec. 4.4.1). When K increases, there is a general trend towards an independent modulation of the firing rates of these two cell populations. Using the average lengths of the gradients for excitatory and PV firing rate fold changes, I quantified how strong the different firing rate modulations are (Sec. 4.4.2). This is especially relevant as we

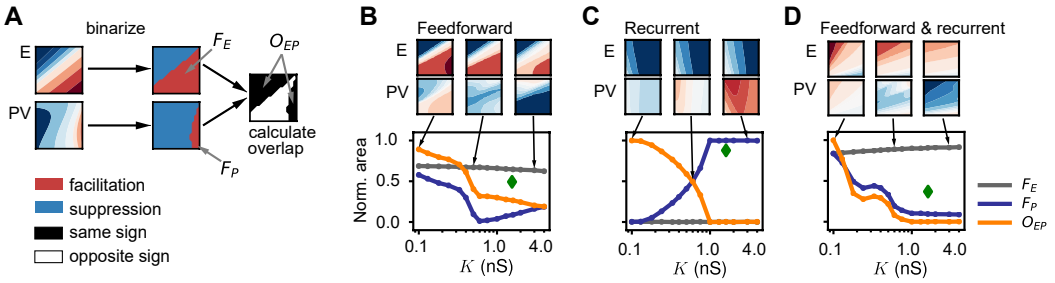

Figure 4.4: Parameter space of facilitation and suppression as a function of SST-feedback K (compare to Fig. 3.5, where I did the same quantification as a function of the coupling scale J). **(A)** Schematic to determine the overlap of facilitating and suppressive response areas of excitatory neurons and PV-interneurons. The thresholded plane of responses (as in Fig. 4.3) for both neuron types is used to compute the total facilitating area and quantify how closely excitatory and PV firing rates follow each other through the overlap of facilitation and suppression. **(B)** Network with feedforward depression only (Fig. 4.3B). Fractional area of facilitation for excitatory neurons (F_E, gray), PV-interneurons (F_P, blue) and the overlap between excitatory and PV response areas (O_{EP}, orange) as a function of SST-feedback strength K. Green diamond shows the value of K used in Fig. 4.3. **(C)** Same as (B) for recurrent potentiation only (Fig. 4.3C). **(D)** Same as (B) for interacting feedforward depression and recurrent potentiation (Fig. 4.3D). Modified from [1].

found a non-ISN with a single subtype of interneurons to enable independent firing rate modulations but at the cost of very small responsiveness of the network. In contrast to this, I found that in the ISN with strong SST-feedback, the network has sufficient responsiveness also in the regimes where PV firing rate fold changes are inverted.

4.4.1 Parameter space of co-modulation of activity during MD

I simulated the complete planes of firing rate fold changes in an ISN with constant $J = 0.1$ while varying K over the range ($K \in [0.1, 4.0]$, Table 4.1). From the planes for all the different values of K, I extracted the fractional areas of facilitation for excitatory and PV neurons (Fig. 4.4). I used the same method for this analysis as in Sec. 3.4.1: I thresholded and then binarized the plane of firing rate fold changes for the excitatory and PV neurons ($\Psi^A(x, y)$, with $A \in \{E, P\}$) at 1, the boundary between facilitation and suppression (Fig. 4.4A). I then calculated the fractional area of facilitation by summing all entries corresponding to facilitation and dividing this sum by the total number of entries. The fractional area of facilitation as a function of SST-feedback K is shown in gray for excitatory neurons (F_E, Fig. 4.4B-D) and in dark blue for the PV-interneurons (F_P). I then calculated the area of co-modulation of excitatory and PV firing rates as the fractional overlap between facilitatory and suppressive areas in the two cell classes. I multiplied each entry in the binarized excitatory response with the corresponding entry in the binarized PV response (element-wise product). When the firing rates of the two cell classes are modulated

in the same direction, the product of the two entries has the value +1 (black region in the combined plane in Fig. 4.4A). If the firing rates of the two cell classes are modulated in opposite directions, the product has the values −1 (white region in the combined plane in Fig. 4.4A). The fractional area of co-modulation is calculated by summing up all positive entries in this product matrix and dividing this by the total number of entries in the plane (O_{EP}, orange, Fig. 4.4B-D).

For small SST-feedback ($K = 0.1$ nS), the facilitatory areas and the area of co-modulation are very similar to the areas in the ISN without SST-feedback in all three different types of plasticity (feedforward, recurrent, and combined feedforward and recurrent, Fig. 4.4B-D, compare to Fig. 3.5B-D). This confirms that only sufficiently strong SST-feedback can lead to the complete inversion of the PV-response. Crucially, the area of co-modulation of excitatory and PV neurons is close to one, showing the tight coordination of the firing rates in the two populations typical of the ISN in which PV-interneurons express the paradoxical effect.

For pure feedforward plasticity, the area of co-modulation monotonically decreases as K increases (O_{EP}, Fig. 4.4B). This monotonic decrease of O_{EP} is almost exclusively driven by the change in the fractional area of facilitation for PV-interneurons (F_P, Fig. 4.4B). Interestingly,F_P does not change monotonically but decays to a minimum at $K \approx 0.6$ and then increases again. The fractional area of facilitation for excitatory neurons (F_E) is almost constant throughout the range of K probed. Combining these three factors, (1) the monotonic decrease of the fractional area of co-modulation, (2) the non-monotonic change of the fractional area of facilitation for PV-interneurons, and (3) the approximate invariance of the fractional area of facilitation for excitatory neurons, shows that the inversion of the response landscape of PV-interneurons happens in a two stage process. Increasing SST-feedback K first leads to a decrease of the fractional area of facilitation for PV-interneurons in the parts of the parameter space where it originally lies in the ISN (and where it overlaps with the region of facilitation for excitatory neurons). Further increasing of K then leads to a reemergence of a facilitatory region for PV firing rate fold changes in a different portion of the parameter space. This reemergence is central for the inversion of the distribution of facilitation and suppression as in Fig. 4.3. The reemergence is also what distinguishes the network with reversed paradoxical effect through SST-feedback from the non-ISN.

For pure recurrent plasticity, a single process leads from complete overlap of the facilitatory and suppressive areas (fractional area of co-modulation of excitatory and PV firing rates $O_{EP} = 1$ for small $K = 0.1$ nS), to no overlap for large K ($O_{EP} = 0$, Fig. 4.4C). Excitatory neurons have zero facilitatory area throughout the range of K; thus, excitatory neurons are always purely suppressed by recurrent plasticity. In contrast, the facilitatory area of PV-neurons (F_P) monotonically increases with K until their rates facilitate in all of the parameter space. For interacting feedforward and recurrent plasticity, I found the same change from co-modulation of excitatory and PV-firing rates in the whole parameter space, to zero co-modulation as the SST-feedback K increases (Fig. 4.4D). F_E is only very little modulated as a function of K, while I find a pronounced decay of F_P as K increases. Thus, also for interacting

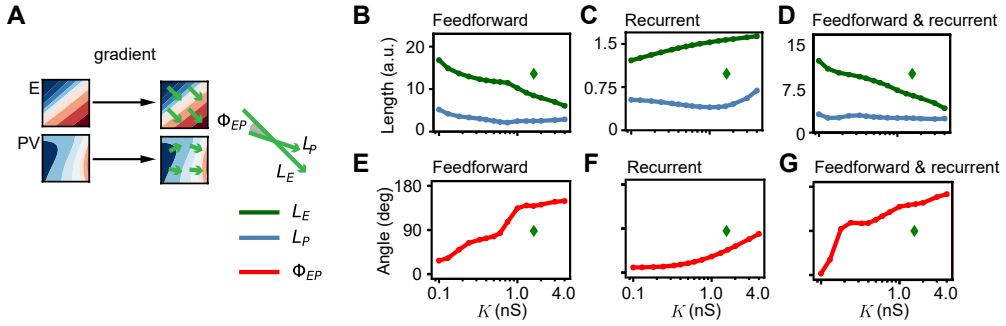

Figure 4.5: Responsiveness of excitatory and inhibitory firing rates to MD-induced plasticity as a function of SST-feedback K in a network with multiple subtypes of interneurons (compare to Fig. 3.6). **(A)** Schematic for the extraction of average gradient lengths and angle between gradients for excitatory and PV responses. I used the length of the gradient as a proxy for the magnitude of responsiveness of each population to MD-induced plasticity. The angle between the gradients denotes the amount of co-modulation of excitatory and PV-interneuron firing rates. **(B)** Network with feedforward depression only. Average gradient length for excitatory responses (L_E, green) and for PV responses (L_P, blue) as a function of K. The green diamond shows the value of K used in Fig. 4.3. **(C)** Same as (B) for recurrent potentiation only. **(D)** Same as (B) for interacting feedforward depression and recurrent potentiation. **(E)** Network with feedforward depression only. Angle between the gradients of excitatory and PV responses (Φ_{EP}) as a function of K. **(F)** Same as (E) for recurrent potentiation only. **(G)** Same as (E) for interacting feedforward depression and recurrent potentiation.

feedforward and recurrent plasticity, the emergence of a large area that allows for independent modulation of the firing rates of excitatory and PV-interneurons is governed by the inversion of the PV-response plane.

4.4.2 Response gradients of the MD-induced firing rate changes

I numerically extracted the average lengths of the gradients of the firing rate fold changes for excitatory and PV neurons and the average angle between those gradients using the same procedures as described in Sec. 3.4.2. The lengths of the gradient of excitatory neurons for all three types of plasticity (feedforward, recurrent and interacting feedforward and recurrent) varied systematically as a function of SST-feedback K, however, in neither of the cases did it change as drastically as when I varied the coupling scale J (L_E, Fig. 4.5B-D, compare to Fig. 3.6B-D). For pure feedforward plasticity, as well as for the interacting feedforward and recurrent plasticity, I found a monotonic decrease of L_E as K increases (Fig. 4.5B and D). In recurrent plasticity, this gradient length increased as a function of K (Fig. 4.5C). However, for all three types of plasticity, the excitatory responsiveness was on the same order of magnitude over the range of K.

The gradient lengths for PV-interneurons (L_P) varied a bit as a function of K but also stayed on the same order of magnitude over the whole range of SST-feedback

K for all the three types of plasticity (Fig. 4.5B-D), even though the distribution of facilitatory and suppressive parameter regions is inverted as K increases.

This inversion of the firing rate modulations of PV-interneurons in response to MD-induced plasticity is also visible as a monotonic increase of average angle between the gradients of excitatory and PV neurons (Φ_{EP}) for all the three types of MD-induced plasticity (Fig. 4.5E-G, compare to Fig. 3.6E-G). In contrast to the non-ISN, where Φ_{EP} was around 90°, for feedforward plasticity (Fig. 3.6E), in the network with two subtypes of interneurons Φ_{EP} reached 180° for strong SST-feedback (Fig. 4.5E). For purely recurrent depression Φ_{EP} only reached up to 90° in the network with two subtypes of interneurons (Fig. 4.5F). This is similar to the 90° in the network with a single subtype of interneurons (Fig. 3.6F), but it seems that the angle in the case with multiple interneuron subtypes reaches only 90° for a different reason: in the network with two subtypes of interneurons, it is due to a non-monotonicity of the gradient in the plane (see Fig. 4.3C, middle, where the facilitation of PV firing rates is not monotonically increasing in the ζ_{EP}-direction, but has a maximum within the parameter range probed), whereas the rate changes of excitation is monotonic also in this network with multple subtypes of interneurons (Fig. 4.3C, left). For combined feedforward and recurrent plasticity, interestingly, the angle reaches 180° for both cases, the network with two subtypes of interneurons where the paradoxical effect is reversed due to SST-feedback (Fig. 4.5G) or in a network with a single subtype of interneurons operating as non-ISN (Fig. 3.6F).

In summary, in an ISN with sufficiently strong SST-feedback, the firing rates of excitatory and PV neurons are independently, even oppositely, modulated in large parts of all the different parameter spaces, in stark contrast to the ISN without SST-feedback. Furthermore, also in contrast to the non-ISN without SST-feedback, this independence could be achieved without sacrificing responsiveness of the recurrent circuitry. Interestingly, the independence of the firing rate modulations is achieved through a selective effect on the responses of PV-interneurons, while excitatory neurons appear to be only modestly affected by an increase in K. The strong effect of SST-feedback emerges even though none of the synapses that either impinge on SST or project from them into the circuit were directly affected.

4.5 Alternative synaptic pathways of deprivation-induced plasticity

Remark: Some of the results of this section (Figs. 4.6 and 4.7) have been published before with identical content in [1].

I investigated two other pathways of plasticity in the network with multiple interneuron subtypes. First, given the strong effect that feedback from SST-interneurons has on the network response to MD-induced plasticity, one question is how plasticity in the synapses providing the SST-feedback affects firing rates in the

circuit. Specifically, since we know from experiments about the changes of recurrent synapses between excitatory and PV neurons, we may wonder whether the firing rate modulations induced by such plasticity of SST-synapses would counteract or fortify the firing rate modulations induced by the experimentally observed plasticity that I have investigated in the previous sections. I found that plasticity of SST-synapses affects firing rates similar to the recurrent plasticity investigated previously, suggesting that these changes would rather fortify the rate modulations induced by plasticity in recurrent synapses between excitatory and PV neurons (Sec. 4.5.1).

Second, experimental investigations show that whisker deprivation (WD) also induces strong plasticity in both feedforward and recurrent pathways, however, through different mechanisms than MD. WD is an experimental paradigm analogous to MD that investigates the developmental plasticity in the primary somatosensory cortex (S1). The sensory deprivation is induced by cutting or plucking a subset of whiskers that provide the direct sensory input to S1. I investigated the firing rate modulations induced by these different mechanisms by WD, which primarily affect the feedforward and recurrent E/I-ratios. The effects of this WD-induced plasticity are similar to the firing rate modulations that emerged from MD-induced plasticity, specifically the interacting feedforward and recurrent plasticity (Sec. 4.5.2).

4.5.1 Potential plasticity in SST-outputs

Little is known about plasticity in synapses from SST-interneurons happening during brief MD during the critical period. However, a decrease of the recurrent E/I-ratio has been measured directly [3]. This decrease of the recurrent E/I-ratio was driven by an increase of inhibition in the recurrent circuit. The experimentally established plasticity of synapses between excitatory and PV neurons is a suitable substrate for this decrease [82]. However, it is possible that synapses from SST cell also contribute. Specifically, a potentiation of synapses from SST onto excitatory neurons would directly contribute by increasing the inhibition impinging onto excitatory neurons. I studied this in the spiking network model by introducing the parameter $\zeta_{ES} > 1$ (Fig. 4.6A). Alternatively, a depression of synapses from SST- onto PV-interneurons could also lead to increased inhibition impinging onto excitatory neurons. Effectively, synapses from SST to PV could be described as disinhibitory. Depression of these can increase the inhibition arriving at excitatory neurons through an increase of the inhibition arriving from the less inhibited PV-interneurons. I studied this in the spiking network model by introducing the parameter $\zeta_{PS} < 1$ (Fig. 4.6A).

Analogous to the investigation of the known synaptic changes induced by MD, I simulated the network on a grid in the two-dimensional (ζ_{PS}, ζ_{ES})-plane (Fig. 4.6B). In all of the parameter space, the firing rates of excitatory neurons are suppressed, confirming the intuition that both of the synaptic changes decrease the recurrent E/I-ratio by an increase of the effective inhibition of excitatory neurons (Fig. 4.6B, left). Note, however, that the suppression of excitatory firing rates is rather modest compared to the strong suppression through plasticity in the reciprocal synapses between excitatory and PV neurons (Fig. 4.3C, left).

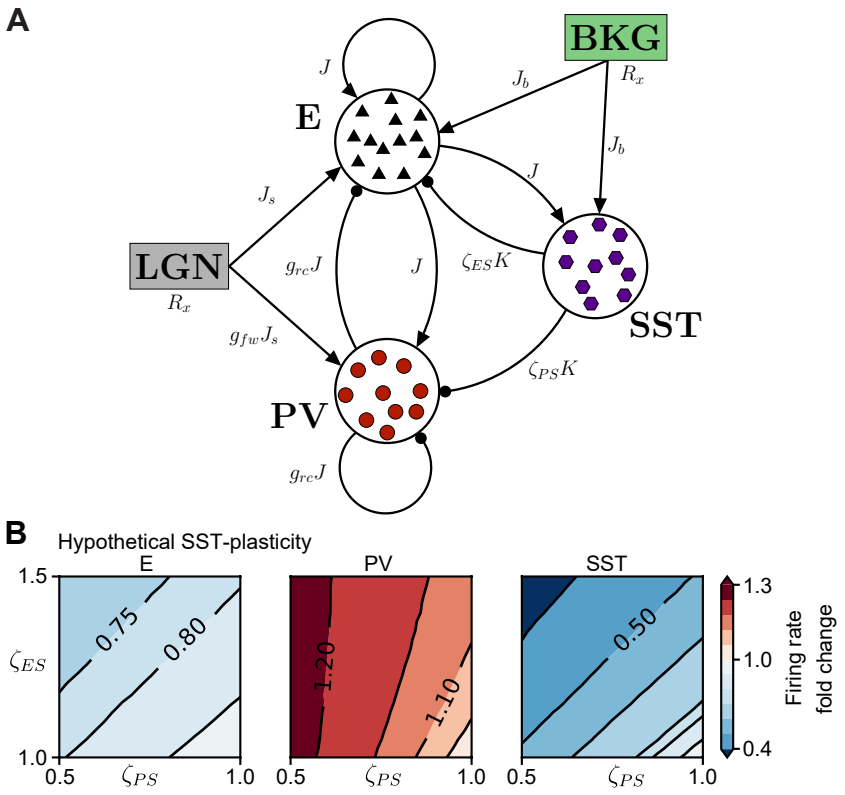

Figure 4.6: Plasticity of outgoing synapses from SST-interneurons affects circuit dynamics similarly to potentiation of recurrent synapses between excitatory and PV-interneurons. (A) Network schematic showing couplings among cells and parameters to model hypothetical MD-induced synaptic changes in synapses from SST-interneurons that could increase recurrent inhibition received by excitatory neurons [3]: depression of synapses from SST- onto PV-interneurons ($\zeta_{PS} < 1$, increasing inhibition through disinhibition of PV-interneurons) and potentiation of synapses from SST-interneurons onto excitatory neurons ($\zeta_{ES} > 1$, directly increasing inhibition of excitatory neurons by SST-interneurons). **(B)** Network firing rate in the (ζ_{PS},ζ_{ES}) plane as fold-change of baseline firing rate (bottom right corner where $\zeta_{PS} = \zeta_{ES} = 1$). The qualitative effect of this hypothetical plasticity closely resembles that of recurrent plasticity (Fig. 4.3C). Modified from [1].

The firing rates of the PV-interneurons are facilitated throughout the parameter space (Fig. 4.6B, middle), reminiscent of the facilitation of PV firing rates through the potentiation of the reciprocal synapses between excitatory and PV neurons (Fig. 4.3C, middle). The firing rates of SST-interneurons are suppressed throughout the space, again tightly following the firing rate modulations of excitatory neurons (Fig. 4.6B, right).

In summary, I found that the qualitative effects emerging from these synaptic changes of synapses outgoing from SST-cells are very similar to the effects of the recurrent plasticity between excitatory and PV neurons. Instead of counteracting any of the effects of the experimentally established recurrent plasticity between excitatory and PV neurons, the potential changes of SST-synapses could work as a synaptic pathway that reinforces them. In the animal, these two forms of recurrent plasticity could act in parallel. Alternatively, the potential plasticity I investigated here could be a redundant synaptic pathway that could be invoked when the plasticity in synapses between excitatory and PV neurons is perturbed. In such a case, plasticity in SST-synapses could be an alternative pathway that governs a similar regulation of activity in the circuit following a perturbation of the sensory experience or take over the same role in the unperturbed circuit.

4.5.2 Whisker-deprivation induced plasticity in the somatosensory cortex

A single day of WD reduces the feedforward responsiveness of PV neurons and excitatory neurons in layer 2/3 of S1 [97]. This reduction of feedforward responsiveness appears very similar to the one seen in layer 4 after brief MD [3]. However, it turns out that the mechanism invoked by brief WD is very different from the one in MD. Gainey and colleagues have shown that the suppression of feedforward responsiveness of PV interneurons stems from a decrease of their intrinsic excitability, rather than from depression of the feedforward synapses onto PV [97]. The decrease of intrinsic excitability is biophysically implemented through an increase of the firing threshold. I implemented this in the model by introducing the variable $\xi_\theta > 0$ (in units mV), which I added to the firing threshold of the PV-interneurons in the spiking network (Fig. 4.7A). In a single leaky integrate-and-fire neuron, such an increase of the firing threshold shifts the neuron's input-output function to the right, meaning that after the increase of the firing threshold, stronger inputs are necessary to drive the neuron above the firing threshold (Fig. 4.7A, inset).

Experiments suggest that this mechanism leads to an increase of the feedforward E/I-ratio in S1 after brief WD, similar to the increase of the feedforward E/I-ratio after brief MD in V1m [95, 96]. Furthermore, the recurrent E/I-ratio seems to be reduced, also in a similar fashion in S1 after brief WD as it is in V1 after brief MD. I implemented this reduction through potentiation of the synapses from PV to excitatory neurons ($\zeta_{EP} > 1$). Experiments only found a trend for potentiation of these synapses that did not reach statistical significance after just one day of WD [97]. However, prolonged WD (6-12 days) induced robust potentiation of these

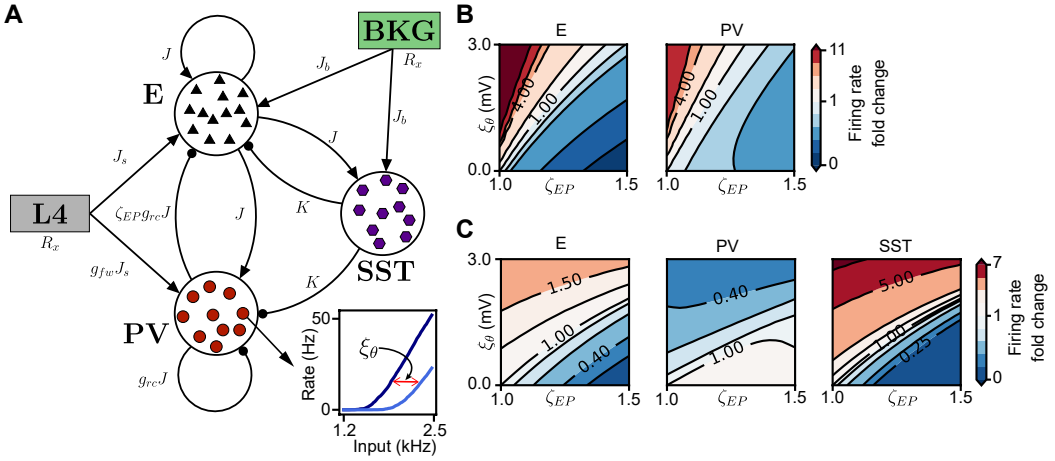

Figure 4.7: Response to whisker-deprivation induced plasticity in the somatosensory cortex (S1) in the ISN with a single and with multiple subtypes of interneurons. (A) Schematic for interacting recurrent and feedforward plasticity in S1: shift in recurrent E/I-ratio through ζ_{EP}, shift in feedforward E/I-ratio through increasing firing threshold of PV-interneurons (ξ_θ). Inset: The f-I curve of a single leaky integrate-and-fire neuron for baseline firing threshold (dark blue) and increased firing threshold (light blue). The excitability of PV-interneurons is decreased via the parameter ξ_θ. **(B)** Network firing rate as fold-change of baseline firing rate in (ζ_{EP}, ξ_θ) plane in the ISN with one type of interneurons (PV, compare to Fig. 3.3D) **(C)** Network firing rate as fold-change of baseline firing rate in (ζ_{EP}, ξ_θ) plane in the ISN with two subtypes of interneurons (PV and SST), with strong SST feedback $K = 1.6$ nS (compare to Fig. 4.3D). Modified from [1].

synapses [99]. I, therefore, combined these two circuit changes and simulated the network on a grid spanning the (ζ_{EP}, ξ_θ)-plane (Fig. 4.7B and C).

Since this is a new type of circuit modification, I first simulated the ISN with only a single subtype of interneurons (PV). The effect of the interacting feedforward and recurrent WD-induced plasticity is qualitatively similar to the effect of interacting feedforward and recurrent MD-induced plasticity (Fig. 4.7B, compare to Fig. 3.3D). The increase of the feedforward E/I-ratio through ξ_θ induces strong facilitation of both excitatory and PV firing rates, while the decrease of the recurrent E/I-ratio through ζ_{EP} suppresses the firing rates. The tightly aligned co-modulation of excitatory and PV firing rates is also preserved in this type of plasticity in an ISN. In an ISN with strong SST-feedback K, the overall structure of the excitatory firing rate modulations remains similar to the ISN without SST-feedback, whereas firing rate modulations of PV-interneurons are inverted both relative to their response in the ISN without SST-feedback and relative to the excitatory response (Fig. 4.7C, compare to Fig. 4.3D).

Even though the increase of the threshold potential of the PV-interneurons is biophysically very different from the depression of feedforward inputs, the effects on the circuit activity are the same. Thus, my modeling predicts that the regulation of activity during early WD, e.g., the first one to two days, could be very similar

to the regulation of activity during brief MD. The dynamical change of firing rates during WD has not been measured *in vivo*, making this a novel prediction from the modeling. The prediction is, however, dependent on the local circuitry. It is also possible that in S1, the firing rates of excitatory and PV neurons are regulated in parallel rather than in a sequence as in V1. Depending on the structure of the firing rate modulations, one could draw conclusions about the role of SST-feedback in the developing S1 and how it compares to V1. Generally, the similarity between the firing rate modulations through the different types of plasticity, specifically the decrease of feedforward input to PV in V1 and the decrease of the intrinsic excitability of PV in S1, could point to a shared principle of firing rate regulation that may be implemented by different means in multiple cortical circuits.

4.6 Mathematical analysis of the network with multiple interneuron subtypes

Remark: Some of the results in this Section (Figs. 4.9, 4.10 and 4.11A-C) and analysis (Eqs. 4.7, 4.8 and 4.25–4.30) have been published before in [1]. Notation has been adapted for consistency and the analysis has been extended.

With multiple subtypes of interneurons in the spiking network, I found that for sufficiently strong SST-feedback, the firing rates of excitatory and PV neurons in the spiking ISN can be modulated independently from each other in response to MD-induced plasticity (Fig. 4.3). This independence is in contrast to the responses in the ISN with a single subtype of interneurons (PV), where the firing rate changes of excitatory and PV neurons closely follow each other (Fig. 3.3). The independence of excitatory and PV firing rates emerges in the ISN through an inversion of the responses of PV-interneurons. From numerical observation, this inversion of PV firing rate changes is related to a reversal of the paradoxical effect, such that, even in the ISN, PV-interneurons respond non-paradoxically upon additional stimulation (Fig 4.1D). In this section, I analyzed the responses of the different populations in a linear rate model that includes PV and SST interneurons. I was able to prove that, indeed, the reversal of the paradoxical effect is the major driving mechanism that underlies the inversion of PV firing rate changes.

I extended the population rate model to include PV- and SST-interneurons and analyzed its stability and steady-state firing rates (Sec. 4.6.1). Specifically, I was able to calculate the exact conditions on SST feedback that lead to a reversal of the paradoxical effect in PV (Sec. 4.6.2). Based on this, I analyzed the linear rate model with MD-induced synaptic weight changes (Sec. 4.6.3). This serves as the basis to analytically explain how the inversion of PV-responses to MD-induced plasticity depends on the strength of SST-feedback and the reversal of the paradoxical effect in Sec. 4.7.

4.6.1 Network structure and stability

The dynamical equation of the linear population rate model with two interneuron subtypes is th same as in the network with a single interneuron subtype (Eqs. 3.16-3.19) with the linear dynamical equation

$$\tau \dot{\boldsymbol{r}} = -\boldsymbol{r} + \boldsymbol{W}\boldsymbol{r} + \boldsymbol{s}. \tag{4.6}$$

Recurrent connections between excitatory and PV neurons are not changed when I introduce SST-interneurons, and are parametrized with the coupling scale w and relative scale for PV-inhibition γ. SST-interneurons receive recurrent excitatory input with weight w and project to the excitatory and PV populations with weight κ. The recurrent connectivity matrix of the population rate model with two subtypes of interneurons in the baseline state, without MD-induced synaptic changes, is analogous to the connectivity of the spiking network (compare to Eq. 4.1):

$$\mathbf{W} = \begin{pmatrix} W_{EE} & W_{EP} & W_{ES} \\ W_{PE} & W_{PP} & W_{PS} \\ W_{SE} & W_{SP} & W_{SS} \end{pmatrix} = \begin{pmatrix} w & -\gamma w & -\kappa \\ w & -\gamma w & -\kappa \\ w & 0 & 0 \end{pmatrix}. \tag{4.7}$$

As in the spiking model, I assumed the same rate r_x for feedforward, thalamic (LGN) and background (BKG) inputs (Fig. 4.8A). The excitatory population receives input from both sources, PV-interneurons receive input only from LGN with relative weight factor g_{fw}, same as in the network with a single subtype of interneurons. SST-interneurons receive input only from BKG:

$$\mathbf{s} = \begin{pmatrix} 2 \cdot r_x \\ g_{fw} r_x \\ r_x \end{pmatrix}. \tag{4.8}$$

I analyzed the stability conditions and the steady-state in this baseline setup, specifically with respect to γ and κ. Then, I also derived the expression for the response of all populations in the model to additional stimulation of PV-interneurons. I expressed the condition for the presence or reversal of the paradoxical effect as a function of the SST-feedback strength κ.

Stability conditions

Stability of the network requires that the real parts of all the eigenvalues of the connectivity matrix in Eq. 4.7 are smaller than one. The eigenvalues of a three-dimensional system are slightly more complicated to derive, however, the underlying symmetry of the network setup aids the derivation. The characteristic polynomial for the matrix $\boldsymbol{W}$ is

$$\lambda^3 + \lambda^2 w(\gamma - 1) + \lambda\kappa = 0 \tag{4.9}$$

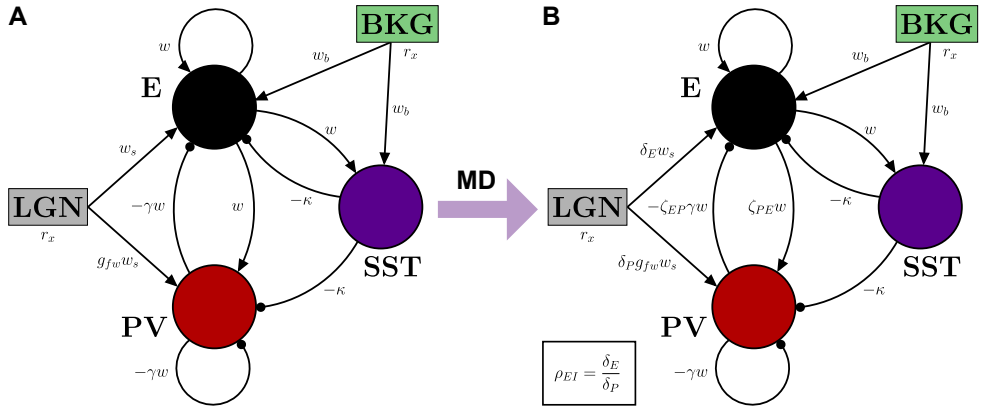

Figure 4.8: Connectivity in the population rate model with two subtypes of interneurons before and after MD. (A) Connectivity in the baseline network before application of MD. **(B)** Connectivity in the network after application of MD.

from which one can factor out one λ (therefore, $\lambda_1 = 0$) and solve for the the roots of the remaining quadratic equation to find the eigenvalues

$$\begin{aligned} \lambda_1 &= 0 \\ \lambda_{2/3} &= -\frac{w(\gamma-1)}{2} \pm \sqrt{\frac{\big(w(\gamma-1)\big)^2}{4} - \kappa w}, \end{aligned} \tag{4.10}$$

where $\lambda_{2/3}$ converge to the eigenvalues of the network without SST-feedback as $\kappa \to 0$. The first term in the square-root in Eq. 4.11 is always positive, with a positive term subtracted that is dependent on the strength of SST-feedback κ. To study this in general for all possible (positive) values of κ, one has to distinguish two cases with either complex conjugate eigenvalues that share the same real part, or with two real eigenvalues.

The eigenvalues $\lambda_{2/3}$ are two distinct real eigenvalues if

$$\kappa w < \frac{\big(w(\gamma-1)\big)^2}{4}. \tag{4.11}$$

In this case, the largest real eigenvalue of $\lambda_{2/3}$ is the one of corresponding to the sum rather than the difference:

$$\lambda_2 = -\frac{w(\gamma-1)}{2} + \sqrt{\frac{\big(w(\gamma-1)\big)^2}{4} - \kappa w}. \tag{4.12}$$

Considering that both $\kappa > 0$ and $w > 0$, and that $w\kappa$ is subtracted in the root, one sees that the real part as a function of κ has an upper bound given by $-w(\gamma-1)$ (when $\kappa = 0$), which corresponds to the largest real eigenvalue of the network with a single subtype of interneurons (Eq. 3.23). Thus, in general, one can already see that κ can only reduce the largest real part. Thus, SST-feedback stabilizes the circuit. If

κ is so big that the eigenvalue becomes a complex number ($\kappa > w(\gamma-1)^2/4$), the real part is $-w(\gamma-1)/2$ (one half of the largest real part in the network with only PV), which is the lower bound of the largest real part for arbitrarily large κ.

To derive the exact stability condition dependent on w, γ and κ, I resorted Eq. 4.12. First, assuming that the eigenvalue λ_2 is real, one finds

$$\sqrt{\frac{\big(w(\gamma-1)\big)^2}{4} - \kappa w} < 1 + \frac{w(\gamma-1)}{2}. \tag{4.13}$$

To derive the general stability condition with arbitrary κ (withing the range of κ for real eigenvalues) is most easily done by squaring both sides. However, one must distinguish two cases, since it is not a priori clear if the right hand side of Eq. 4.13 is positive or negative. However, the case that the left hand side of Eq. 4.13 is negative:

$$1 + \frac{w(\gamma-1)}{2} < 0 \tag{4.14}$$

implies

$$1 < -\frac{w(\gamma-1)}{2} \tag{4.15}$$

in which case the network can not be stable for any value of κ, because then the lower bound of the real parts for arbitrary κ is larger than stability permits. With both sides in Eq. 4.13 positive, dynamical stability implies as relation between w, γ and κ:

$$\kappa > 1 - \gamma - \frac{1}{w} \tag{4.16}$$

or, expressed as condition that γ and κ have to fulfill together:

$$\kappa + \gamma > 1 - \frac{1}{w}, \tag{4.17}$$

where $\kappa + \gamma$ describes the overall, summed level of inhibition in the circuit. Interestingly, the condition that the sum $\kappa + \gamma$ has to fulfill for stability is the same as the condition on γ in the network with a single subtype of interneurons (Eq. 3.24).

We therefore find that any $\kappa > 0$ does not destabilize the circuit, but further stabilizes it. One can formalize this also for the different operating regimes. For the non-ISN ($w < 1$) the condition Eq. 4.17 is always fulfilled for any $\kappa \geq 0$ (and any $\gamma \geq 0$), because the right hand side in Eq. 4.16 is negative. The network is stable without any inhibitory feedback. In the ISN, there are two limiting cases: At the transition to ISN ($w = 1$), stability requires $\kappa > -\gamma$, which is also fulfilled for arbitrary inhibitory feedback from the SST population ($\kappa \geq 0$) and for non-zero inhibitory feedback from the PV population ($\gamma > 0$, though this can be arbitrary small). Second, for $w \to \infty$, the specific condition on κ becomes

$$\kappa > 1 - \gamma. \tag{4.18}$$

Specifically, if PV-inhibition in the circuit is strong enough to ensure stability ($\gamma > 1$ for $w \to \infty$), no SST-feedback is required for stability also in this case. If the

PV-inhibition is not sufficient to ensure stability, a certain level of SST-feedback is necessary for stabilization.

The eigenvalues Eq. 4.11 are complex conjugate if:

$$\kappa w > \frac{\big(w(\gamma-1)\big)^2}{4}. \tag{4.19}$$

Since this defines a lower bound on SST-feedback, this is generally the relevant case for stronger SST-feedback. The real part of both eigenvalues is:

$$\Re\big(\lambda_{2/3}\big) = -\frac{w(\gamma-1)}{2}. \tag{4.20}$$

Interestingly, stability in this case of large κ is independent of κ. However, as in the case with a single subtype of interneurons, there is a specific condition on γ given a certain size of the overall coupling scale w:

$$\begin{aligned} 1 &> -\frac{w(\gamma-1)}{2} \\ \Rightarrow \gamma &> 1-\frac{2}{w} \end{aligned} \tag{4.21}$$

which is very similar to the stability condition in the network with a single subtype of interneurons ($\gamma > 1 - 1/w$, Eq. 3.24). However, the condition here in Eq. 4.22 is weaker, requiring less strong PV-inhibition in the circuit for dynamical stability. Thus, even though the stability condition itself is independent of κ in the case of complex conjugate eigenvalues, stability requires less PV-inhibition thanks to sufficient inhibition from SST. However, interestingly this also implies that it is still necessary to have a certain level of PV-inhibition for dynamical stability even with very strong inhibition from SST. Even though SST-feedback overall has a stabilizing effect, the way the SST-interneurons are coupled in the circuit makes them insufficient to stabilize the dynamics on their own, without any contribution from PV-interneurons, for arbitrary w.

Steady state firing rate

The steady-state firing rate $\mathbf{r}_{ss} = (\mathbb{I} - \mathbf{W})^{-1} \cdot \mathbf{s}$ with the recurrent connectivity in Eq. 4.7 is:

$$\begin{pmatrix} r^E_{ss} \\ r^P_{ss} \\ r^S_{ss} \end{pmatrix} = \frac{1}{\eta} \begin{pmatrix} 1+\gamma w & -\gamma w & -\kappa \\ w(1-\kappa) & 1-w+w\kappa & -\kappa \\ w(1+\gamma w) & -\gamma w^2 & 1-w+\gamma w \end{pmatrix} \begin{pmatrix} 2\cdot r_x \\ g_{fw} r_x \\ r_x \end{pmatrix}, \tag{4.22}$$

where

$$\eta = \det(\mathbb{I} - \mathbf{W}) = 1 - w + \gamma w + \kappa w = 1 - w\big(1-(\gamma+\kappa)\big). \tag{4.23}$$

From the stability condition on the total inhibition $\gamma + \kappa$ in the circuit (Eq. 4.17) it follows that this determinant η is positive for any stable network. To calculate this,

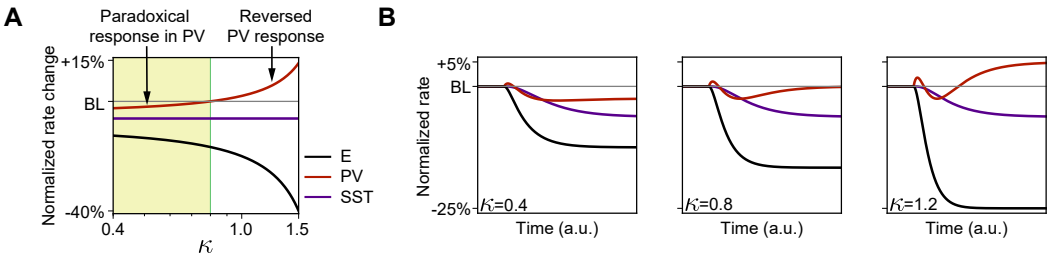

Figure 4.9: Reversal of the paradoxical effect in PV-interneurons through SST-feedback (compare to Fig. 3.9). **(A)** Relative change of steady-state rate of PV-interneurons induced by current injection to PV-interneurons as a function of SST-feedback κ. The rate of each population is normalized to its baseline firing rate before additional current injection (BL). **(B)** Dynamics of the linear population rate model following onset of step current to PV-interneurons for $\kappa = 0.4$ (left), $\kappa = 0.8$ (middle) and $\kappa = 1.2$ (right). Colors are the same as in (A). Modified from [1].

one can simply substitute the expression $\kappa + \gamma$ by its lower bound $1 - 1/w$ and find as lower bound for η:

$$\eta > 1 - w\left(1 - \left(1 - \frac{1}{w}\right)\right) = 0. \tag{4.24}$$

4.6.2 Reversal of the paradoxical effect

From the general steady state solution (Eq. 4.22) I calculated the response to an additional drive to the PV-population for all three populations in analogy to the network with a single subtype of interneurons (Eqs. 3.30 and 3.31). The rate changes of the populations in response to a perturbation ξ that is injected to PV are:

$$\begin{aligned}\begin{pmatrix}\Delta r^E_{ss}\\ \Delta r^P_{ss}\\ \Delta r^S_{ss}\end{pmatrix} &= \frac{1}{\eta}\begin{pmatrix}1+\gamma w & -\gamma w & -\kappa\\ w(1-\kappa) & 1-w+w\kappa & -\kappa\\ w(1+\gamma w) & -\gamma w^2 & 1-w+\gamma w\end{pmatrix}\begin{pmatrix}0\\ \xi\\ 0\end{pmatrix}\\ &= \frac{1}{\eta}\begin{pmatrix}-\gamma w\xi\\ (1-w+w\kappa)\xi\\ -\gamma w^2\xi\end{pmatrix}.\end{aligned} \tag{4.25}$$

Since $\eta = \det(\mathbb{I} - \mathbf{W}) > 0$ when the network is stable (Eq. 4.24), η can be left aside for the analysis of the directions of the rate changes following PV-perturbation:

$$\begin{aligned}\Delta r^E_{ss} &\propto -\gamma w\xi\\ \Delta r^P_{ss} &\propto (1-w+w\kappa)\xi\\ \Delta r^S_{ss} &\propto -\gamma w^2\xi.\end{aligned} \tag{4.26}$$

As in the network with a single interneuron subtype, the excitatory neurons' response is always in the opposite direction to ξ. The firing rate of the excitatory population

is suppressed when an additional current is injected to the PV population and facilitated when the external current to the PV population is reduced (or if a negative current is injected). This directionality is independent of w, γ and, crucially, independent of κ. Note, however, that the amount of rate changes depends on all of these parameters, also on κ that influences η (Eq. 4.24). The activity change of the SST population is proportional to the rate change of the excitatory population, with a proportionality factor w. Thus, the SST population always expresses a form of the paradoxical effect, as its rate changes opposite to the current injected to PV. Interestingly, this happens in the absence of direct inhibition of the SST population by the PV population and is driven by a withdrawal of recurrent excitation, as is the paradoxical effect in the ISN with a single subtype of interneurons.

The response of the PV population is proportional to $(1 - w + w\kappa)\xi$. Therefore, the direction of rate changes in the PV depends not only on w, as it did without SST-feedback (Eq. 3.31), but also on κ. Assuming the network is a non-ISN, the PV population always responds with a change of activity in the same direction as ξ, since $w < 1$, for general $\kappa > 0$. However, in the ISN, where $w > 1$, the direction of the response of PV depends on the specific relation between w and κ. It responds with a change of activity in the opposite direction to ξ, that is, shows the paradoxical effect, when the condition is satisfied:

$$1 - w + w\kappa < 0 \quad \Longleftrightarrow \quad \kappa < \frac{w-1}{w}. \tag{4.27}$$

Thus the paradoxical effect in the PV population can be reversed with sufficiently strong feedback from SST-cells. More generally: for any $w > 1$, which induces a paradoxical effect in PV, there exists a sufficiently large κ that leads to a reversal of the paradoxical effect and thus a non-paradoxical response of the PV population. This reversal of the paradoxical effect is shown in Fig. 4.9A (red) for a specific example ($w = 5$). The steady-state rate after injection of a positive current into PV is suppressed for small κ but facilitated for large κ.

Interestingly, the transient response of the PV-population always shows a dynamical signature of the paradoxical effect (Fig 4.9B, red). In the network with small $\kappa = 0.4$, the PV firing rate increases directly after the additional stimulus is delivered and then decreases through the withdrawal of recurrent excitation (black curve). The PV firing rate ultimately settles in a decreased firing rate, which is the paradoxical effect (Fig 4.9B, left). This dynamics is very similar in the ISN with a single subtype of interneurons leading to the paradoxical effect (Fig 3.9B, right). For large $\kappa = 1.2$, the activity first transiently increases, driven by the external stimulus, then decreases, because of the withdrawal of recurrent excitation, but then increases again and ultimately settles in a steady-state activity higher than the steady-state before injection of the external stimulus (Fig 4.9B, right). Thus, the paradoxical effect reverses. The increase of the PV firing rate that leads to the reversal state after the transient decrease is caused by a withdrawal of recurrent inhibition from the SST population, whose firing rate decreases throughout the range of κ (purple). Through this, the network overall (in terms of the summed

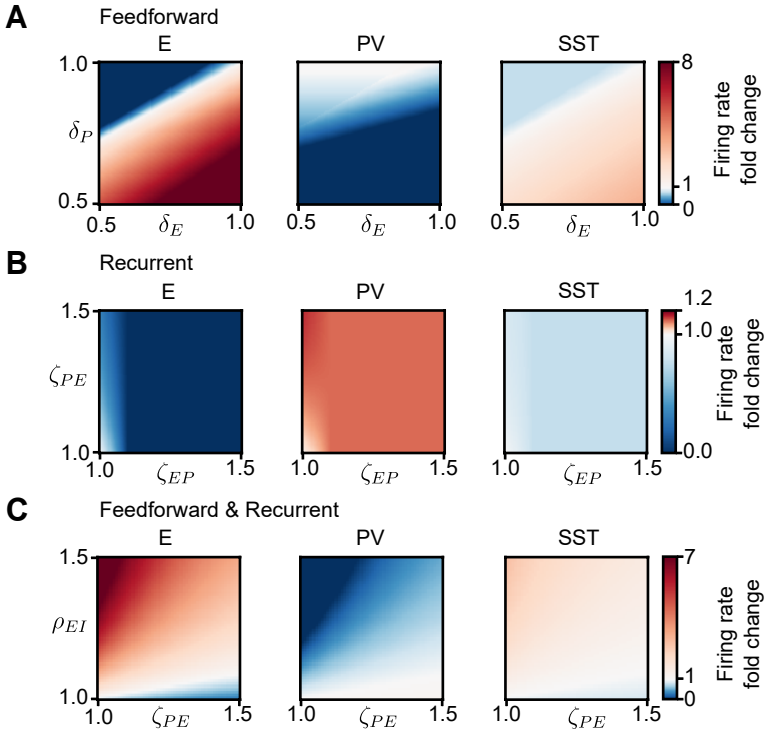

Figure 4.10: Response to synaptic changes induced by brief MD in a linear model of the network with two subtypes of interneurons (PV and SST) 1: response planes (compare to Fig. 4.3B-D for correspondence with the spiking case). **(A)** Network firing rate for an ISN ($w = 5$) in the (δ_E, δ_P) plane (feedforward depression) with strong SST-feedback ($\kappa = 1.2$) as fold-change of baseline firing rate for excitatory (left), PV (middle) and SST populations (right). **(B)** Same as (A) for recurrent potentiation ((ζ_{EP}, ζ_{PE}) plane). **(C)** Same as (A) for interacting feedforward and recurrent plasticity: decrease of the recurrent E/I-ratio through ζ_{PE}, increase of the feedforward E/I-ratio as $\rho_{EI} = \delta_E/\delta_P$. Modified from [1].

response of all interneuron populations) still shows the paradoxical effect, even when the PV population responds non-paradoxically.

4.6.3 Analysis of MD-induced steady-state activity changes

Including the MD-induced synaptic changes in the recurrent connectivity matrix of the population rate model with two subtypes of interneurons, it becomes:

$$\mathbf{W}^{md} = \mathbf{W}(\zeta_{EP}, \zeta_{PE}) = \begin{pmatrix} w & -\zeta_{EP}\gamma w & -\kappa \\ \zeta_{PE} w & -\gamma w & -\kappa \\ w & 0 & 0 \end{pmatrix}. \tag{4.28}$$

With MD affecting only LGN-inputs, while background input is unchanged, the feedforward connectivity after applying MD reads:

$$\mathbf{s}^{md} = \mathbf{s}(\delta_E, \delta_P) = \begin{pmatrix} (\delta_E + 1) r_x \\ \delta_P g_{fw} r_x \\ r_x \end{pmatrix}. \tag{4.29}$$

I have proven that the additional SST-feedback has a stabilizing effect on the dynamics of the network compared to the network with only a single subtype of interneurons, expressed as weaker conditions on γ that need to be satisfied as $\kappa > 0$ (Eqs. 4.16 and 4.22). Furthermore, in the network with only a single subtype of interneurons, I have also proven that the network after MD-induced plasticity in the recurrent synapses remains stable for arbitrary amounts of recurrent potentiation (Eqs. 3.37 and 3.38). Therefore, I assumed stability in the network with multiple interneuron subtypes after induction of recurrent potentiation.

General steady state firing rate fold changes following from MD-induced synaptic changes

The general solution of the steady state firing rates including MD-changes in both feedforward and recurrent synapses is

$$\begin{aligned} \begin{pmatrix} r_{ss}^E \\ r_{ss}^P \\ r_{ss}^S \end{pmatrix} &= \frac{1}{\eta^{md}} \begin{pmatrix} 1 + \gamma w & -\zeta_{EP}\gamma w & -\kappa + \gamma w \kappa(\zeta_{EP} - 1) \\ w(\zeta_{PE} - \kappa) & 1 - w + w\kappa & w\kappa(1 - \zeta_{PE}) - \kappa \\ w(1 + \gamma w) & -\zeta_{EP}\gamma w^2 & 1 - w + \gamma w \ + \gamma w^2(\zeta_{PE}\zeta_{EP} - 1) \end{pmatrix} \\ &\quad \cdot \begin{pmatrix} (\delta_E + 1) r_x \\ \delta_P g_{fw} r_x \\ r_x \end{pmatrix}, \end{aligned} \tag{4.30}$$

where

$$\begin{aligned} \eta^{md} &= \det\left(\mathbb{I} - \mathbf{W}^{md}\right) \\ &= 1 - w + \gamma w + \kappa w - \big(\zeta_{EP} - 1\big)\kappa\gamma w^2 + \big(\zeta_{EP}\zeta_{PE} - 1\big)\gamma w^2. \end{aligned} \tag{4.31}$$

This determinant is a combination of the determinant without MD-induced changes (η in Eq. 4.23) and two terms that are added that depend on MD-induced plasticity. To better distinguish, I used index *bl* for the determinant of the recurrent connectivity without MD-induced plasticity:

$$\begin{aligned} \eta^{bl} &= 1 - w + \gamma w + \kappa w \\ \eta^{md} &= \eta^{bl} - \big(\zeta_{EP} - 1\big)\kappa\gamma w^2 + \big(\zeta_{EP}\zeta_{PE} - 1\big)\gamma w^2. \end{aligned} \tag{4.32}$$

From Eq. 4.30, the excitatory, PV and SST firing rates as a function of all four parameters for MD-induced plasticity follow

$$\begin{aligned}
r_{ss}^{E,md} &= \frac{r_x}{\eta^{md}} \cdot \Big\{ \big(1+\gamma w\big)\big(\delta_E+1\big) - \zeta_{EP}\gamma w \delta_P g_{fw} - \kappa + \gamma w \kappa \big(\zeta_{EP}-1\big) \Big\} \\
r_{ss}^{P,md} &= \frac{r_x}{\eta^{md}} \cdot \Big\{ w\big(\zeta_{PE}-\kappa\big)\big(\delta_E+1\big) + \big(1-w+w\kappa\big)\delta_P g_{fw} - \kappa + w\kappa\big(1-\zeta_{PE}\big) \Big\} \\
r_{ss}^{S,md} &= \frac{r_x}{\eta^{md}} \cdot \Big\{ w\big(1+\gamma w\big)\big(\delta_E+1\big) - \zeta_{EP}\gamma w^2 \delta_P g_{fw} \\
&\qquad + \Big(1-w+\gamma w+\gamma w^2(\zeta_{PE}\zeta_{EP}-1)\Big) \Big\}. \qquad (4.33)
\end{aligned}$$

I studied the changes of activity after MD-induction relative to the baseline state before MD-induction. Numerically solving the firing rate fold changes shows very good agreement with the spiking network, including the inversion of PV firing rate fold changes (Fig. 4.10, compare to Fig. 4.3B-D for correspondence with the spiking case). As in the case with a single subtype of interneurons in the previous chapter, I analytically studied the difference between baseline rates and rates after MD, instead of the ratio that is used for the firing rate fold changes. Using the difference, the condition for facilitation of firing rates is:

$$\Xi_{ij}^{A} = R_{md}^{A}(x,y) - R_{bl}^{A} > 0, \qquad (4.34)$$

with the baseline rate before MD-induced plasticity:

$$\begin{aligned}
r_{ss}^{E,bl} &= \frac{r_x}{\eta^{bl}} \cdot \Big\{ 2 \cdot \big(1+\gamma w\big) - \gamma w g_{fw} - \kappa \Big\} \\
r_{ss}^{P,bl} &= \frac{r_x}{\eta^{bl}} \cdot \Big\{ 2w\big(1-\kappa\big) + \big(1-w+w\kappa\big) g_{fw} - \kappa \Big\} \\
r_{ss}^{S,bl} &= \frac{r_x}{\eta^{bl}} \cdot \Big\{ 2w\big(1+\gamma w\big) - \gamma w^2 g_{fw} + \big(1-w+\gamma w\big) \Big\}. \qquad (4.35)
\end{aligned}$$

I used these expressions to investigate the inversion of the firing rate fold changes in the PV-population and how this inversion relates to the reversal of the paradoxical effect, especially the condition on κ leads to this reversal (Eq. 4.27).

4.7 Analysis of the inversion of PV-responses through SST-feedback

Remark: Some of the results in this section have been published before in [1]. The notation has been adapted for consistency, and the analysis has been extended.

I studied the conditions for facilitation of the firing rates in the different populations for the two types of plasticity, indexed as ffw for feedforward depression

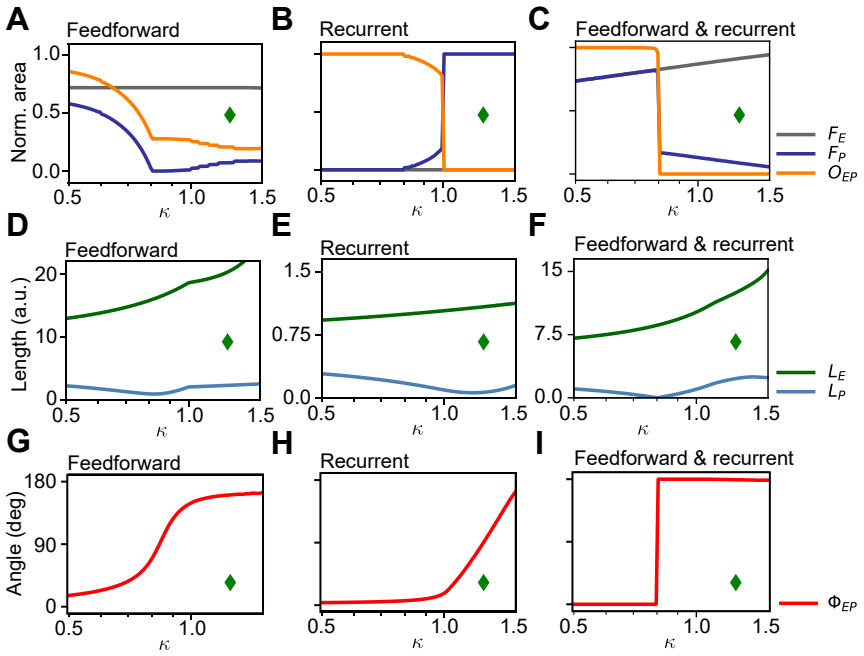

Figure 4.11: Response to synaptic changes induced by brief MD in a linear model of the network with two subtypes of interneurons (PV and SST) 2: distribution of faciliation and gradients. (A) Network with feedforward depression only. Fractional area of facilitation for excitatory neurons (F_E, gray), PV-interneurons (F_P, blue) and the overlap between excitatory and PV response areas (O_{EP}, orange) as a function of SST-feedback κ. **(B)** Same as (A) for recurrent potentiation only. **(C)** Same as (A) for interacting feedforward and recurrent plasticity. **(D)** Network with feedforward depression only. Average length of the gradient for excitatory neurons (L_E, green) and PV-interneurons (L_P, light blue) as a function of SST-feedback κ. **(E)** Same as (D) for recurrent potentiation only. **(F)** Same as (D) for interacting feedforward and recurrent plasticity. **(G)** Network with feedforward depression only. Average angle between the gradients of excitatory PV neurons (Φ_{EP}, red) as a function of SST-feedback κ. **(H)** Same as (G) for recurrent potentiation only. **(I)** Same as (G) for interacting feedforward and recurrent plasticity. The green diamond in (A-I) shows the value of κ used in Fig. 4.10. Panels (A-C) reproduced from [1]. For correspondence to the spiking case, compare (A-C) with Fig. 4.4B-D and (D-I) with Fig. 4.5B-G.

(Sec. 4.7.1) and rec for recurrent potentiation (Sec. 4.7.2). The focus of this analysis is the effect of SST-feedback in the ISN. To understand the inversion of the PV firing rate fold changes in the ISN, it is crucial to study the conditions for the emergence of a facilitatory area. However, for feedforward depression, it is not sufficient to prove if facilitation of firing rates can exist for the PV population or if it cannot exist (as it was for the distinction between the ISN and non-ISN in the previous chapter). The inversion of the PV-response through SST-feedback rather leads to the fact that facilitation exists in completely disparate regions of the parameter space before and after inversion (that is, for weak SST-feedback κ vs. strong SST-feedback). For recurrent potentiation, it is sufficient to prove that facilitation of the PV firing rates can only exist above a certain strength of SST-feedback κ in the ISN, since no facilitation can be found for this plasticity in the ISN with a single subtype of interneurons.

4.7.1 Conditions on facilitation for feedforward depression

Excitatory and SST populations

From the firing rate change due to MD-induced feedforward depression (Eq. 4.33 with $\zeta_{EP} = \zeta_{PE} = 1$), I derived the condition to find facilitation of excitatory firing rates

$$\begin{aligned} \Xi^E_{\text{ffw}} &= r^{E,\text{ffw}}_{ss} - r^{E,bl}_{ss} \\ &= \frac{r_x}{\eta^{bl}} \cdot \Big\{ \big(1+\gamma w\big)\big(\delta_E + 1\big) - \gamma w \delta_P g_{fw} - \kappa \Big\} \\ &\quad -\frac{r_x}{\eta^{bl}} \cdot \Big\{ 2 \cdot \big(1+\gamma w\big) - \gamma w g_{fw} - \kappa \Big\} > 0 \\ \Leftrightarrow \quad &\Big(\big(1+\gamma w\big)\big(\delta_E - 1\big) - \gamma w g_{fw} (\delta_P - 1) \Big) > 0. \end{aligned} \tag{4.36}$$

Interestingly, with a focus on the induction of facilitation or suppression, the terms that depend on SST-feedback κ cancel or can be removed (specifically η^{bl}). Thus, the condition for facilitation remains the same as in the network with a single subtype of interneurons (including its dependence on the coupling scale w). The area in which facilitation is induced stays constant (F_E, Fig. 4.11A, compare to Fig. 4.4B). Nonetheless, the amount of firing rate modulation depends on κ through η^{bl}, as is visible in the length of the gradients (L_E, Fig. 4.11D, compare to Fig. 4.5B).

One can derive the boundary as a condition on δ_P that is necessary to induce facilitation. From the κ-independent condition Eq. 4.36 the same boundary as in the network without SST-feedback follows (see Eq. 3.50)

$$\delta_P < 1 - (1-\delta_E)\frac{1+\gamma w}{g_{fw}\gamma w}. \tag{4.37}$$

Replacing the inequality with equality, I plotted this boundary in Fig. 4.12B (left).

For SST-interneurons I derived the condition to find facilitation:

$$\begin{aligned}\Xi^S_{\text{ffw}} &= r^{S,\text{ffw}}_{ss} - r^{S,bl}_{ss} \\ &= \frac{r_x}{\eta^{bl}} \cdot \Big(w\big(1+\gamma w\big)\big(\delta_E+1\big) - \gamma w^2 \delta_P g_{fw} + \big(1-w+\gamma w\big)\Big) \\ &\quad -\frac{r_x}{\eta^{bl}} \cdot \Big(2w\big(1+\gamma w\big) - \gamma w^2 g_{fw} + \big(1-w+\gamma w\big)\Big) > 0 \\ \Leftrightarrow \quad &\Big(w\big(1+\gamma w\big)\big(\delta_E-1\big) - \gamma w^2 g_{fw}\big(\delta_P-1\big) > 0, \end{aligned} \tag{4.38}$$

which is independent of κ, just like Ξ^E (Eq. 4.36). Moreover, one can rearrange this to find that the relative rate modulation of SST-interneurons is directly proportional to the one of excitatory neurons (Eq. 4.36):

$$\Xi^S_{\text{ffw}} = w \cdot \Xi^E_{\text{ffw}}. \tag{4.39}$$

Therefore, the condition to find either facilitation or suppression in SST-interneurons is the same as in excitatory neurons, and everywhere in the parameter space of feedforward depression, SST firing rates follow the firing rates of excitatory neurons. From the way MD affects the model circuit, a simple explanation is that none of the direct input synapses onto SST-interneurons are affected by MD. With these synaptic changes, MD can only indirectly affect SST firing rates through modulation of the firing rate of the presynaptic excitatory population.

PV population

From the firing rate change due to MD-induced feedforward depression (Eq. 4.33 with $\zeta_{EP} = \zeta_{PE} = 1$) for PV-interneurons as the condition to find facilitation follows:

$$\begin{aligned}\Xi^P_{\text{ffw}} &= r^{P,\text{ffw}}_{ss} - r^{P,bl}_{ss} \\ &= \frac{r_x}{\eta^{bl}} \cdot \Big\{w\big(1-\kappa\big)\big(\delta_E+1\big) + \big(1-w+w\kappa\big)\delta_P g_{fw} - \kappa\Big\} \\ &\quad -\frac{r_x}{\eta^{bl}} \cdot \Big\{2w\big(1-\kappa\big) + \big(1-w+w\kappa\big)g_{fw} - \kappa\Big\} \\ \Leftrightarrow \quad &\Big(w\big(1-\kappa\big)\big(\delta_E-1\big) + \big(1-w+w\kappa\big)g_{fw}(\delta_P-1)\Big) > 0. \end{aligned} \tag{4.40}$$

The activity changes relative to baseline have κ-dependent terms both modulating the response to δ_E and δ_P. To derive the δ_P necessary for facilitation for a given δ_E, I first rearranged the condition (always writing terms such that negative signs are explicit):

$$(1-w+w\kappa)\delta_P > (1-\delta_E)\frac{w(1-\kappa)}{g_{fw}} + (1-w+w\kappa). \tag{4.41}$$

To further simplify through dividing by $(1-w+w\kappa)$ we need to consider a number of cases that switch the direction of the inequality because of division with a negative

term. However, replacing the inequality by equality, where we do not need to consider these cases, the boundary between facilitation and suppression is:

$$\delta_P = (1 - \delta_E)\frac{w(1-\kappa)}{g_{fw}(1 - w + w\kappa)} + 1. \tag{4.42}$$

I plotted this boundary for multiple values of SST-feedback κ (Fig. 4.12B, right). As a function of κ, the boundary seems to be rotated in the plane. However, one needs to consider the inequality in order to know on which side of the boundary the facilitatory part of the parameter space is.

The sign of $(1 - w + w\kappa)$ in the ISN ($w > 1$) depends on the relation between w and κ. This sign needs to be considered before division in Eq. 4.41. First, with $\kappa = 0$, the term is negative because the network is ISN. The size of κ necessary to make this term positive is $\kappa = (w-1)/w$, which is exactly the value of κ that leads to the reversal of the paradoxical effect (Eq. 4.27). If $\kappa < (w-1)/w$ then $(1 - w + w\kappa) < 0$ and division leads to:

$$\delta_P < 1 - (1 - \delta_E)\frac{w(1-\kappa)}{g_{fw}(w - 1 - w\kappa)}, \tag{4.43}$$

where I have written negative signs explicitly and all individual terms are positive (specifically: $(1 - w + w\kappa) = -(w - 1 - w\kappa)$). Note that in this range of κ for sure $1 - \kappa > 0$ (taking into account that $(w-1)/w \leq 1$ is true for all $w > 0$).

The condition to find facilitation in Eq. 4.43 can be fulfilled by feedforward depression and imposes an upper bound on δ_P, similar to the ISN without SST-feedback (Eq. 3.56).

In an intermediate range of κ, two conditions on κ need to be considered: if $(w-1)/w < \kappa$ and $\kappa < 1$: then $(1 - w + w\kappa) > 0$ while also $(1 - \kappa) > 0$, thus the condition becomes

$$\delta_P > 1 + (1 - \delta_E)\frac{w(1-\kappa)}{g_{fw}(1 - w + w\kappa)} \tag{4.44}$$

which cannot be fulfilled by feedforward depression. Thus, in this range of κ in the ISN there is no facilitation of PV firing rates. This is shown in the numerical solution: it is constantly zero for intermediate κ in this range (F_P, Fig. 4.11A). In the example shown, I used $w = 5$, thus the range of κ goes from 0.8 to 1. Note that for very strong coupling in the network, as $w \to \infty$, this range shrinks to zero as $(w-1)/w \to 1$.

Finally, for $\kappa > 1$ then $(1 - w + w\kappa) > 0$ and $(1 - \kappa) < 0$, thus we divide by $(1 - w + w\kappa)$ without resorting it and write negative signs explicit with $(\kappa - 1) > 0$:

$$\delta_P > 1 - (1 - \delta_E)\frac{w(\kappa-1)}{g_{fw}(1 - w + w\kappa)}. \tag{4.45}$$

This is more similar to the condition in Eq. 4.43, but with flipped direction. Thus, one finds a parameter range for $\delta_P < 1$ that allows facilitation, but facilitation imposes a lower bound on the depression of feedforward inputs to PV rather than an

upper bound. That we have for $\kappa > 1$ the same condition as for small $\kappa < (w-1)/w$, but with the direction of the inequality reversed, is the analytical description of the inversion of PV-responses through SST-feedback.

In the non-ISN, the term $(1 - w + w\kappa)$ is positive for arbitrary κ. Division yields

$$\delta_P > 1 + (1 - \delta_E)\frac{w(1-\kappa)}{g_{fw}(1 - w + w\kappa)}. \tag{4.46}$$

However, if this condition can be fulfilled by feedforward depression depends on the sign of $1 - \kappa$: if $\kappa < 1$ the condition Eq. 4.46 cannot be fulfilled in feedforward depression because it would require $\delta_P > 1$. Thus, in the non-ISN with weak SST-feedback, feedforward depression cannot induce facilitation, consistent with the result for the non-ISN with a single subtype of interneurons (equivalent to $\kappa = 0$). If $\kappa > 1$ it can be fulfilled. Writing signs explicitly Eq. 4.46 becomes

$$\delta_P > 1 - (1 - \delta_E)\frac{w(\kappa-1)}{g_{fw}(1 - w + w\kappa)}. \tag{4.47}$$

Thus, for strong enough feedback from SST-interneurons, PV rates can be facilitated above baseline also in the non-ISN. Note however that this condition again imposes a lower bound on δ_P.

4.7.2 Conditions on facilitation for recurrent potentiation

In the numerical solution, I found for the excitatory population that the area of facilitation in response to recurrent potentiation constantly stays at zero throughout the range of κ (Fig. 4.11B). In the PV-population, I found a reversal from a purely suppressive effect of recurrent potentiation in the ISN to a purely facilitatory effect for sufficiently large κ. This reversal is similar to the one in feedforward depression. First, a facilitatory area emerges at the value of κ at which the paradoxical effect reverses (Eq. 4.27). Second, the whole parameter plane is facilitatory when $\kappa > 1$ (Fig. 4.11B). Using the analytical description of the population rates, I derived the conditions in κ to find this reversal in the PV-population. I found that facilitation of the rate of the PV population becomes indeed possible in ζ_{PE} when κ is large enough to reverse the paradoxical effect, and that facilitation is present in both dimensions (ζ_{EP} and ζ_{PE}) when $\kappa > 1$. I focused on the ISN in this analysis since facilitation is already possible in the non-ISN without SST-feedback.

Excitatory and SST populations

The the condition to find facilitation of the firing rate of the excitatory population after recurrent potentiation (with $\delta_E = \delta_P = 1$) is:

$$\begin{aligned}
\Xi^E_{\text{rec}} &= r^{E,\text{rec}}_{ss} - r^{E,bl}_{ss} \\
&= \frac{r_x}{\eta^{md}} \cdot \Big\{2\big(1 + \gamma w\big) - \zeta_{EP}\gamma w g_{fw} - \kappa + \gamma w\kappa\big(\zeta_{EP} - 1\big)\Big\} \\
&\quad -\frac{r_x}{\eta^{bl}} \cdot \Big\{2\big(1 + \gamma w\big) - \gamma w g_{fw} - \kappa\Big\} > 0.
\end{aligned} \tag{4.48}$$

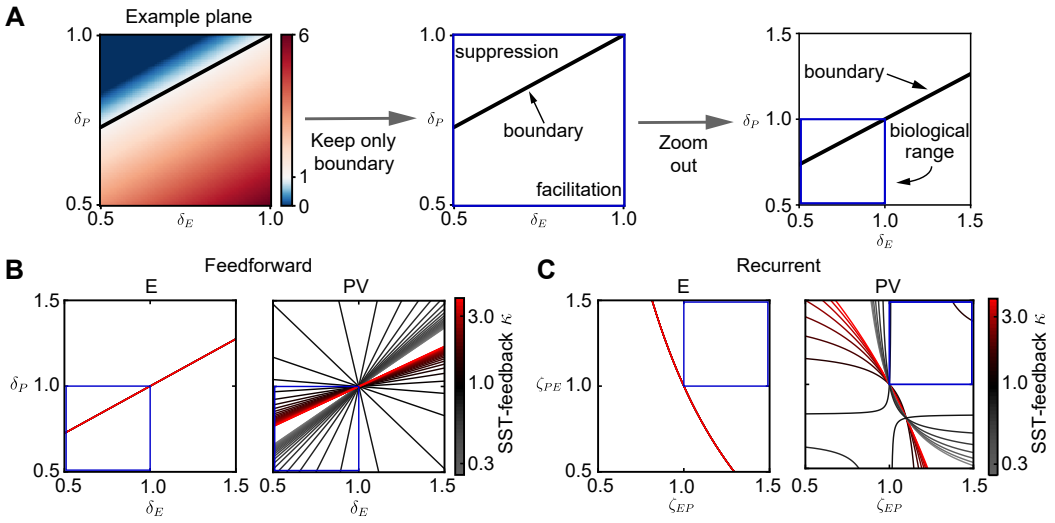

Figure 4.12: The boundary between facilitation and suppression in the ISN with multiple interneuron subtypes (compare to Fig. 3.11). **(A)** Extraction of the boundary between facilitation and suppression. The boundary is defined through the parameter combinations for which rate after MD is the same as in baseline (white areas, left). I only kept the boundary for further visualization and dismissed the firing rate fold changes (middle, boundary analytically derived). I also included non-biological parameter ranges to see how the boundary behaves when no facilitatory area within the biological range of MD-induced plasticity can be found. Blue box shows biological range (feedforward depression: $\delta_E \leq 1$ and $\delta_P \leq 1$, recurrent potentiation: $\zeta_{EP} \geq 1$ and $\zeta_{PE} \geq 1$). **(B)** Boundary between facilitation and suppression for feedforward depression in the network with two subtypes of interneurons as a function of the SST-feedback κ for excitatory (left, based on Eq. 4.37) and PV neurons (right, based on Eq. 4.42). **(C)** Same as (B) for recurrent potentiation, based on Eq. 4.51 for excitatory (left) and Eq. 4.60 for PV (right).

Using the relation between η_{md} and η_{bl} from Eq. 4.32, I derived the expression for ζ_{PE} at the boundary between facilitation and suppression for given ζ_{EP} as

$$\zeta_{PE} = \frac{1}{\zeta_{EP}} + A_E \frac{\zeta_{EP} - 1}{\zeta_{EP}} \tag{4.49}$$

with

$$A_E = \kappa + \frac{(1 - w + \gamma w + \kappa w)(\kappa - g_{fw})}{w\big(2 - \kappa + \gamma w(2 - g_{fw})\big)}. \tag{4.50}$$

In full generality, the solution in Eq. 4.49 is very complex. To simplify, I replace the variable g_{fw} and use it as a constant with the value I have used for the numerical solution of the population rate model and in the spiking network $g_{fw} = 2$. With this, Eq. 4.49 simplifies to:

$$\zeta_{PE} = \frac{1}{\zeta_{EP}} + \frac{1 - w + \gamma w}{w} \cdot \frac{\zeta_{EP} - 1}{\zeta_{EP}}. \tag{4.51}$$

The boundary is invariant under changes of κ, shown in Fig. 4.12C (left). Therefore, the same conditions for facilitation and suppression hold as in the network without SST-feedback.

To get a global sense for the general case, especially if facilitation is possible in the ISN with multiple interneurons subtypes for a given κ, I simplified the expression in Eq. 4.48 without resolving the explicit boundary, as in Sec. 3.7 (Eqs. 3.63-3.68) using the substitution $\zeta_{EP} = 1 + \varepsilon_{EP}$. With this, Eq. 4.48 can be rearranged:

$$\Xi^{E}_{\text{rec}} = r^{E,\text{rec}}_{ss} - r^{E,bl}_{ss} = \left(\frac{\eta^{bl}}{\eta^{md}} - 1\right) r^{E,bl}_{ss} - \varepsilon_{EP}\frac{\gamma w r_x}{\eta^{md}}\big(g_{fw} - \kappa\big). \tag{4.52}$$

For a biologically reasonable solution, the firing rate in the baseline state must be positive. Thus, assuming $\eta^{bl} < \eta^{md}$, the first term in this simplified expression of the firing rate changes is negative. Contrary to the case without SST-feedback (Eq. 3.68), the term subtracted from this is not generally positive. However, as long as $\kappa < g_{fw}$, it is positive, since then a negative term is subtracted from a positive term on the right hand side of Eq. 4.52. With g_{fw} well above one (inspired by biological measurements I used $g_{fw} = 2$), this only modestly constrains κ with respect to the inversion of PV which requires $\kappa > 1$. Thus, SST-feedback can fully invert PV-responses without inducing facilitation of the excitatory firing rate in response to recurrent potentiation.

The condition to find facilitation of the firing rate of the SST population after recurrent potentiation is

$$\begin{aligned}
\Xi^{S}_{\text{rec}} &= r^{S,\text{rec}}_{ss} - r^{S,bl}_{ss} \\
&= \frac{r_x}{\eta^{md}}\Big\{2w(1+\gamma w) - \zeta_{EP}\gamma w^2 g_{fw} + 1 - w + \gamma w + \gamma w^2(\zeta_{PE}\zeta_{EP} - 1)\Big\} \\
&\quad -\frac{r_x}{\eta^{bl}}\Big\{2w(1+\gamma w) - \gamma w^2 g_{fw} + 1 - w + \gamma w\Big\} > 0.
\end{aligned} \tag{4.53}$$

Contrary to the feedforward case, the response of the SST population to recurrent potentiation is not clearly identifiable as the excitatory solution multiplied with w (Eq. 4.39), but also contains more complex terms. Specifically, it contains a mixture term with the product $\zeta_{PE}\zeta_{EP}$. However, it is possible to derive that the boundary between facilitation and suppression is the same in the SST population as it is in the excitatory population.

Using the relation between η_{md} and η_{bl} from Eq. 4.32, I derived the expression for ζ_{PE} at the boundary between facilitation and suppression for given ζ_{EP} as

$$\zeta_{PE} = \frac{1}{\zeta_{EP}} + A_S\frac{\zeta_{EP} - 1}{\zeta_{EP}}, \tag{4.54}$$

where from direct resorting of the SST rate changes in Eq. 4.53 first follows

$$A_S = \frac{\kappa w\big((1+\gamma w)(g_{fw} - 2)\big) + (\kappa - g_{fw}(1 - w + \gamma w)}{w\big(2 - \kappa + \gamma w(2 - g_{fw})\big)}. \tag{4.55}$$

This A_S has already the same denominator as the fraction in A_E in Eq. 4.50, but there are two major discrepancies: the solution of A_E furthermore includes a κ added to the fraction, and the fraction has a different numerator. To bring A_S into a form comparable to A_E, I added a zero to A_S of the form:

$$0 = \kappa - \kappa \frac{w\big(2 - \kappa + \gamma w(2 - g_{fw})\big)}{w\big(2 - \kappa + \gamma w(2 - g_{fw})\big)}. \tag{4.56}$$

Adding this zero to A_S, I derived the numerator (leaving out the added κ for now):

$$\begin{aligned} &-\kappa w\big(2 - \kappa + \gamma w(2 - g_{fw})\big) + \kappa w(1 + \gamma w)(g_{fw} - 2) + (\kappa - g_{fw})(1 - w + \gamma w) \\ &= (\kappa - g_{fw})(1 - w + \gamma w + \kappa w) \end{aligned} \tag{4.57}$$

and thus I find A_S (Eq. 4.55) to be:

$$A_S = \kappa + \frac{(1 - w + \gamma w + \kappa w)(\kappa - g_{fw})}{w\big(2 - \kappa + \gamma w(2 - g_{fw})\big)}. \tag{4.58}$$

So $A_S = A_E$ (latter in Eq. 4.50), and since the equation for the boundary for the SST population (Eq. 4.54) has the same form as the equation for the excitatory population (Eq. 4.49), the boundary between facilitation and suppression in the excitatory and SST populations is the same. From this follows that SST firing rates will always change in the same direction as the excitatory firing rates.

PV population

The condition to find facilitation of the firing rate of the PV population after recurrent potentiation is

$$\begin{aligned} \Xi^P_{\text{rec}} &= r^{P,\text{rec}}_{ss} - r^{P,bl}_{ss} \\ &= \frac{r_x}{\eta^{md}} \cdot \Big\{2w\big(\zeta_{PE} - \kappa\big) + \big(1 - w + w\kappa\big)g_{fw} - \kappa + w\kappa\big(1 - \zeta_{PE}\big)\Big\} \\ &\quad - \frac{r_x}{\eta^{bl}} \cdot \Big\{2w\big(1 - \kappa\big) + \big(1 - w + w\kappa\big)g_{fw} - \kappa\Big\} > 0. \end{aligned} \tag{4.59}$$

From this I derived the boundary between facilitation and suppression using the same route as for the excitatory and SST populations

$$\zeta_{PE} = \frac{\kappa(\zeta_{EP} - 1) - A_P + 1}{\zeta_{EP} - A_P} \tag{4.60}$$

with

$$A_P = \frac{(2 - \kappa)(1 - w + \gamma w + \kappa w)}{\gamma w\big(2w(1 - \kappa) + (1 - w + \kappa w)g_{fw} - \kappa\big)}. \tag{4.61}$$

This boundary has a complex behavior (Fig. 4.12C, right), primarily because it has a singularity where $\zeta_{EP} = A_P$. The term A_P grows linearly in κ, with values in

the range $A_P \in [0.5, 2]$ for $\kappa \in [0, 3]$. This range of A_P overlaps with the biological range for potentiation $\zeta_{EP} > 1$, which I numerically choose as $\zeta_{EP} \in [1, 1.5]$. The inversion of the PV firing rate changes, from purely suppressive in the ISN with small κ to purely facilitatory in the ISN with large κ, mainly happens through a combination of (1) a shift of the singularity and (2) a change of the curvature of the two branches of the boundary in Eq. 4.60. This is most clearly visible in the switch in the shape of the curves from convex (gray tones, small κ) to concave (red tones, large κ) in Fig. 4.12C (right). Right to and above the gray curves, PV firing rates suppress, whereas to the right and above the red curves, they facilitate, indicating the inversion from a purely suppressive effect of recurrent potentiation in the ISN with small κ to a purely facilitatory effect in the ISN with large κ.

To more clearly express the emergence of facilitation, and especially its relation to the reversal of the paradoxical effect happening at $\kappa = (w-1)/w$ and the full inversion in feedforward depression happening at $\kappa = 1$, I studied the rate changes without resolving the explicit boundary. I simplified the expression in Eq. 4.59 as in Sec. 3.7 (Eqs. 3.63-3.68). With the substitutions $\zeta_{EP} = 1 + \varepsilon_{EP}$ and $\zeta_{PE} = 1 + \varepsilon_{PE}$, I first calculated:

$$\zeta_{EP}\zeta_{PE} - 1 = \varepsilon_{EP} + \varepsilon_{PE} + \varepsilon_{EP}\varepsilon_{PE} \tag{4.62}$$

with which I derived as the condition to find facilitation ($\Xi^P_{\text{rec}} > 0$, Eq. 4.59):

$$\begin{aligned} 0 \;<\; & \gamma w\big[\kappa\varepsilon_{EP} - (\varepsilon_{EP} + \varepsilon_{PE} + \varepsilon_{EP}\varepsilon_{PE})\big]\big[2w(1-\kappa) + (1 - w + \kappa w)g_{fw} - \kappa\big] \\ & + \varepsilon_{PE}(1 - w + \gamma w + \kappa w)(2 - \kappa). \end{aligned} \tag{4.63}$$

Replacing the variable g_{fw} with the numerical values I have used ($g_{fw} = 2$) to simplify Eq. 4.63 I furthermore find as condition to find facilitation:

$$0 < \gamma w\big[\kappa\varepsilon_{EP} - (\varepsilon_{EP} + \varepsilon_{PE} + \varepsilon_{EP}\varepsilon_{PE})\big] + \varepsilon_{PE}(1 - w + \gamma w + \kappa w). \tag{4.64}$$

Finally, analyzing the two dimensions of recurrent potentiation independently this condition yields

$$\Xi^P_{\text{rec}} > 0 \begin{cases} \text{if:} \;\; \kappa > \dfrac{w-1}{w} & \text{for} \;\; \varepsilon_{EP} = 0 \;\text{ and }\; \varepsilon_{PE} > 0 \\ \text{if:} \;\; \kappa > 1 & \text{for} \;\; \varepsilon_{EP} > 0 \;\text{ and }\; \varepsilon_{PE} = 0. \end{cases} \tag{4.65}$$

Therefore, if $\kappa > (w-1)/w$, potentiation of connections from the excitatory to the PV population ($\zeta_{PE} = 1 + \varepsilon_{PE}$) facilitates the firing rate of the PV population. Facilitation in this dimension thus emerges exactly when κ is large enough to reverse the paradoxical effect (Eq. 4.27). If $\kappa > 1$, also potentiation of connections from the PV to the excitatory population ($\zeta_{EP} = 1 + \varepsilon_{EP}$) facilitates the firing rate of the PV population. Since $(w-1)/w \leq 1$ for all $w > 0$, the firing rate of the PV population facilitates for both dimensions of recurrent potentiation when $\kappa > 1$.

Taken together, in this simplified analysis, I can prove that facilitation becomes first possible if $\kappa > (w-1)/w$, and is present in both dimensions of the recurrent potentiation if $\kappa > 1$, consistent with the numerical solution of the facilitatory area

in recurrent potentiation (F_P in Fig. 4.11B). Note, however, that this does not imply potentiation in the whole (ζ_{EP}, ζ_{PE})–plane. For κ close to the transition ($\kappa \approx 1$), the linear model predicts that strong potentiation of both connections suppresses firing rates (black curve in upper right corner Fig 4.12C, right). This is an effect of a naive solution of the purely linear case, which does not consider thresholding such that all rates are positive, also after recurrent potentiation. For strong potentiation of both connections, excitatory rates in the naive model become negative. This suppression does not appear when the positivity of firing rates is properly taken into account (Fig 4.10B, middle). Thus, taking the threshold into account, the facilitatory effect of both parameters of recurrent potentiation leads to a purely facilitatory response in the whole plane or recurrent potentiation ($F_P = 1$) for $\kappa \geq 1$ (Fig 4.11B).

4.8 Independent and coordinated firing rate changes with a single and with two subtypes of interneurons

To quantify the potential of the ISN with strong SST-feedback to produce independent or opposite firing rate changes in excitatory and PV neurons, specifically in relation to the biological data, I have used a Monte-Carlo approach. This approach makes use of the systematic study of the effects of feedforward and recurrent plasticity in the spiking network with a single subtype of interneurons (Fig. 3.3) and with two subtypes of interneurons (Fig. 4.3). I compared the predictions in the ISN with strong SST-feedback and in the ISN without SST-feedback in the spiking network (Fig. 4.13). To compare these predictions to the experimental results, I also extracted the experimental mean and variability from [54] using the open-source software *WebPlotDigitizer* [88].

I hypothesized that the changes of correlations in the spiking of LGN neurons [76] first affect feedforward synapses, followed by a delayed effect on the recurrent synapses. Therefore, to produce the firing rates on the first day of MD, with suppressed firing rates of PV while excitatory rates stay at baseline, I selected parameters for feedforward depression that keep excitatory rates around their baseline (Fig. 4.3B, white region in the left plane). Specifically, I used the simulated grid of firing rate fold changes for excitatory neurons ($\Psi^A(x, y)$, Eq. 3.10) and extracted the range for δ_E and δ_P that leads to normalized excitatory rates around the experimentally measured mean of 1.05, allowing a deviation of $\pm 15\%$. I obtained an area in the space of feedforward depression that produces excitatory rates scattering around the target value. I then generated 70 random points in this defined area and simulated the network with these randomly generated parameters. With strong SST-feedback, these parameters indeed give rise to suppressed firing rates in the PV-interneurons, for which were not selected for. Overall I found a good agreement of the firing rate configuration for the first day of MD compared to the experimental values (Fig. 4.13, left). In the ISN without SST-feedback, these

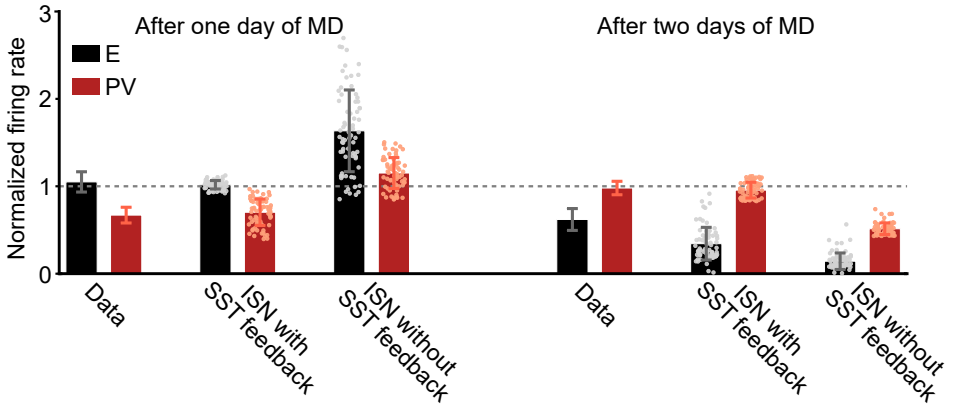

Figure 4.13: The ISN with strong SST-feedback can capture the independent modulation of excitatory and inhibitory firing rates observed during brief MD. Firing rates of excitatory (black) and PV (red) neurons after one day of MD (left histograms) and after two days of MD (right histograms), normalized by the firing rates before MD-induction. Data are extracted from [54] and reproduced in the spiking network in the ISN regime that includes strong feedback from SST and in an ISN without SST-feedback (Methods). Experimental data are mean ± SEM, simulation data are mean ± SD. Reproduced from [1].

values for feedforward depression lead to facilitation of the rates for both excitatory and PV-interneurons.

In a second step, I generated parameters of the recurrent potentiation that can recover PV firing rates back to baseline on the second day of MD. For each (δ_E, δ_P)–pair used for the first day of MD, I randomly searched one (ζ_{EP}, ζ_{PE})–pair that facilitates PV-firing rates from their suppressed state to the experimentally measured value around baseline of 0.99 on the second day of MD, again allowing ±15% deviation. Because recurrent potentiation is purely suppressive in the excitatory population, this naturally leads to suppression of the excitatory firing rates and, therefore, a firing rate configuration resembling the firing rates on the second day of MD *in vivo* (Fig. 4.13, right). In the ISN without SST-feedback, these parameters strongly suppressed both excitatory rates and PV rates.

In conclusion, I could reproduce the temporally distinct modulation of firing rates in an ISN with strong SST-feedback using a simple Monte-Carlo procedure. In the ISN without SST-feedback, the excitatory and PV firing rates are not modulated independently but potentiate together on the first day of MD and suppress together on the second day of MD. This confirms the numerical and analytical results on the role of SST-feedback to enable independent modulation of the firing rates in an ISN.

4.9 Discussion

I have shown that the introduction of a second subtype of interneurons, SST in addition to PV, selectively inverts the responses of PV-interneurons in the ISN given

sufficiently strong feedback from SST in both the spiking network (Fig. 4.3) and in a linear rate model, taking positivity of firing rates into account (Fig. 4.10). Since the firing rates of the excitatory and PV populations are tightly coordinated in the ISN without SST-feedback (Chapter 3), this inversion through strong SST-feedback allows for independent modulation of the excitatory and PV firing rates, as measured *in vivo*, in large parts of the parameter spaces of MD-induced plasticity. Contrary to this, the ISN with no or weak SST-feedback allows independent modulations of the excitatory and PV firing rates only in tightly constrained parts of the parameter space. The large freedom of values for feedforward and recurrent plasticity that allow a distinct, even opposite, modulation of the firing rates of excitatory and PV neurons makes the ISN with multiple subtypes of interneurons a natural model to explain the independent modulation *in vivo* given sufficiently strong feedback from the SST-interneurons (Fig. 4.13).

The inversion of PV responses that leads to this independent modulation emerges from a reversal of the paradoxical effect in PV interneurons, which I demonstrated numerically in the spiking network (Fig. 4.1D) and analytically in the population rate description of the network (Fig. 4.9). A prediction from this is that the independent modulation of firing rates during MD happens in a network in which PV interneurons do not express the paradoxical effect. This is similar to the prediction from the non-ISN in the previous chapter, which also enables the independent modulation through a non-paradoxical response in PV even without SST-feedback. However, an interesting difference between the predictions from the non-ISN without SST-feedback and the ISN with strong SST-feedback is the mechanism underlying the non-paradoxical response of PV. In the non-ISN, it depends on a weak excitatory to excitatory coupling. In the ISN with multiple subtypes of interneurons, the non-paradoxical response of PV can also happen with strong excitatory to excitatory coupling but depends on strong feedback from SST onto the circuit. In this case, the discrepancy between the experimentally identified paradoxical effect in PV in adult animals and the model prediction for the developing V1 could be explained by a strong, transient SST-feedback onto PV-interneurons. While strong SST-feedback which decreases during development was found in experiments [152], it remains to be determined if this SST-feedback is strong enough to lead to a non-paradoxical response of PV in the developing cortex.

A related prediction from the model with multiple interneuron subtypes is that SST-interneurons tightly follow the excitatory rate changes in response to MD-induced plasticity. From my modeling and analytical results, I predict that SST-interneurons are regulated in the same temporal order as excitatory neurons during the first two days of MD. So far, the rates of SST during MD have not been investigated *in vivo*. In the extracellular recordings used for the long-term measurement in [54, 89], SST-interneurons are hard to identify. Different morphological subtypes would either be identified as regular spiking (and thus be pooled with the excitatory neurons) or fast-spiking (and thus be pooled with PV) [24]. Potentially, chronic calcium imaging recordings could solve this [80].

Furthermore, the plasticity of synapses involving SST-interneurons during sensory deprivation is unknown. One study measured the plasticity of synapses that involve regular-spiking non-pyramidal units, which are potentially SST-interneurons, during brief MD [93]. These experiments found that the amplitude of connections from regular-spiking non-pyramidal interneurons (RSNP) onto pyramidal excitatory neurons increases, while the connection probability from RSNP to pyramidal neurons decreases. How these two factors would exactly trade of warrants further investigation. However, the experiments did not involve molecular labeling, thus it is unclear whether the plasticity measured indeed involves SST-interneurons. Furthermore, these experiments were conducted in younger animals before the critical period, and it has been shown that such pre-critical period plasticity involves different mechanisms than critical period plasticity [83, 85] (Sec. 2.2 and Table 2.1). In Sec. 4.5.1 I have investigated a related potential form of recurrent plasticity in the synapses outgoing from SST-interneurons that is consistent with experimental measurements of an increase of recurrent inhibition impinging onto excitatory neurons [3]. I found that the effects on the firing rates induced by this plasticity of SST-synapses are comparable to the firing rate changes induced by synaptic changes in the recurrent circuit between excitatory and PV neurons (Fig. 4.6). The prediction is that this hypothesized plasticity could serve as a redundant pathway that either enforces the effects of the plasticity in connections between excitatory and PV neurons or could fully take over its role in case of a disruption of the plasticity in these connections, e.g., in a pathological condition [153, 154].

Finally, I also investigated plasticity induced by brief whisker deprivation (WD, Fig. 4.7). In experiments, WD induces plasticity in different circuit elements (intrinsic excitability rather than feedforward inputs). However, despite this difference, the regulation of rates appears very similar when applying interacting feedforward and recurrent plasticity in the model network that represents either the MD or the WD changes (compare Figs. 4.3D and 4.7). The activity changes *in vivo* during brief MD have not been measured so far, but my modeling would predict a comparable regulation of firing rates (see also [95]). This prediction may also hint at a common principle at work in both the somatosensory and visual cortices when their sensory inputs are disrupted by deprivation.

5 Spike-timing-dependent plasticity in networks with multiple subtypes of interneurons

Experimental work supports the idea that spike-timing-dependent plasticity (STDP) mechanisms underlie activity-dependent adaptation of synaptic connectivity in a variety of neural circuits [155–158]. The strength of a synapse undergoing STDP depends on the timing of presynaptic and postsynaptic spikes within a few milliseconds of precision [155]. In recurrent networks, the synaptic connectivity and the spiking statistics of the neurons in the circuit shape each other reciprocally [34]. This mutual dependence of connectivity and activity present in recurrent networks with synapses undergoing STDP made progress on understanding these systems hard, especially in highly recurrent networks that require inhibition to stabilize activity [159]. Recent modeling work, however, could uncover how the interaction of multiple plasticity mechanisms can lead to recurrent circuits with self-stabilizing connectivity and activity [30, 31, 160–162], and used such circuits to explain developmental processes and computations in the cortex [90, 163] (see also [4] for review).

These studies of recurrent, balanced networks with multiple, interacting plasticity mechanisms opened the possibility to train specific network connectivity that is self-stabilizing through spontaneous activity, but also flexibly responds to structured external stimulation [30]. Yet, the diversity of interneuron subtypes has not been incorporated into such models. I showed in the previous chapters how fundamental the role of this diversity can be in shaping the activity of cortical networks in response to developmental plasticity. Therefore, I extended previous research on balanced networks with plastic synapses to incorporate the diversity of interneuron subtypes, specifically PV- and SST-interneurons. I have investigated the plasticity in this circuit using a common theoretical framework to study learning-related and developmental plasticity: the training and stabilization of *Hebbian assemblies* [26]. Such assemblies are groups of highly interconnected neurons that share a common stimulus specificity, for example, neurons with similar orientation preference in V1, which are strongly connected in rodents [50,51]. The strong connectivity among these neurons emerges during experience-dependent development [27] (but see also [28]).

A consequence of the strong connectivity between groups of neurons embedded in a network is that they spontaneously co-activate. This underlies the stabilization of the connectivity through ongoing long-term potentiation [30]. The spontaneous activity is highly correlated with evoked responses. In cortical circuits, evidence for

the existence of such assemblies was found [164]. It has also been established that they can indeed emerge through Hebbian plasticity *in vivo* [29].

I will first introduce the model structure that I developed to study the emergence and modification of Hebbian assemblies in networks with two interneurons subtypes (Sec. 5.1). Specifically, I introduce the procedures for training the network and the plasticity mechanisms. Beyond the current literature, I have not only introduced additional interneuron subtypes, but I have also used two different input streams (LGN and BKG, as in the previous model for MD). This opens new ways of training assemblies in these networks, e.g., cooperation or multiple forms of competition between the two input sources. I investigated the network connectivity that emerges through multiple training procedures with varying competition between LGN- and BKG-inputs and studied the structure of spontaneous spiking activity that emerges (Sec. 5.2). I found that strongly connected Hebbian assemblies can also emerge with multiple interneuron subtypes when the LGN and BKG inputs cooperate (Sec. 5.2.1). These assemblies spontaneously reactivate, stabilizing the connectivity. If the two input streams compete with each other, neurons within an assembly are more weakly connected, but still, a identifiable connectivity structure emerges (Sec. 5.2.2). However, the differences in the connectivity within and between assemblies neurons strongly influences if and how strongly the neurons in the assembly spontaneously co-activate. Therefore, the structure of the spontaneous activity is highly dependent on the relationship between inputs from LGN and BKG. I derive the conditions for training structured connectivity in a network with multiple interneuron subtypes and interacting input streams. Finally, I briefly discuss the results in Sec. 5.3.

5.1 Modeling approach

5.1.1 Network structure and plasticity

I simulated large scale, sparse networks of leaky integrate-and-fire neurons. The neurons are connected through conductance-based synapses and follow the dynamics described in Eqs. 3.3-3.5. Parameters of single neurons are listed in Table 5.1. Similar to the work in the previous chapter, I studied a model with one excitatory population and two subtypes of interneurons (PV and SST). The recurrent synaptic weights between neurons of the different cell classes follow the schematic (Fig. 5.1A):

$$\boldsymbol{W}_{\text{rec}} = \begin{pmatrix} W^{EE} & W^{EP} & W^{ES} \\ W^{PE} & W^{PP} & W^{PS} \\ W^{SE} & W^{SP} & W^{SS} \end{pmatrix} = \begin{pmatrix} \mathbf{J_{EE}} & \mathbf{J_{EP}} & \mathbf{J_{ES}} \\ J & g_{rc}J & g_{rc}J \\ J/2 & 0 & 0 \end{pmatrix}, \tag{5.1}$$

where the connection weights written in boldface are plastic and follow connection-type specific STDP rules (see below). The remaining connection weights are static (do not undergo plasticity) and stay at their initial values throughout the simulation (Table 5.2). With the initial parameters, spiking activity in the network is asynchronous and irregular (Fig. 5.1B).

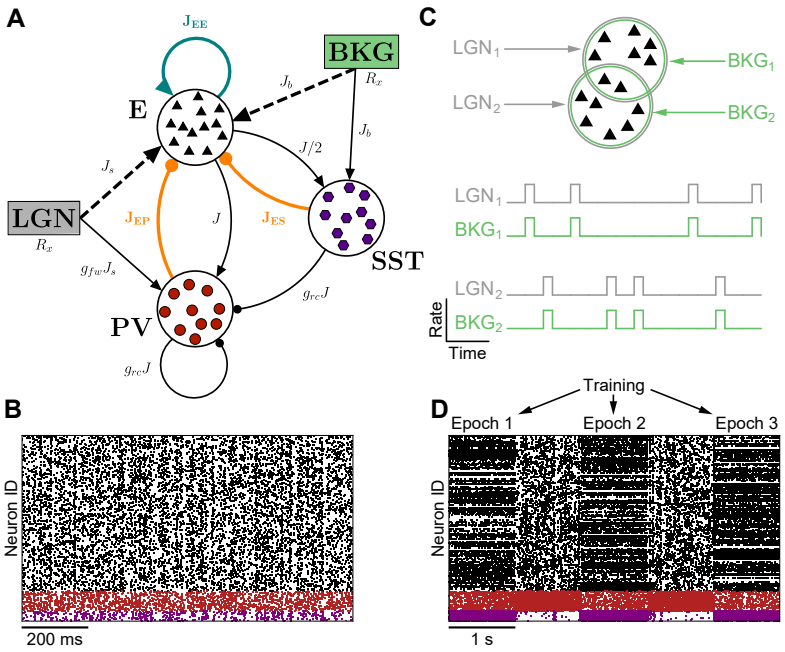

Figure 5.1: Network structure and training with multiple subtypes of interneurons and STDP. **(A)** Network schematic showing synaptic connections among neurons. Triplet STDP and normalization in excitatory to excitatory recurrent connections (J_{EE}, teal), inhibitory STDP in PV to excitatory and SST to excitatory connections (J_{EP} and J_{ES}, orange). Remaining connections are static. Input connections from LGN and from BKG to excitation are structured (dashed lines, see also C). **(B)** Raster plot of activity before training, showing asynchronous irregular activity. Only a subset of the total populations is plotted from excitatory (black), PV (red) and SST (purple) neurons. **(C)** Structured inputs to the network. Top: assemblies are defined as groups of neurons that receive the same inputs. Bottom: During training, neurons of one assembly receive increased inputs for one second, after which another randomly chosen group is driven. Example case with the same structure for LGN and BKG inputs. **(D)** Raster plot of activity during training. Shown are five seconds with three training epochs of randomly chosen assemblies and two breaks to allow network activity to settle to its resting state with baseline input. Colors same as in (B).

Excitatory plasticity

Plasticity in the model changes synapses within defined bounds (Table 5.2). Recurrent connections between excitatory cells (J_{EE}) are governed by two different plasticity mechanisms: as a spike timing-dependent plasticity I used triplet STDP [165] (Fig. 5.2, left). Furthermore, as a mechanism that controls the overall excitatory-to-excitatory coupling strength, I used a heterosynaptic normalization of the total incoming excitatory weight onto excitatory neurons [105].

Similar to the more common pair-based STDP, in triplet STDP weights of single neurons are updated dependent on the timing of pairs of spikes in the pre- and postsynaptic cells (pre–post or post–pre). However, weight updates furthermore take triplets of spikes into account (e.g. post–pre–post or pre–post–pre). The weight updates are based on multiple spike detectors in the single neuron that are updated either by a pre- or postsynaptic spike, and decay with a spike detector-specific time constant, implementing a memory of the timespan since the last relevant spike occurred. Triplet STDP is based on four spike detectors: r_1 and r_2 that trace presynaptic spikes, and o_1 and o_2 that trace postsynaptic spikes. When a respective spike occurs, at time point t^{pre} or t^{post}, the respective spike detector is updated by $+1$. When no spike occurs, the value of the spike detector exponentially decays to zero with its specific timescale. The dynamics of all the four spike detectors in the triplet STDP reads:

$$\begin{aligned}
\tau_+ \frac{dr_1(t)}{dt} &= -r_1(t) \qquad \text{if:} \quad t \neq t^{pre} \\
r_1(t) &\to r_1(t) + 1 \quad \text{if:} \quad t = t^{pre} \\
\\
\tau_x \frac{dr_2(t)}{dt} &= -r_2(t) \qquad \text{if:} \quad t \neq t^{pre} \\
r_2(t) &\to r_2(t) + 1 \quad \text{if:} \quad t = t^{pre} \\
\\
\tau_- \frac{do_1(t)}{dt} &= -o_1(t) \qquad \text{if:} \quad t \neq t^{post} \\
o_1(t) &\to o_1(t) + 1 \quad \text{if:} \quad t = t^{post} \\
\\
\tau_y \frac{do_2(t)}{dt} &= -o_2(t) \qquad \text{if:} \quad t \neq t^{post} \\
o_2(t) &\to o_2(t) + 1 \quad \text{if:} \quad t = t^{post}.
\end{aligned} \tag{5.2}$$

The values of these spike detectors determine the size of the weight update when a pre- or postsynaptic spike occurs, giving rise to the specific form of the plasticity function over the spike-time differences in the pre- and postsynaptic neurons (Fig. 5.2, left). The amplitudes A_y^x set an overall scale of the respective weight update: A_2^+ and A_2^- are the amplitudes for spike pair-based potentiation ($+$) and depression ($-$), respectively, and A_3^+ and A_3^- are the amplitudes for spike triplet-based potentiation

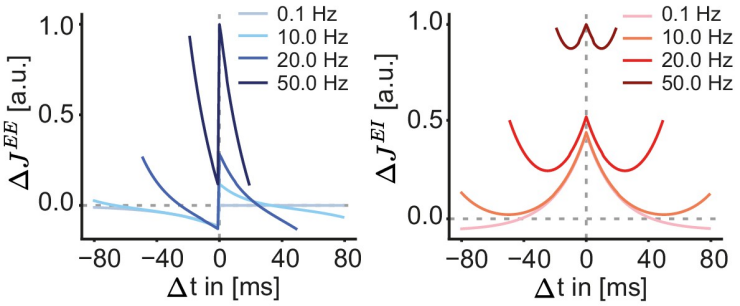

Figure 5.2: Functions for modeling excitatory and inhibitory STDP. Left: triplet STDP used for excitatory synapses onto excitatory neurons (J_{EE}, Fig. 5.1A). Curves show the weight changes ΔJ_{EE} induced by 60 spike pairs at the given time delay (Δt). Colors of the curves denote the frequency of consecutive pairs. Right: same measure for the symmetric STDP used for inhibitory synapses onto excitatory neurons (ΔJ_{EI}). I used the same plasticity rule for synapses from PV onto excitatory neurons (J_{EP}, Fig. 5.1A) and from SST onto excitatory neurons (J_{ES}, Fig. 5.1A). Figure from [163].

and depression (Table 5.2). The weight of a synapse $w(t)$ is potentiated when the postsynaptic neuron spikes:

$$w(t) \longrightarrow w(t) + r_1(t)\left[A_2^+ + A_3^+ o_2(t-\tau_d)\right] \qquad \text{if:} \quad t = t^{post}, \tag{5.3}$$

where τ_d is a small positive constant, and the weight update uses $o_2(t-\tau_d)$ such that it is based on the spike detector before its update by the postsynaptic spike. This can be the simulation time step or, in continuous time, any $\tau_d > 0$ [165].

The synapse is depressed when the presynaptic neuron spikes:

$$w(t) \longrightarrow w(t) - o_1(t)\left[A_2^- + A_3^- r_2(t-\tau_d)\right] \qquad \text{if:} \quad t = t^{pre}. \tag{5.4}$$

Note that these updates and the corresponding spike detectors are all for a specific, single synapse $w_{ij}(t)$ connecting neuron j to neuron i, but for clarity, I omitted indices. The dynamics are the same for all excitatory synapses. The triplet STDP is well suited to study the emergence of Hebbian assemblies because it promotes the weight increase in bidirectional connections, contrary to a purely pair-based STDP [47, 166].

Learning through purely Hebbian plasticity suffers from an intrinsic instability. The synapse between two neurons potentiates when a postsynaptic spike follows closely two a presynaptic spike. However, in the case of excitatory synapses, this potentiation makes future coincidences more likely as the increased weight leads to a stronger postsynaptic depolarization. This positive feedback-loop of weight increases can lead to runaway potentiation of excitatory weights. This can lead to a network where all synapses go to their upper bound. To prevent this potentiation of all synapses, and to induce some competition between the synapses, I thus furthermore used a heterosynaptic plasticity that homeostatically regulates the total incoming

excitatory weight impinging on an excitatory neuron [167]. This is implemented as a normalization where from each synapse onto one neuron the average deviation from the initial total incoming weight is subtracted:

$$w_{ij}(t) \longrightarrow w_{ij}(t) - \frac{\sum_{j=1}^{N_E} \Big(w_{ij}(t) - w_{ij}(0) \Big)}{k_{in}^{(i)}}, \tag{5.5}$$

where N_E is the number of excitatory neurons (Table 5.1) and $k_{in}^{(i)}$ is the number of incoming excitatory synapses that the excitatory neuron i receives. This normalization keeps the sum of all excitatory synaptic weights onto each excitatory neuron at the initial level. It is applied as a hard reset to the target value every 20 ms. Through this, we can also ensure that the network stays in the same operating regime throughout the training and during spontaneous activity.

Inhibitory plasticity

Normalization induces competition between synapses and aids in imprinting structured connectivity. But since it only constrains the mean of the input and allows a wide range of variance of the input weights to one neuron, it is not fit to homeostatically control firing rates to stay at a target level. A particular form of inhibitory plasticity is a mechanism that enables such firing-rate homeostasis [168]. In my model, the weights of inhibitory connections from PV- and from SST-interneurons onto excitatory neurons both follow the same STDP rule, proposed by Vogels, Sprekeler and colleagues [168]. The rule is based on two spike detectors for pre- and postsynaptic spikes (x_r and x_o, respectively). Each spike detector is updated by $+1$ when a pre- or postsynaptic spike occurs, and the values of both spike detectors decay exponentially to zero with a common timescale τ_{inh} (Fig. 5.2, right):

$$\begin{aligned}
\tau_{inh}\frac{dx_r(t)}{dt} &= -x_r(t) \qquad \text{if:} \quad t \neq t^{pre} \\
x_r(t) &\rightarrow x_r(t) + 1 \quad \text{if:} \quad t = t^{pre} \\
\\
\tau_{inh}\frac{dx_o(t)}{dt} &= -x_o(t) \qquad \text{if:} \quad t \neq t^{post} \\
x_o(t) &\rightarrow x_o(t) + 1 \quad \text{if:} \quad t = t^{post}.
\end{aligned} \tag{5.6}$$

The inhibitory weight is potentiated when the postsynaptic excitatory neuron spikes:

$$w(t) \longrightarrow w(t) + \eta x_r \qquad \text{if:} \quad t = t^{post}, \tag{5.7}$$

where η is the learning rate. After every postsynaptic spike, the inhibitory weight is increased, and this increase is the higher the more presynaptic spikes occurred. When a presynaptic inhibitory interneuron spikes, the weight is updated as:

$$w(t) \longrightarrow w(t) + \eta(x_o - 2\rho_0\tau_{inh}) \quad \text{if:} \quad t = t^{pre}. \tag{5.8}$$

This update defines the target of homeostasis through the inhibitory STDP, with $2\rho_0\tau_{inh}$ as the target firing rate of the postsynaptic excitatory neuron. If the postsynaptic spike detector (x_o) is below $2\rho_0\tau_{inh}$, the weight is depressed (Fig. 5.2, right). This is the case when the time between the last postsynaptic spike and the occurring presynaptic spike is long, indicative of a low postsynaptic firing rate. If the postsynaptic rate was high, and x_o is high, the inhibitory weight potentiates.

Connectivity in the network develops with all these three plasticity rules acting in parallel. Hebbian triplet STDP leads to potentiation of excitatory synapses between neurons with correlated spiking. In contrast to STDP that is only based on pairs of spikes, the triplet STDP also allows the simultaneous strengthening of reciprocal synapses between two cells, necessary for the emergence of Hebbian assemblies [166]. Normalization of excitatory synapses prevents runaway potentiation of excitatory synapses and induces competition between synapses that connected highly correlated neurons vs. synapses that connect weakly correlated neurons. Inhibitory STDP homeostatically regulates firing rates in the network and ensures inhibition that stabilizes the dynamics of the network. I used structured external inputs to train connectivity.

5.1.2 Inputs and training

Input connections from LGN and BKG are static in the model. Excitatory neurons receive inputs from the LGN and BKG sources, while PV-interneurons are exclusively driven by strong LGN-inputs. SST receive exclusively BKG inputs (Fig. 5.1A).

All neurons receive independent spike-trains with Poisson statistics, where spike trains from both the LGN and BKG have a rate R_x in the baseline state throughout the simulation (i.e., before, during, and after training). Training is implemented through subsequent activations of assemblies. When an assembly is activated during training, the rates of both inputs from LGN and BKG to an excitatory neuron in one assembly are increased by R_{act} (receiving a total rate of $R_x + R_{act}$, Table 5.1). External inputs to all inhibitory neurons of either subtype are increased during such stimulation epochs in a balanced fashion, such that the relation of the mean inputs to the overall excitatory population and the inhibitory populations is preserved. Thus, during stimulation epochs, the rate of external input to the interneurons is:

$$R_{ext}^{P/S} = R_x + \frac{R_{act}}{M}, \tag{5.9}$$

where $M = 20$ is the number of assemblies in the excitatory population. I defined an assembly by randomly selecting 5% of the excitatory neurons for all M groups (Fig. 5.1C, top). The members of the assemblies defined through LGN-inputs are chosen randomly for each assembly and independently across assemblies. This procedure leads to some random overlap of the assemblies. Most neurons belong to one single assembly, some belong to multiple assemblies, while others belong to none.

During training, single assemblies are activated for one second. After a break of one second, where no group is specifically driven, and only baseline input is delivered,

Table 5.1: Neuron parameters in the network with multiple interneuron subtypes undergoing STDP.

Symbol	Value/	Description
C_m	300 pF	Membrane capacitance
g_L	15 nS	Leak conductance
E_L	-66 mV	Leak reversal potential
V_θ	-52 mV	Threshold potential
V_r	-60 mV	Post-spike reset potential
E_E	0 mV	Excitatory reversal potential
E_P, E_S	-75 mV	Inhibitory reversal potential from PV (P) and SST (S) interneurons
τ_{ref}	5 ms	Absolute refractory period
τ_{syn}	2 ms	Synaptic conductance time constant
N_E	4000	Number of excitatory neurons
N_P	1000	Number of PV-interneurons
N_S	500	Number of SST-interneurons
p	0.1	Connection probability between two cells
M	20	Number of assemblies in the network
ε	0.05	Probability of any excitatory neuron to belong to any assembly
R_x	2 kHz	Rate of the external input outside a stimulation phase
R_{act}	2 kHz	Rate increase of the external input during a stimulation phase

Table 5.2: Connection weights and plasticity parameters in the network with multiple interneuron subtypes undergoing STDP. Top: Synapse weights, both initial and the range for learning if applicable. Middle: Parameters for the triplet STDP of excitatory synapses. Bottom: Parameters for the symmetric inhibitory STDP.

Symbol	Value/ Range	Description
Weights		
J_{EE}	$[0.1, 10]$ nS	Interval of weights between excitatory (E) neurons, undergoing triplet STDP and normalization
J_{EE}^{init}	1 nS	Initial weight of connections among E
J_{EP}	$[0.1, 150]$ nS	Interval of PV$\rightarrow E$ inhibition, undergoing inhibtory STDP
J_{EP}^{init}	15 nS	Initial weight of connections PV$\rightarrow E$
J_{ES}	$[0.1, 105]$ nS	Interval of SST$\rightarrow E$ inhibition, undergoing inhibtory STDP
J_{ES}^{init}	15 nS	Initial weight of connections SST$\rightarrow E$
J_{PE}	1 nS	Weights of $E \rightarrow$PV connections, not undergoing STDP
J_{SE}	0.5 nS	Weights of $E \rightarrow$PV connections, not undergoing STDP
J_{PP}, J_{PS}	15 nS	Weights of PV$\rightarrow$PV and SST$\rightarrow$PV
triplet STDP		
τ_+	16.8 ms	Time-constant of the presynaptic spike detector, trace for spike-pair detection (r_1)
τ_-	33.7 ms	Time-constant of the postsynaptic spike detector, trace for spike-pair detection (o_1)
τ_x	101.0 ms	Time-constant of the presynaptic spike detector, trace for spike-triplet detection (r_2)
τ_y	125.0 ms	Time-constant of the postynaptic spike detector, trace for spike-triplet detection (o_2)
A_2^+	7.5^{-10} nS	Amplitude of spike-pair based potentiation
A_2^-	7.0^{-3} nS	Amplitude of spike-pair based depression
A_3^+	9.3^{-3} nS	Amplitude of spike-triplet based potentiation
A_3^-	2.3^{-4} nS	Amplitude of spike-triplet based depression
inhibitory STDP		
τ_I	20 ms	Time-constant of the pre- and and the postsynaptic spike detector
η	1	Learning rate
ρ_0	1 Hz	Target rate of the postsynaptic cells

another randomly chosen assembly receives increased input (Fig. 5.1C, bottom, and Fig. 5.1D). This is repeated until all assemblies were driven once, defining one trial. The training consists of 15 trials, with a total of 300 stimulation epochs and 300 breaks in-between. The whole simulation consists of three stages. (1) In an initial period before training, I let the network activity settle for five seconds. (2) I then trained the networks for 600 seconds. (3) After training, the network is simulated for another 600 seconds to collect data about the spiking statistics and development of the connectivity without training.

In a first step, LGN and BKG assemblies are the same, with the same neurons contained and input-rate increase at the same times (Sec. 5.2.1). This is equivalent to highly cooperative feedforward inputs from LGN and feedback (BKG) inputs from higher-order areas (comparable, e.g., to perfect predictions of the neurons that an upcoming sensory stimulus will drive). From a pure modeling perspective, it is equivalent to train with just one structured input source. In a second step, I varied the spatial and temporal relationship between the two input structures (Sec. 5.2.2). The spatial relation should be understood in terms of the neurons belonging to one group. I compared the case with the exact same groups for LGN and BKG inputs (as outlined above) with a case where groups from LGN and BKG are chosen independently, such that the overlap of the groups is at chance level. For both of these cases of spatial overlap (perfect spatial correlation or no spatial correlation, i.e., random overlap), I also varied the temporal relation. When defining the same groups from LGN and BKG, we may drive them simultaneously to induce strong cooperation. In contrast, one can also drive them at independently chosen random times to induce temporal competition. I study all four combinations of spatial and temporal perfect correlations vs. no correlations.

All simulations are performed using custom code from the Lab written in *Julia*, originally adapted for multiple interneuron subtypes by Marcel Jüngling. I analyzed connectivity and activity in the network using *Python*.

5.2 Training Hebbian assemblies in a network with multiples subtypes of interneurons

5.2.1 Emergence of Hebbian assemblies in networks with multiple subtypes of interneurons

When LGN and BKG inputs cooperate, the excitatory strengths (or weights) within the assemblies strongly increase during training (Fig. 5.3A, black). Due to normalization of the total excitatory strength each neuron receives (Eq. 5.5), the average excitatory connection weights between neurons in different assemblies decrease (Fig. 5.3A, gray), while the overall mean excitation is kept constant (Fig. 5.3A, blue). Note that the decrease of inter-assembly weights is much less pronounced because the decrease affects connections from all neurons onto the neurons of one assembly, i.e., from all neurons in the other 19 assemblies, and the neurons that do not belong

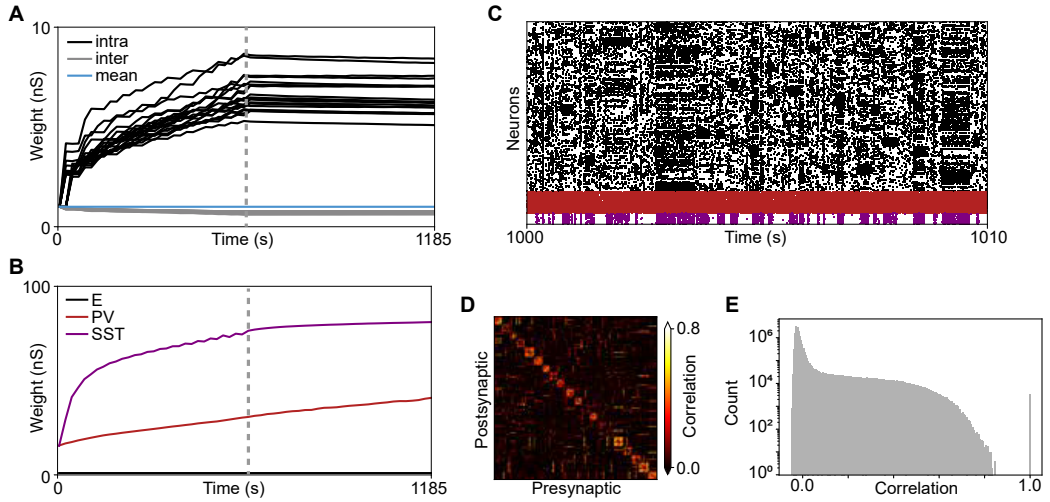

Figure 5.3: Trained Hebbian assemblies in a network with multiple interneuron subtypes (PV and SST). **(A)** Mean connection strengths between neurons within an assembly (intra, black), between neurons in different asseblies (inter, gray) and overall average excitatory strength (mean, blue). Light-gray dashed line: end of training. **(B)** Average inhibitory connection strength onto all excitatory neurons from PV-interneurons (red) and from SST-interneurons (purple). **(C)** Raster plot of activity after training of excitatory neurons (black), PV-interneurons (red) and SST-interneurons (purple). **(D)** Matrix of cross-correlations between assembly neurons, sorted by assembly membership. **(E)** Histogram of pairwise-correlations of spike trains of excitatory neurons, logarithmic y-scale.

to any assembly. When training ends, connection strengths do not quickly return to the level from before training but remain strong with only a modest reduction. Thus, stable assemblies can be imprinted into the network.

Inhibitory projections from either subtype strongly potentiate during training (Fig. 5.3B). This potentiation is driven by the high rates of excitatory neurons during stimulation epochs, although the rates are, on average, very close to the target rate. Interestingly, projections from SST-interneurons steeply increase to a high-level during training and then saturate (Fig. 5.3B, purple), while projections from PV-interneurons increase more slowly but keep increasing throughout the simulation (Fig. 5.3B, red). In a test-simulation running for longer, I find that also PV-projection strengths saturate after long enough time, without affecting the conclusions presented here.

The continued increase of PV-projection weights after the end of training is driven by transient increases of the spontaneous rates. These emerge from spontaneous reactivations of the imprinted assemblies, which also stabilize the excitatory connectivity within assemblies (Fig. 5.3C). Biologically, these reactivations of imprinted assemblies resemble the spontaneous correlations between cortical neurons which are correlated also during stimulus evoked activation [164]. Grouping the neurons by

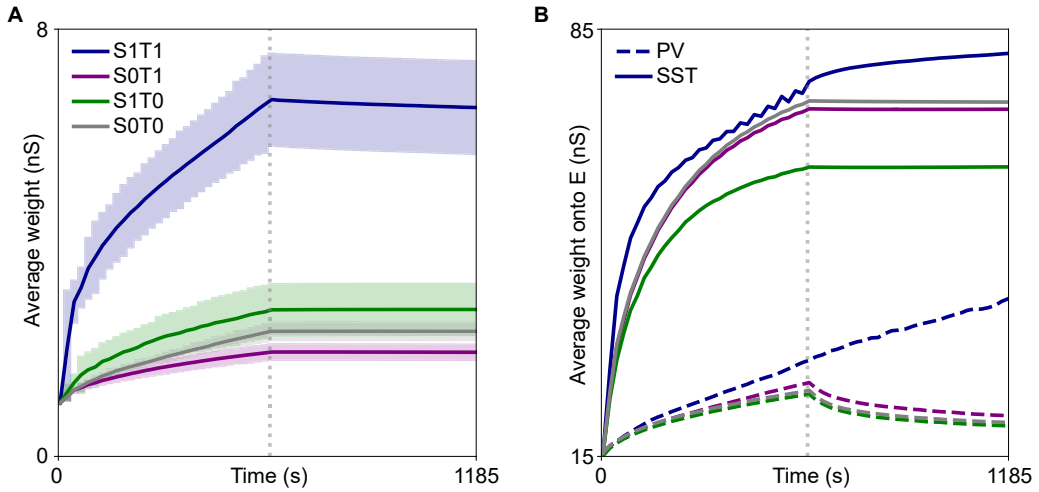

Figure 5.4: Emerging connectivity with competing assembly structures. **(A)** Mean connection strength of neurons within an assembly, averaged over all LGN-assemblies. Shaded area: standard deviation of the mean connection strengths within the different assemblies. Different training cases with perfect spatial and temporal correlations of the LGN and BKG input groups (S1T1, blue, compare with Fig. 5.3A), with no spatial correlations but perfect temporal correlations (S0T1, purple), with perfect spatial correlations but no temporal correlations (S1T0, green) and with neither spatial nor temporal correlations (S0T0, gray). Note that I only show intra-assembly weights here. Inter-assembly weights decrease by a small amount according to the heterosynaptic normalization (as shown in Fig. 5.3A). **(B)** Average inhibitory connection strength onto all excitatory neurons from PV-interneurons (dashed lines) and from SST-interneurons (full lines). Colors same as in (A).

their stimulus preferences, in the model represented by assembly membership, the matrix of instantaneous cross-correlations has a block-diagonal shape (Fig. 5.3D).

As a consequence of these reactivations of assemblies, inducing high correlations between neurons within the same assembly while neurons in different assemblies fire primarily asynchronous, the distribution of correlations has a long tail towards few strong correlations (Fig. 5.3E). However, it is not bimodal as one might expect from the matrix in Fig. 5.3D.

5.2.2 Competing inputs perturb the emergence of reliable spontaneous reactivations

In the next step, I studied the emerging connectivity and activity in the network when the two input streams are not perfectly correlated and, therefore, potentially compete rather than cooperate. Competition between the assembly structures from the LGN- and the BKG-inputs can be induced in two ways, spatial or temporal. I chose the spatial or temporal correlations either at zero (no spatial overlap of assemblies; temporal co-activation of them at chance level) or one (identical assemblies; identical

times of activation). Removing only temporal correlation of inputs leads to variable combinations of assemblies belonging to the same set that are driven simultaneously during training. Removing only the spatial correlations leads to fixed associations of independently defined assemblies. Removing both spatial and temporal correlations leaves no comprehensible relation between the two input streams apart from the internal consistency within each input stream. Including the case in the previous section, which I also consider for comparison here, I thus have four combinations. To denote these different cases, I used the shortcuts: S1T1, S0T1, S1T0, and S0T0, where S denotes the spatial component and T the temporal component, with the numbers pointing to the correlation value.

Training the network in these different scenarios, I found that some increase of the connection strengths among neurons within an assembly is visible in all cases (Fig. 5.4A). I calculated the mean of excitatory connections within each of the assemblies (as in Fig. 5.3A), and then plotted the mean and standard deviation of these assembly means. I did this calculation with respect to the assemblies defined by the LGN-inputs. It is equivalent for the assemblies defined by BKG-inputs.

Although the excitatory weights increase within assemblies in all cases, the case S1T1 stands out in developing by far the strongest connections (Fig. 5.4A, blue). Overall, the other three cases show comparable increases of these strengths, with some slight quantitative differences. From the other three cases, the case S1T0 develops the strongest intra-assembly weights (Fig. 5.4A, green), while S0T1 develops the weakest (Fig. 5.4A, purple). Thus, it appears that the spatial overlap is more important for the development of strongly connected Hebbian assemblies than the consistency of their temporal co-activation. Interestingly, when independent assemblies are driven at independent times (S0T0), the weights are in between these cases, suggesting that removing the temporal correlation from spatially uncorrelated assemblies reduces their competition (Fig. 5.4A, gray).

The development of the inhibitory projections from SST-interneurons does not show any qualitative differences, though the overall level of these weights is different for the different cases. In all of them, the potentiation of SST-projections onto excitatory neurons is largely established by the end of the training (gray dashed line) and remains constant or saturates soon after the end of training (Fig. 5.4B, full lines). In contrast, inhibitory projections from PV-interneurons show some dynamics also after the training. Specifically, again the case S1T1 is distinct with PV-weights further increasing after training (Fig. 5.4B, blue dashed), while in all other three cases, the PV-weights decrease after the training ended (Fig. 5.4B, green, purple and gray dashed).

This decrease of PV-weights might be connected to differences in the spontaneous activity after training. While in the S1T1-case, very strong reactivations of the imprinted assemblies dominate the spontaneous activity (Fig. 5.5A, left, see also Fig. 5.3C), the rasters of all other three cases appear much more homogeneous, largely asynchronous irregular (Fig. 5.5B-D, left), with only some occasional reactivations in the cases S0T1 (Fig. 5.5B) and S1T0 (Fig. 5.5C). Explicitly plotting the rates of the different assemblies confirms the impression from the raster (Fig. 5.5A-D, right).

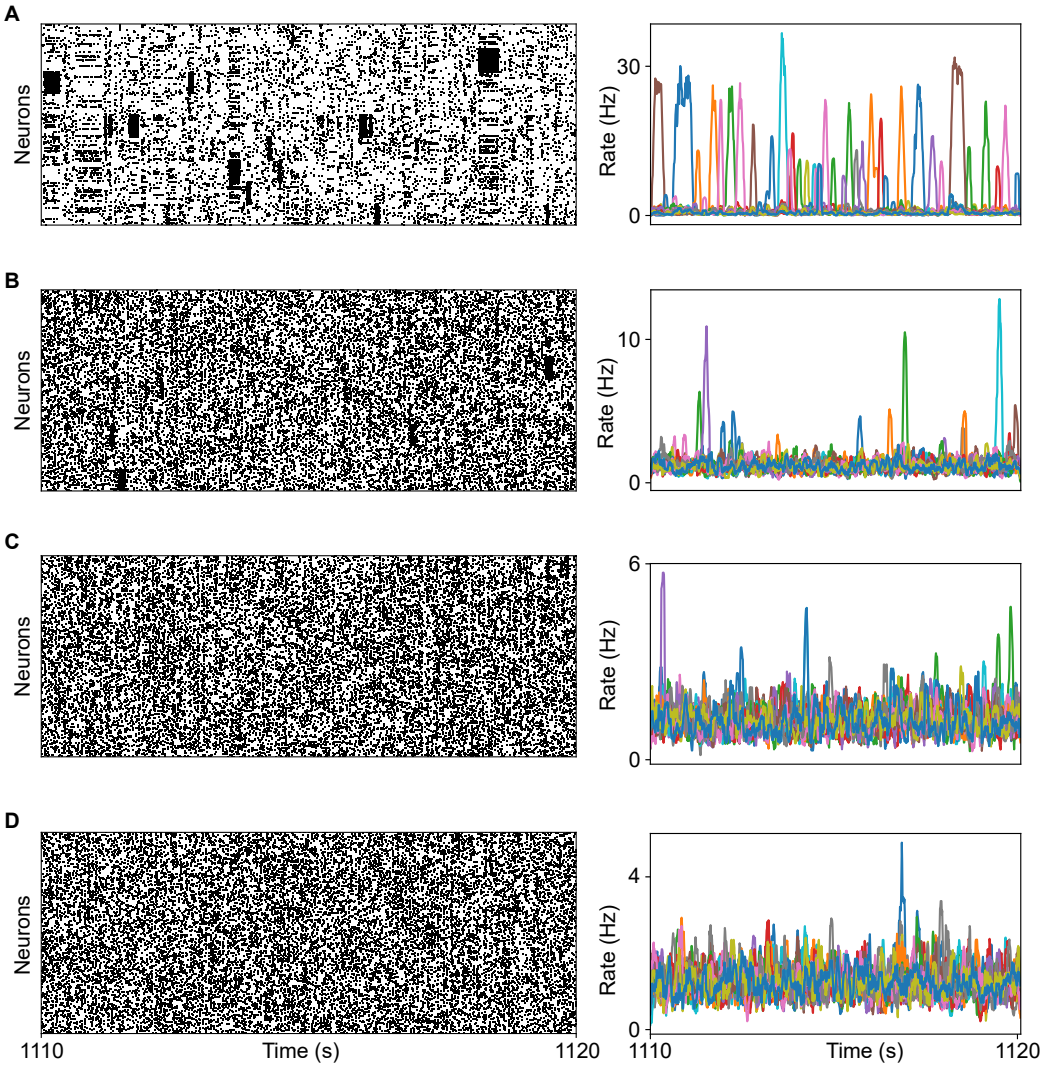

Figure 5.5: Emerging activity with competing assembly structures. **(A)** Left: Raster plot of spontaneous activity of excitatory neurons from nine assemblies (sub-selected for better visibility) in the network with perfect spatial and temporal correlations of the LGN and BKG input groups (S1T1). Right: instantaneous rates of all different assemblies during spontaneous activity, colors indicate different assemblies. Assemblies shown are chosen according to LGN-inputs. Assemblies according to BKG-inputs behave equivalently. **(B)** Same as (A) in the network with no spatial correlations but perfect temporal correlations (S0T1). **(C)** Same as (A) in the network with perfect spatial correlations but no temporal correlations (S1T0). **(D)** Same as (A) in the network with neither spatial nor temporal correlations (S0T0). Note the different ranges of rates in the different cases (right plots). In the S1T1-case (A, right), single assemblies strongly activate up to 30 Hz and are clearly distinguished from random fluctuations, while in the S0T0-case (D, right), rates only fluctuate a few Hertz around a steady rate.

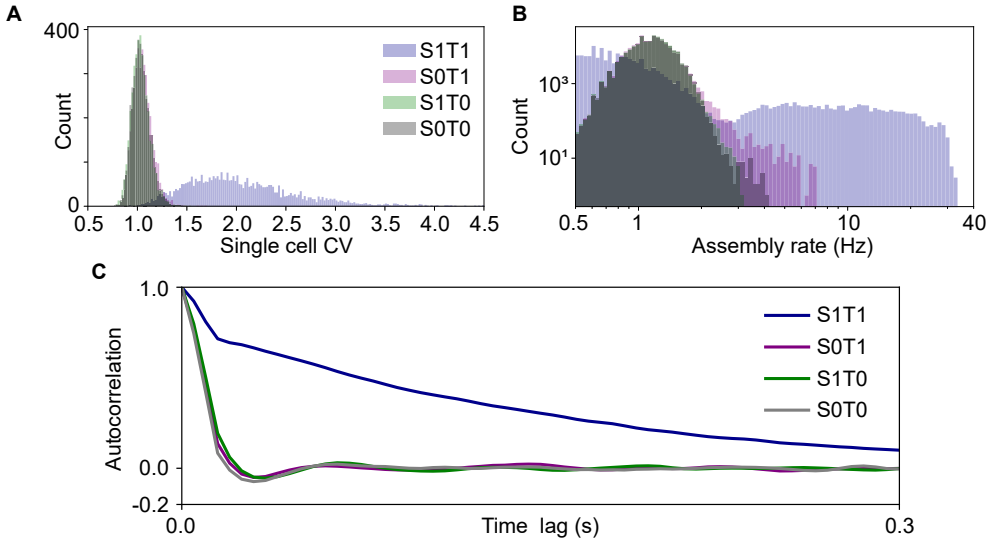

Figure 5.6: Statistics of spontaneous activity with competing assembly structures. **(A)** Histogram of the coefficients of variation (CV's) of the single neurons in the network with perfect spatial and temporal correlations of the LGN and BKG input groups (S1T1, blue), in the network with no spatial correlations but perfect temporal correlations (S0T1, purple), in the network with perfect spatial correlations but no temporal correlations (S1T0, green) and in the network with neither spatial nor temporal correlations (S0T0, gray). The three histograms S0T1, S1T0 and S0T0 lie on top of each other and have much smaller CV than S1T1. **(B)** Histogram of the smoothened assembly rates (compare right panels in Fig. 5.5). Both x- and y-scales are logarithmic. Colors same as in (A). **(C)** Autocorrelation-spectrum of the smoothened assembly rates in the different training scenarios. Colors same as in (A). The lines for S0T1, S1T0 and S0T0 lie very close to each other and show a much faster decay of the autocorrelation than S1T1.

I have plotted the mean rates of the different assemblies, derived from the spike trains binned with a time window of 5 ms and convolved with a box-shaped filter of 15 ms width.

Interestingly, the reactivations are more pronounced in S0T1 than in S1T0, although the mean of the excitatory weights in S0T1 is smaller than in S1T0. When the assemblies are defined independently for LGN- and BKG-inputs, but pairs of LGN- and BKG-assemblies are co-activated consistently (S0T1), the excitatory neurons are effectively embedded in larger assemblies. Thus, even though the mean is smaller in this case, the total excitation impinging on each neuron from within this larger, combined assembly can still be stronger. Furthermore, inhibition from SST-interneurons follows a different trend: it is smaller in S0T1, where assemblies are more strongly co-activated, than it is in the S1T0-case. SST-inhibition likely also contributes to the suppression of spontaneous rate increases that occur with the assembly reactivations. Together, these two factors could explain the more pronounced spontaneous reactivations in S0T1.

To analyze the properties of the emerging activity in more detail, I first calculated the coefficient of variation (CV) of the single neurons' inter-spike intervals in the different cases (Fig 5.6A, left). To calculate the CV, I extract all inter-spike intervals of one neuron i and calculate the mean (μ_i) and standard deviation (σ_i) of the histogram of all inter-spike intervals. The CV is the ratio of the standard deviation and the mean (σ_i/μ_i). Doing this for each excitatory neuron independently, I extracted a histogram of the CVs in the network. In the asynchronous irregular regime, the spike trains of single neurons have Poissonian statistics, which implies that the CV of the inter-spike intervals in a single neuron is close to one[1]. If a neuron fires in a more bursting fashion, that is, it has long breaks between spikes at some times and very fast sequences of spikes at other times, the CV of the inter-spike intervals is above one. The presence of assemblies that switch between a baseline state of low rate, and an activated state with high rates (visible in Fig. 5.5A), could induce such a super-Poissonian CV. Indeed, in the S1T1-case, the CVs of the different neurons in the network distribute around a value clearly above one (2.058 ± 0.578, mean±standard deviation of the histogram of CVs, Fig 5.6A, blue). In contrast, the lack of substantial reactivations of assemblies in the other three cases entails that the CVs of the neurons take values close to one, without clearly visible quantitative differences between the different cases (mean ± standard deviation of the histograms of the CVs for the different cases: S0T1: 1.053 ± 0.098; S1T0: 1.033 ± 0.092; S0T0: 1.036 ± 0.091).

I further confirmed the underlying bimodal rate distribution using the histogram of the assembly rates. Taking the convolved rates (as shown in Fig 5.5, right) and extracting the histogram of the different rate values each assembly takes over time, I find bimodality in the S1T1-case, while in the other cases, the rates have a unimodal distribution (Fig 5.6B).

The bimodal rate distribution is a consequence of the bistability of the assembly activations. The activated and the inactivated state of an assembly are stable states of this assembly [138]. Assemblies in the network would stay in one of these two states indefinitely unless they are pushed towards the other. In the finite-size networks I simulated, this push emerges from the fluctuations within the network, thus inducing a state of metastability, with abrupt jumps between the activated and inactivated states emerging spontaneously during ongoing, spontaneous activity [169]. Such metastable dynamics induces a slow autocorrelation timescale [138]. I found a marked increase of the autocorrelation timescale of the convolved firing rates in the S1T1-case (Fig 5.6C, blue) in comparison to all other cases, which have virtually indistinguishable timescales and show a much faster decay of the autocorrelation of the rates.

In conclusion, even though the connectivity shows signs of strengthened connectivity underlying Hebbian assemblies in all the different training cases, the evaluation

[1] An alternative measure is the Fano-factor of the spike-counts in a time-window of a given size, which is also one for Poissonian statistics. This is, for example, used in [138] to quantify variability of rates in a network with Hebbian assemblies in different conditions.

of the activity only confirms a dichotomy between a scenario in which feedforward inputs from the LGN and feedback inputs from BKG strongly cooperate (S1T1) and the other cases where they either compete in the spatial or in the temporal domain, or in both.

5.3 Discussion

Little is known about plasticity in highly recurrent spiking networks with multiple interneuron subtypes. I found that stable, self-activating assemblies can emerge in a network with multiple interneuron subtypes through training (Fig. 5.3), extending previous work carried out with a single subtype of interneurons [30, 32, 160]. The inhibitory STDP serves as a controller of the excitatory firing rates. Therefore, in the network where synapses from two interneuron subtypes follow this STDP rule, the rate is regulated by two, potentially competing, controllers. This is a new aspect of this model, and it is not clear *a priori* if stable connectivity and activity can be achieved when both PV and SST synapses undergo STDP that is tailored to homeostatically control firing rates such as the symmetric rule I used [168] (Fig. 5.2). To avoid an artificially unstable situation, I used the same target rate for synapses from both interneuron populations. This is also in accordance with the observation that excitatory neurons have a firing rate set-point [89]. Based on this, I would argue that the target rate should be defined postsynaptically, at the level of the excitatory neuron, because the alternative would be that a single interneuron somehow signals a target rate for each of its synapses made onto different excitatory neurons. With such matching target firing rates, stable activity is reached in the network.

I also followed this approach, to use the same plasticity for synapses of either inhibitory subtype, in order to reduce the difference between PV- and SST-interneurons to their coupling in the circuit and, crucially, the sources of their external inputs (Fig. 5.1). However, even with the same target rate, the interaction of the two controllers (plastic synapses from PV- and SST-interneurons) could induce instability of the connection weights, which could, for example, oscillate, especially if the controllers act at different timescales. In the simulations, I see stable average inhibition emerging. In future work, however, it might be vital to investigate this problem more carefully, as it may be necessary for biological reasons to investigate different parameters of the STDP in synapses from PV- and from SST-interneurons, or even different plasticity rules in the two cell classes [170–172].

In conjunction with the diversity of interneurons, also the diversity of input streams driving cortical neurons becomes important, since interneuron subtypes participate in subtype-specific ways in these different signaling streams [61]. In my model, I introduced two input streams representing feedforward inputs from the LGN and a general input stream of feedback information (BKG). I found pronounced self-activating assemblies only in the case when these two input streams cooperate, such that the same groups of neurons are driven at the same points in time during training (Fig. 5.5). Competition between the two input sources reduces the average

connection strengths within assemblies. Connections between neurons within the same assembly (either defined from LGN- or BKG-inputs) still potentiate above the baseline level, but this potentiation is considerably weaker than when the two input sources cooperate (Fig. 5.4).

In the cases that include competition, I could not find marked statistical signatures of spontaneous self-activations of the imprinted assemblies (Fig. 5.6). A direction one could take from here is to study the emerging structure with more subtle competition, for example, using intermediate cases between perfect overlap and complete statistical independence of the assemblies driven by LGN and BKG. It could also be necessary to use more subtle statistical measures to distinguish the cases I have studied so far and to scrutinize their dynamical differences. My investigations of the CV's of single cells in the network and the temporal statistics of the rates of assemblies only showed a dichotomy between the case where the two inputs streams cooperate with a perfect spatial and temporal correlation, and all other cases where the input streams either compete in space or in time (Fig. 5.6).

In recent work, Lagzi and colleagues investigated how the presence of PV- and SST-interneurons influences the formation and maintenance of assemblies with plasticity observed in mouse orbitofrontal cortex, and compared this to a circuit that only includes PV-interneurons [172]. The experimental results in this study showed that synapses from PV-interneurons have plasticity according to the symmetric STDP rule, whereas synapses from SST-interneurons follow an antisymmetric STDP in this part of the frontal cortex. In their model with multiple subtypes, they introduced assembly specific SST-interneurons and global feedback from PV-interneurons. Based on the plasticity rules and the induced correlation structure following from the connectivity, Lagzi and colleagues could show that SST-interneurons develop lateral inhibition between the assemblies, while PV-interneurons provide homeostasis of firing rates. Interestingly, the lateral inhibition from SST-interneurons drove a differentiation of the PV-population into PV-subpopulations with outputs tuned to different assemblies [172].

In my approach, I did not specify SST- or PV-interneurons to belong to any specific assembly. I rather studied how the "collision" of LGN- and BKG-inputs in the excitatory population shapes the emerging assembly structure, both in terms of connectivity and spontaneous activity. The aspect of how feedforward inputs, representing sensory drive, and background inputs with assembly structure that could, e.g., represent predictions of sensory inputs from higher-order cortical or thalamic centers, interact, was not investigated in [172]. In future work it would be most interesting to merge these aspects in a single model, where the single interneuron subtypes could receive assembly-specific inputs that conform with the sources of their primary inputs (LGN vs. BKG), and investigate how connectivity motifs of local or lateral inhibition and disinhibition may emerge and shape spontaneous dynamics and processing capabilities of the network.

6 Discussion and outlook

In my research, I investigated the interplay of strong recurrent connections in networks, especially the role of the operating regime, the multiplicity of interneuron subtypes, and synaptic plasticity. I found that understanding this interplay is vital to apprehend how deprivation-induced plasticity shapes activity in visual cortex, and especially identified the paradoxical effect as one of the fundamental mechanisms. The role of the paradoxical effect for computations in adults has already been recognized (reviewed in [133]). However, it has not been investigated in the developing cortex with its exceptionally high plasticity in response to sensory experience. In this chapter, I first summarize the main results and their relation to previous results from the literature, with a focus on the paradoxical effect and the interaction of multiple interneuron subtypes (Sec. 6.1). I will then outline the limitations of my model concerning some important biological mechanisms which I have not integrated into the network model and possible implications of these mechanisms (Sec. 6.2). I will then discuss future directions to investigate the role diverse interneurons play in shaping cortical computations, specifically PV and SST, and experience-dependent plasticity (Sec. 6.3). Finally, I conclude with some thoughts on the larger context into which this work fits (Sec. 6.4).

6.1 Interneuron diversity and deprivation-induced plasticity

The focus during my doctoral research was the regulation of activity in cortical circuits undergoing sensory deprivation and how the activity regulation can be generated by deprivation-induced plasticity (Chapters 3 and 4). The existing connectivity of the model networks strongly interacts with the specific synaptic changes taken from the experimental study of the visual cortex during monocular deprivation (MD), an important paradigm for the study of plasticity during experience-dependent brain development. Additionally, I also considered the effects of plasticity in the somatosensory cortex induced by whisker deprivation (WD). Through simulation and analysis, I was able to uncover how the paradoxical effect with a single or with multiple subtypes of interneurons governs the possible responses in circuits to experimentally identified plasticity:

1. In an ISN with a single subtype of interneurons, the rate changes generated by MD-induced plasticity are tightly coordinated between the excitatory and PV-interneurons. This tight coordination impedes the independent modulation of their firing rates observed *in vivo*, where inhibitory rates drop on the first

day of MD and recover on the second, whereas excitatory rates stay at their baseline level on the first day and only drop on the second day [54].

2. In a non-ISN with a single subtype of interneurons, independent modulation of firing rates is possible. However, the weak recurrent coupling causes much-reduced responsiveness of this network to the MD-induced synaptic changes.

3. Sufficiently strong feedback from SST-interneurons in an ISN with two subtypes of interneurons leads to a reversal of the paradoxical effect in PV-interneurons and an inversion of the responses of PV-interneurons to MD-induced plasticity. This means that excitatory and PV firing rates can be modulated in opposite directions and can thus be temporally modulated in the same sequence as measured experimentally after MD.

4. Analysis of a linear population-rate model proves that the coordination between excitatory and PV neurons is driven by the emergence of the paradoxical effects. Opposite modulation becomes possible through the reversal of the paradoxical effect in PV-interneurons through SST-feedback.

5. Furthermore, this analysis shows that the reversal of the paradoxical effect is always possible in a stable network, with finite SST-feedback strength sufficient in networks with arbitrarily strong couplings among excitatory and PV neurons.

6. Although brief WD invokes plasticity mechanisms different from those invoked by brief MD, my model shows similar regulation of activity in both cases. Specifically, I found tight coordination of excitatory and PV firing rates in an ISN with a single subtype of interneurons and an inversion of the PV response with strong SST-feedback. Thus, I predict that the firing rates during ongoing WD are modulated in S1 in the same temporal sequence as firing rates in V1 during MD.

Based on these results, I can make testable predictions about the circuitry and population responses during MD. First, I predict that PV should respond non-paradoxically to stimulation during the critical period to robustly enable the independent modulation of excitatory and PV firing rates. This could be directly tested in animals during the critical period using optogenetic stimulation in the same way as in adult animals [135, 136]. Further analysis would need to clarify if such a potential non-paradoxical response of PV-interneurons would be caused by the operating regime, that is, if the cortical circuit is non-ISN during the critical period, or if it is caused by strong SST-feedback in an ISN. The distinction could be made by specifically investigating the inhibitory currents that impinge on an excitatory cell before and after optogenetic stimulation of interneurons [130] (see also Appendix A.2). Given that in the adult circuit PV-interneurons do show the paradoxical effect, either case would be a transient phenomenon of the developing circuit. Interestingly,

strong SST-feedback onto PV-interneurons has been experimentally identified in the developing cortex during the first postnatal week [152].

The network model shows close coordination of the activity changes in excitatory and SST neurons in response to deprivation-induced plasticity. Therefore, the model predicts that the same time-course of suppression and recovery in excitatory and SST neurons would be observed during brief MD. In recordings with extracellular electrodes, SST cannot be isolated easily, as their spike shape falls either into the fast-spiking class (overlapping with PV-interneurons) or the regular spiking class (overlapping with excitatory neurons), dependent on the morphological subtype of SST-interneurons [24]. Potentially, chronic calcium imaging could be used to observe the activity of labeled SST-interneurons during MD [80].

The comparison to the effects of plasticity induced by WD showed that the regulation of activity is very similar in both types of deprivation-induced plasticity. My model, therefore, predicts that activity during brief WD could be regulated similarly to MD. This similarity of activity changes upon WD and MD suggests a shared principle in both types of deprivation that serves the regulation of activity and the adaptation to changes of the sensory environment.

My research builds on existing results about the paradoxical effect in inhibition stabilized networks, especially [130] and related work [134, 136, 173]. My work contributes multiple novel aspects to our understanding of the role of the paradoxical effect and the interactions between multiple interneuron subtypes in cortical circuits:

- It is not *a priori* clear that the operating regime or the paradoxical effect should play a role in the dynamics of rate regulation uncovered by experiments during brief MD. The relations I found strengthen the view that paradoxical effects and the form of inhibitory stabilization present in the circuit are a fundamental property of the dynamics of neural circuits, appearing in multiple biological contexts.

- The response to a simultaneous perturbation of feedforward inputs to excitatory and inhibitory neurons has not been addressed by previous work on the paradoxical effect. I found that the general characteristic of the paradoxical effect that excitatory and inhibitory rates change in the same direction upon stimulation of the inhibitory neurons gives an intuitive understanding of the coordination of excitatory and PV firing rates. I have shown how this principle leads to a tight alignment in the whole plane of feedforward changes.

- Sensory deprivation severely affects recurrent connections in the circuit. Previous work on the paradoxical effect focused on the effects of perturbations of external (feedforward) inputs to inhibitory neurons. How the paradoxical effect impacts the response to these recurrent changes in the network has not been addressed by previous work. Interestingly, the paradoxical effect drives a similar alignment between excitatory and inhibitory rates in the recurrent as in the feedforward synaptic changes.

- My analysis includes a general understanding of the interaction of the overall coupling scale, the strength of PV-connections, and SST-feedback, that goes beyond the basic understanding of the paradoxical effect for a single specific inhibitory connectivity based on experimental data.

Inspired by the fundamental impact of interneuron diversity on the effects of deprivation-induced plasticity, I furthermore explored the emergence and maintenance of Hebbian assemblies in networks with multiple interneuron subtypes and the effects of these assemblies on the emerging spiking statistics (Chapter 5). The characteristics of the connectivity and spiking of the emerging assemblies when the network is trained with a single input structure are consistent with previous work on plasticity in large-scale spiking networks with a single subtype of interneurons [30, 32, 160]. However, competition between multiple input structures that serve as a model of the different input streams profoundly changes the emerging connectivity as well as spiking. Different interneuron subtypes engage in these different input streams in subtype-specific ways. This model requires further investigation, but it already shows that the interaction of multiple input streams in networks with multiple interneuron subtypes shapes circuits in specific ways. The consequences of this interaction for the circuit structure could underly different processing capabilities of networks in different regions that, on a coarse level, start with similar structures, but are shaped by different signals in experience-dependent development, for example, primary and secondary sensory areas.

6.2 Including further biological aspects of cortical circuits with diverse interneuron subtypes

Introducing the diversity of interneurons is one step to integrate the real complexity of cortical circuits into models. Adding SST-feedback to the prevalent models with a single subtype of interneurons fundamentally altered the response to MD-induced plasticity. Depending on the research question, there are several aspects of the biological circuit to be considered. Starting from my results and other results from the literature, crucial factors for future work on the effects of sensory deprivation could be: (1) the subcellular specificity of interneuronal targeting, especially of PV- and SST-interneurons, (2) additional interneuron subtypes, prominently VIP-positive interneurons, and (3) the implementation of plasticity mechanisms, such as spike-timing-dependent plasticity, homeostatic scaling, or intrinsic plasticity, that continuously shape the circuit. I briefly outline these aspects and their potential for future research.

6.2.1 Subcellular specificity of interneuronal targeting

Synapses from PV-interneurons preferentially target the soma or the axon-initial segment, whereas SST-interneurons target the apical dendrite. I have studied networks of point neurons, represented by a membrane potential that is equally

affected by all synapses independent of their potentially different positions on the postsynaptic neuron. The introduction of two interneuron subtypes could make a neuron structure that differentiates between the soma and the dendrite vital. While synapses directly on the soma or axon initial segment are in a position to strongly suppress the output of the excitatory neurons, dendritic synapses interfere already at the level of the inputs from other excitatory neurons that impinge on the dendrites. In the simplest scenario, the different placement of the synapses primarily influences the effective synaptic weight, which would be accounted for in my model through the variation of the different synaptic weight parameters (e.g., J for the coupling scale and K for SST-feedback). More likely, however, the active integration properties of dendrites will lead to more complex phenomena that shape the interaction of the activity with synaptic changes induced by deprivation [174]. Active dendrites induce strong nonlinearities in the integration and response of single neurons [175]. Although the balanced state entails linear responses of the network [113] (see also Sec. 2.5.1), nonlinear synaptic integration can lead to nonlinear network responses also in the balanced state [176]. Recent modeling work has shown that the nonlinearity in recurrent networks induced by nonlinear synapses allows networks to switch the operating regime between ISN and non-ISN without long-term changes of the synaptic weights [177]. Integrating synaptic nonlinearities in networks with diverse interneuron subtypes, including the subcellular specificity of their targeting in the context of MD, is, therefore, an interesting direction for future studies.

6.2.2 Increasing interneuron diversity

PV and SST are two from a much larger set of interneuron subtypes, albeit overall the two most numerous in cortex [24]. Apart from the much more elaborate distinction into morphologically defined subtypes, which entails multiple subtypes also for each of the interneuron subtypes defined by genetic markers, the third large class of molecularly identified cortical interneurons expresses the serotonin-receptor 5HT3a (abbreviated as 5HT3aR). Out of these, especially VIP-interneurons have been experimentally implied in cortical computations, e.g., in mediating disinhibitory higher-order feedback during locomotion or due to the context a signal is embedded into [62, 178]. This disinhibitory feedback is due to the fact that VIP-interneurons primarily engage in reciprocal inhibition with SST-interneurons but do not seem to directly project to PV-interneurons or excitatory neurons (Sec. 2.1.2). Thus, their activation inhibits SST-interneurons and through this likely disinhibit excitatory neurons. However, the exact effects in complex circuits still require further investigation.

A number of models with multiple interneuron subtypes have been set up introducing three classes of interneurons, PV, SST and VIP [130, 136, 173, 179]. I have followed a route of introducing the diversity of interneurons successively, building on a network with a single subtype, and scrutinizing the role of the addition of a second one (see also [172, 180, 181]). This enabled a detailed understanding of the changes of network responses that are brought about by a simple case of interneuron

diversity. The additional introduction of VIP-interneurons would be an important next step and could be done building on my results with careful consideration of the interaction of PV- and SST-interneurons already present in the circuit.

In the context of my research, it would be most interesting to investigate how the introduction of VIP affects the presence or absence of the paradoxical effect in PV-interneurons and how this plays out in the response to MD-induced plasticity. It could be the case that VIP-interneurons amplify the effects of SST, in which case my results for the role of interneuron diversity could stand largely unmodified. However, VIP-feedback that inhibits SST-interneurons could also interfere with the reversal of the paradoxical effect in PV through SST-feedback that I have found. A general investigation would require to study under which conditions VIP-interneurons prevent the reversal of the paradoxical effect due to SST-feedback. VIP-feedback could still be highly permissive to allow the reversal, as I have found it with two subtypes of interneurons. However, likely some conditions that prevent this reversal could exist in a circuit with three subtypes of interneurons. It should be scrutinized if these conditions can be fulfilled in the developing brain. If such conditions exist, this would not necessarily refute my results. Rather, an interesting scenario that enriches the repertoire of responses could emerge. The reciprocal inhibition between SST- and VIP-interneurons could give rise to a form of winner-take-all competition that allows a switch between different states in which either SST- or VIP-feedback dominate the response properties of the circuit [182], allowing the circuit to switch between states. How MD affects such a dynamical interaction of the subtypes needs to be determined.

Finally, it is important to point out the limitation through the parametric simplification I have used for the connections between excitatory neurons and diverse interneurons. I defined synaptic connections, their strength, and probability between the different populations in the network by a minimal number of parameters that allows to independently investigate the roles of the different interneuron subtypes. Biological networks are unlikely to follow such simple schematics and thus likely require systematic investigation, including more degrees of freedom in the connectivity. However, I believe that the study of simplified models can produce understanding that applies to a variety of neuronal circuits across different brain areas. The parametric study in a few dimensions allows us to abstract from the details of any particular circuit and to gradually uncover principles of the neural dynamics and computations emerging in brains that may be applicable across brain regions and species.

6.2.3 Including ongoing plasticity with biologically inspired mechanisms to study the trajectory of activity regulations

An important direction to extend my research would be to combine interneuron diversity and biologically inspired plasticity mechanisms, such as STDP in excitatory and inhibitory synapses, with a focus on MD. The deliberate modification of weights of non-plastic synapses in my model allowed me to explore the effects of plasticity

in an unbiased way, with any combination of weight changes in any of the different synapses (Chapters 3 and 4). Using this freedom of exploration, I specifically investigated how the rate modulations found *in vivo* can be achieved with the highest robustness, without a requirement for fine-tuning of the specific parameters. When using biologically inspired learning mechanisms, the configurations of the weight changes are much less under control.

Such mechanisms have been investigated with a single subtype of interneurons by Wu et al. [90]. Wu and colleagues focused on the long-term homeostasis of firing rates without modeling the detailed regulation of rates during the first two days of MD. They also found that correlations drop during early MD and are homeostatically regulated back to their baseline after prolonged MD, restoring not only average levels of correlation but also detailed correlation structure [90]. Extending my work, it would be interesting to study the specific trajectory of modulations of the rates and correlations during the first two days of MD and how this could be generated by biologically inspired mechanisms. The relations between the potential regulations of activity and the interaction between multiple interneuron subtypes that I investigated could be combined with the insights from [90]. Specifically, Wu and colleagues found that homeostasis of rates depends on intrinsic plasticity of excitatory neurons, while the recovery of correlation structure depends on synaptic scaling.

Ma et al. [91] used biologically inspired plasticity mechanisms to generate the sequence of firing rate changes. They found that the plasticity parameters to achieve this are tightly constrained. Thus, along the lines of my research to find the network structure that allows the opposite modulation of firing rates in wide ranges of the parameter spaces of MD-induced plasticity, finding a network that more robustly could generate the experimentally observed regulation of rates than the one investigated by Ma et al. with ongoing plasticity seems desirable. The counter-proposal would be that a multitude of interacting plasticity mechanisms exactly achieves such fine-tuning of synaptic changes through their interactions [31].

The combination of previous work, especially [90, 91], with my results could guide the formulation of a model of the plasticity underlying the regulation of activity during early MD with biologically inspired plasticity mechanisms in a network with multiple interneuron subtypes. I have studied the action of biologically motivated plasticity mechanisms in a network with multiple subtypes of interneurons (Chapter 5). The learning of Hebbian assemblies in this network is not fundamentally different from the learning in a network with a single subtype of interneurons [30], specifically in the case where feedforward and background inputs drive the same groups of neurons at the same time (Sec. 5.2.1). Biologically-inspired, ongoing plasticity in a network with a defined operating regime could be used to investigate how plasticity leads to distinct rate regulation and homeostatic recovery. Beyond firing rates, it would be most interesting to understand how interneuron diversity affects the regulation of correlations.

6.3 Feedforward and feedback processing in local cortical circuits

6.3.1 The effects of deprivation on signal processing in cortical circuits

I have investigated how activity levels can generally be regulated in V1m through the MD-induced synaptic changes. A different direction would be to investigate the functional role of these synaptic changes. Specifically, how may the processing of incoming signals in a circuit change after MD-induced plasticity? My model showed that the increase of the feedforward E/I-ratio increases excitatory firing rates. Thus, a hypothesis would be that the plasticity of feedforward synapses could, despite its depressive nature, amplify the signals arriving in V1m after the eye is closed because it increases the feedforward E/I-ratio. Recurrent plasticity of the inhibitory pathways would then have a major homeostatic role, such that the combination of feedforward and recurrent plasticity allows amplification of signals while average activity levels are kept at normal levels.

Eyelid suture likely decreases the visual information due to blurred vision through the closed eyelid. This should not be misunderstood as a decrease of the average activity level of LGN-cells, as this is not the case [76]. An experimental hint at how this blurred vision affects visual signals is the decrease of the correlations between neurons in LGN after eyelid suture [76]. Correlations between spikes of thalamic neurons contain relevant information about the properties of visual stimuli [183], and, especially at the population level, such changes of the correlations can have large effects on the information encoded and on the best strategy to decode this information [184]. Also, within V1, average firing rates show little dependence on the general presence or absence of visual input as they are similar during prolonged periods of light and of darkness [185]. Transitions between light and darkness only lead to transient deviations. Similar to the effect of eyelid-closure in LGN, darkness decreased pairwise correlations compared to the light-condition. Both of these results point at a role of pairwise correlations in the encoding of visual scenes and their disruption through the removal of visual stimulation either through MD or in darkness.[1]

One way to investigate processing in this context is to study rate fluctuations rather than average activity levels that contribute to pairwise correlations. For example, *balanced amplification* is a fast, transient form of amplification of signals that emerges in the ISN-regime with strong recurrent excitation [137]. In an ISN with a single subtype of interneurons, strong and fast amplification of excitatory input pulses, or any brief fluctuation of the inputs to the excitatory population in favor of excitation, can be achieved, but it has not yet been investigated with multiple interneuron subtypes. Specifically, it would be interesting to investigate

[1] Interestingly, intra-ocular application of TTX *increased* pairwise-correlations in LGN in [76]. This could reflect a dominance of intrinsically generated activity when visual input is completely lacking compared to the case where it is reduced. The correlations of purely intrinsically generated activity could be a side effect of recurrent coupling or shared higher-order inputs.

how the reversal of the paradoxical effect in PV-interneurons affects the strength of amplification. More generally, it remains yet to be investigated how the interaction of multiple interneuron subtypes, in general, affects the amplitude and timescale of the amplification. In this framework, also the effects of MD-induced plasticity on the signal amplification could be investigated.

The recurrent inhibitory plasticity acting during MD could also contribute to a sharpening of the tuning of responses, both in healthy development with normal visual experience and after deprivation, when stimuli are blurred. In the study of MD-induced plasticity, Maffei and colleagues discovered a plasticity mechanism where potentiation of inhibitory synapses is induced by the combination of depolarization of the postsynaptic excitatory cell and firing of the presynaptic inhibitory cell, and this mechanism most likely underlies the potentiation of inhibition after brief MD [82]. The induction of this plasticity is furthermore prevented by coincident spiking of the pre- and postsynaptic cells. Thus, it is not primarily dependent on the relative timing of spikes as in STDP but resembles a voltage-dependent synaptic plasticity rule [186, 187]). Functionally, the induction protocol suggests that this plasticity can sharpen responses of excitatory neurons by increasing inhibition on cells that are weakly driven by some stimulus, such that they are depolarized below threshold but not enough to spike consistently. These weakly driven cells would be strongly suppressed after induction of the inhibitory plasticity until their spiking responses are completely suppressed. In contrast, the consistently spiking cells, those that are strongly selective to this stimulus, would not be suppressed. Together, these two processes would lead to sharply tuned excitatory responses. The blurring of the sensory signals seems a powerful paradigm to invoke this plasticity even more strongly than in normal visual experience.

6.3.2 Control of the mode of signal processing by specific interneuron subtypes

I have started to investigate the structure emerging in circuits with multiple interneurons subtypes shaped by multiple interacting plasticity mechanisms. A biological question one could investigate in this framework is how PV and SST interact to control the activity of assemblies with respect to the reliability of cortical responses. Recent experimental studies have started to uncover the role of interneuron subtypes in this context [188–191].

Agetsuma and colleagues showed that optogenetic suppression of PV-interneurons increases the overlap of different assemblies activated by different stimuli [188]. When PV-interneurons are suppressed, the responses to different stimuli are less distinguishable, leading to a decreased reliability in the representation of stimuli. These experiments did not directly investigate the role of SST, however, other experiments have shown that SST strongly suppress PV in different cortical areas [192, 193]. The suppression of PV through SST could have similar effects on the circuit dynamics *in vivo* as the optogenetic suppression of PV. Indeed, Rikhye and colleagues showed in a different context that the suppression of PV through

optogenetic activation of SST is sufficient in their case to produce the same effects as the direct optogenetic suppression of PV [189]. They investigated the direct interplay between the two interneuron subtypes in the control of trial-to-trial reliability. Optogenetic suppression of PV-interneurons increases the trial-to-trial reliability in response to natural movies, and the suppression of PV through optogenetic activation of SST yields the same effect.

These results point at an intricate interplay of PV and SST in controlling the mode of output. The circuit could be imagined as being either in a mode of well-distinguished but unreliable (over trials) assemblies, representing distinct stimulus patterns (high PV-activity), or in a mode of overlapping ensembles, less suited to distinguish patterns, but reliable over repeated encounters of the same pattern (high SST activity, with suppressed PV activity). The different influences of PV and SST on the discrimination of assemblies on the one hand, and the trial-to-trial reliability on the other, could resemble two modes of the faithful representation of the visual world: (1) strong discrimination between (similar) stimuli, which is achieved with non-overlapping ensembles emerging from PV activation, and (2) the reliable identification across encounters of the same visual scene (which will vary on each encounter), achieved with similar responses across trials produced by SST activation.

Why cortical circuits may implement these modes can be hypothesized from the input sources of the two inhibitory populations: PV are strongly driven by direct feedforward input, so their activity is tied to the dominance of thalamic drive. Such dominance of thalamic drive could push the network in a state of accurate representation, resembled by well-distinguished ensembles responding to different stimuli. SST in turn are strongly driven by lateral inputs, providing context to the local patch of visual input. Such contextual information could push the network in a state where it encodes a visual scene less accurately but more reliably [194]. Reliability across trials is relevant if the animal has prior information that contains what the behavioral consequences of a certain encounter should be, consistent with the observations that SST also receive higher-order feedback. Potentially, it is less important how exactly a scene looks when contextual information is available because this is already partially known, but rather what it means across encounters.

To study this in a model, one could compare the spontaneous activity and driven responses during activation and inactivation of either of the interneuron subtypes. The overlap of ensembles in these different conditions is accessible through the correlation structure of the network. Reliability could be accessed using an approach presented in [195], where it was shown that even in a network that globally appears chaotic and unreliable, intermittent periods of reliable single-cell firing coexist with periods of unreliable firing.

6.3.3 Control of tuning of single neurons by specific interneuron subtypes

The investigation of the effects of PV and SST activity in the control of cortical ensembles in a structured spiking network could also be further used to shed light

on the controversial issue of tuning of interneuron responses and the specificity of their connections. Both PV- and SST-interneurons have been found to have either tuned or untuned responses [196–198]. The same holds true for the connections PV-interneurons receive from local excitatory cells [199–202].

Through investigating the development of *Hebbian assemblies* in a network containing both of these interneuron classes, we may be able to unravel some of the conditions under which connection specificity in interneurons could emerge and what the consequences of activating specific interneurons would be in different circuits with tuned or untuned interneurons. Kuhlman and colleagues found a specific developmental effect in the responses of PV-interneurons [203]: while the PV-interneurons have high orientation tuning at eye-opening, this tuning is lost during the subsequent development. Such untuned inhibition is well suited to control the exclusive activation of only a single assembly at a time, which will suppress all others through a spread of PV inhibition across excitatory cells of different tuning. That PV-interneurons are delivering this type of inhibition is also suggested by the results in [188].

Interestingly, Cottam and colleagues measured that SST-activation could unmask a tuning of PV-interneurons which is not visible during normal activity in the circuit [193]. The resulting tuned activation of PV-interneurons could further strengthen the effects of their suppression found optogenetically, as it opens opportunities for excitatory cells from multiple assemblies to become active simultaneously. Tuning of cells, their coupling into assemblies and the conditions for the developmental emergence of both these properties are most likely highly interwoven and may be best investigated from a comparative perspective across species [204].

6.4 Final remarks

Developmental plasticity dependent on experience is crucial for the adaptation of animal behavior to a changing environment, and perturbation of sensory experience exposes the mechanisms underlying this adaptation. Critical periods of heightened plasticity governed by distinct mechanisms play an important role in the development of sensory processing, and maybe also in human psychiatric disorders [205]. Critical period plasticity influences, for example, the E/I-ratio in cortical circuits [3,94]. This suggests a potential connection to the development of autism spectrum disorders, as one hypothesis about the physiological changes underlying autism is an altered E/I-ratio, recently studied by comparing multiple mouse models of autism phenomenology [206]. The alterations of the E/I-ratio could be related to autism-associated genes that influence homeostatic plasticity [207]. A critical period of heightened plasticity that shares some similarities with the developmental critical period plasticity also seems to follow a stroke and could be important for, at least partial, recovery [208] (but see [209]).

Understanding critical period plasticity thus plays an important role in understanding computations in the healthy brain and how these are changed in disease.

The plasticity during critical periods shapes excitatory and inhibitory connections in feedforward and recurrent circuits. My research has shown that the operating regime, defined by the level of recurrent excitation and the role that inhibition plays in stabilizing the network, crucially shapes how activity in cortical circuits is controlled by critical period plasticity based on the presence or absence of the paradoxical effect. The rich interactions between multiple interneuron subtypes add a further layer to this as they allow a decoupling of the paradoxical effect from the operating regime in certain interneuron subtypes, potentially driven by transient inhibitory circuitry in the developing cortex. Accordingly, the inhibitory circuitry itself is highly plastic during the critical period and shaped by the sensory experience of the animal. These insights contribute to an unfolding rich panorama of inhibitory interactions in the developmental emergence of functional cortical circuits.

A Appendix

A.1 Stability in an excitatory-inhibitory network with general connectivity

The connectivity of a general population-rate network with one excitatory and one inhibitory population is:

$$\boldsymbol{W} = \begin{pmatrix} w_{EE} & -w_{EI} \\ w_{IE} & -w_{II} \end{pmatrix}, \tag{A.1}$$

where w_{AB} is the strength of the connection from population B to population A. Since the Jacobian of this system is $\mathcal{J} = \boldsymbol{W} - \mathbb{I}$, we can calculate the conditions for stability from the eigenvalues of W. Stability requires that the largest real part is < 1. We find the eigenvalues of $\boldsymbol{W}$ from solving the characteristic polynomial of this general connectivity matrix

$$\det(\boldsymbol{W} - \lambda\mathbb{I}) = \lambda^2 - \lambda(w_{EE} - w_{II}) - w_{EE}w_{II} + w_{EI}w_{IE} \tag{A.2}$$

to be

$$\begin{aligned} \lambda_{1/2} &= \frac{w_{EE} - w_{II}}{2} \pm \sqrt{\frac{(w_{EE} - w_{II})^2}{4} + w_{EE}w_{II} - w_{EI}w_{IE}} \\ &= \frac{w_{EE} - w_{II}}{2} \pm \sqrt{\frac{(w_{EE} + w_{II})^2}{4} - w_{EI}w_{IE}}. \end{aligned} \tag{A.3}$$

Whether the eigenvalues are two real eigenvalues or a conjugate pair of complex numbers, stability always requires that at least the fraction outside the square-root in Eq. A.3 is smaller than one, which implies:

$$w_{EE} < 2 + w_{II}. \tag{A.4}$$

Note that this condition is equivalent to the stability condition that the trace of the Jacobian is negative. If the eigenvalues are two complex conjugate values ($w_{EI}w_{IE} > \frac{(w_{EE}+w_{II})^2}{4}$), this is already all that needs to be fulfilled for stability. Note, however, that the assumption of conjugate eigenvalues rests on a condition on $w_{EI}w_{IE}$, which describes the strength of the reciprocal loop between excitation and inhibition. The eigenvalues are complex conjugate when this is strong. An interesting implication of the condition in Eq. A.4 is that for $w_{EE} > 2$ inhibition of inhibition ($w_{II} > w_{EE} - 2 > 0$) is required for stability. In [114] it was pointed out that the role of such disinhibitory connections was not clear at the time, while it was known that such connections are present in the cortex.

If the two eigenvalues are real, the largest eigenvalue is λ_1 and stability requires:

$$\lambda_1 = \frac{w_{EE} - w_{II}}{2} + \sqrt{\frac{(w_{EE} + w_{II})^2}{4} - w_{EI}w_{IE}} < 1. \tag{A.5}$$

Resorting this and squaring both sides[1] we get:

$$\begin{aligned} \sqrt{\frac{(w_{EE} + w_{II})^2}{4} - w_{EI}w_{IE}} &< \frac{2 - w_{EE} + w_{II}}{2} \\ \Rightarrow (w_{EE} + w_{II})^2 - 4w_{EI}w_{IE} &< \left(2 - w_{EE} + w_{II}\right)^2 \\ \Rightarrow w_{EE} - w_{II} + w_{EE}w_{II} - 1 &< w_{EI}w_{IE}. \end{aligned} \tag{A.6}$$

Resorting the left hand side in the last line in this Eq. A.6, it follows:

$$(w_{EE} - 1)(w_{II} + 1) < w_{EI}w_{IE}. \tag{A.7}$$

This relation, which is equivalent to the condition that the determinant of the Jacobian is positive, proves that for $w_{EE} < 1$ no reciprocal connections between excitation and inhibition are necessary to provide dynamical stability because then the left-hand side in Eq. A.7 is negative and is thus fulfilled for all $w_{AB} \geq 0$ (Note that the condition $(w_{II} > w_{EE} - 2)$ is always fulfilled in this case). For $w_{EE} > 1$, reciprocal connections between excitation and inhibition are necessary to provide dynamical stability. Therefore, also for general connectivity, the network is an ISN for w_{EE} with suitable inhibitory connectivity. The paradoxical effect generally emerges at the value of recurrent excitatory coupling where inhibition becomes necessary for dynamical stability (Eq. 2.9).

A.2 Paradoxical change of the total inhibitory current in the network with multiple subtypes of interneurons

Litwin-Kumar and colleagues provided the original insight into generalized paradoxical effects with multiple subtypes of interneurons [130][2]. They found that the rate changes of any single interneuron subtype (or any subpopulation of interneurons, see [134]) is not sufficient to decide if the network under consideration is a non-ISN or an ISN. What distinguishes the two operating regimes is the total inhibitory current received by the excitatory population: in the ISN, this inhibitory current changes in the same direction as the excitatory rate. In the non-ISN, it changes in the opposite direction. This holds for arbitrary many interneuron subtypes,

[1] One can proof that both sides in the second step, when I square, are positive. It follows from the condition that $\lambda_1 < 1$ requires at least $w_{EE} < 2 + w_{II}$. This is the same condition as $0 < 2 - w_{EE} + w_{II}$. Thus, the inequality does not change its direction upon squaring.

[2] Note that this generalization of the paradoxical effect is different from the one I developed. In my work, I have provided an account of the paradoxical effect beyond the original work by Tsodyks et al. [114] by (a) considering two-dimensional perturbations and (b) applying such perturbation in both feedforward and, crucially, also recurrent synaptic pathways.

and arbitrary connectivity [130]. The criterion is equivalent to the rate changes of the inhibitory population in networks with a single subtype of interneurons, as in the classical paradoxical effect investigated by Tsodyks and colleagues [114]. However, with multiple subtypes of interneurons, the different subpopulations can change their rate in any direction because the operating regime only constrains the weighted sum of the rate changes. The individual rate changes are weighted with the respective connection weight in the connectivity matrix. After a brief recall of the general result from [130], I derive these currents for the specific model I have studied with two subtypes of interneurons (Sec. 4.6.2), and show that the currents indeed change paradoxically also when the paradoxical effect in PV-interneurons is reversed through SST-feedback.

As a reminder for the general case (see also Sec. 2.5.3), I denote the elements of the connectivity matrix from the different interneuron populations X onto the excitatory population E with W_{EX}, without specifying the number (say m) and biological interpretation of the interneuron subtypes. The linear dynamics following a perturbation is:

$$\tau \frac{d\boldsymbol{\Delta} r}{dt} = \big(\boldsymbol{W} - \mathbb{I}\big)\boldsymbol{\Delta} r + \boldsymbol{b} \tag{A.8}$$

with the general connectivity matrix

$$\boldsymbol{W}_{\text{general}} = \begin{pmatrix} W_{EE} & -W_{EX_1} & \cdots & -W_{EX_m} \\ \vdots & & \ddots & \vdots \\ W_{X_m E} & & \cdots & -W_{X_m X_m} \end{pmatrix}. \tag{A.9}$$

We consider an arbitrary perturbation targeting the interneurons

$$\boldsymbol{b} = (0, b_1, b_2, \cdots, b_m)^T. \tag{A.10}$$

The excitatory rate change and inhibitory current into the excitatory neurons follow from multiplying the steady-state of Eq. A.8 with the vector $\boldsymbol{e}_1 = (1, 0, \cdots, 0)^T$:

$$(W_{EE} - 1)\Delta r_E^{ss} = \sum_{i=1}^{m} W_{EX_i} \Delta r_{X_i}^{ss} = \Delta I_{E \leftarrow I}, \tag{A.11}$$

where the $\Delta r_{X_i}^{ss}$ are the changes of the steady-state rates of the different interneuron populations and Δr_E^{ss} the steady-state rate change of the excitatory population. The generalized notion of the paradoxical effect follows from the fact that the inhibitory current is equal to the excitatory rate change multiplied by $1 - W_{EE}$. The excitatory population is stable without recurrent inhibition for $W_{EE} < 1$ (non-ISN) and unstable for $W_{EE} > 1$ (ISN).

In my model I have used the connectivity matrix

$$\boldsymbol{W} = \begin{pmatrix} W_{EE} & W_{EP} & W_{ES} \\ W_{PE} & W_{PP} & W_{PS} \\ W_{SE} & W_{SP} & W_{SS} \end{pmatrix} = \begin{pmatrix} w & -\gamma w & -\kappa \\ w & -\gamma w & -\kappa \\ w & 0 & 0 \end{pmatrix} \tag{A.12}$$

for a network with two subtypes of interneurons. The parameter w defines the operating regime in this network, while suitably chosen γ and κ guarantee stability for arbitrary w. To prove consistency with the previous results, I have calculated the inhibitory current following an external input to PV-interneurons. Furthermore, I have calculated the responses to the two other possible perturbations: a current injection to only SST-interneurons and a current injection to both interneuron subtypes.

A.2.1 Current after perturbation of PV-interneurons

The steady-state rates after injection of an additional current to PV-interneurons change as (see Eq. 4.25):

$$\begin{aligned}\begin{pmatrix}\Delta r_E^{ss}\\ \Delta r_P^{ss}\\ \Delta r_S^{ss}\end{pmatrix} &= \frac{1}{\eta}\begin{pmatrix}1+\gamma w & -\gamma w & -\kappa\\ w(1-\kappa) & 1-w+w\kappa & -\kappa\\ w(1+\gamma w) & -\gamma w^2 & 1-w+\gamma w\end{pmatrix}\begin{pmatrix}0\\ \xi\\ 0\end{pmatrix}\\ &= \frac{1}{\eta}\begin{pmatrix}-\gamma w\xi\\ (1-w+w\kappa)\xi\\ -\gamma w^2\xi\end{pmatrix}, \end{aligned} \tag{A.13}$$

where $\eta = \det(\mathbb{I}-\mathbf{W}) > 0$. Thus, this determinant does not influence the direction of the rate changes. As discussed in detail in Sec. 4.6.2, sufficiently strong SST-feedback can reverse the direction of rate change of the PV-population in an ISN. The change of the inhibitory current ($\Delta I_{E\leftarrow I}$) after stimulation of the PV-interneurons with the connectivity in Eq. A.12 is:

$$\begin{aligned}\Delta I_{E\leftarrow I} &= W_{EP}\Delta r_P^{ss} + W_{ES}\Delta r_S^{ss}\\ &= \frac{1}{\eta}\big(\gamma w\cdot(1-w+w\kappa)\xi - \kappa\cdot\gamma w^2\xi\big)\end{aligned} \tag{A.14}$$

which simplifies to:

$$\Delta I_{E\leftarrow I} = \frac{\gamma w(1-w)}{\eta}\xi. \tag{A.15}$$

This derived current behaves consistently with the results from [130]: in the ISN ($w > 1$), the current change has the opposite sign than the external current ξ, which means it changes in the same direction as the excitatory rate (e.g., decrease when we inject a positive current to PV). In the non-ISN ($w < 1$), the current change has the same sign, changing in the direction opposite to the excitatory rate. Therefore, also in my model, the direction of the current change depends on the operating regime of the circuit. The direction of the current change is independent of κ and behaves paradoxically regardless of whether PV-interneurons show the paradoxical effect without (or with weak) SST-feedback, or if it is reversed through strong SST-feedback.

A.2.2 Perturbation of SST-interneurons and both interneuron populations simultaneously

We can furthermore derive the rate changes and the corresponding changes of the inhibitory currents, following the injection of an additional current to (a) SST-interneurons only and (b) to both subtypes.

Injection of an external current to the SST-interneurons causes the rate changes

$$\begin{aligned}\begin{pmatrix}\Delta r_E^{ss}\\ \Delta r_P^{ss}\\ \Delta r_S^{ss}\end{pmatrix} &= \frac{1}{\eta}\begin{pmatrix}1+\gamma w & -\gamma w & -\kappa\\ w(1-\kappa) & 1-w+w\kappa & -\kappa\\ w(1+\gamma w) & -\gamma w^2 & 1-w+\gamma w\end{pmatrix}\begin{pmatrix}0\\0\\\xi\end{pmatrix}\\ &= \frac{1}{\eta}\begin{pmatrix}-\kappa\xi\\ -\kappa\xi\\ (1-w+\gamma w)\xi\end{pmatrix}. \end{aligned} \tag{A.16}$$

Thus, PV-interneurons always respond with a negative rate change to this injection, independent of the operating regime and the strnegth of SST-feedback (as long as it is non-zero). The direction of the SST-interneurons' rate change depends on the operating regime and the role of PV-interneurons in stabilizing the network. Specifically, stabilization of the dynamics through PV-interneurons requires $\gamma > (w-1)/w$, and therefore the SST-interneurons respond non-paradoxically to an external rate perturbation when PV-interneurons are sufficient to stabilize the network independent of the strength of SST-feedback. Note that general stability of the network requires $\gamma + \kappa > (w-1)/w$, but that the derivation of this condition is only possible if γ is non-zero.

Following this perturbation of the SST-interneurons, the inhibitory current changes as:

$$\begin{aligned}\Delta I_{E\leftarrow I} &= W_{EP}\Delta r_P^{ss} + W_{ES}\Delta r_S^{ss}\\ &= \frac{1}{\eta}\big(\gamma w\cdot\kappa\xi + \kappa\cdot(1-w+\gamma w)\xi\big)\\ &= \frac{\kappa(1-w)}{\eta}\xi, \end{aligned} \tag{A.17}$$

where the direction of the current change depends on the operating regime, consistent with [130].

Injection of an external current to both the PV- and the SST-interneurons causes the rate changes:

$$\begin{aligned}\begin{pmatrix}\Delta r_E^{ss}\\ \Delta r_P^{ss}\\ \Delta r_S^{ss}\end{pmatrix} &= \frac{1}{\eta}\begin{pmatrix}1+\gamma w & -\gamma w & -\kappa\\ w(1-\kappa) & 1-w+w\kappa & -\kappa\\ w(1+\gamma w) & -\gamma w^2 & 1-w+\gamma w\end{pmatrix}\begin{pmatrix}0\\\xi\\\xi\end{pmatrix}\\ &= \frac{1}{\eta}\begin{pmatrix}-(\gamma w+\kappa)\xi\\ (1-w+w\kappa-\kappa)\xi\\ (1-w+\gamma w-\gamma w^2)\xi\end{pmatrix}. \end{aligned} \tag{A.18}$$

The excitatory rate changes in the opposite direction to ξ, but the directions of the firing rate changes of the inhibitory populations depend on the interplay of the different parameters of inhibitory feedback and the operating regime. The behavior is best visible by bringing the inhibitory rate changes into the forms (removing η):

$$\begin{aligned} \Delta r_P^{ss} &\propto \xi(1-w)(1-\kappa) \\ \Delta r_S^{ss} &\propto \xi(1-w)(1+\gamma w) \end{aligned} \tag{A.19}$$

which show that the rate changes of the SST-interneurons only depends on the operating regime, whereas the PV firing rate changes depend on both the operating regime and, as in the case of a current injection only to PV-interneurons, the strength of SST-feedback. For $\kappa < 1$, PV-interneurons respond paradoxically in the ISN, while $\kappa > 1$ leads to a reversal of the paradoxical effect. Interestingly, in the non-ISN, the value of SST-feedback can also generate a paradoxical effect in PV-interneurons as the rate changes become opposite to ξ for $\kappa > 1$. This is consistent with the findings from [136], where it was shown that the presence of a paradoxical effect does not necessarily imply that the network is an ISN.

The inhibitory current to excitatory neurons follows:

$$\begin{aligned} \Delta I_{E\leftarrow I} &= W_{EP}\Delta r_P^{ss} + W_{ES}\Delta r_S^{ss} \\ &= \frac{1}{\eta}\Big(\gamma w \cdot (1-w+w\kappa-\kappa)\xi + \kappa \cdot (1-w+\gamma w-\gamma w^2)\xi\Big)\Big) \\ &= \frac{(\gamma w+\kappa)(1-w)}{\eta}\xi \end{aligned} \tag{A.20}$$

consistent with the generalized paradoxical effect in [130]: the inhibitory current changes in the same direction as ξ in the non-ISN and opposite direction to ξ in the ISN.

A.3 Variations of the spiking network with two subtypes of interneurons

A.3.1 Adding background inputs to PV-interneurons

My original setup in chapter 4 was based on the idea that SST-interneurons receive inputs from other sources than the LGN, and that PV-interneurons do not receive inputs from these sources. E.g. SST integrate inputs over larger areas in cortex than PV [59, 210], receive higher-order cortical feedback and neuromodulatory inputs [61, 68], and inputs from higher-order thalamic nuclei [69]. BKG-input to excitatory neurons can also be considered to stem from one or multiple of these sources. However, that BKG projects to both excitatory and SST neurons does not imply in any way that neurons in the two populations receive shared, correlated input or the like. Both excitatory and SST neurons are driven by uncorrelated spike trains with Poissonian statistics. PV-interneurons likely also receive some

higher-order feedback and neuromodulatory inputs, although the extent is not clear. In this context, it appears important to test the effect of background inputs to PV.

Additional input to PV-interneurons strongly suppresses firing rates in all populations. Therefore, concurrently with the introduction of BKG-inputs to PV, I increased the inputs to excitatory neurons. The most important feature of the BKG input in my model remains: that it is not affected by MD-induced plasticity. The new vector of input weights with input from background to PV and increased background input to excitatory neurons is (compare to Eq. 4.4):

$$\mathbf{W}_{\text{ffw}} = \begin{pmatrix} J_s + 2J_b \\ g_{fw}J_s + J_b \\ 0 + J_b \end{pmatrix}. \tag{A.21}$$

I found that adding background inputs to PV-interneurons does not affect the response to MD-induced synaptic changes in the feedforward, recurrent, and interacting feedforward and recurrent cases (Fig. A.1, compare to Fig. 4.3). As long as the firing rates are at appropriate levels before the induction of MD-induced plasticity, I find the inversion of the PV response planes also when PV receive the background.

A.3.2 Connection from PV- to SST-interneurons

Direct projections from PV to SST interneurons have been found to be very weak in V1 [56]. The measured interaction strength suggests no relevant synaptic connectivity, with comparable size to the connections from SST onto other SST, where the consensus seems that synaptic contacts between these are absent [24]. Accordingly, in my primary model, I have set both SST to SST and PV to SST connections to zero, as in many model networks of V1 that include multiple subtypes of interneurons [130, 173] and in similar models for A1 [179, 180]. However, Jiang et al. [53] report connections between PV-interneurons and SST-interneurons, although with smaller probability than the classically modeled connections (like SST to PV, or PV to PV). Based on these results, Mahrach and colleagues [136] investigated the effects of this connection from PV to SST on paradoxical effects in a V1 circuit model, but find that it does not affect excitatory and PV responses qualitatively.

Since connections from PV to SST have been confirmed experimentally in [53], I also tested the effect of adding them in my model with two subtypes of interneurons. The matrix describing the recurrent connectivity in this case is (compare to Eq. 4.1):

$$\mathbf{W} = \begin{pmatrix} J & g_{rc}J & K \\ J & g_{rc}J & K \\ J & g_{rc}J/2 & 0 \end{pmatrix}, \tag{A.22}$$

where all parameters have the same values as in chapter 4 ($J = 0.1$, $g_{rc} = 8$, $K = 1.6$, see Table 4.1). Based on the smaller connection probability reported in [53], I chose the PV to SST synaptic strengths smaller than the other PV output synapses.

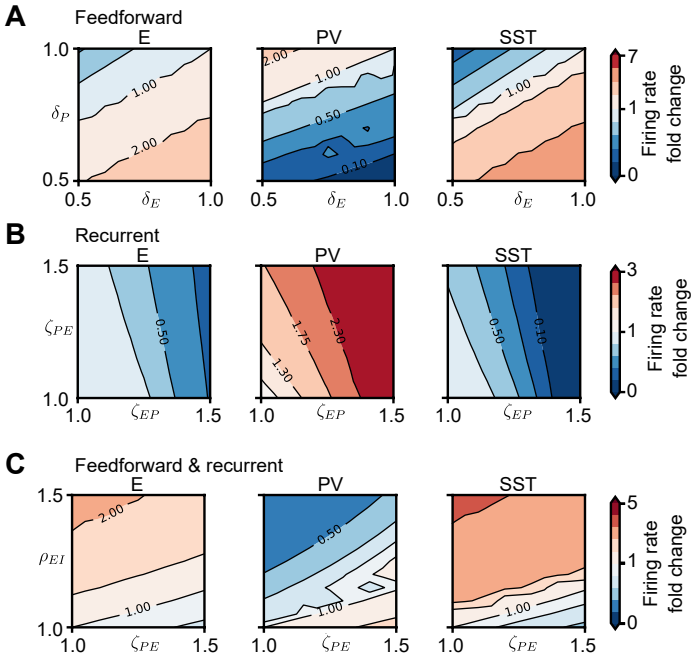

Figure A.1: Qualitative structure of firing rate fold-changes is preserved when BKG inputs also drive PV-interneurons. **(A)** Network firing rate in the (δ_E, δ_P) plane as fold-change of baseline firing rate (baseline in top right corner) for excitatory (left panel), PV (middle) and SST neurons (right) in the network with strong SST feedback ($K = 1.6$ nS). **(B)** Network firing rate in the (ζ_{EP}, ζ_{PE}) plane as fold-change of baseline firing rate (baseline in bottom left corner). **(C)** Network firing rate in the (ζ_{PE}, ρ_{EI}) plane as fold-change of baseline firing rate (baseline in bottom left corner).

Nonetheless, even this weight substantially affects the baseline activity, so I expect it to be sufficient to test the effects of this connection on the response of the network.

The additional inhibition from PV to SST can strongly suppress baseline firing rates in all populations, though most strongly in SST. The resulting low firing rates of SST are problematic concerning facilitation of PV firing rates following MD-induced plasticity, since the facilitation of PV firing rates is based on a release of the PV-interneurons from inhibition by SST-interneurons. If SST-rates are too low, this release cannot happen. Thus, I adapted the inputs to the model to increase firing rates. As in the previous section, I increased the background input to excitatory neurons (using a factor of 1.3), which effectively adjusted overall activity levels. I modified the vector of input weights to be (compare to Eq. 4.4):

$$\mathbf{W}_{\text{ffw}} = \begin{pmatrix} J_s + 1.3J_b \\ g_{fw}J_s + 0 \\ 0 + J_b \end{pmatrix}. \tag{A.23}$$

Firing rate changes following MD-induced synaptic changes remain qualitatively the same in this network when inhibition from PV to SST is added (Fig A.2, compare

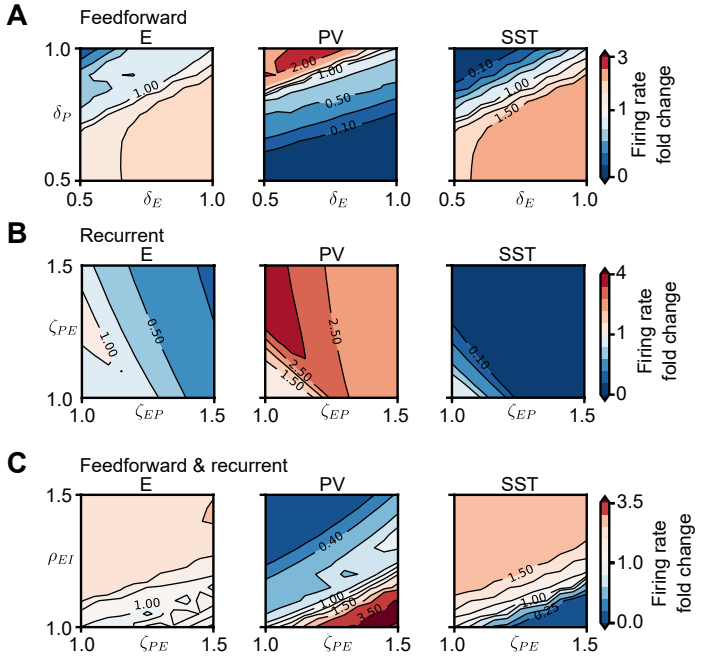

Figure A.2: Qualitative structure of firing rate fold-changes is preserved when PV-interneurons inhibit SST-interneurons. **(A)** Network firing rate in the (δ_E, δ_P) plane as fold-change of baseline firing rate (baseline in top right corner) for excitatory (left panel), PV (middle) and SST neurons (right) in the network with strong SST feedback ($K = 1.6$ nS). **(B)** Network firing rate in the (ζ_{EP}, ζ_{PE}) plane as fold-change of baseline firing rate (baseline in bottom left corner). **(C)** Network firing rate in the (ζ_{PE}, ρ_{EI}) plane as fold-change of baseline firing rate (baseline in bottom left corner).

to Fig. 4.3). Thus, consistent with the findings from [136], I found that addition of this connection does not affect my conclusions about the role of SST-feedback on the response of PV-interneurons.

Bibliography

[1] L. M. A. Richter and J. Gjorgjieva, "Interneuron subtypes enable independent modulation of excitatory and inhibitory firing rates after sensory deprivation," *bioRxiv*, 2021.

[2] ——, "A circuit mechanism for independent modulation of excitatory and inhibitory firing rates after sensory deprivation," *Proc Natl Acad Sci USA*, vol. 119, p. e2116895119, 2022.

[3] N. J. Miska, L. M. A. Richter, B. A. Cary, J. Gjorgjieva, and G. G. Turrigiano, "Sensory experience inversely regulates feedforward and feedback excitation-inhibition ratio in rodent visual cortex," *eLife*, vol. 7, p. e38846, 2018.

[4] L. M. A. Richter and J. Gjorgjieva, "Understanding neural circuit development through theory and models," *Curr Opin Neurobiol*, vol. 46, pp. 39–47, 2017.

[5] M. Rabinovich, R. Huerta, and G. Laurent, "Transient dynamics for neural processing," *Science*, vol. 321, pp. 48–50, 2008.

[6] D. Sussillo, "Neural circuits as computational dynamical systems," *Curr Opin Neurobiol*, vol. 25, pp. 156–163, 2014.

[7] J. S. Espinosa and M. P. Stryker, "Development and plasticity of the primary visual cortex," *Neuron*, vol. 75, pp. 230–249, 2012.

[8] L. C. Katz and C. J. Shatz, "Synaptic activity and the construction of cortical circuits," *Science*, vol. 274, pp. 1133–1138, 1996.

[9] T. N. Wiesel, "Postnatal development of the visual cortex and the influence of environment," *Nature*, vol. 299, pp. 583–591, 1982.

[10] M. Fagiolini, T. Pizzorusso, N. Berardi, L. Domenici, and L. Maffei, "Functional postnatal development of the rat primary visual cortex and the role of visual experience: dark rearing and monocular deprivation," *Vision Res*, vol. 34, pp. 709–720, 1994.

[11] T. K. Hensch, "Critical period plasticity in local cortical circuits," *Nat Rev Neurosci*, vol. 6, pp. 877–888, 2005.

[12] T. N. Wiesel and D. H. Hubel, "Single-cell responses in striate cortex of kittens deprived of vision in one eye," *J Neurophysiol*, vol. 26, pp. 1003–1017, 1963.

[13] D. H. Hubel and T. N. Wiesel, "The period of susceptibility to the physiological effects of unilateral eye closure in kittens," *J Physiol*, vol. 206, pp. 419–436, 1970.

[14] N. Brunel, "Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons," *J Comput Neurosci*, vol. 8, pp. 183–208, 2000.

[15] F. Wolf, R. Engelken, M. Puelma-Touzel, J. D. F. Weidinger, and A. Neef, "Dynamical models of cortical circuits," *Curr Opin Neurobiol*, vol. 25, pp. 228–236, 2014.

[16] L. F. Abbott, "Theoretical neuroscience rising," *Neuron*, vol. 60, pp. 489–495, 2008.

[17] A. D. Huberman, M. B. Feller, and B. Chapman, "Mechanisms underlying development of visual maps and receptive fields," *Annu Rev Neurosci*, vol. 31, pp. 479–509, 2008.

[18] C. Bargmann and E. Marder, "From the connectome to brain function," *Nat Methods*, vol. 10, p. 483–490, 2013.

[19] E. Marder, "Variability, compensation, and modulation in neurons and circuits," *Proc Natl Acad Sci USA*, vol. 108, pp. 15 542–15 548, 2011.

[20] J. H. Kirchner and J. Gjorgjieva, "Emergence of local and global synaptic organization on cortical dendrites," *Nat Commun*, vol. 12, p. 4005, 2021.

[21] J. Gjorgjieva, R. A. Mease, W. J. Moody, and A. L. Fairhall, "Intrinsic neuronal properties switch the mode of information transmission in networks," *PLoS Comput Biol*, vol. 10, p. e1003962, 2014.

[22] K. D. Miller, J. B. Keller, and M. P. Stryker, "Ocular dominance column development: analysis and simulation," *Science*, vol. 245, pp. 605–615, 1989.

[23] R. Brette, "Philosophy of the spike: Rate-based vs. spike-based theories of the brain," *Front Syst Neurosci*, vol. 9, 2015.

[24] R. Tremblay, S. Lee, and B. Rudy, "GABAergic interneurons in the neocortex: from cellular properties to circuits," *Neuron*, vol. 91, pp. 260–292, 2016.

[25] B. Rudy, G. Fishell, S. Lee, and J. Hjerling-Leffler, "Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons," *Dev Neurobiol*, vol. 71, pp. 45–61, 2011.

[26] D. O. Hebb, *The organization of behavior: A neuropsychological theory.* New York: Wiley & Sons, 1949.

[27] H. Ko, L. Cossell, C. Baragli, J. Antolik, C. Clopath, S. B. Hofer, and T. D. Mrsic-Flogel, "The emergence of functional microcircuits in visual cortex," *Nature*, vol. 496, pp. 96–100, 2013.

[28] H. Ko, T. D. Mrsic-Flogel, and S. B. Hofer, "Emergence of feature-specific connectivity in cortical microcircuits in the absence of visual experience," *J Neurosci*, vol. 34, pp. 9812–9816, 2014.

[29] L. Carrillo-Reid and R. Yuste, "Playing the piano with the cortex: role of neuronal ensembles and pattern completion in perception and behavior," *Curr Opin Neurobiol*, vol. 64, pp. 89–95, 2020.

[30] A. Litwin-Kumar and B. Doiron, "Formation and maintenance of neuronal assemblies through synaptic plasticity," *Nat Commun*, vol. 5:5319, 2014.

[31] F. Zenke, E. J. Agnes, and W. Gerstner, "Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks," *Nat Commun*, vol. 6, p. 6922, 2015.

[32] S. Sadeh, C. Clopath, and S. Rotter, "Emergence of functional specificity in balanced networks with synaptic plasticity," *PLoS Comput Biol*, vol. 11(6), p. e1004307, 2015.

[33] D. C. Van Essen, C. H. Anderson, and D. J. Felleman, "Information processing in the primate visual system: an integrated systems perspective," *Science*, vol. 255, pp. 419–423, 1992.

[34] G. K. Ocker, Y. Hu, M. A. Buice, B. Doiron, K. Josíc, R. Rosenbaum, and E. Shea-Brown, "From the statistics of connectivity to the statistics of spike times in neuronal networks," *Curr Opin Neurobiol*, vol. 46, pp. 109–119, 2017.

[35] S. Eckmann, L. Klimmasch, B. E. Shi, and J. Triesch, "Active efficient coding explains the development of binocular vision and its failure in amblyopia," *Proc Natl Acad Sci USA*, vol. 117, pp. 6156–6162, 2020.

[36] P. Berens and T. Euler, "Neuronal diversity in the retina," *Neuroforum*, vol. 23, pp. A93–A101, 2017.

[37] T. Baden, P. Berens, M. R. Rosón, M. Bethge, and T. Euler, "The functional diversity of retinal ganglion cells in the mouse," *Nature*, vol. 529, pp. 345–350, 2016.

[38] X. Y. Ji, B. Zingg, L. Mesik, Z. Xiao, L. I. Zhang, and H. W. Tao, "Thalamocortical innervation pattern in mouse auditory and visual cortex: laminar and cell-type specificity," *Cereb Cortex*, vol. 26, pp. 2612–2625, 2016.

[39] N. J. Priebe and A. W. McGee, "Mouse vision as a gateway for understanding how experience shapes neural circuits," *Front Neural Circuits*, vol. 8, p. 123, 2014.

[40] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *J Physiol*, vol. 148, pp. 574–591, 1959.

[41] M. Kaschube, M. Schnabel, S. Löwel, D. M. Coppola, L. White, and F. Wolf, "Universality in the evolution of orientation columns in the visual cortex," *Science*, vol. 330, pp. 1113–1116, 2010.

[42] F. Scala, D. Kobak, S. Shan, Y. Bernaerts, S. Laturnus, C. R. Cadwell *et al.*, "Layer 4 of mouse neocortex differs in cell types and circuit organization between sensory areas," *Nat Commun*, vol. 10, p. 4174, 2019.

[43] T. Binzegger, R. J. Douglas, and K. A. C. Martin, "A quantitative map of the circuit of cat primary visual cortex," *J Neurosci*, vol. 24, pp. 8441–8453, 2004.

[44] G. M. Shepherd and T. B. Rowe, "Neocortical lamination: insights from neuron types and evolutionary precursors," *Front Neuroanat*, vol. 11, p. 100, 2017.

[45] V. Braitenberg and A. Schüz, *Cortex: Statistics and Geometry of Neuronal Connectivity*. Berlin Heidelberg: Springer-Verlag, 1998.

[46] S. Song, P. J. Sjöström, M. Reigl, S. B. Nelson, and D. B. Chklovskii, "Highly nonrandom features of synaptic connectivity in local cortical circuits," *PLoS Biol*, vol. 3, 2005.

[47] N. R. Tannenbaum and Y. Burak, "Shaping neural circuits by high-order synaptic interactions," *PLoS Comput Biol*, vol. 12, p. e1005056, 2016.

[48] H. Eichenbaum, "Barlow versus Hebb: When is it time to abandon the notion of feature detectors and adopt the cell assembly as the unit of cognition?" *Neurosci Lett*, vol. 680, pp. 88–93, 2018.

[49] A. L. Fairhall, "The receptive field is dead. long live the receptive field?" *Curr Opin Neurobiol*, vol. 25, pp. ix–xii, 2014.

[50] L. Cossell, M. F. Iacaruso, D. R. Muir, R. Houlton, E. N. Sader, H. Ko *et al.*, "Functional organization of excitatory synaptic strength in primary visual cortex," *Nature*, vol. 518, pp. 399–403, 2015.

[51] H. Ko, S. B. Hofer, B. Pichler, K. A. Buchanan, P. J. Sjöström, and T. D. Mrsic-Flogel, "Functional specificity of local synaptic connections in neocortical networks," *Nature*, vol. 473, pp. 87–90, 2011.

[52] A. D. Lien and M. Scanziani, "Tuned thalamic excitation is amplified by visual cortical circuits," *Nat Neurosci*, vol. 16(9), pp. 1315–1323, 2013.

[53] X. Jiang, S. Shen, C. R. Cadwell, P. Berens, F. Sinz, A. S. Ecker *et al.*, "Principles of connectivity among morphologically defined cell types in adult neocortex," *Science*, vol. 350, p. aac9462, 2015.

[54] K. B. Hengen, M. E. Lambo, S. D. van Hooser, D. B. Katz, and G. G. Turrigiano, "Firing rate homeostasis in visual cortex of freely behaving rodents," *Neuron*, vol. 80, pp. 335–342, 2013.

[55] J. Urban-Ciecko and A. L. Barth, "Somatostatin-expressing neurons in cortical networks," *Nat Rev Neurosci*, vol. 17, pp. 401–409, 2016.

[56] C. K. Pfeffer, M. Xue, M. He, Z. J. Huang, and M. Scanziani, "Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons," *Nat Neurosci*, vol. 16, pp. 1068–1076, 2013.

[57] A. M. Packer and R. Yuste, "Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: A canonical microcircuit for inhibition?" *J Neurosci*, vol. 31, pp. 13 260–13 271, 2011.

[58] M. Xue, B. V. Atallah, and M. Scanziani, "Equalizing excitation-inhibition ratios across visual cortical neurons," *Nature*, vol. 511, pp. 596–600, 2014.

[59] H. Adesnik, W. Bruns, H. Taniguchi, Z. J. Huang, and M. Scanziani, "A neural circuit for spatial summation in visual cortex," *Nature*, vol. 490, pp. 226–231, 2012.

[60] H. Nienborg, A. Hasenstaub, I. Nauhaus, H. Taniguchi, Z. J. Huang, and E. M. Callaway, "Contrast dependence and differential contributions from somatostatin- and parvalbumin-expressing neurons to spatial integration in mouse V1," *J Neurosci*, vol. 33, pp. 11 145–11 154, 2013.

[61] R. Batista-Brito, E. Zagha, J. M. Ratliff, and M. Vinck, "Modulation of cortical circuits by top-down processing and arousal state in health and disease," *Curr Opin Neurobiol*, vol. 52, pp. 172–181, 2018.

[62] Y. Fu, J. M. Tucciarone, J. S. Espinosa, N. Sheng, D. P. Darcy, R. A. Nicoll *et al.*, "A cortical circuit for gain control by behavioral state," *Cell*, vol. 156, pp. 1139–1152, 2014.

[63] S.-H. Lee, A. C. Kwan, S. Zhang, V. Phoumthipphavong, J. G. Flannery, S. C. Masmanidis *et al.*, "Activation of specific interneurons improves V1 feature selectivity and visual perception," *Nature*, vol. 488, pp. 379–383, 2012.

[64] N. R. Wilson, C. A. Runyan, F. L. Wang, and M. Sur, "Division and subtraction by distinct cortical inhibitory networks *in vivo*," *Nature*, vol. 488, pp. 343–348, 2012.

[65] B. V. Atallah, W. Bruns, M. Carandini, and M. Scanziani, "Parvalbumin expressing interneurons linearly transform cortical responses to visual stimuli," *Neuron*, vol. 73, pp. 159–170, 2012.

[66] S.-H. Lee, A. C. Kwan, and Y. Dan, "Interneuron subtypes and orientation tuning," *Nature*, vol. 508, pp. E1–E4, 2014.

[67] G. Ma, Y. Liu, L. Wang, Z. Xiao, K. Song, Y. Wang *et al.*, "Hierarchy in sensory processing reflected by innervation balance on cortical interneurons," *Sci Adv*, vol. 7, p. eabf5676, 2021.

[68] I. Yavorska and M. Wehr, "Somatostatin-expressing inhibitory interneurons in cortical circuits," *Front Neural Circuits*, vol. 10, p. 76, 2016.

[69] G. Pouchelon, D. Dwivedi, Y. Bollmann, C. K. Agba, Q. Xu, A. M. Mirow *et al.*, "The organization and development of cortical interneuron presynaptic circuits are area specific," *Cell Rep*, vol. 37, p. 109993, 2021.

[70] A. van Ooyen, "Using theoretical models to analyse neural development," *Nat Rev Neurosci*, vol. 12, p. 311–326, 2011.

[71] J. Cang and D. A. Feldheim, "Developmental mechanisms of topographic map formation and alignment," *Annu Rev Neurosci*, vol. 36, pp. 51–77, 2013.

[72] J. B. Ackman, T. J. Burbridge, and M. C. Crair, "Retinal waves coordinate patterned activity throughout the developing visual system," *Nature*, vol. 490, pp. 219–225, 2012.

[73] F. Siegel, J. A. Heimel, J. Peters, and C. Lohmann, "Peripheral and central inputs shape network dynamics in the developing visual cortex in vivo," *Curr Biol*, vol. 22, pp. 253–258, 2012.

[74] J. Shen and M. T. Colonnese, "Development of activity in mouse visual cortex," *J Neurosci*, vol. 36, pp. 12 259–12 275, 2016.

[75] Y. Gu and J. Cang, "Binocular matching of thalamocortical and intracortical circuits in the mouse visual cortex," *eLife*, vol. 5, p. e22032, 2016.

[76] M. L. Linden, A. J. Heynen, R. H. Haslinger, and M. F. Bear, "Thalamic activity that drives visual cortical plasticity," *Nat Neurosci*, vol. 12(4), pp. 390–392, 2009.

[77] M. Y. Frenkel and M. F. Bear, "How monocuar deprivation shifts ocular dominance in visual cortex of young mice," *Neuron*, vol. 44, pp. 917–923, 2004.

[78] J. T. Trachtenberg, C. Trepel, and M. P. Stryker, "Rapid extragranular plasticity in the absence of thalamocortical plasticity in the developing primary visual cortex," *Science*, vol. 287, pp. 2029–2032, 2000.

[79] Y. Yazaki-Sugiyama, S. Kang, H. Câteau, T. Fukai, and T. K. Hensch, "Bidirectional plasticity in fast-spiking GABA circuits by visual experience," *Nature*, vol. 462, pp. 218–221, 2009.

[80] T. Rose, J. Jaepel, M. Hübener, and T. Bonhoeffer, "Cell-specific restoration of stimulus preference after monocular deprivation in the visual cortex," *Science*, vol. 352, pp. 1319–1322, 2016.

[81] M. E. Lambo and G. G. Turrigiano, "Synaptic and intrinsic homeostatic mechanisms cooperate to increase L2/3 pyramidal neuron excitability during a late phase of critical period plasticity," *J Neurosci*, vol. 33, pp. 8810–8819, 2013.

[82] A. Maffei, K. Nataraj, S. B. Nelson, and G. G. Turrigiano, "Potentiation of cortical inhibition by visual deprivation," *Nature*, vol. 443, pp. 81–84, 2006.

[83] S. Lefort, A. C. Gray, and G. G. Turrigiano, "Long-term inhibitory plasticity in visual cortical layer 4 switches sign at the opening of the critical period," *Proc Natl Acad Sci USA*, vol. 110, p. E4540–E4547, 2013.

[84] M. Nahmani and G. G. Turrigiano, "Deprivation induced strengthening of presynaptic and postsynaptic inhibitory transmission in layer 4 of visual cortex during the critical period," *J Neurosci*, vol. 34, pp. 2571–2582, 2014.

[85] A. Maffei, M. E. Lambo, and G. G. Turrigiano, "Critical period for inhibitory plasticity in rodent binocular V1," *J Neurosci*, vol. 30, pp. 3304–3309, 2010.

[86] B. T. Barrett, A. Bradley, and P. V. McGraw, "Understanding the neural basis of amblyopia," *The Neuroscientist*, vol. 10, pp. 106–117, 2004.

[87] J. Toldi, I. Rojik, and O. Fehér, "Neonatal monocular enucleation-induced cross-modal effects observed in the cortex of adult rat," *Neuroscience*, vol. 62, pp. 105–114, 1994.

[88] A. Rohatgi, "WebPlotDigitizer," 2011.

[89] K. B. Hengen, A. Torrado Pacheco, J. N. McGregor, V. H. S. D., and G. G. Turrigiano, "Neuronal firing rate homeostasis is inhibited by sleep and promoted by wake," *Cell*, vol. 165, pp. 180–191, 2016.

[90] Y. K. Wu, K. B. Hengen, G. G. Turrigiano, and J. Gjorgjieva, "Homeostatic mechanisms regulate distinct aspects of cortical circuit dynamics," *Proc Natl Acad Sci USA*, vol. 117, pp. 24 514–24 525, 2020.

[91] Z. Ma, G. G. Turrigiano, R. Wessel, and K. B. Hengen, "Cortical circuit dynamics are homeostatically tuned to criticality in vivo," *Neuron*, vol. 104, pp. 655–664.e4, 2019.

[92] L. Wang, M. Kloc, Y. Gu, S. Ge, and A. Maffei, "Layer-specific experience dependent rewiring of thalamocortical circuits," *J Neurosci*, vol. 33, pp. 4181–4191, 2013.

[93] A. Maffei, S. B. Nelson, and G. G. Turrigiano, "Selective reconfiguration of layer 4 visual cortical circuitry by visual deprivation," *Nat Neurosci*, vol. 7, pp. 1353–1359, 2004.

[94] S. J. Kuhlman, N. D. Olivas, E. Tring, T. Ikrar, X. X., and J. T. Trachtenberg, "A disinhibitory microcircuit initiates critical period plasticity in the visual cortex," *Nature*, vol. 501, pp. 543–546, 2013.

[95] M. A. Gainey and D. E. Feldman, "Multiple shared mechanisms for homeostatic plasticity in rodent somatosensory and visual cortex," *Phil Trans R Soc B*, vol. 372, p. 20160157, 2017.

[96] L. Li, M. A. Gainey, J. E. Goldbeck, and D. E. Feldman, "Rapid homeostasis by disinhibition during whisker map plasticity," *Proc Natl Acad Sci USA*, vol. 111, pp. 1616–1621, 2014.

[97] M. A. Gainey, J. W. Aman, and D. E. Feldman, "Rapid disinhibition by adjustment of PV intrinsic excitability during whisker map plasticity in mouse S1," *J Neurosci*, vol. 38, pp. 4749–4761, 2018.

[98] C. B. Allen, T. Celikel, and D. E. Feldman, "Long-term depression induced by sensory deprivation during cortical map plasticity *in vivo*," *Nat Neurosci*, vol. 6, pp. 291–299, 2003.

[99] D. R. C. House, J. Elstrott, E. Koh, J. Chung, and D. E. Feldman, "Parallel regulation of feedforward inhibition and excitation during whisker map plasticity," *Neuron*, vol. 72, pp. 819–831, 2011.

[100] T. Toyoizumi and K. D. Miller, "Equalization of ocular dominance columns induced by an activity-dependent learning rule and the maturation of inhibition," *J Neurosci*, vol. 29, pp. 6514–6525, 2009.

[101] T. Toyoizumi, H. Miyamoto, Y. Yazaki-Sugiyama, N. Atapour, T. K. Hensch, and K. D. Miller, "A theory of the transition to critical period plasticity: Inhibition selectively suppresses spontaneous activity," *Neuron*, vol. 80, pp. 51–63, 2013.

[102] J. Bono and C. Clopath, "Synaptic plasticity onto inhibitory neurons as a mechanism for ocular dominance plasticity," *PLoS Comput Biol*, vol. 15, p. e1006834, 2019.

[103] B. S. Blais, H. Z. Shouval, and L. N. Cooper, "The role of presynaptic activity in monocular deprivation: Comparison of homosynaptic and heterosynaptic mechanisms," *Proc Natl Acad Sci USA*, vol. 96, pp. 1083–1087, 1999.

[104] B. S. Blais, L. N. Cooper, and H. Z. Shouval, "Effect of correlated lateral geniculate nucleus firing rates on predictions for monocular eye closure versus monocular retinal inactivation," *Phys Rev E*, vol. 80, p. 061915, 2009.

[105] M. Chistiakova, N. Bannon, M. Bazhenov, and M. Volgushev, "Heterosynaptic plasticity: multiple mechanisms and multiple roles," *The Neuroscientist*, vol. 20, pp. 483–498, 2014.

[106] J. Triesch, A. D. Vo, and A.-S. Hafner, "Competition for synaptic building blocks shapes synaptic plasticity," *eLife*, vol. 7, p. e37836, 2018.

[107] E. L. Bienenstock, L. N. Cooper, and P. W. Munro, "Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex," *J Neurosci*, vol. 2, pp. 32–48, 1982.

[108] C. D. Rittenhouse, H. Z. Shouval, M. A. Paradiso, and M. F. Bear, "Monocular deprivation induces homosynaptic long term depression in visual cortex," *Nature*, vol. 397, pp. 347–350, 1999.

[109] N. S. Desai, S. B. Nelson, and G. G. Turrigiano, "Activity-dependent regulation of excitability in rat visual cortical neurons," *Neurocomputing*, vol. 26-27, pp. 101–106, 1999.

[110] G. G. Turrigiano, K. R. Leslie, N. S. Desai, L. C. Rutherford, and S. B. Nelson, "Activity-dependent scaling of quantal amplitude in neocortical neurons," *Nature*, vol. 391, pp. 892–896, 1998.

[111] W. L. Shew and D. Plenz, "The functional benefits of criticality in the cortex," *The Neuroscientist*, vol. 19, pp. 88–100, 2013.

[112] C. van Vreeswijk and H. Sompolinsky, "Chaos in neuronal networks with balanced excitatory and inhibitory activity," *Science*, vol. 274, pp. 1724–1726, 1996.

[113] ——, "Chaotic balanced state in a model of cortical circuits," *Neural Comput*, vol. 10, pp. 1321–1371, 1998.

[114] M. V. Tsodyks, W. E. Skaggs, T. J. Sejnowski, and B. L. McNaughton, "Paradoxical effects of external modulation of inhibitory interneurons," *J Neurosci*, vol. 17, pp. 4382–4388, 1997.

[115] W. R. Softky and C. Koch, "The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs," *J Neurosci*, vol. 13, pp. 334–350, 1993.

[116] M. N. Shadlen and W. T. Newsome, "The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding," *J Neurosci*, vol. 18, pp. 3870–3896, 1998.

[117] M. R. Cohen and A. Kohn, "Measuring and interpreting neuronal correlations," *Nat Neurosci*, vol. 14, p. 811–819, 2011.

[118] B. Doiron, A. Litwin-Kumar, R. Rosenbaum, G. K. Ocker, and K. Josić, "The mechanics of state-dependent neural correlations," *Nat Neurosci*, vol. 19, p. 383–393, 2016.

[119] T. W. Troyer and K. D. Miller, "Physiological gain leads to high isi variability in a simple model of a cortical regular spiking cell," *Neural Comput*, vol. 9, no. 5, pp. 971–983, 1997.

[120] Y. Ahmadian and K. D. Miller, "What is the dynamical regime of cerebral cortex?" *Neuron*, vol. 109, pp. 3373–3391, 2021.

[121] A. Renart, J. de la Rocha, P. Bartho, L. Hollender, N. Parga, A. Reyes, and K. D. Harris, "The asynchronous state in cortical circuits," *Science*, vol. 327, pp. 587–590, 2010.

[122] G. Hennequin, T. P. Vogels, and W. Gerstner, "Non-normal amplification in random balanced neuronal networks," *Phys Rev E*, vol. 86, p. 011909, 2012.

[123] S. Ostojic, "Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons," *Nat Neurosci*, vol. 17(4), pp. 594–600, 2014.

[124] O. Harish and D. Hansel, "Asynchronous rate chaos in spiking neuronal circuits," *PLoS Comput Biol*, vol. 11, p. e1004266, 2015.

[125] D. Hansel and C. van Vreeswijk, "The mechanism of orientation selectivity in primary visual cortex without a functional map," *J Neurosci*, vol. 32, pp. 4049–4064, 2012.

[126] S. Sadeh, S. Cardanobile, and S. Rotter, "Mean-field analysis of orientation selectivity in inhibition-dominated networks of spiking neurons," *SpringerPlus*, vol. 3, p. 148, 2014.

[127] S. Denève and C. K. Machens, "Efficient codes and balanced networks," *Nat Neurosci*, vol. 19, pp. 375–382, 2016.

[128] R. J. Douglas, C. Koch, M. Mahowald, K. A. C. Martin, and H. H. Suarez, "Recurrent excitation in neocortical circuits," *Science*, vol. 269, pp. 981–985, 1995.

[129] S. Sadeh and S. Rotter, "Orientation selecitvity in inhibition dominated networks of spiking neurons: Effects of single neuron properties and network dynamics," *PLoS Comput Biol*, vol. 11, p. e1004045, 2015.

[130] A. Litwin-Kumar, R. Rosenbaum, and B. Doiron, "Inhibitory stabilization and visual coding in cortical circuits with multiple interneuron subtypes," *J Neurophysiol*, vol. 115, pp. 1399–1409, 2016.

[131] H. Ozeki, I. M. Finn, E. S. Schaffer, K. D. Miller, and D. Ferster, "Inhibitory stabilization of the cortical network underlies visual surround suppression," *Neuron*, vol. 62, pp. 578–592, 2009.

[132] H. Adesnik, "Synaptic mechanisms of feature coding in the visual cortex of awake mice," *Neuron*, vol. 95, pp. 1147–1159, 2017.

[133] S. Sadeh and C. Clopath, "Inhibitory stabilization and cortical computation," *Nat Rev Neurosci*, vol. 22, pp. 21–37, 2021.

[134] S. Sadeh, R. A. Silver, T. Mrsic-Flogel, and R. D. Muir, "Assessing the role of inhibition in stabilizing neocortical networks requires large-scale perturbation of the inhibitory population," *J Neurosci*, vol. 37, pp. 12 050–12 067, 2017.

[135] A. Sanzeni, B. Akitake, H. C. Goldbach, C. E. Leedy, N. Brunel, and M. H. Histed, "Inhibition stabilization is a widespread property of cortical networks," *eLife*, vol. 9, p. e54875, 2020.

[136] A. Mahrach, G. Chen, N. Li, C. van Vreeswijk, and D. Hansel, "Mechanisms underlying the response of mouse cortical networks to optogenetic manipulation," *eLife*, vol. 9, p. e49967, 2020.

[137] B. K. Murphy and K. D. Miller, "Balanced amplification: A new mechanism of selective amplification of neural activity patterns," *Neuron*, vol. 61, pp. 635–648, 2009.

[138] A. Litwin-Kumar and B. Doiron, "Slow dynamics and high variability in balanced cortical networks with clustered connections," *Nat Neurosci*, vol. 15(11), pp. 1498–1505, 2012.

[139] G. Hennequin, Y. Ahmadian, D. B. Rubin, D. B. Lengyel, and K. D. Miller, "The dynamical regime of sensory cortex: Stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability," *Neuron*, vol. 98, pp. 846–860, 2018.

[140] A. Sanzeni, M. H. Histed, and N. Brunel, "Response nonlinearities in networks of spiking neurons," *PLoS Comput Biol*, vol. 16, p. e1008165, 2020.

[141] L. A. Khibnik, K. K. A. Cho, and M. F. Bear, "Relative contribution of feedforward excitatory connections to expression of ocular dominance plasticity in layer 4 of visual cortex," *Neuron*, vol. 66, pp. 493–500, 2010.

[142] G. B. Smith, A. J. Heynen, and M. F. Bear, "Bidirectional synaptic mechanisms of ocular dominance plasticity in visual cortex," *Phil Trans R Soc B*, vol. 364, pp. 357–367, 2009.

[143] H. Sompolinsky, A. Crisanti, and H. J. Sommers, "Chaos in random neural networks," *Phys Rev Lett*, vol. 61, pp. 259–262, 1988.

[144] K. C. Wood, J. M. Blackwell, and M. N. Geffen, "Cortical inhibitory interneurons control sensory processing," *Curr Opin Neurobiol*, vol. 46, pp. 200–207, 2017.

[145] H. Hu, J. Gan, and P. Jonas, "Fast-spiking, parvalbumin$^+$ GABAergic interneurons: From cellular design to microcircuit function," *Science*, vol. 345, p. 1255263, 2014.

[146] D. Kim, H. Jeong, J. Lee, J. Ghim, E. S. Her, S.-H. Lee, and M. W. Jung, "Distinct roles of parvalbumin- and somatostatin-expressing interneurons in working memory," *Neuron*, vol. 92, pp. 902–915, 2016.

[147] X.-J. Wang, J. Tegnér, C. Constantinidis, and P. S. Goldman-Rakic, "Division of labor among distinct subtypes of inhibitory neurons in a cortical microcircuit of working memory," *Proc Natl Acad Sci USA*, vol. 101, pp. 1368–1373, 2004.

[148] J. Veit, R. Hakim, M. P. Jadi, T. J. Sejnowski, and H. Adesnik, "Cortical gamma band synchronization through somatostatin interneurons," *Nat Neurosci*, vol. 20, pp. 951–959, 2017.

[149] A. Adler, R. Zhao, M. E. Shin, R. Yasuda, and W.-B. Gan, "Somatostatin-expressing interneurons enable and maintain learning-dependent sequential activation of pyramidal neurons," *Neuron*, vol. 102, pp. 202–216, 2019.

[150] S. J. B. Butt, J. A. Stacey, Y. Teramoto, and C. Vagnoni, "A role for GABAergic interneuron diversity in circuit development and plasticity of the neonatal cerebral cortex," *Curr Opin Neurobiol*, vol. 43, pp. 149–155, 2017.

[151] I. Scheyltjens and L. Arckens, "The current status of somatostatin-interneurons in inhibitory control of brain function and plasticity," *Neural Plast*, p. 8723623 (20), 2016.

[152] S. N. Tuncdemir, B. Wamsley, F. Stam, F. J. Osakada, M. Goulding, E. M. Callaway, B. Rudy, and G. Fishell, "Early somatostatin interneuron connectivity mediates the maturation of deep layer cortical circuits," *Neuron*, vol. 89, pp. 521–535, 2016.

[153] A. Shamir, O.-B. Kwon, I. Karavanova, D. Vullhorst, E. Leiva-Salcedo, M. J. Janssen, and A. Buonanno, "The importance of the NRG-1/ErbB4 pathway for synaptic plasticity and behaviors associated with psychiatric disorders," *J Neurosci*, vol. 32, pp. 2988–2997, 2012.

[154] Y. Sun, T. Ikrar, M. F. Davis, N. Gong, X. Zheng, Z. D. Luo *et al.*, "Neuregulin-1/ErbB4 signaling regulates visual cortical plasticity," *Neuron*, vol. 92, pp. 160–173, 2016.

[155] G. Bi and M. Poo, "Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength and postsynaptic cell type," *J Neurosci*, vol. 18, pp. 10 464–10 472, 1997.

[156] H. Markram, J. Lübke, M. Frotscher, and B. Sakmann, "Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs," *Science*, vol. 275, pp. 213–215, 1997.

[157] J. A. D'amour and R. C. Froemke, "Inhibitory and excitatory spike-timing-dependent plasticity in the auditory cortex," *Neuron*, vol. 86, pp. 514–528, 2015.

[158] Y.-X. Fu and Y. Dan, "Spike timing and visual cortical plasticity," in *Excitatory-Inhibitory Balance: Synapses, Circuits, Systems*, T. K. Hensch and M. Fagiolini, Eds. New York: Kluwer Academic / Plenum Publishers, 2003, ch. 16, pp. 255–267.

[159] A. Morrison, A. Aertsen, and M. Diesmann, "Spike-timing-dependent plasticity in balanced random networks," *Neural Comput*, vol. 19, pp. 1437–1467, 2007.

[160] T. Miconi, J. L. McKinstry, and G. M. Edelman, "Spontaneous emergence of fast attractor dynamics in a model of developing primary visual cortex," *Nat Commun*, vol. 7, p. 13208, 2015.

[161] G. K. Ocker, A. Litwin-Kumar, and B. Doiron, "Self-organization of microcircuits in networks of spiking neurons with plastic synapses," *PLoS Comput Biol*, vol. 11, p. e1004458, 2015.

[162] G. K. Ocker and B. Doiron, "Training and spontaneous reinforcement of neuronal assemblies by spike timing plasticity," *Cereb Cortex*, vol. 29, pp. 937–951, 2019.

[163] A. Schulz, C. Miehl, M. J. Berry II, and J. Gjorgjieva, "The generation of cortical novelty responses through inhibitory plasticity," *eLife*, vol. 10, p. e65309, 2021.

[164] J.-e. K. Miller, I. Ayzenshtat, L. Carrillo-Reid, and R. Yuste, "Visual stimuli recruit intrinsically generated cortical ensembles," *Proc Natl Acad Sci USA*, vol. 111, pp. E4053–E4061, 2014.

[165] J.-P. Pfister and W. Gerstner, "Triplets in a model of spike timing-dependent plasticity," *J Neurosci*, vol. 26, pp. 9673–9682, 2006.

[166] L. Montangie, C. Miehl, and J. Gjorgjieva, "Autonomous emergence of connectivity assemblies via spike triplet interactions," *PLoS Comput Biol*, vol. 16, 2020.

[167] M. Chistiakova, N. Bannon, J.-Y. Chen, M. Bazhenov, and M. Volgushev, "Homeostatic role of heterosynaptic plasticity: models and experiments," *Front Comput Neurosci*, vol. 9, p. 89, 2015.

[168] T. P. Vogels, H. Sprekeler, F. Zenke, C. Clopath, and W. Gerstner, "Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks," *Science*, vol. 334, pp. 1569–1573, 2011.

[169] G. La Camera, A. Fontanini, and L. Mazzucato, "Cortical computations via metastable activity," *Curr Opin Neurobiol*, vol. 58, pp. 37–45, 2019.

[170] M. E. Horn and R. A. Nicoll, "Somatostatin and parvalbumin inhibitory synapses onto hippocampal neurons are regulated by distinct mechanisms," *Proc Natl Acad Sci USA*, vol. 115, pp. 589–594, 2018.

[171] M. Udakis, V. Pedrosa, S. E. L. Chamberlain, C. Clopath, and J. R. Mellor, "Interneuron-specific plasticity at parvalbumin and somatostatin inhibitory synapses onto CA1 pyramidal neurons shapes hippocampal output," *Nat Commun*, vol. 11, p. 4395, 2020.

[172] F. Lagzi, M. C. Bustos, A.-M. Oswald, and B. Doiron, "Assembly formation is stabilized by parvalbumin neurons and accelerated by somatostatin neurons," *bioRxiv*, 2021.

[173] L. C. Garcia del Molino, G. Y. Yang, J. F. Mejias, and X.-J. Wang, "Paradoxical response reversal of top-down modulation in cortical circuits with three interneuron types," *eLife*, vol. 6, p. e29742, 2017.

[174] M. Larkum, J. Zhu, and B. Sakmann, "A new cellular mechanism for coupling inputs arriving at different cortical layers," *Nature*, vol. 398, pp. 338–341, 1999.

[175] P. Poirazi, T. Brannon, and B. W. Mel, "Arithmetic of subthreshold synaptic summation in a model CA1 pyramidal cell," *Neuron*, vol. 37, pp. 977–987, 2003.

[176] G. Mongillo, D. Hansel, and C. van Vreeswijk, "Bistability and spatiotemporal irregularity in neuronal networks with nonlinear synaptic transmission," *Phys. Rev. Lett.*, vol. 108, p. 158101, 2012.

[177] Y. K. Wu and F. Zenke, "Nonlinear transient amplification in recurrent neural networks with short-term plasticity," *eLife*, vol. 10, p. e7126, 2021.

[178] A. J. Keller, M. Dipoppa, M. M. Roth, M. S. Caudill, A. Ingrosso, K. D. Miller, and M. Scanziani, "A disinhibitory circuit for contextual modulation in primary visual cortex," *Neuron*, vol. 108, pp. 1181–1193, 2020.

[179] K. V. Kuchibhotla, J. V. Gill, G. W. Lindsay, E. S. Papadoyannis, R. E. Field, T. A. H. Sten *et al.*, "Parallel processing by cortical inhibition enables context-dependent behavior," *Nat Neurosci*, vol. 20, pp. 62–71, 2017.

[180] Y. Park and M. N. Geffen, "A circuit model of auditory cortex," *PLoS Comput Biol*, vol. 16, p. e1008016, 2020.

[181] H. Bos, A.-M. Oswald, and B. Doiron, "Untangling stability and gain modulation in cortical circuits with multiple interneuron classes," *bioRxiv*, 2020.

[182] L. Hertäg and H. Sprekeler, "Amplifying the redistribution of somato-dendritic inhibition by the interplay of three interneuron types," *PLoS Comput Biol*, vol. 15, pp. 1–29, 2019.

[183] G. B. Stanley, J. Jin, Y. Wang, G. Desbordes, Q. Wang, M. J. Black, and J.-M. Alonso, "Visual orientation and directional selectivity through thalamic synchrony," *Journal of Neuroscience*, vol. 32, pp. 9073–9088, 2012.

[184] B. Averbeck, P. E. Latham, and A. Pouget, "Neural correlations, population coding and computation," *Nat Rev Neurosci*, vol. 7, pp. 358–366, 2006.

[185] A. Torrado Pacheco, E. I. Tilden, S. M. Grutzner, B. J. Lane, Y. K. Wu, K. B. Hengen *et al.*, "Rapid and active stabilization of visual cortical firing rates across light-dark transitions," *Proc Natl Acad Sci USA*, vol. 116, pp. 18 068–18 077, 2019.

[186] C. Clopath, L. Büsing, E. Vasilaki, and W. Gerstner, "Connectivity reflects coding: A model of voltage based STDP with homeostasis," *Nat Neurosci*, vol. 13, pp. 344–352, 2010.

[187] V. Pedrosa and C. Clopath, "Voltage-based inhibitory synaptic plasticity: network regulation, diversity, and flexibility," *bioRxiv*, 2020.

[188] M. Agetsuma, J. P. Hamm, K. Tao, S. Fujisawa, and R. Yuste, "Parvalbumin-positive interneurons regulate neuronal ensembles in visual cortex," *Cereb Cortex*, vol. 28, pp. 1831–1845, 2018.

[189] R. V. Rikhye, M. Yildirim, M. Hu, V. Breton-Provencher, and M. Sur, "Reliable sensory processing in mouse visual cortex through cooperative interactions between somatostatin and parvalbumin interneurons," *J Neurosci*, vol. 41, pp. 8761–8778, 2021.

[190] B. M. Krause, C. A. Murphy, D. J. Uhlrich, and M. I. Banks, "PV+ cells enhance temporal population codes but not stimulus-related timing in auditory cortex," *Cereb Cortex*, vol. 29, pp. 627–647, 2019.

[191] M. M. Karnani, J. Jackson, I. Ayzenshtat, J. Tucciarone, K. Manoocheri, W. G. Snider, and R. Yuste, "Cooperative subnetworks of molecularly similar interneurons in mouse neocortex," *Neuron*, vol. 90, pp. 86–100, 2016.

[192] H. Xu, H. Jeong, R. Tremblay, and B. Rudy, "Neocortical somatostatin-expressing GABAergic interneurons disinhibit the thalamorecipient layer 4," *Neuron*, vol. 77, pp. 155–167, 2013.

[193] J. C. H. Cottam, S. L. Smith, and M. Häusser, "Target-specific effects of somatostatin-expressing interneurons in neocortical visual processing," *J Neurosci*, vol. 33(50), pp. 19 567–19 578, 2013.

[194] M. Pecka, Y. Han, E. Sader, and T. D. Mrsic-Flogel, "Experience-dependent specialization of receptive field surround for selective coding of natural scenes," *Neuron*, vol. 84, pp. 457–469, 2014.

[195] G. Lajoie, K. K. Lin, and E. Shea-Brown, "Chaos and reliability in spiking networks with temporal drive," *Phys Rev E*, vol. 87, p. 052901(13), 2013.

[196] A. M. Kerlin, M. L. Andermann, V. K. Berezovskii, and R. C. Reid, "Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex," *Neuron*, vol. 67, pp. 858–871, 2010.

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

[198] W. Ma, B. Liu, Y. Li, Z. J. Huang, L. I. Zhang, and H. W. Tao, "Visual representations by cortical somatostatin inhibitory neurons - selective but with weak and delayed responses," *J Neurosci*, vol. 30(43), pp. 14 731–14 379, 2010.

[199] Y. Yoshimuara and E. M. Callaway, "Fine-scale specificity of cortical networks depends on inhibitory cell type and connectivity," *Nat Neurosci*, vol. 8(11), pp. 1552–1559, 2005.

[200] D. D. Bock, W.-C. A. Lee, A. M. Kerlin, M. L. Andermann, G. Hood, A. W. Wetzel *et al.*, "Network anatomy and *in vivo* physiology of visual cortical neurons," *Nature*, vol. 471, pp. 177–182, 2011.

[201] S. B. Hofer, H. Ko, B. Pichler, J. Vogelstein, H. Ros, H. Zeng *et al.*, "Differential connectivity and response dynamics of excitatory and inhibitory neurons in visual cortex," *Nat Neurosci*, vol. 14, pp. 1045–1052, 2011.

[202] D. E. Wilson, G. B. Smith, A. L. Jacob, T. Walker, J. Dimidschtein, G. Fishell, and D. Fitzpatrick, "GABAergic neurons in ferret visual cortex participate in functionally specific networks," *Neuron*, vol. 93, pp. 1058–1065, 2017.

[203] S. J. Kuhlman, E. Tring, and J. T. Trachtenberg, "Fast-spiking interneurons have an initial orientation bias that is lost with vision," *Nat Neurosci*, vol. 14, pp. 1121–1123, 2011.

[204] M. Hemberger, L. Pammer, and G. Laurent, "Comparative approaches to cortical microcircuits," *Curr Opin Neurobiol*, vol. 41, pp. 24–30, 2016.

[205] O. Marín, "Developmental timing and critical windows for the treatment of psychiatric disorders," *Nat Med*, vol. 22, pp. 1229–1238, 2016.

[206] M. W. Antoine, T. Langberg, P. Schnepel, and D. E. Feldman, "Increased excitation-inhibition ratio stabilizes synapse and circuit excitability in four autism mouse models," *Neuron*, vol. 101, pp. 648–661.e4, 2019.

[207] V. Tatavarty, A. Torrado Pacheco, C. Groves Kuhnle, H. Lin, P. Koundinya, N. J. Miska *et al.*, "Autism-associated Shank3 is essential for homeostatic compensation in rodent V1," *Neuron*, vol. 106, pp. 769–777, 2020.

[208] T. H. Murphy and D. Corbett, "Plasticity during stroke recovery: from synapse to behaviour," *Nat Rev Neurosci*, vol. 10, pp. 861–872, 2009.

[209] M. Nahmani and G. G. Turrigiano, "Adult cortical plasticity following injury: Recapitulation of critical period mechanisms?" *Neuroscience*, vol. 283, pp. 4–16, 2014.

[210] M. Beierlein, J. R. Gibson, and B. W. Connors, "Two dynamically distinct inhibitory networks in layer 4 of the neocortex," *J Neurophysiol*, vol. 90, no. 5, pp. 2987–3000, 2003.

Acknowledgements

I would like to thank my supervisor Julijana Gjorgjieva for her constant and passionate support and guidance throughout my PhD. I have learned invaluable lessons about the processes of scientific research from her, from designing a project to presenting the results and everything that has to happen in-between. And I want to express my gratitude to her for the incredible opportunities to learn from other scientists and share my results on multiple occasions throughout my PhD, especially during a specialized workshop at the Janelia Research Campus, during the Cajal Course in Computational Neuroscience, and through a visit at Brandeis University.

Special thanks to our collaborators at Brandeis University, Gina Turrigiano and Nathaniel Miska, with whom to discuss monocular deprivation, homeostatic plasticity, and the individual idiosyncrasies of rats and mice was a great pleasure. I have learned a lot in this collaboration, and I am proud that I could contribute to our collaborative work.

I also want to thank the members of my thesis advisory committee, Matthias Kaschube, and Johannes Letzkus. The insightful discussions we had during our meetings were of great value to the progress of my research. Furthermore, I also want to thank Christian Müller for his knowledgeable advice on neurophysiology across the tree of life and his ideas during many informal discussions.

For the many interesting discussions and the uncountable instances in which I could find help when facing a scientific problem, I furthermore want to thank all members of the *Computation in neural circuits* group at the Max Planck Institute for Brain Research. Special thanks to Sebastian Onasch, Jan Kirchner, and Yue Kris Wu who provided feedback on my thesis. A big thank you also to Marcel Jüngling. I am grateful for the exciting work we could do together on STDP, and I thank him for his input on my own work in this direction. Also, a big thank you to Kai Röth, especially for his support during the finalization of my thesis.

Finally, I want to thank Anna for her understanding when deadlines came close and her constant effort to keep me connected with the world outside of science, and many other things. Without her support, I could not have pursued my research the way I did. Last but not least, I want to thank my Family, Michael, Raphael, David, and Elissa, along with their spouses Beate, Daniela, and Kathrin, and my uncle Joachim, who all got involved in one way or the other in the progress of my research. I dedicate this thesis to the memory of my mother, Michaela.