Stability of dynamics and memory in the balanced state

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Paul Manz
aus
Wilhelmshaven

Bonn, 2022

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter:	Prof. Dr. Raoul-Martin Memmesheimer
2. Gutachter:	Prof. Dr. Markus Diesmann
Tag der Promotion:	17.02.2023
Erscheinungsjahr:	2024

Abstract

Computational modeling of neural circuits has successfully explained the observed irregular and asynchronous activity in the brain as the result of a dynamical balance of excitatory and inhibitory inputs to individual neurons. In this balanced state, the activity of each neuron is governed by fluctuations rather than the mean of net input resulting in irregular spiking. The balanced state raises some questions concerning stability, however: First, the high dimensional and irregular activity seems to imply chaos, which may be at odds with stably representing information as sequences of spike times. A second issue arises when synaptic strengths are subject to activity-dependent change called synaptic plasticity: it is not clear how the stability of synaptic weight patterns, which are believed to encode memories, are maintained during irregular activity.

We first study the dynamical stability and phase space structure of balanced networks of inhibitory neurons with external excitatory input. We consider two types of neurons: standard leaky and novel antileaky integrate-and-fire neurons, which accelerate toward the threshold. We determine the voltage probability distributions and self-consistent firing rates of networks with both neuron types. Further, we compute the full spectrum of Lyapunov exponents (LEs) and the covariant Lyapunov vectors (CLVs). While networks with only leaky integrate-and-fire neurons are dynamically stable, we find that there is approximately one positive LE for each antileaky integrate-and-fire neuron in a network, indicating chaos. A simple mean-field approach, which can be justified by properties of the CLVs, explains this finding. As an application, we propose a spike-based computing scheme where our networks serve as computational reservoirs, and their different stability properties yield different computational capabilities.

We then study how strongly interconnected groups of neurons, called assemblies, which may encode memories, can remain stable during balanced state activity. Hebbian plasticity, which strengthens the connections of neurons that receive correlated input can reinforce connections within existing assemblies but is unstable on its own. Previous models of assemblies require additional mechanisms of fast homeostatic plasticity, often with biologically implausible timescales, to stabilize Hebbian plasticity. We provide a model of neuronal assembly generation and maintenance purely based on spike-timing-dependent plasticity (STDP) between excitatory neurons. It uses stochastically spiking neurons and STDP that depresses connections of uncorrelated neurons. We find that assemblies do not grow beyond a certain size, because temporally imprecise spike correlations dominate plasticity in large assemblies. We also demonstrate that assemblies in our model can generate and maintain prominent and stable overlap structures. Our model can furthermore exhibit representational drift, where assemblies over a slow timescale exchange neurons with each other. Finally, the model indicates that assembly size is inversely related to the density of connectivity.

Acknowledgements

I am greatly indebted to my supervisor Raoul-Martin Memmesheimer for introducing me to computational neuroscience and giving me the opportunity to work in this exciting and fascinating field of research, for his patient guidance and commitment to answering all my questions throughout these years. I would also like to thank Markus Diesmann for agreeing to be my second reviewer.

I am grateful to Sven Goedeke for fruitful collaboration in Chapter 3, useful comments on the manuscript of Chapter 4, for proofreading this manuscript, and for lots of other helpful advice. I am also grateful to Christian Klos for ongoing collaboration, technical advice, and having been an enjoyable office mate. Thanks also goes to my other long-term colleagues Felipe Kalle Kossio, Wilhelm Braun, Paul Züge, Simon Altrogge, and all other present and former members of the Neural Network Dynamics and Computation group for many stimulating discussions both at work and during lunch breaks, for all the helpful comments on my work, and for creating a pleasant work environment.

I would also like to thank the other members of the Institute of Genetics for accepting our theory-heavy group into their fold and for having provided us with interim offices when we got started. Special thanks goes to Simone Christian and Stefan Klein for all their valuable assistance.

After more than a decade the time has now finally come for me to leave the University of Bonn. I would therefore very much like to thank all the lecturers in physics and their assistants who have helped me appreciate the beauty of the subject and all others involved in making the system work. I hope many more generations of physicists will get to enjoy such high standards of education.

Finally, I would like to thank my friends, my family, and Urmi Ninad for all the warmth and generosity.

List of Publications

[1] **P. Manz**, S. Goedeke, and R.-M. Memmesheimer
Dynamics and computation in mixed networks containing neurons that accelerate towards spiking
Phys. Rev. E **100** (2019) 042404

[2] **P. Manz** and R.-M. Memmesheimer
Purely STDP-based assembly dynamics: stability, learning, overlaps, drift and aging
PLOS CB **19** (2023) e1011006

Contents

Abstract iii

Acknowledgements v

List of Publications vii

1 Introduction 1

2 Fundamentals 5
- 2.1 Biological neuron models . . . 5
- 2.2 Leaky integrate-and-fire neurons . . . 6
- 2.3 Synapses . . . 7
  - 2.3.1 Current-based synapses . . . 7
  - 2.3.2 Synaptic plasticity . . . 8
- 2.4 Balanced state . . . 10
- 2.5 Hawkes model . . . 12
  - 2.5.1 Cross-correlations . . . 12
- 2.6 Dynamical stability . . . 13
  - 2.6.1 Lyapunov exponents . . . 13
  - 2.6.2 Numerical computation of Lyapunov exponents . . . 14
  - 2.6.3 Covariant Lyapunov vectors . . . 16
  - 2.6.4 Numerical Computation of CLVs . . . 17

3 Dynamics and computation in mixed networks containing neurons that accelerate towards spiking 19
- 3.1 Introduction . . . 21
- 3.2 Mixed networks of neurons with concave and convex rise function . . . 22
- 3.3 Network firing rate and membrane potential distributions . . . 25
- 3.4 Growth of dynamical perturbations . . . 29
  - 3.4.1 Mean-field approach . . . 29
  - 3.4.2 Network single spike Jacobian . . . 33
  - 3.4.3 Volume contraction . . . 34
  - 3.4.4 Numerical computation of the Lyapunov spectrum . . . 34
- 3.5 Stable and unstable directions . . . 35
  - 3.5.1 Lyapunov vectors and perturbation growth . . . 35

3.5.2 Stable and unstable directions in mixed networks . . . 36
3.6 Computations with precisely timed spikes . . . 38
3.6.1 Network architecture and task design . . . 38
3.6.2 Switchable temporal XOR/AND . . . 39
3.6.3 Detect or ignore input time differences . . . 41
3.7 Discussion . . . 43
3.8 Appendices . . . 45
3.8.A Voltage probability distribution of LIF neurons . . . 45
3.8.B Mean-field Lyapunov exponents . . . 47
3.8.C Network single spike Jacobian . . . 49
3.8.D Volume contraction . . . 49
3.8.E Dependence of the Lyapunov spectrum on indegree and network size . . . 50
3.8.F Computing covariant Lyapunov vectors . . . 52

4 Purely STDP-based assembly dynamics: stability, learning, overlaps, drift and aging 53
4.1 Introduction . . . 54
4.2 Materials and methods . . . 55
4.2.1 Poisson neurons . . . 55
4.2.2 Spike-timing-dependent plasticity . . . 57
4.3 Results . . . 58
4.3.1 Spontaneous assembly formation . . . 58
4.3.2 Plasticity in homogeneously connected assemblies . . . 59
4.3.3 Storing new assemblies . . . 62
4.3.4 Overlapping Assemblies . . . 63
4.3.5 Drifting assemblies . . . 65
4.3.6 Aging . . . 66
4.4 Discussion . . . 68
4.5 Appendices . . . 70
4.5.A Time-averaged weight change in fully connected assemblies . . . 70
4.5.B Asymptotic behavior of motif contributions to time-averaged plasticity . . . 73
4.5.C Parameters . . . 77
4.5.D Additional figures . . . 78

5 Summary and outlook 81

Bibliography 85

CHAPTER 1

Introduction

Understanding the brain is a monumental task, given the complexity of its structure and its extensive repertoire of functions. It will require work on a wide range of scales, from biochemistry and molecular biology to the scale of entire brain areas. The past century has seen significant progress in understanding the fundamental functional units of the brain: neurons, specialized cells that communicate via short electrical pulses called action potentials or spikes through junctions called synapses.

One of the major areas of ongoing work in neuroscience is to explain how large networks of interconnected neurons can give rise to the brain's ability to process sensory information and use it to make decisions and control behavior. While experiments in systems neuroscience can simultaneously sample the spiking activity of ever greater numbers of neurons, on the theoretical side, computational neuroscience serves to describe this data, create models to elucidate the mechanisms generating it, and provide interpretations for the functional reasons behind these mechanisms. One of the most important modes of operation in computational neuroscience is the study of mathematical models of populations of interconnected neurons, which we call neural networks[1]. Mathematically, they are dynamical systems that can be described by discrete maps and ordinary differential equations. The theory of dynamical systems is well-established in physics and mathematics, making computational neuroscience an interdisciplinary field of research.

A prime example of successful use of the computational modeling approach is the explanation for the experimentally observed irregularity of activity found, for instance, in neurons in the cortex. Given that each cortical neuron receives signals from many other neurons, one would expect that these signals sum to a smooth and steady input resulting in regular periodic instead of irregular output spiking. The solution to this apparent paradox is the balanced state, in which inputs that excite and inhibit neurons cancel each other out on average, resulting in a mean input that is not sufficient to drive spiking by itself [3–5]. Instead, spiking is driven by fluctuations around the mean, resulting in the observed irregular state. Notably, a simple discrete-time model, in which neurons are represented by binary units with an on-state (spiking) and an off-state (not spiking), was sufficient to demonstrate this effect. This highlights how often simple models, which ignore many of the biological intricacies of real neurons can be powerful enough to represent the dynamical features under consideration. Simple

[1] Throughout this thesis, we discuss biological neural networks, which model the activity of brain circuits. These need to be distinguished from artificial neural networks used in artificial intelligence, which are inspired by but often bear little resemblance to the former. We always refer to biological neural networks when using the term "neural network".

models have the two-fold advantage of making the systems computationally and analytically more tractable. This helps us to deepen our theoretical understanding of the observed phenomenon.

In this thesis, we set out to advance our understanding of stability in balanced state networks. We define stability as the ability of a system to maintain or restore its properties in the presence of perturbing influences. This becomes especially relevant for the balanced state: since it is governed by fluctuations, perturbations are intrinsic to the system, and properties of importance must be resistant to them. In the brain, stability is necessary to preserve information. We focus here on two levels of information: the first level is short-term information encoded in the spiking activity of neurons which can be used in the present and the second level is long-term information stored in patterns of synaptic strengths which can be recalled at later times.

In Chapter 3 we address the subject of dynamical stability: irregular activity in high-dimensional nonlinear dynamical systems such as in balanced state networks is often associated with chaos, a phenomenon ubiquitous in nature characterized by strong sensitivity of a dynamical trajectory to its initial conditions [6]. In a chaotic neural network, incoming information is quickly lost since any intrinsic noise present will have an outsize effect on the network's state at a later time; the same input would lead to entirely different outputs, making the network output unreliable and unsuitable for computation. On the other hand, too much stability may limit a network's ability to discriminate between similar inputs, as the network's response would converge to the same output leading to stereotyped behavior. To deal with this tradeoff, it is therefore often assumed that networks should operate at the "edge of chaos" [7]. Previous studies have proven, numerically and analytically, that networks of pulse-coupled inhibitory leaky integrate-and-fire (LIF) neurons are not sensitive to small changes in initial conditions despite their irregular spike patterns [8, 9]. Furthermore, introducing recurrent excitatory interactions to these networks leads to a gradual shift to chaos. In contrast to a gradual emergence of chaos, in Chapter 3 we introduce a novel type of model neuron that is sufficient to induce chaos if even one such neuron is present in a network that would otherwise be stable. This type of neuron, dubbed "XIF", represents neurons that are dynamically balanced at a level where the voltage would start accelerating toward the threshold, a feature the standard leaky integrate-and-fire neuron does not account for. We study mixed networks of both standard LIF and the novel XIF neurons; with this, we pay tribute to the fact that networks in the brain consist of different neuron types, inhibitory neurons being particularly diverse. We investigate these mixed networks using Lyapunov exponents, which quantify sensitivity to initial conditions [10]. To understand how single neuron properties give rise to chaos, we compare the results of the numerical computation with that of a mean-field ansatz we developed for our networks. To further understand the relationship of the neuron types with stable and unstable phase space directions, we investigate the covariant Lyapunov vectors of the system which indicate the directions with more or less sensitivity to initial conditions. We then use the insights we gained to develop a scheme for computation with precise spike timings that harnesses the properties of both neuron types.

In Chapter 4, we study the stability of assemblies encoding long-term memory. To allow accurate recall years after their creation, long-term memories need to be structurally encoded in brain circuits. The primary candidate for such a structure is patterns of synaptic strengths. In particular, groups of neurons can have strengthened interconnections allowing them to coactivate: we call these groups memory assemblies. The creation of memory assemblies is made possible by synaptic plasticity: the ability of synapses to undergo lasting change depending on the activity of the neurons they connect. Since plasticity depends on the activity and the activity is driven by fluctuations, we also expect fluctuations in the weights. We, therefore, expect some mechanism to stabilize the memories

against these random fluctuations; otherwise, memories would quickly fade away as random noise in the connections accumulates over time. Hebbian plasticity, in which the synaptic strength between correlated neurons increases [11], provides precisely such a mechanism: the correlations induced by existing memories induce plasticity, reinforcing these memories. Hebbian plasticity alone is unstable, however: unchecked it causes individual weights to grow beyond bounds on the one hand and assemblies to expand to more and more neurons on the other hand. Both of these problems lead to abnormally high firing rates and therefore prevent the formation of stable memories. To address the first problem, we can assume a physiological upper bound for the strength of individual synapses. For the second problem, some more or less ad hoc homeostatic plasticity mechanisms have been introduced to induce competition between weights [12–14]. Homeostatic plasticity with the time scale needed to prevent pathological growth of assemblies is biologically implausible, however [15]. Here we instead propose a plasticity rule depending on the relative timings of spikes that is by itself intrinsically stable. Our analysis shows that with our spike-timing-dependent plasticity rule, the typical scale of time lags between spikes in larger assemblies prevents further growth. We then test if our model can overcome some other problems and challenges related to assemblies with homeostatic plasticity. One such problem is the difficulty of having neurons present in multiple assemblies [16–18]. Such overlaps between assemblies may encode associations between different memory items as they make it more likely for the activation of one memory to trigger another. While memories in the brain are stable over long timespans, they need not be represented by the same set of neurons over time, but representations may drift instead [19]. In a hypothesized mechanism, neurons may, on a slow time scale, lose their connections to an assembly and connect to another, independent of any external signal [13]. This gradual exchange of neurons may cause an assembly to eventually only share a chance-level amount of neurons with its original configuration. The process can, however, happen slowly enough that we can reliably track an assembly's identity over time – despite the drift in neural representation, the underlying memory is stable. We show that such representational drift is possible in our assembly model, facilitated by the irregularity of the balanced state. Finally, the model predicts that when the number of synaptic connections in the brain decreases, as observed in aging [20, 21], the size of the neuron ensembles underlying memories increases. This process may render memories in the aging brain more robust and prominent but also less specific.

CHAPTER 2

Fundamentals

This chapter introduces fundamental concepts of computational neuroscience, and their basis in neurobiology, mainly following the textbooks by Dayan and Abbott [22] and Gerstner et al. [23]. In addition, we introduce some known computational techniques that we employ in the following chapters. We begin by giving a brief overview of the mechanisms governing biological neurons to motivate and establish the leaky integrate-and-fire neuron as a model for spiking neurons and the current-based synapse as a model for interactions between them. Next, we discuss the balanced state in which the average excitation and inhibition each neuron receives cancel out. It is characterized by asynchronous and irregular spiking, motivating us to introduce the stochastically spiking Hawkes neuron model. We then discuss synaptic plasticity, particularly spike-timing-dependent plasticity (STDP), and derive some basic results of STDP in networks of Hawkes neurons. Finally, we introduce concepts from the ergodic theory of dynamical systems, namely Lyapunov exponents and covariant Lyapunov vectors, mostly following the textbook by Pikovsky and Politi [10]. We then present algorithms for their numerical computation by Benettin et al. [24] and Ginelli et al. [25].

2.1 Biological neuron models

The brain's constituent cells can be divided into two main classes: neurons and glial cells. The latter establishes a support system for neurons and play, if at all, only a subordinate role in the actual computations. The primary information processing units of the brain are then neurons, cells that are specialized to transmit electrical signals. While there are many types of neurons with different morphologies across different species, we focus here on the types of neurons found in the mammalian cortex. These are structured into three parts: axon, soma, and dendrites. Dendrites are branching filaments that gather incoming signals from other neurons and transmit them to the soma. The soma is the neuron's cell body, containing its nucleus and performing the most critical metabolic functions. It is crucial for generating action potentials from inputs it receives from the dendrites. When the soma generates an action potential, it travels along the axon, a long fiber that connects via synapses to the dendrites of other neurons.

Differences in ion concentration of intra- and extracellular fluid induce a characteristic voltage between the inside and outside of a neuron's cell membrane. These ion concentrations are controlled by active and passive ion channels in the cell membrane whose behavior depends on the voltage and

the ion concentrations themselves. While the membrane voltage may not be the same in all of a neuron's regions, in order to describe its dynamics, it is often sufficient to model the voltage by a single variable. Neuron models that neglect the spatial extent in this way are called point neurons, and we restrict ourselves to these for the remainder of this thesis. Taking each ion concentration and their respective channels into account, we can describe a point neuron using a set of ordinary differential equations. This way, we can model the most important aspect of single-neuron dynamics, the generation of action potentials, or spikes: if the membrane voltage exceeds a specific voltage, called the threshold, a positive feedback loop is put in motion causing a rapid increase in membrane voltage followed by a rapid decrease to a level usually below the resting potential. This phenomenon has a characteristic voltage trajectory independent of the preceding activity. It is, therefore, not the shape of the spikes themselves that carries information but rather their rate or timing.

A series of spike times of a neuron is called a spike train, which we often express as a sum of Dirac delta distributions,

$$S(t) = \sum_k \delta(t - t_k), \tag{2.1}$$

for spike times t_k. We can think of S as a density of spikes for a given neuron in a given trial.

While the term firing rate is ubiquitous in neuroscience, it may refer to one of several different concepts: the convolution of the spike train $S(t)$ with a moving time window, the spike count over a long duration divided by the duration, or the trial or ensemble average of a spike train. In Chapter 3, we study static networks, for which we compute the long-term average firing rates. On the other hand, in Chapter 4, the networks themselves are subject to change, and firing rates refer to trial-averaged rates in that chapter.

2.2 Leaky integrate-and-fire neurons

Because of the stereotyped nature of spikes, a detailed biophysical model of a neuron's ion channel dynamics is often not necessary. Instead, it is often sufficient to model a neuron's membrane as a leaky capacitor, as shown in Fig. 2.1: the resistor represents ions leaking through passive channels, which in the absence of external currents restores the voltage V in the membrane to its resting value V_0, which is maintained by ion pumps in the membrane. These are represented by a battery connected to the capacitor. Together, the leak resistance R and the capacity C of the membrane give rise to a membrane time constant $\tau = RC$. In our convention we set $R = 1$, such that voltages and currents have the same units. Taken together, the subthreshold dynamics given an external current $I(t)$ obeys a single linear ordinary differential equation,

$$\tau \dot{V}(t) = -(V(t) - V_0) + I(t). \tag{2.2}$$

Since the duration of spikes themselves is much shorter than the characteristic time scale of network dynamics, we model them as instantaneous: We introduce an explicit rule into our model that instantaneously sets V to V_{reset} whenever it reaches the firing threshold V_{thr}. This rule, together with Eq. (2.2) defines the leaky integrate-and-fire neuron (LIF), a simple yet powerful spiking neuron model. In Chapter 3 we study networks containing a modified version of the LIF neuron in particular with regard to the dynamical stability of such networks.

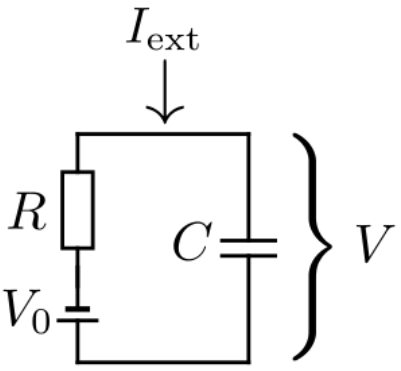

Figure 2.1: Circuit diagram for the subthreshold dynamics of a leaky integrate-and-fire neuron. External current I_{ext} charges the membrane, shown as a capacitor with capacity C. Ion pumps slowly, via a leak resistance R restore the voltage to the neuron's resting potential V_0.

2.3 Synapses

Neurons communicate with each other through connecting structures called synapses. The primary type of synapse in the mammalian brain is the chemical synapse which consists of a presynaptic terminal and a postsynaptic bouton separated by a gap called the synaptic cleft. A spike of the presynaptic neuron triggers a release of neurotransmitters into the synaptic cleft that then bind to receptors of the postsynaptic bouton. These neurotransmitters can activate ion channels and, depending on their type, induce a positive or negative current into the postsynaptic neuron. In the former case, the synapse is called excitatory; in the latter case, it is called inhibitory. Neurons typically release the same kind of neurotransmitters at all their presynaptic terminals and thus act either excitatorily or inhibitorily on all of their postsynaptic partners, an observation called Dale's law [26]. Thus we can classify neurons as excitatory or inhibitory depending on their effect on postsynaptic neurons. The effect of a synaptic input is proportional to the synapse's strength which in our models is given by a single real variable. We can then portray the network structure as a weighted directed graph, fully described by a square matrix W, called weight matrix. In our convention, the entry W_{ij} denotes the strength of the connection of neuron j to neuron i.

2.3.1 Current-based synapses

A presynaptic spike causes a rapid change in a postsynaptic neuron's membrane conductance g. The change in g induces a current into the neuron resulting in a shift in its membrane potential called postsynaptic potential (PSP). The current-based synapse is a simplified model in which the effect of a spike on a postsynaptic neuron does not depend on the neuron's voltage. We can describe the action of such a synapse by a time-dependent current I_{ij} into the postsynaptic neuron i, which we write as

$$\tau_{\text{syn}} \frac{\mathrm{d}I_{ij}(t)}{\mathrm{d}t} = -I_{ij}(t) + W_{ij}\tau_{\text{syn}} S_j(t), \tag{2.3}$$

for a synapse of strength W_{ij} and a presynaptic spike train $S_j(t)$. In Chapter 3 we simplify synaptic currents by taking the limit $\tau_{\text{syn}} \to 0$ while keeping the overall induced electric charge constant. In

this limit, the postsynaptic current takes the shape of a delta distribution. With Eq. (2.2) this means that a presynaptic spike induces a jump in the voltage of the postsynaptic neuron; the synaptic weight scales the height of the jump.

In Chapter 4, we use synaptic currents analogous to Eq. (2.3). With the Hawkes model we use there, instead of charging or discharging a postsynaptic neuron's membrane voltage, the currents temporally change the neuron's spiking probability.

2.3.2 Synaptic plasticity

Animals need to be able to adapt to changing environments. Such adaptation includes learning to recognize previously unknown stimuli, changing the behavioral response to known stimuli, and learning to generate new behaviors. Consequently, brain circuits need to be able to change their response patterns to incoming information. A vital feature subject to change, or plasticity, is the strength of individual synapses. These undergo transient changes on a fast timescale, called short-term plasticity (STP), and changes on a slower timescale which may last indefinitely, called long-term potentiation (LTP) and long-term depression (LTD), respectively. Here, we focus on LTP and LTD and disregard STP.

If plasticity is a response to incoming information, it needs to depend on the neural activity which encodes the information. The idea of activity-dependent plasticity was first proposed by Hebb, who famously suggested that if a presynaptic neuron repeatedly plays a causal role in the spiking of a postsynaptic neuron, the strength of the synapse under consideration will increase [11]. Conversely, if presynaptic spikes repeatedly fail to elicit spikes in the postsynaptic neuron, the considered synapse will weaken. Hebb then postulated that such plasticity could elicit the organization of networks into cell assemblies, groups of neurons that are interconnected in a way that facilitates their collective reactivation. These assemblies could, when reactivated, perform actions or recall learned percepts or concepts. To date, the precise nature and functional properties of neuronal assemblies are still a matter of ongoing debate [27, 28].

Experiments have shown a lasting increase in synaptic strength if presynaptic spiking induces postsynaptic spiking [29], and, in addition, that the sign and strength of LTD depend on the relative timing of pre-and postsynaptic spikes [30, 31] leading to the proposal of spike-timing-dependent plasticity (STDP), in which changes in synaptic strength are a function of the spike times of pre- and postsynaptic neurons. The most common model of STDP considers only the time differences of pairs of pre- and postsynaptic spikes. These models capture some key features of plasticity, like receptive field formation; however, they fail to reproduce the dependence on the frequency of spike pairs [31], and the results of experiments with spike-triplet protocols [32–34]. These results can be recovered with higher order STDP rules such as triplet rules [35]. We focus only on pair-based STDP in which each spike pairing of the pre- and postsynaptic spike trains contributes to the synaptic strength via a function $F(\Delta t)$ of the time lag between pre- and postsynaptic spike, called the STDP function or STDP window. At each spike time, plasticity acts additively on the pre- and postsynaptic weights of the spiking neuron, with amplitudes given by F; this update rule can be written as:

$$\frac{\mathrm{d}}{\mathrm{d}t}W_{ij}(t) = \sum_{t_{k_i}, t_{k_j} \leq t} \left(F(t_{k_i} - t_{k_j})\delta(t - t_{k_i}) + F(t_{k_i} - t_{k_j})\delta(t - t_{k_j}) \right). \tag{2.4}$$

Fig. 2.2 shows an experimentally observed symmetric STDP window found in the CA3 region of the

hippocampus by triggering series of spike pairs with different time lags and observing the change in EPSP amplitude. In simulations Eq. (2.4) involves evaluating $F(\Delta t)$ at each spike time of neuron i

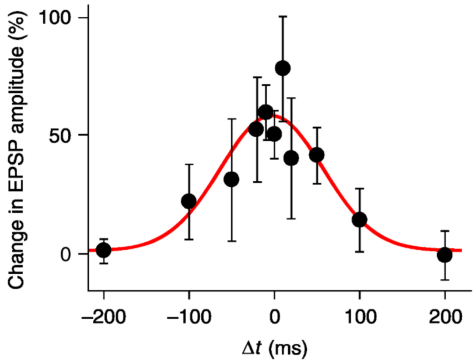

Figure 2.2: Experimentally observed STDP function in CA3 region, showing the relative change of LTP strength with respect to the time interval of pre- and postsynaptic spikes. Figure adapted from [36].

for all spikes of neuron j and vice-versa. This becomes unfeasible for longer simulations of large networks. We can lift this computational burden if we express the STDP window as a superposition of decaying exponentials with time constants $\tau_{+,\alpha}$ and $\tau_{-,\beta}$ and amplitudes $A_{+,\alpha}$ and $A_{-,\beta}$, which is often a natural choice to fit experimental data. Splitting F into a causal ($\Delta t > 0$) and an anticausal ($\Delta t < 0$) part we write

$$F(\Delta t) = \begin{cases} \sum_{\alpha} A_{+,\alpha}\, \mathrm{e}^{-\Delta t/\tau_{+,\alpha}}, & \Delta t > 0 \\ \sum_{\beta} A_{-,\beta}\, \mathrm{e}^{\Delta t/\tau_{-,\beta}}, & \Delta t < 0 \end{cases} . \tag{2.5}$$

We then introduce auxiliary variables $x_{i,\alpha}(t)$ and $y_{i,\beta}(t)$ for the exponentials in F, such that

$$\tau_{+,\alpha} \frac{\mathrm{d}x_{i,\alpha}(t)}{\mathrm{d}t} = -x_{i,\alpha}(t) + A_{+,\alpha} \sum_k \delta(t - t_i^k) \tag{2.6}$$

$$\tau_{-,\beta} \frac{\mathrm{d}y_{i,\beta}(t)}{\mathrm{d}t} = -y_{i,\beta}(t) + A_{-,\beta} \sum_k \delta(t - t_i^k). \tag{2.7}$$

We can then write the update to W_{ij} as

$$\frac{\mathrm{d}}{\mathrm{d}t} W_{ij}(t) = \sum_{\alpha} x_{j,\alpha}(t) \sum_k \delta(t - t_i^k) + \sum_{\beta} y_{i,\beta}(t) \sum_k \delta(t - t_j^k). \tag{2.8}$$

Thus, at each spike of neuron i the coupling W_{ij} is updated according to the sum of the causal synaptic traces of neuron j and at each spike of neuron j it is updated with the sum of the anticausal traces of neuron i. While these traces may come with biophysical interpretations, such as calcium entry due to backpropagating action potentials [37], we here treat them as auxiliary variables of a phenomenological plasticity rule.

2.4 Balanced state

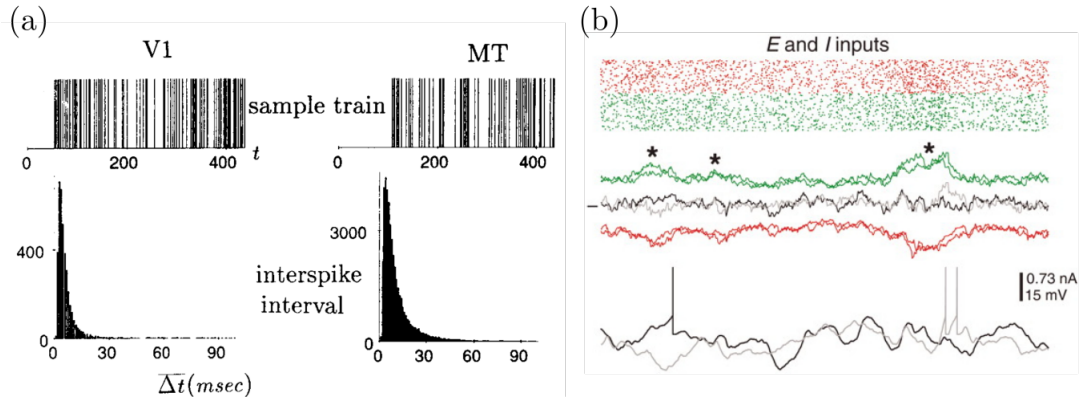

Figure 2.3: Irregular spiking in the balanced state. (a): Upper: irregular spike trains in primary visual and motor cortices of awake, behaving macaque monkeys. Lower: distributions of interspike intervals in these areas showing approximately exponential shapes, in agreement with those of Poisson point processes. (b): Asynchronous and irregular spiking in densely connected simulated network: Two neurons with shared inputs. Top: activity of excitatory (green) and inhibitory (red) presynatpic populations. Middle: excitatory, inhibitory and net (black and gray) inputs currents. Bottom: membrane voltages of the two neurons. The balance of excitation and inhibition decorrelated the shared inputs and leads to asynchronous activity between the two neurons. Panel (a) adapted from [38], panel (b) adapted from [39].

Neurons in the mammalian cortex typically receive inputs from a large number (in the order of 1000-10000 [40]) of other neurons in the same network. For self-consistency, the output rate of a neuron should have the same scale as the rates of the incoming ones; it thus seems reasonable to assume that the strength of each input connection is small, i.e., that it scales with $1/K$, with K the typical network indegree of a neuron. Assuming that these inputs are independent, the standard deviation of the summed inputs then scales with $1/\sqrt{K}$ while the total input remains constant. For large K, this implies that activity is driven by the mean input, leading one to predict that firing in the cortex is regular and periodic. Experiments, in contrast, have shown that the mammalian cortex exhibits highly irregular spiking activity: the spikes of a single neuron appear at seemingly random intervals [38, 41]. Formally, the spike trains resemble a Poisson process in which the time intervals between spikes are exponentially distributed, with a coefficient of variation close to 1, see Fig. 2.3a.

Intrinsic stochasticity of spike generation in individual neurons cannot account for this irregularity as neurons respond fairly reliably to the same time-varying input [42–44]. The solution to this apparent paradox lies in postulating that the inhibitory and excitatory inputs to each neuron are large but balanced, such that the mean current is not enough to drive the voltage up to the threshold. Spiking is instead driven by fluctuations around the mean input current which causes the observed irregular activity [4]. Networks can self-organize to such a balanced state for a wide range of network parameters if both excitatory and inhibitory weights are strong ($\propto 1/\sqrt{K}$), as shown in [5]. This model consists of excitatory and inhibitory populations of binary neurons with thresholds Θ_E and Θ_I recurrently connected to themselves and each other. At each time step the neurons' activity variable is set to 1 (spiking) if their inputs exceed their threshold or 0 (not spiking) otherwise. Both populations receive large constant external input $\sqrt{K}\Lambda_{0,E}$ and $\sqrt{K}\Lambda_{0,I}$. Connections between two neurons occur with a

probability of K/N, with strengths $1/\sqrt{K}J_{AB}$ where $A, B \in \{E, I\}$ denoting excitatory and inhibitory populations respectively. The model admits a mean-field theory that becomes exact in the large N limit where neurons receive mean input

$$\begin{aligned} u_I &= \sqrt{K}\left(\Lambda_{0,I} + J_{IE}m_E - J_{II}m_I\right) - \Theta_I \\ u_E &= \sqrt{K}\left(\Lambda_{0,E} + J_{EE}m_E - J_{EI}m_I\right) - \Theta_E. \end{aligned} \tag{2.9}$$

Here m_A is the average firing rate of population A. In the large K limit, if the terms in the brackets add up to a positive value, neurons will fire at their maximum firing rate, whereas if the terms are negative, neurons will be quiescent. Such unbalanced solutions can be rendered inconsistent with appropriate constraints on J_{AB}. With these constraints, the activity must settle to a state where the terms in the brackets almost cancel, resulting in mean input that is not sufficient to drive spiking. Then, fluctuations rather than the mean input drive spiking, and spiking will be irregular. A necessary condition for the mean-field solution of Eq. (2.9) is that the correlation of a neuron's activity with its outgoing connections is small. This holds for random networks with sparse connectivity ($K/N \to 0$), as in [5]. As shown in [39] however, it is also possible to achieve an asynchronous and irregular state in dense networks with $K/N = \text{const}$: if inhibition is sufficiently fast and strong, it can cancel correlations that arise due to shared inputs, see Fig. 2.3b. Whereas both [5] and [39] use the limit $N \to \infty$, the authors of [45, 46] have shown that this limit is not required to achieve decorrelation: the presence of inhibitory feedback which is ubiquitous in balanced networks is sufficient to suppress correlations. The theoretical underpinnings of the balanced state are thus valid also in the biological realm with $K \approx 10^4$ and $N \approx 10^6$. After being initially demonstrated in networks of binary networks, [5], the balanced state has subsequently been studied with leaky integrate-and-fire neurons [47, 48]. Since then, experiments have demonstrated the existence of E-I balance in multiple brain areas, such as the cortex [49–51] and the hippocampus [52].

The existence of the balanced state seems puzzling: it implies that most spikes only serve to neutralize other spikes. Since each spike comes with a metabolic cost, this seems wasteful, given that the brain consumes a sizeable proportion of an animal's available energy. Therefore, the balanced state likely has some functional advantage that justifies this drawback. One such advantage may be higher responsiveness: if many neurons' membrane potentials are balanced at a level close to spiking, they can be quickly brought to spiking when external inputs change. Furthermore, a balanced network can have a much wider dynamical range, since an excitation dominated network might quickly reach saturation firing rates in response to increasing input. Another notable feature of balanced state networks is the linearity of population activity with respect to the input even if individual neurons are highly nonlinear. This feature may also contribute to wider dynamical range.

In Chapter 3, we investigate networks of recurrently coupled inhibitory spiking neurons that receive a constant excitatory current. Such a setup can also generate a balanced state where the mean recurrent inhibitory input cancels the constant excitatory drive. Then, in Chapter 4, we simply assume that the system is in the fluctuation-driven balanced state and not explicitly model inhibition. Based on the balanced state assumption, we model each excitatory neuron's spiking activity as an inhomogeneous Poisson process with an intensity dependent on the input it receives from other neurons in the network.

2.5 Hawkes model

If we take the balanced state as given, we can propose a neuron model whose spiking activity is irregular by design. We model spiking as resulting from a stochastic point process rather than being deterministically determined by integrating inputs up to a threshold. Consider an inhomogeneous Poisson process in which the probability to emit a spike in an infinitesimal time interval $[t, t + \mathrm{d}t]$ is given by an intensity function $\lambda(t)\mathrm{d}t$. We choose the intensity function to have a constant finite baseline value λ_0, representing spontaneous background activity due to external inputs. Spiking input into a neuron has the effect of transiently raising the probability of a spike by a synaptic kernel function $a(t)$, scaled by the synaptic weight. We define $a(t)$ as exponentially decaying, in analogy to the synaptic currents Eq. (2.3),

$$a(t) = \frac{\Theta(t)}{\tau_\mathrm{s}} \mathrm{e}^{-\frac{t}{\tau_\mathrm{s}}}, \tag{2.10}$$

where $\Theta(t)$ is the Heaviside function, and τ_s is a synaptic time constant. The intensity $\lambda_i(t)$ given inputs at times t_k^m has the compact expression

$$\lambda_i(t) = \lambda_i^0 + \sum_k W_{ik} \sum_m a(t - t_k^m). \tag{2.11}$$

This type of self-exciting stochastic process goes back to Hawkes [53, 54]; outside of neuroscience, it has seen use in seismology [55], mathematical finance [56] or even the study of gang violence [57]. Inhibitory neurons are not explicitly needed to maintain the irregular spiking activity in networks of Hawkes neurons. Since they are not otherwise needed, in Chapter 4 we do not include inhibitory neurons and implicitly assume that their presence is giving rise to the irregularity.

2.5.1 Cross-correlations

We define the cross-correlation function $C_{ij}(t, \tau)$ between two neurons as the joint probability of neuron i emitting a spike at time t and neuron j emitting a spike at time $t + \tau$. In our notation, we can write this as

$$C_{ij}(t, \tau) = \langle S_i(t + \tau) S_j(t) \rangle. \tag{2.12}$$

The angular brackets here denote averages over the statistical ensemble; since the spike trains $S_i(t)$ result from inhomogeneous Poisson processes, their ensemble average is given by their intensity

$$\langle S_i(t) \rangle = \lambda_i(t). \tag{2.13}$$

If neurons i and j spike as uncorrelated homogeneous Poisson processes, the joint spiking probability is simply the product of the individual probabilities, i.e., the intensities:

$$C_{ij}(t, \tau) = \langle S_i(t + \tau) \rangle \langle S_j(t) \rangle = \lambda_i \lambda_j. \tag{2.14}$$

If neurons i and j are Hawkes neurons, and there is a (static) synapse from neuron j to neuron i with strength w, then the effect of a spike of neuron j on neuron i will appear as a conditional probability,

$$\langle S_i(t + \tau) | S_j(t) \rangle = \lambda_i^0 + w a(\tau), \tag{2.15}$$

and the cross-correlation (assuming no other interactions) becomes [58]

$$C_{ij}(t,\tau) = \langle S_i(t+\tau), S_j(t)\rangle \tag{2.16}$$
$$= \langle S_i(t+\tau)|S_j(t)\rangle \langle S_j(t)\rangle \tag{2.17}$$
$$= \lambda_i^0\lambda_j^0 + wa(\tau)\lambda_j^0. \tag{2.18}$$

For an interconnected network of neurons, the conditional probabilities will not just include direct interactions between pairs of neurons but also more complex interactions involving other neurons. If the weight matrix W of such a network is static, we can write the cross-correlations in frequency space as [53]

$$\tilde{C}(\omega) = 2\pi\delta(\omega)\boldsymbol{r}\boldsymbol{r}^T + \left(\mathbb{1} - \tilde{a}(\omega)W\right)^{-1} D \left(\mathbb{1} - \tilde{a}(-\omega)W^T\right)^{-1}. \tag{2.19}$$

Here r is the vector of static time-averaged firing rates, that results from the network dynamics, and we use the non-unitary definition of the Fourier transform, $\tilde{g}(\omega) = \int_{-\infty}^{\infty} \mathrm{d}t\ \mathrm{e}^{-i\omega t} g(t)$.

2.6 Dynamical stability

In this section, we discuss measures that characterize the stability of a dynamical system. In practice, these measures provide information about the behavior of a dynamical trajectory in the presence of small perturbations. To make statements about stability regardless of the chosen trajectory, we need the dynamical system to be ergodic. In physical terms, ergodicity states that the statistical properties of a trajectory over long durations are equivalent to the statistical properties of an ensemble of trajectories at one point in time. It implies that time averages of functions of a sample trajectory are independent of the choice of initial conditions and interchangeable with the ensemble average.

2.6.1 Lyapunov exponents

We start with an intuitive explanation of the largest Lyapunov exponent, which measures the average rate of convergence or divergence of two neighboring trajectories that differ only by an infinitesimal displacement. We use a discrete-time dynamical system[1] defined by states $\boldsymbol{U}(t) \in \mathbb{R}^N$, a map $F : \mathbb{R}^N \to \mathbb{R}^N$, and the relation

$$\boldsymbol{U}(t+1) = \boldsymbol{F}(\boldsymbol{U}(t)). \tag{2.20}$$

The linearized evolution of a small perturbation $\boldsymbol{u}(t)$ of a trajectory $\boldsymbol{U}(t)$ obeys

$$\boldsymbol{u}(t+1) = \frac{\mathrm{d}\boldsymbol{F}}{\mathrm{d}\boldsymbol{U}}(\boldsymbol{U}(t))\boldsymbol{u}(t) =: J(t)\boldsymbol{u}(t), \tag{2.21}$$

where $J(t)$ is called the Jacobian of the map at time t. We are interested in the long-term evolution of perturbations. Given an initial perturbation $u(0)$ at $t = 0$ we write

$$\boldsymbol{u}(t) = \prod_{k=0}^{t-1} J(k)\boldsymbol{u}(0) =: H(t)\boldsymbol{u}(0). \tag{2.22}$$

[1] As will be shown in the second chapter, the networks of integrate-and-fire neurons with instantaneous synaptic currents can be formulated as a discrete map. Nevertheless, what follows also applies similarly to continuous systems.

The distance between the original and the perturbed trajectory is given by the norm of the perturbation, which, using Eq. (2.22), evolves as

$$\|\boldsymbol{u}\|^2(t) = \boldsymbol{u}(0)^\intercal H^\intercal(t) H(t) \boldsymbol{u}(0) =: \boldsymbol{u}(0)^\intercal M(t) \boldsymbol{u}(0). \tag{2.23}$$

The leading Lyapunov exponent is the time-averaged growth rate of the logarithm of this perturbation,

$$\lambda_1 = \lim_{t\to\infty} \frac{1}{t} \ln \frac{\left\|H(t)\boldsymbol{u}(0)\right\|}{\left\|\boldsymbol{u}(0)\right\|}. \tag{2.24}$$

In other words, if the limit in Eq. (2.24) exists, a generic perturbation will expand or contract exponentially with the inverse time constant given by the first Lyapunov exponent λ_1. The existence of the limit follows from Oseledets' multiplicative ergodic theorem [59] if the dynamical system is ergodic. More generally, the theorem guarantees the existence of the limit

$$\lim_{t\to\infty} M(t)^{\frac{1}{2t}} = P \tag{2.25}$$

where P has positive eigenvalues $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_N$. This allows us to generalize the notion of the leading Lyapunov exponent by defining the spectrum of Lyapunov exponents as

$$\lambda_i = \log \mu_i. \tag{2.26}$$

We can see that this definition matches the one in Eq. (2.24) for the first LE. The other LEs extend the idea of perturbation growth to the growth of volumes in tangent space: the first k Lyapunov exponents describe the time-averaged rate of expansion or contraction of the volume $V_k(t)$ of a parallelopiped spanned by k perturbation vectors:

$$\lim_{t\to\infty} \frac{1}{t} \frac{V_k(t)}{V_k(0)} = \prod_{i=1}^{k} \mathrm{e}^{\lambda_i t} V_k(0). \tag{2.27}$$

The LEs also describe the growth of perturbations depending on their direction: A perturbation grows with the smallest LE, λ_N, if it is part of a one-dimensional subspace of the phase space, $\mathbb{D}_N$. It grows with λ_{N-1} if it is part of $\mathbb{D}_{N-1} \setminus \mathbb{D}_N$ where $\mathbb{D}_{N-1}$ is a two-dimensional subspace containing $\mathbb{D}_N$. The phase space is then structured with subsequent orders of $\mathbb{D}_i$ such that $\mathbb{D}_{i+1} \subset \mathbb{D}_i$ and $\dim(\mathbb{D}_i) = N + 1 - i$, where $\mathbb{D}_1$ is the full phase space. Perturbations in $\mathbb{D}_i \setminus \mathbb{D}_{i+1}$ will grow according to λ_i.

We finally note that the LEs are independent of initial conditions and invariant under coordinate transformations, which is often helpful in their computation.

2.6.2 Numerical computation of Lyapunov exponents

The matrix P in Eq. (2.25) becomes ill-conditioned for large t, since its eigenvalues cover a range that grows exponentially with t. It is therefore not suitable for the numerical computation of the LEs. Instead, to compute the Lyapunov spectrum, we employ the relationship of the LEs with the time-averaged expansion of volumes, Eq. (2.27), using the algorithm by Benettin et al. [24]. The idea is to split the computation into small steps where in each step we evolve an orthonormal tangent basis

with the Jacobian, obtain the volume expansions and contractions, and reorthogonalize.

Starting with m orthogonal vectors, which we express as columns of an $N \times m$ matrix Q, we apply the Jacobian $J(t)$ to obtain the linearized evolution of Q which we call $\tilde{Q}$.

$$\tilde{Q}(t+1) = J(t)Q(t). \tag{2.28}$$

$\tilde{Q}$ admits a unique decomposition into the product of an $N \times m$ orthogonal matrix which we take to be $Q(t+1)$ and an upper-triangular matrix R with positive diagonal elements, called QR decomposition:

$$\tilde{Q}(t) = Q(t)R(t) \tag{2.29}$$

The volume expansion is encoded in the determinant of R, given by the product of its diagonal elements. Since the expansion rate is given by the product of the exponentials of the LEs, we can write:

$$\frac{1}{t}\sum_i \log R_{ii} = \sum_i \lambda_i, \tag{2.30}$$

for sufficiently large t. If we consider the upper-left $(N-1) \times (N-1)$ block of R and the orthogonal $N \times (N-1)$ matrix consisting of the first $N-1$ columns of Q, we obtain the QR-decomposition of the first $N-1$ columns of $\tilde{Q}$. Its volume transform which is given by the first $N-1$ Lyapunov exponents differs from the volume transform of the N-dimensional volume by the factor R_{NN} – hence we have $\frac{1}{t}\log R_{NN} = \lambda_N$. We repeat this argument for each dimension to arrive at

$$\log R_{ii} = \lambda_i. \tag{2.31}$$

We can split this computation into smaller parts. We generalize the long-term Jacobians $H(t)$ in Eq. (2.22) to cover the tangent evolution between two specific points in time:

$$H(t_j, t_i) = \prod_{t=t_i}^{t_j} J(t), \tag{2.32}$$

such that $H(t, t_0) = H(t, t_1)H(t_1, t_0)$. Then, on the one hand we have

$$H(t, t_0)Q(t_0) = \tilde{Q}(t, t_0) = Q(t, t_0)R(t, t_0), \tag{2.33}$$

and on the other hand

$$\begin{aligned} H(t, t_0)Q(t_0) &= H(t, t_1)H(t_1, t_0)Q(t_0) \\ &= H(t, t_1)\tilde{Q}(t_1, t_0) \\ &= H(t, t_1)Q(t_1, t_0)R(t_1, t_0) \\ &= \tilde{Q}(t, t_1)R(t_1, t_0) \\ &= Q(t, t_1)R(t, t_1)R(t_1, t_0). \end{aligned} \tag{2.34}$$

Because the product of upper-triangular matrices is upper-triangular, Eq. (2.34) also has the form of a QR decomposition. Since the QR decomposition is unique we can identify $Q(t, t_0)$ in Eq. (2.33) with

$Q(t, t_1)$ in Eq. (2.34) and we obtain

$$R(t, t_0) = R(t, t_1)R(t_1, t_0). \tag{2.35}$$

We can therefore split the computation into small steps such that

$$\lambda_i = \frac{1}{t} \sum_k \log R_{ii}(t_k, t_{k-1}). \tag{2.36}$$

We provide a pseudocode for the full Lyapunov spectrum in Algorithm 1.

Algorithm 1 Algorithm for Lyapunov spectrum (Benettin)

```
U ← U_0
Q ← Q_0
for t = 1 to T do
    J ← dF/dU
    Q ← J · Q
    U ← F(U)
    Q, R ← QR(Q)
    for i = 1 to N do
        λ_i ← λ_i + log R_ii
    end for
end for
for i = 1 to N do
    λ_i ← λ_i/T
end for
```

2.6.3 Covariant Lyapunov vectors

Assuming a nondegenerate Lyapunov spectrum, we can span the tangent space at each point with a set of N vectors $\boldsymbol{v}_i$ that are covariant with the tangent dynamics,

$$J(t)\boldsymbol{v}_i(t) \propto \boldsymbol{v}_i(t+1), \tag{2.37}$$

and grow or contract according to their corresponding LEs:

$$\left|\boldsymbol{v}_i(\pm t)\right| = \mathrm{e}^{\pm t\lambda_i}\boldsymbol{v}_i(0), \tag{2.38}$$

for large t. They are called covariant Lyapunov vectors (CLVs), and their existence also follows from Oseledets' theorem. Eqs. (2.37) and (2.38) admit an analogy to eigenvectors and eigenvalues: An eigenvector is mapped to a vector proportional to itself with a proportionality constant given by its corresponding eigenvalue. This is not possible for the Jacobians of a dynamical system, since they map between different tangent spaces and thus there is no notion of a vector being mapped to itself. Instead, a CLV covaries with J and grows or contracts on average according to a fixed LE.

From Eqs. (2.37) and (2.38) it follows that any perturbation, expressed as a linear combination of a set of CLVs, $\{\boldsymbol{v}_i\}$, will asympotically align with the CLV corresponding to the largest LE, called the

forward most expanding direction:

$$\lim_{t\to\infty} H(t) \sum_{\{i\}} a_i \boldsymbol{v}_i(0) \propto \boldsymbol{v}_{\underset{i\in\{i\}}{\operatorname{argmax}}\,\lambda_i}(t). \tag{2.39}$$

Analogously a perturbation that is evolved backward in time will align with the backward most expanding direction, i.e., the CLV corresponding to the smallest LE. The map F need not be invertible for this. It is sufficient to trace an already computed trajectory back in time, and consider the backward tangent dynamics of that trajectory. We can then understand the subspaces $\mathbb{D}_i$ from Section 2.6.1 as those subspaces spanned by the CLVs with LE less or equal to λ_i. The CLVs, thus, provide us with a structure of the local phase space given by stable and unstable directions.

2.6.4 Numerical Computation of CLVs

The authors of [25] have described an efficient numerical algorithm to compute the CLVs of a dynamical system that makes use of the CLVs' backward and forward alignment properties, see Eq. (2.39). The algorithm starts by evolving a system of orthonormal vectors forward in time and reorthogonalizing it after each step, such that the ith vector aligns with the subspace of the first i CLVs. Then the vectors are evolved backward in time while still restricted to their respective subspaces such that they align with the backward most expanding direction in that subspace, i.e. the ith CLV.

As with the LEs, we begin with a set of N random orthogonal basis vectors $\boldsymbol{q}_i$, which are arranged column-wise as a matrix Q. By successively applying the Jacobian and performing QR-decompositions this basis, after sufficient time, converges to the Gram-Schmidt (GS) basis $G(t) = \big(\boldsymbol{g}_1(t), \ldots, \boldsymbol{g}_N(t)\big)$, which only depends on the phase space point $\boldsymbol{U}(t)$. The first GS basis vector, having evolved unrestricted, will be aligned with $\boldsymbol{v}_1$ and the second one, which has the constraint of being orthogonal to the first one, will be in the subspace spanned by $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, and so on. In other words, we can write

$$\boldsymbol{g}_i = \sum_j^i d_j \boldsymbol{v}_j. \tag{2.40}$$

We next evolve the system for a sufficiently long time up until $t = T_2$ while at each step computing and saving both the GS matrices and the upper triangular matrices $R(t)$ that result from the QR decompositions. We then take a set of random vectors $w = (\boldsymbol{w}_1, \ldots, \boldsymbol{w}_N)$ at $t = T_2$ with $\boldsymbol{w}_i$ in the subspace spanned by the first i GS basis vectors. We express them in the GS basis as

$$\boldsymbol{w}_i = \sum_k^i C_{ki}\, \boldsymbol{g}_k, \tag{2.41}$$

or in matrix form as

$$W = GC. \tag{2.42}$$

Our goal is then to evolve the vectors $\boldsymbol{w}_i$ backward in time along the trajectory U has taken in the forward evolution. This will cause them to align with the backward most expanding direction in their

respective subspaces, i.e. $\boldsymbol{v}_i$, cf. Eq. (2.40). For the backward evolution, we use

$$\begin{aligned} W(t+1) &= J(t)W(t) \\ &= J(t)G(t)C(t) \\ &= G(t+1)R(t+1)C(t) \\ \implies G(t+1)C(t+1) &= G(t+1)R(t+1)C(t) \\ \implies C(t+1) &= R(t+1)C(t) \\ \implies C(t-1) &= R^{-1}(t)C(t) \end{aligned} \tag{2.43}$$

We can then use the GS basis saved at $t = t_1$ to recover the CLVs from C using the basis transform Eq. (2.42). Pseudocode is shown in Algorithm 2.

Algorithm 2 Algorithm for CLVs (Ginelli)

$\boldsymbol{U} \leftarrow \boldsymbol{U}_0$
$Q \leftarrow \mathbb{1}_N$
for $t = 1$ to T_1 **do**
  $J \leftarrow \mathrm{d}\boldsymbol{F}/\mathrm{d}\boldsymbol{U}$
  $Q \leftarrow J \cdot Q$
  $\boldsymbol{U} \leftarrow \boldsymbol{F}(\boldsymbol{U})$
  $Q, \_ \leftarrow \mathrm{QR}(Q)$
end for
$G(0) \leftarrow Q$
for $t = 1$ to T_2 **do**
  $J \leftarrow \mathrm{d}\boldsymbol{F}/\mathrm{d}\boldsymbol{U}$
  $G(t) \leftarrow J \cdot G(t-1)$
  $\boldsymbol{U} \leftarrow \boldsymbol{F}(\boldsymbol{U})$
  $G(t), R(t) \leftarrow \mathrm{QR}(G(t))$
end for
$C(T_2) \leftarrow \mathrm{random\_upper\_triangular}(N, N)$
for $t = T_2$ to 2 **do**
  $C(t-1) \leftarrow R^{-1}(t)C(t)$
  $\mathrm{normalize}(C(t-1))$
end for
for $t = 1$ to T_3 **do**
  $v_{i,j}(t) \leftarrow \sum_k G(t)^{\mathsf{T}}_{ik} C(t)_{kj}$
end for

CHAPTER 3

Dynamics and computation in mixed networks containing neurons that accelerate towards spiking

Contribution statement

This chapter is based on the following article published in Physical Review E:

[1] **P. Manz**, S. Goedeke, and R.-M. Memmesheimer
Dynamics and computation in mixed networks containing neurons that accelerate towards spiking
Phys. Rev. E **100** (2019) 042404

In this article, we study the dynamical properties of recurrent spiking networks that contain two types of inhibitory integrate-and-fire neurons: the standard LIF neuron discussed in Section 2.2 and a novel type, called XIF, with negative leak current. These neuron types differ significantly in their subthreshold dynamics; however, they can both collectively give rise to a balanced state with irregular and asynchronous spiking, regardless of the proportions of each neuron type in the network. We investigate the influence of the two neuron types on the network dynamics, phase space structure, and computational capabilities. Notably, we find that already a single XIF neuron can qualitatively change the network dynamics and that mixed networks may combine the computational capabilities of ones with only one type of neuron.

In Section 3.2 we describe the neuron models, particularly introducing the anti-leaky XIF neuron. We assume fully inhibitory networks with constant external excitatory drive. With this, we describe and compare the subthreshold dynamical properties of both neuron types, defining their free firing rate in the absence of recurrent inhibition. We show how networks with LIF and XIF neurons can give rise to a balanced state with characteristically irregular and asynchronous spike patterns if the synapses of the XIF neurons are endowed with a simple voltage dependence. My contribution to this section was devising the voltage dependence of the XIF neurons, performing the simulations, creating

the figures, and writing large parts of the text.

Section 3.3 contains a self-consistent approximation of the network firing rate of networks with LIF and XIF neurons. We have built upon the shot-noise approach by Richardson [60] where Poisson-like input spikes causing finite-size jumps are assumed and adopted it to our neuron models. In particular, we needed to modify the approach to accommodate for the voltage dependence of the synapses of XIF neurons. For this section, I took part in deriving the analytical expressions, performed all numerical computations and analyses, created the figure, Fig. 3.3, and wrote parts of the text.

Section 3.4 begins with a mean-field approach for the computation of the Lyapunov exponents (LEs), see Section 2.6.1, of networks of LIF and XIF neurons using the firing rates obtained in Section 3.3. This mean-field Lyapunov spectrum provides a good approximation to the numerically computed LEs and exactly matches its sum. We find that there is approximately one positive LE for each XIF neuron. In particular, while a network containing only LIF neurons is stable [8], adding only a single XIF neuron is sufficient to render the network dynamics chaotic. My contributions were deriving and computing the mean-field Lyapunov spectrum, proving the exactness of the sum of the mean-field exponents, numerically computing the Lyapunov spectrum, creating all figures, and writing large parts of the text.

In Section 3.5 we study the stable and unstable directions of mixed networks via the covariant Lyapunov vectors (CLVs), see Section 2.6.3, with particular regard to the neuron types. We find that stable and unstable directions correspond in good approximation to the directions of LIF and XIF neurons, confirming our assumptions made for the mean-field Lyapunov spectrum. My contributions to this section were performing all simulations and numerical computations and analyses, creating all figures, and writing large parts of the text.

Section 3.6 applies the insights from the preceding sections on computations with precisely timed spikes [61] where our mixed networks serve as dynamical reservoirs. The section discusses two computational tasks: In the first, a network solves a temporal XOR or AND task depending on context. In the second task, a network needs to either be sensitive or insensitive to small differences in input timings, depending on the input. Notably, the second task requires knowledge of the network's CLVs. My contribution to this section was preparing the network state and computation of the CLVs for the second task, and writing parts of the section's text.

Appendix 3.8.A contains further analysis concerning the dependence of the results in Section 3.3 on network size and indegree. I performed the simulations, created the figure, and wrote large parts of the text. Appendices 3.8.B and 3.8.D contain the derivations of the mean-field approach for the Lyapunov spectrum and its sum, which are described in Section 3.4 and which I performed. Appendix 3.8.E discusses the dependence of the Lyapunov spectrum on indegree and network size, for which I did all simulations, created the figure, and wrote large parts of the text.

The article is presented here with minor editorial changes and the formatting adapted to match this thesis.

Code availability

The code required to reproduce the results of this chapter has been make publicly available at `https://github.com/axionmonodromy/lyap-xif-lif`.

3.1 Introduction

Biological neural networks consist of a large variety of interconnected neurons, which communicate via short stereotypical electrical pulses called action potentials or spikes. After a neuron has generated a spike, this travels along the axon and is transmitted to other neurons at synaptic contacts. The electrical membrane potential of the receiving neuron is then changed by an excitatory or inhibitory current pulse. Sufficiently many excitatory inputs in turn lead to spike generation in a receiving neuron. Many biological neural networks generate irregular and asynchronous spiking. This is likely caused by a dynamically balanced network state, in which the average inhibitory and excitatory input current to each neuron sum to a value that is insufficient for frequent spike generation [3–5, 62]. Spikes are caused by fluctuations in the inputs and the resulting spiking dynamics appear random and irregular.

Irregular dynamics are often chaotic, implying that the dynamics are sensitive to perturbations: initially small ones can strongly grow with time, which results in ultimately large quantitative differences between perturbed and unperturbed trajectories. A powerful tool to quantify this sensitivity and therewith the local phase space structure are the Lyapunov exponents (LEs) and associated with them the covariant Lyapunov vectors (CLVs) [10, 63]. The sign of the largest LE indicates whether the system is chaotic and its magnitude equals the long-term average growth or decay rate of generic infinitesimal perturbations. The spectrum of LEs describes the long-term average evolution of volumes spanned by tangent vectors and the change of infinitesimal perturbations in non-generic directions, which are specified by the CLVs. To each LE, there is a CLV. The size of a perturbation in the CLV's direction changes with an average rate of plus or minus the corresponding LE for long-term forward or backward time evolution, respectively. The CLVs thereby indicate the directions of the unstable and stable manifolds along a trajectory. Furthermore, the spectrum of LEs can be used to derive dynamical quantities such as the Kaplan-Yorke fractal dimension of a chaotic attractor [64].

In our study, we consider purely inhibitory networks of current-based, oscillating integrate-and-fire type neurons with post-synaptic currents of infinitesimally short duration and instantaneous reset. It has been shown numerically [65, 66] and analytically [8, 9] that if such networks contain only leaky integrate-and-fire (LIF) neurons, the networks' irregular balanced state dynamics are stable against infinitesimal and small finite size perturbations and are thus not chaotic but a realization of stable chaos [67, 68]. The dynamics ultimately converge to a periodic orbit; the durations of the preceding irregular transients, however, grow exponentially with system size. The stability of the network dynamics is robust against introducing excitatory connections and considering synaptic currents of finite temporal extent [9, 66] and there is a smooth transition to chaos upon increasing the number of excitatory connections and the duration of synaptic currents. The computational abilities of the stable precise spiking dynamics have not yet been explored, even though the specific structure of the phase space, which is composed of "flux tubes", may be beneficial and exploitable [69].

LIF neurons incorporate a leak current as found in biological neurons [22]. This increases linearly with increasing membrane potential and leads to dissipation (contraction of phase space volume) in the subthreshold dynamics. When driven by a constant depolarizing input current, the membrane potential therefore has negative second derivative; the neuron has a purely concave so-called rise function. In the considered class of networks, this implies the stability of the microscopic dynamics if only LIF neurons are present [8, 9]. In biological neurons as well as in neuron models that explicitly model spike generation, such as the quadratic and the exponential integrate-and-fire neuron [23], the membrane potential accelerates towards a spike for larger membrane potentials. The rise function thus has a convex part. Ref. [70] showed that networks of quadratic integrate-and-fire neurons that are

otherwise similar to those considered in refs. [8, 9, 66] exhibit chaos. Furthermore, ref. [70] computed the spectrum of LEs and quantities that are derivable from them, as well as the statistics of the first CLV, which points into the directions to which a generic perturbation vector aligns in the long term.

Motivated by the above results and by the fact that there are many different types of cortical inhibitory interneurons [71], in the present study we investigate the impact of inserting a different type of neuron, with non-concave rise function, into inhibitory networks of LIF neurons. To be specific, we insert "anti-leaky" integrate-and-fire (XIF) neurons with purely convex rise function. We choose the letter "X" in the abbreviation to highlight this convexity and the expansion of phase space volume by the flow of the subthreshold dynamics. XIF neurons may be interpreted as a model for a class of biological neurons whose membrane potential lingers in a region where it accelerates towards spiking. Simultaneously, these neurons maintain similar analytical tractability as their leaky counterparts because of their mostly linear subthreshold dynamics. We describe our neuron and network models in detail in the next section. Thereafter, self-consistent firing rates and membrane potential probability distributions for both types of neurons are analytically derived, assuming Poisson input with finite size spike impacts. We then consider the dynamical stability properties and local phase space structures of the network dynamics, computing the entire spectra of LEs both numerically and analytically in a mean-field approximation. We also compute their CLVs to investigate how the stable and unstable directions are related to the different neuron types within the network.

Finally, we consider computations in pure and mixed networks of the considered types and show how the richer phase space structure in mixed ones can be exploited. For this, we propose a reservoir computer based entirely on precisely timed spikes. Reservoir computing has been introduced several times at different levels of elaborateness and in different flavors, in machine learning and in neuroscience [72–75]. A reservoir computer consists of a high dimensional, nonlinear dynamical system, the reservoir or liquid, and a comparably simple readout. The reservoir "echoes" the input in a complicated, nonlinear way; it acts like a random filter bank with finite memory as each of its units generates a nonlinearly filtered version of the current input and its recent past while forgetting more remote inputs [72, 74–76]. The simple, often linear readout can then be trained to extract the desired results, while the reservoir is static. In our scheme, the output neuron is spiking and thus nonlinear, the desired outputs are trains of precisely timed spikes. The learning thus requires different approaches than learning of conventional continuous targets; gradient-descent based methods [77] fail due to the discontinuity at the threshold as well as methods that require errors to be small but finite [78]. A number of algorithms have been suggested to learn precisely timed spikes [61, 79–85], mostly using heuristic approaches. For our readout neuron, we can use the Finite Precision Learning scheme [61]. It has been shown to generically converge if the input-output relation is realizable at all, which explains its numerically found superior learning abilities [83].

3.2 Mixed networks of neurons with concave and convex rise function

We consider a recurrent network with N neurons. The kth spike of neuron j, which is sent at time t_{jk}, generates a postsynaptic current pulse $h_i(V_i^-)W_{ij}\delta\left(t - t_{jk}\right)$ in neuron i. Here $W_{ij} \leq 0$ is the weight of the inhibitory connection and $h_i(V_i^-)$ is a possible voltage-dependent modulation, which depends on the membrane potential of neuron i just before input arrival, given by the left-hand side limit

$V_i^- = V_i(t^-) = \lim_{\varepsilon \searrow 0} V_i(t-\varepsilon)$. We assume that all excitatory inputs to neuron i can be gathered into a constant excitatory external input current $I_i^{\text{ext}} > 0$ and that the remaining explicitly modeled recurrent inhibition is fast [8, 69, 86]. We further assume that there is a leak term with prefactor $\gamma_i \neq 0$. Taken together, we model the subthreshold membrane potential dynamics of neuron i by

$$\dot{V}_i = -\gamma_i V_i + I_i^{\text{ext}} + h_i(V_i^-) \sum_{j=1}^{N} W_{ij} \sum_k \delta\left(t - t_{jk}\right). \tag{3.1}$$

When V_i reaches the spike threshold at time t, $V_i{}^- = V_{\text{th}} > 0$, it is reset, $V(t) = V_{\text{re}} = 0$, and a spike is emitted. This, in turn, generates in a postsynaptic neuron l a current pulse as introduced above, which causes V_l to decrease in jump-like manner from V_l^- to $V_l^- + h_l(V_l^-)W_{li}$. The rise function, i.e., the membrane potential dynamics with $V_i(0) = 0$ in absence of recurrent inhibitory input [87, 88], reads

$$V_i(t) = \frac{I_i^{\text{ext}}}{\gamma_i}\left[1 - \exp(-\gamma_i t)\right]. \tag{3.2}$$

It is concave for $\gamma_i > 0$ and convex for $\gamma_i < 0$. There are two types of neurons in our networks: LIF neurons with dissipation and concave rise function, which obey Eq. (3.1) with $\gamma_i > 0$, and anti-leaky XIF neurons with convex rise function, which obey Eq. (3.1) with $\gamma_i < 0$, see Fig. 3.1. The membrane potential dynamics of an LIF neuron has a globally attracting fixed point at $V_{\infty,i} = I_i^{\text{ext}}/\gamma_i$, if there is no threshold for spike generation and no inhibitory input. We assume $V_{\infty,i} > V_{\text{th}}$, so neurons without inhibitory input periodically spike and reset. For our study it is sufficient to endow the LIF neurons with a simple, current-based synapse model, setting $h_i(V_i^-) = 1$. A coarse approximation of the membrane potential dynamics without threshold and neglecting input fluctuations yields $\dot{\bar{V}}_i = -\gamma_i \bar{V}_i + I_i^{\text{ext}} + I_i^{\text{inh}}$, where I_i^{inh} is the average inhibitory input current. In the balanced state, its attractor at $\bar{V}_{\infty,i} = \left(I_i^{\text{ext}} + I_i^{\text{inh}}\right)/\gamma_i$ is below or close to the spike threshold, such that spikes are always or typically generated by input fluctuations, more specifically by periods of less than average inhibition.

In the absence of inhibitory input XIF neurons have an unstable, repelling fixed point at $V_{-\infty,i} = I_i^{\text{ext}}/\gamma_i < 0$. If the membrane potential starts above this separatrix, it increases exponentially towards the threshold. When it reaches there, the neuron spikes, its membrane potential resets to zero, increases towards the threshold again and so forth: XIF neurons oscillate and spike periodically for any $I_i^{\text{ext}} > 0$, if there is no inhibitory input. If the membrane potential starts below the separatrix, it decreases exponentially to $-\infty$. Also in the presence of recurrent inhibitory inputs an XIF neuron is unrecoverably switched off once its membrane potential falls below $I_i^{\text{ext}}/\gamma_i$, since the inputs only decrease the membrane potential further. Averaging over the inhibitory inputs as before yields an effective separatrix at $\bar{V}_{-\infty,i} = \left(I_i^{\text{ext}} + I_i^{\text{inh}}\right)/\gamma_i$. Membrane potentials falling below it have a tendency to further decrease, causing the neuron to effectively switch off. This can be also seen from the phase response curve of XIF neurons, which gets steeper for negative phases, in contrast to that of LIF neurons which becomes flatter, see Fig. 3.1c. In other words, in XIF neurons an incoming inhibitory input at a low potential still above the separatrix (and thus at a low phase) has a larger effect in the sense that it delays the next spiking more than the same input arriving at a higher potential. As a consequence, we observe in networks containing XIF neurons with purely current-based input [$h(V_i) = 1$] that many of these neurons are first effectively and then unrecoverably switched off, if the

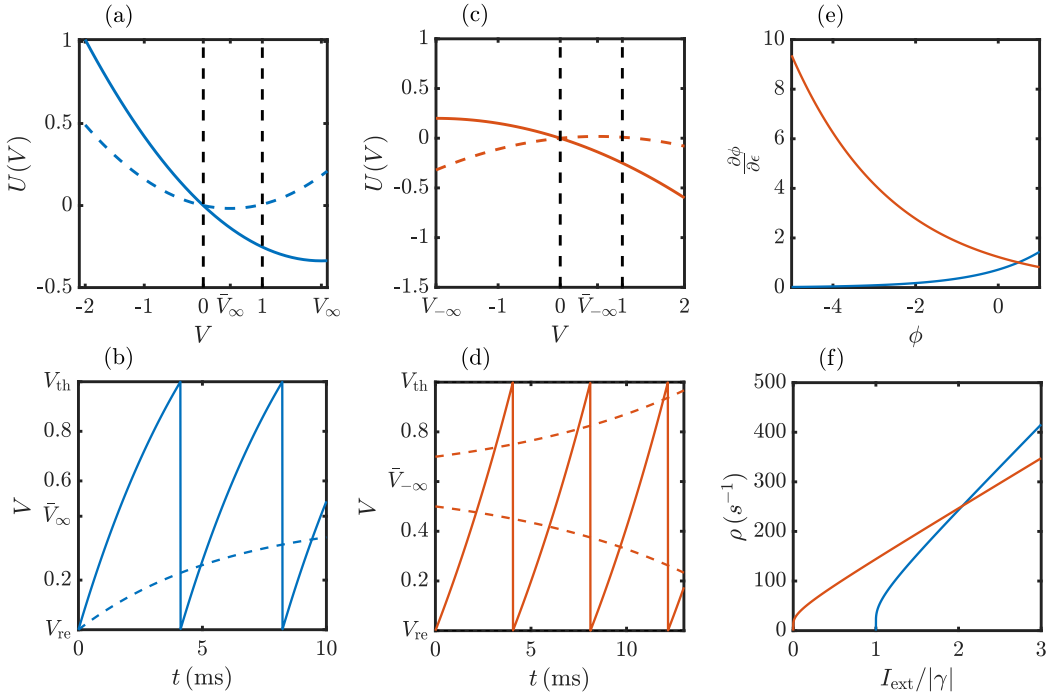

Figure 3.1: LIF and XIF neuron dynamics with constant input. Blue and red indicate LIF and XIF neurons, respectively. Solid curves in (a-d) indicate excitatory constant input only, dashed lines inclusion of an average inhibitory input current I^{inh} (the cutoff for XIF inputs is neglected). (a,c) Potential function $U(V)$ of the membrane potential (voltage) V; V follows the negative gradient of U, $\dot{V}(t) = -U'(V(t))$, if there is no threshold. (a) U for an LIF neuron is an upward parabola; V tends to the stable fixed point at U's minimum (at V_∞ or $\bar{V}_\infty$), if there is no threshold. (c) U for an XIF neuron is a downward parabola; V tends to $-\infty$ or $+\infty$ when starting left or right of U's maximum (at $V_{-\infty}$ or $\bar{V}_{-\infty}$), if there is no threshold. A monotonically decreasing potential function U between reset potential and threshold (left and right vertical dashed lines) indicates mean driven periodic spiking (solid curves in (a,c)). In the balanced state the spiking is fluctuation driven with (a) U's minimum below threshold for LIF neurons and (c) U's maximum above reset for XIF neurons (dashed curves). (b,d) Example trajectories for LIF and XIF dynamics including threshold and reset. Without inhibition, V is periodically driven over the threshold and reset. Once averaged inhibition is included, the LIF voltage (b) converges to the subthreshold fixed point at $\bar{V}_\infty$, while the XIF voltage (d) is repelled from $\bar{V}_{-\infty}$. (e) Infinitesimal phase response curves [89–91]. Inputs to LIF (XIF) neurons have a smaller (larger) spike delaying effect, the lower V is. (f) Rates of free LIF and XIF neurons at different strengths of the normalized external drive. Neuron parameters and (if applicable) values for excitatory drive and average inhibitory input are as in our network simulations.

network dynamics are irregular and the inhibitory inputs are therefore strongly fluctuating. In order to prevent this biologically implausible phenomenon, we introduce a voltage dependence

$$h(V_i^-) = \Theta\left(V_i^- - V_{\text{cutoff}}\right) \tag{3.3}$$

of the input coupling strength, where Θ is the Heaviside theta function. Inhibitory inputs arriving at a membrane potential lower than V_{cutoff} then do not induce a further decrease. This provides a simple conductance-based model for the synapses, where the driving force of the current vanishes below $V_i = V_{\text{cutoff}}$ and is constant above. We assume $V_{-\infty,i} < V_{\text{cutoff}} + W_{ij}$ for all j to exclude unrecoverable switching off and $V_{\text{cutoff}} \leq V_{\text{re}}$. We exemplarily checked that the overall network dynamics and their stability properties remain qualitatively unchanged, if we also endow the LIF neurons with these synapses.

For simplicity, we choose the parameters of all LIF and of all XIF neurons identical, i.e., $\gamma_i = \gamma_{\text{LIF}}$, $I_i = I_{\text{LIF}}$, etc., if neuron i is an LIF neuron, and $\gamma_i = \gamma_{\text{XIF}}$, $I_i = I_{\text{XIF}}$, etc., if neuron i is an XIF neuron. The spike threshold and reset potentials are $V_{\text{th}} = 1$ and $V_{\text{re}} = 0$, independent of the neuron type. We set $V_{\text{cutoff}} = V_{\text{re}}$ to avoid any effective switching off of XIF neurons. Coupling strengths are homogeneous, $W_{ij} = W$ if the coupling is present. To keep the number of relevant parameters small, we further choose $I^{\text{ext}}_{\text{LIF}}/\gamma_{\text{LIF}} = V_{\infty,\text{LIF}} = -V_{-\infty,\text{XIF}} = -I^{\text{ext}}_{\text{XIF}}/\gamma_{\text{XIF}}$. The additional choice $\gamma_{\text{XIF}} = -\gamma_{\text{LIF}}$ leads already in absence of recurrent inhibition to a higher spike rate $\rho_{\text{free,XIF}}$ in XIF neurons, since

$$\rho_{\text{free,XIF}} = \frac{-\gamma_{\text{XIF}}}{\ln\left(\frac{-V_{-\infty,\text{XIF}}+V_{\text{th}}}{-V_{-\infty,\text{XIF}}}\right)}, \tag{3.4}$$

whereas in LIF neurons,

$$\rho_{\text{free,LIF}} = \frac{\gamma_{\text{LIF}}}{\ln\left(\frac{V_{\infty,\text{LIF}}}{V_{\infty,\text{LIF}}-V_{\text{th}}}\right)}. \tag{3.5}$$

As a consequence, we observe that in a mixed network the XIF suppress the LIF neurons, which become quiescent. Using the analytical results of the next section, we therefore rescale γ_{LIF} such that the spike rates in both populations are identical. Further, we fix the neurons' indegree to the same number K, implying that $\sum_j W_{ij}$ is identical for each neuron i. This reduces quenched noise [92] and avoids strong differences in average spike rates and switched off neurons.

With the described network model setup, we observe balanced states of asynchronous irregular spiking activity for any ratio of neurons with concave and convex rise function, see Fig. 3.2 for an illustration.

3.3 Network firing rate and membrane potential distributions

Mean-field theories have been developed in statistical physics [93] and are frequently used in computational neuroscience, see, for example, refs. [48, 92, 94, 95]. The basic idea is to average the interactions in a high-dimensional system to obtain for each element an effective action, which is not influenced by this element anymore. One can thereby reduce a high-dimensional problem to low-dimensional ones. In this section we analytically determine the steady-state firing rate and the voltage probability densities for LIF and XIF neurons in mixed networks using a mean-field approximation. We use the results to obtain neuron parameters that lead to the same average firing

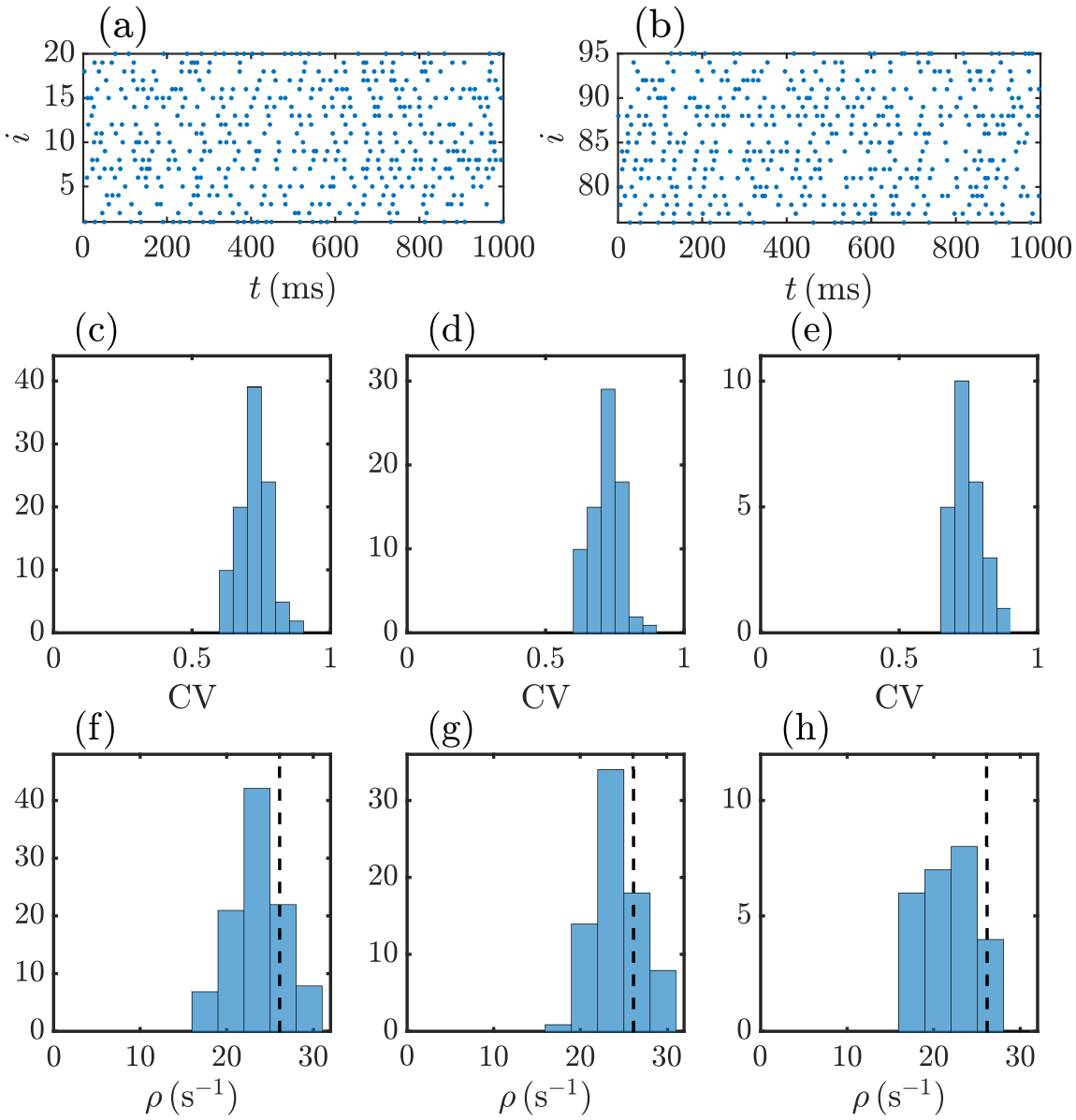

Figure 3.2: Mixed networks of LIF neurons with concave rise function and of XIF neurons with convex rise function can exhibit a balanced state with asynchronous irregular activity in both types of neurons. (a,b) Spiking activity for a subset of the LIF (a) and the XIF (b) neurons in a network with 75 LIF and 25 XIF neurons ($N = 100$). (c-e) Distribution of coefficients of variation of inter-spike intervals for all neurons (c) and for LIF (d) and XIF (e) neurons separately. (f-h) Distribution of the average spike rates of all neurons (f) and of LIF (g) and XIF (h) neurons separately. The analytically derived rate $\rho \approx 26.1\,\mathrm{s}^{-1}$ (Eqs. (3.18),(3.19)) is indicated by a black dashed vertical line. We use $\gamma_{\mathrm{XIF}} = -0.1\,\mathrm{ms}^{-1}$, $\gamma_{\mathrm{LIF}} = 0.169\,\mathrm{ms}^{-1}$, $V_{\infty,\mathrm{LIF}} = -V_{-\infty,\mathrm{XIF}} = 2$ and a randomly connected network with fixed indegrees $K = 50$ and non-zero synaptic strength $W_{ij} = W = -0.2$.

rates for both neuron types and thus to homogeneous firing rates in the entire network. In addition we employ the firing rates to analytically approximate the Lyapunov spectrum of the network dynamics using a mean-field approach in Section 3.4.1.

We approximate the superposed input spike trains to a neuron by a Poisson spike train with a given rate, i.e. we assume that all input spikes are sent independently of each other. A common approach is to additionally consider the limit of a large number of small inputs. The neuron dynamics can then be approximated by a diffusion process, which allows to compute firing rates and membrane potential distributions [96, 97]. This diffusion approximation assumes that the inputs have (infinitesimally) small amplitude and arrive at (infinitely) high rate. Here we use a shot noise approach, which accounts for the finite input rate and size of individual inputs [96, 97], in the recent formulation of refs. [60, 86]. This allows to more accurately obtain the firing rates and membrane potential distributions. In particular, the fact that in our networks the voltage probability density does not go to zero at threshold is reflected. We shortly review the approach for LIF neurons [60, 86, 98] and then extend it to XIF neurons with the voltage-dependent coupling Eq. (3.3).

The shot-noise approach (like the diffusion approximation) is based on the continuity equation for the voltage probability density $p(V,t)$. For our neuron models it reads

$$\frac{\partial p}{\partial t} + \frac{\partial j}{\partial V} = \sigma_{\text{inh}} + \sigma_{\text{reset}}, \tag{3.6}$$

where $j(V,t) = \dot{V}(V)p(V,t)$ is the drift probability current with velocity $\dot{V}(V) = -\gamma V + I^{\text{ext}}$. $\sigma_{\text{inh}}(V,t)$ and $\sigma_{\text{reset}}(V,t)$ are source terms incorporating the effects of inputs and resets of the neuron's membrane potential V.

For the LIF neuron without the voltage-dependent input, inhibitory input spikes arriving when the considered neuron is at a voltage V give rise to a sink at V, whereas spikes arriving when the neuron is at a voltage $V - W > V$ give rise to a source at V. We therefore have a first source term

$$\sigma_{\text{inh}}(V,t) = r(t)\left[p(V-W,t) - p(V,t)\right] \tag{3.7}$$

with the rate $r(t)$ of input spikes. We note that refs. [60, 86, 98] include this term in the probability current. The second source term is due to the spike and reset mechanism of the neuron model. Its threshold and reset act as Dirac delta sink and source at the corresponding discrete voltages,

$$\sigma_{\text{reset}}(V,t) = \rho(t)\left[\delta(V - V_{\text{re}}) - \delta(V - V_{\text{th}})\right]. \tag{3.8}$$

This term is proportional to the instantaneous firing rate $\rho(t)$ of the stochastic neuron dynamics or, in other words, to the probability current through the threshold [$\rho(t) = j(V_{\text{th}}, t) \geq 0$].

We investigate stationary network dynamics, which are described by constant r and ρ and time-independent $p(V)$. For these Eq. (3.6) reduces to the linear delay differential equation (or differential-difference equation)

$$\frac{d}{dV}p(V)\left(-\gamma V + I^{\text{ext}}\right) = r\left[p(V-W) - p(V)\right] + \rho\left[\delta(V - V_{\text{re}}) - \delta(V - V_{\text{th}})\right]. \tag{3.9}$$

Dividing Eq. (3.9) by $\rho > 0$ yields an equation for the rescaled density $q(V) = p(V)/\rho$, which is independent of the unknown steady-state firing rate ρ. This equation can be integrated for example with the method of steps [99]. The integration starts with the "initial conditions" $q(V) = 0$ for

$V > V_{\text{th}}$ and thus $q(V_{\text{th}}) = 1/\left(-\gamma V_{\text{th}} + I^{\text{ext}}\right)$ slightly below V_{th}. The normalization of $p(V)$ allows us to compute ρ via

$$\frac{1}{\rho} = \int_{-\infty}^{\infty} q(V) dV. \tag{3.10}$$

To obtain an analytic expression for ρ, one applies a bilateral Laplace transform $\tilde{f}(s) = \int_{-\infty}^{\infty} f(V) e^{sV} dV$. We can focus on $s \geq 0$; $\tilde{q}(0)$ yields ρ^{-1}. The Laplace transform of the rescaled Eq. (3.9) results in a linear first-order ordinary differential equation for $\tilde{q}(s)$,

$$\frac{d}{ds}\tilde{q}(s) = \left[\frac{I^{\text{ext}}}{\gamma} + \frac{r\left(\mathrm{e}^{Ws} - 1\right)}{\gamma s}\right]\tilde{q}(s) + \frac{\mathrm{e}^{V_{\text{re}}s} - \mathrm{e}^{V_{\text{th}}s}}{\gamma s}. \tag{3.11}$$

It can be solved by variation of constants. The solution of the homogeneous equation is

$$Z_0(s) = A\mathrm{e}^{\Psi(s)} \tag{3.12}$$

with an arbitrary constant A and

$$\begin{aligned}\Psi(s) &= \frac{I^{\text{ext}}}{\gamma}s + \frac{r}{\gamma}\int_0^s \frac{\mathrm{e}^{Wu} - 1}{u} du \\ &= \frac{I^{\text{ext}}}{\gamma}s + \frac{r}{\gamma}\left[\mathrm{Ei}(Ws) - \log(-Ws) - \Gamma\right].\end{aligned} \tag{3.13}$$

Here, $\mathrm{Ei}(x)$ is the exponential integral $\mathrm{Ei}(x) = -\int_{-x}^{\infty} \frac{\mathrm{e}^{-t}}{t} dx$ and Γ is the Euler-Mascheroni constant. The solution of the full equation then reads

$$\tilde{q}(s) = \mathrm{e}^{\Psi(s)}\left[\tilde{q}(0) - \int_0^s \mathrm{e}^{-\Psi(u)} \frac{\mathrm{e}^{V_{\text{th}}u} - \mathrm{e}^{V_{\text{re}}u}}{\gamma u} du\right]. \tag{3.14}$$

Since the support of $q(V)$ is bounded from above by V_{th}, $\tilde{q}(s) = \int_{-\infty}^{\infty} q(V)\mathrm{e}^{Vs} dV \leq \mathrm{e}^{V_{\text{th}}s}/\rho_0$. To balance the faster exponential growth $\sim \exp(I^{\text{ext}}s/\gamma)$ of its prefactor $\exp\left[\Psi(s)\right]$, the bracket on the right hand side of Eq. (3.14) needs to vanish for large s. We thus have

$$\tilde{q}(0) = \int_0^{\infty} \mathrm{e}^{-\Psi(u)} \frac{\mathrm{e}^{V_{\text{th}}u} - \mathrm{e}^{V_{\text{re}}u}}{\gamma u} du = \frac{1}{\rho}. \tag{3.15}$$

For an XIF neuron without voltage-dependent synapses there is no stationary membrane potential probability density $p(V)$. This is because for any time $t > 0$ there is a finite probability that the membrane potential of a neuron jumps below I^{ext}/γ and thereafter tends to minus infinity. In contrast, for an XIF neuron with the voltage dependence Eq. (3.3), $p(V)$ exists and we may use the same approach as for the LIF neuron to determine it together with the firing rate. Since membrane potentials do not drop below $V_{\text{cutoff}} + W$, we focus on the interval $[V_{\text{cutoff}} + W, V_{\text{th}}]$, where $p(V)$ can be nonzero. The couplings' voltage dependence enters the source term σ_{inh} in Eq. (3.6): If V is below V_{cutoff},

incoming spikes have no effect and the sink term due to them vanishes. Eq. (3.7) therefore changes to

$$\sigma_{\text{inh}}(V,t) = r(t)\left[p(V-W,t) - h(V)p(V,t)\right], \tag{3.16}$$

where we used that $V - W \geq V_{\text{cutoff}}$ in the relevant voltage range such that a modification of the source term is unnecessary. The stationary continuity equation becomes

$$\frac{d}{dV}p(V)\left(-\gamma V + I^{\text{ext}}\right) = r\left[p(V-W) - h(V)p(V)\right] + \rho\left[\delta(V - V_{\text{re}}) - \delta(V - V_{\text{th}})\right], \tag{3.17}$$

which can be rescaled and integrated using the method of steps to obtain $q(V)$, ρ and $p(V)$ as before. The nonlinear prefactor $h(V)$, however, impedes the derivation of ρ via Laplace transform.

We apply the above results to find mixed networks in which LIF and XIF neurons have similar firing rates. Eq. (3.15) provides a map G_{LIF} from the input to the output rate, $G_{\text{LIF}}(r) = \rho$. Eq. (3.17) implicitly defines such a map G_{XIF} for XIF neurons. The firing rate ρ of the LIF and XIF neurons in the desired mixed network needs to solve both self-consistency Eqs.

$$G_{\text{LIF}}(K\rho) = \rho, \tag{3.18}$$

$$G_{\text{XIF}}(K\rho) = \rho, \tag{3.19}$$

with the neurons' indegree K. We employ Eq. (3.19) to compute ρ for XIF neurons. Thereafter, we adapt the parameters of Eq. (3.18) such that the same ρ becomes a solution. Specifically, we solve for γ_{LIF}, keeping the other parameters fixed.

Fig. 3.3 compares the voltage densities $p(V)$ and rates ρ obtained from the shot noise approach with those of an LIF and an XIF neuron that receive input spike trains as they are generated in the recurrent network of Fig. 3.2. There is a pronounced discrepancy between the densities and rates for an LIF neuron for $K = 50$ and small N, because both the individual (see Fig. 3.2a-e) and the superposed input spike trains in these dense networks are more regular than Poisson spike trains. Removing spatial correlations for example by increasing N reduces the discrepancy, see Fig. 3.3b-d and Appendix 3.8.A for further analysis. Such input spike trains reduce the variance of the voltage and generate a $p(V)$ that is more concentrated around the value $\left(I_i^{\text{ext}} + I_i^{\text{inh}}\right)/\gamma_i$, where I_i^{inh} is the average inhibitory input current as discussed in Section 3.2. For the XIF neuron, the input spike train statistics has less impact on $p(V)$. Presumably, this is because voltage excursions due to input fluctuations are anyways suppressed by the voltage dependence of the input strength (for potentials near V_{cutoff}) and by the drive towards threshold (for larger potentials). We note that the assumption of Poisson input spike trains is the only approximation in the chosen approach, such that sampled membrane potential distributions of neurons with Poisson input match the analytical ones up to the sampling noise as shown in Appendix 3.8.A.

3.4 Growth of dynamical perturbations

3.4.1 Mean-field approach

After obtaining the spike rates and membrane potential distributions using a statistical mean-field theory, we investigate the mixed network dynamics from a dynamical systems perspective. We first analytically determine the Lyapunov spectrum using again a mean-field approach. It focuses on the

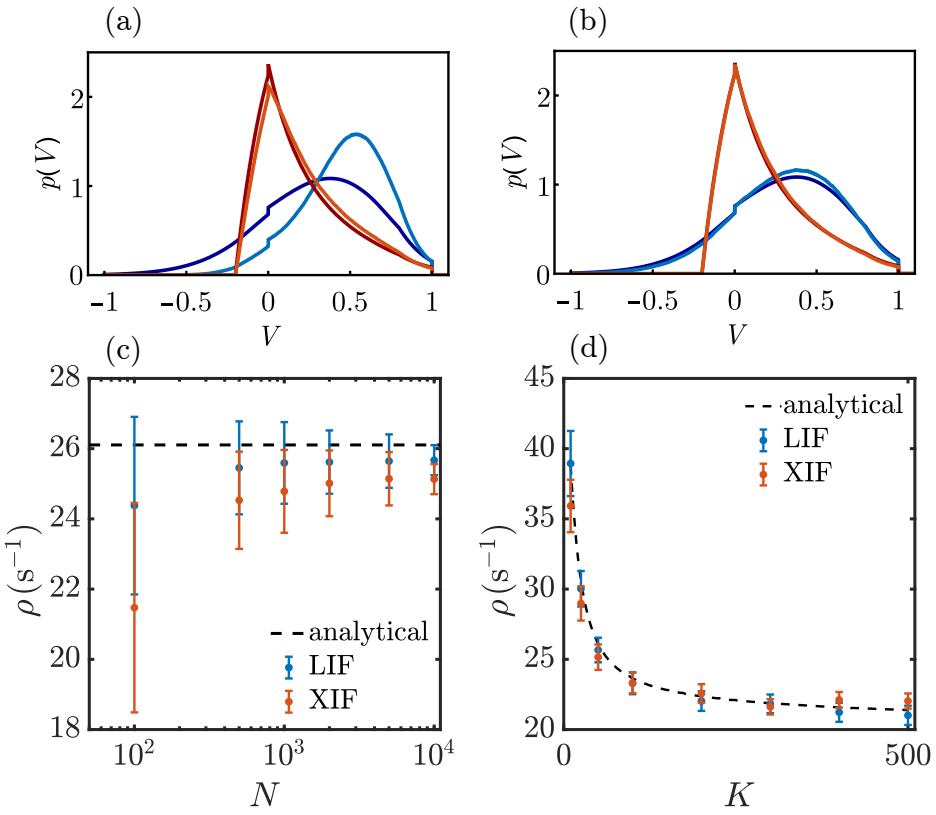

Figure 3.3: Analytically and numerically estimated voltage probability densities and spike rates of LIF and XIF neurons. (a,b) Voltage probability densities for networks of (a) $N = 100$ and (b) $N = 10000$ neurons. The dark blue and dark red curves show the analytical results Eqs. (3.9) and (3.17) of $p(V)$ for LIF and XIF neurons, where ρ is obtained self-consistently from Eqs. (3.18) and (3.19). The light blue and light red curves show representative numerically sampled voltage densities of an LIF and an XIF neuron, where the input spike trains are superpositions of simultaneous output spike trains of K neurons in the recurrent network. (c,d) Spike rates of neurons in networks of (c) different size N and indegree $K = 50$ and (d) size $N = 10000$ and different indegree K. In (d) the presynaptic weights are scaled with $1/K$ such that their sum is independent of K. Numerically measured average spike rates of LIF and XIF neurons in the different networks are shown by blue and red dots. Error bars display the standard deviations of the rate distributions. Analytical results obtained from Eqs. (3.18) and (3.19) are displayed by dashed black lines. Remaining parameters are as in Fig. 3.2.

evolution of perturbations to a single neuron and treats the input from other neurons as external. Specifically, we disregard perturbations of the rest of the network including those generated by the considered neuron's changed spiking. Inputs thus arrive at the same times in the perturbed and in the unperturbed system and do not change the neuron's perturbation. Fig. 3.4 illustrates this and compares the resulting evolution of a perturbation of an XIF neuron and of an LIF neuron: The perturbation of the XIF neuron gradually increases as long as it is not spiking, while that of the LIF neuron decreases. Conversely, in the XIF neuron spiking and resetting reduces perturbations, while it increases them in the LIF neuron; compare the values of the (finite size) distance $\left|\delta V(t)\right| = |\tilde{V}(t) - V(t)|$ between two neighboring trajectories $\tilde{V}(t)$ and $V(t)$ in Fig. 3.4a,b before and after a spike event has taken place in both the perturbed and the unperturbed dynamics. To assess the influence of these two processes, we first note that in a freely oscillating neuron they need to cancel each other such that perturbations persist on average and the LE is zero. We then note that the inhibitory inputs do not affect perturbations but prolong the subthreshold evolution between spikes. Its impact therefore dominates, and perturbations in XIF neurons grow over time, while they shrink in LIF neurons. This does not depend on the specifics of the LIF and XIF dynamics but is a consequence of the curvature of the rise function and the inhibitory inputs.

In Appendix 3.8.B, we make the gained intuitive understanding precise by quantifying the growth of perturbations and the resulting LE. For this, we describe the dynamics by a sequence of discrete maps from the state at a time (infinitesimally) shortly after generation of a spike to the state at a time shortly after generation of the next spike. The discrete time dynamics of small perturbations are then given by the "single spike Jacobians" [69, 70] $J(k)$. For the effective single neuron dynamics here, they reduce to scalar factors

$$J(k) = \frac{\partial V(t_{k+1}^+)}{\partial V(t_k^+)} = \exp\left(\frac{\gamma}{\rho_{\text{free}}} - \gamma(t_{k+1} - t_k)\right) \tag{3.20}$$

with the free firing rate ρ_{free} (Eq. (3.4) or (3.5)) of the neuron.

The growth rate of perturbations and thus the mean-field LE are given by the long-term average of Eq. (3.20),

$$\begin{aligned}\lambda_{\text{mf}} &= \lim_{L\to\infty} \frac{1}{t_L} \sum_{k=0}^{L-1} \ln\left|J_{\text{mf}}(k)\right| \\ &= -\gamma\left(1 - \frac{\rho}{\rho_{\text{free}}}\right).\end{aligned} \tag{3.21}$$

This expression confirms the intuitive understanding that without perturbed inputs the growth rate depends (i) on the growth rate during subthreshold evolution and (ii) on the prevalence of subthreshold evolution ($\rho < \rho_{\text{free}}$) or spike sending ($\rho > \rho_{\text{free}}$) relative to the free neuron case. In particular, without input we have $\lambda_{\text{mf}} = 0$ and if the neuron is silenced $\lambda_{\text{mf}} = -\gamma$. In our inhibitory networks we have $\rho < \rho_{\text{free}}$ such that $\lambda_{\text{mf}} > 0$ for XIF and $\lambda_{\text{mf}} < 0$ for LIF neurons. In networks in the balanced state, the actual spike rate is much smaller than the spike rate of a neuron if only excitation is present. Since in our networks the latter equals the spike rate of the freely oscillating neuron, we have $\rho/\rho_{\text{free}} \ll 1$. Thus the mean-field approach indicates that the growth rate of perturbations is mainly given by the subthreshold growth. The mean-field approach further indicates that a single XIF neuron renders the entire network dynamics unstable and that the number of unstable directions equals the number of

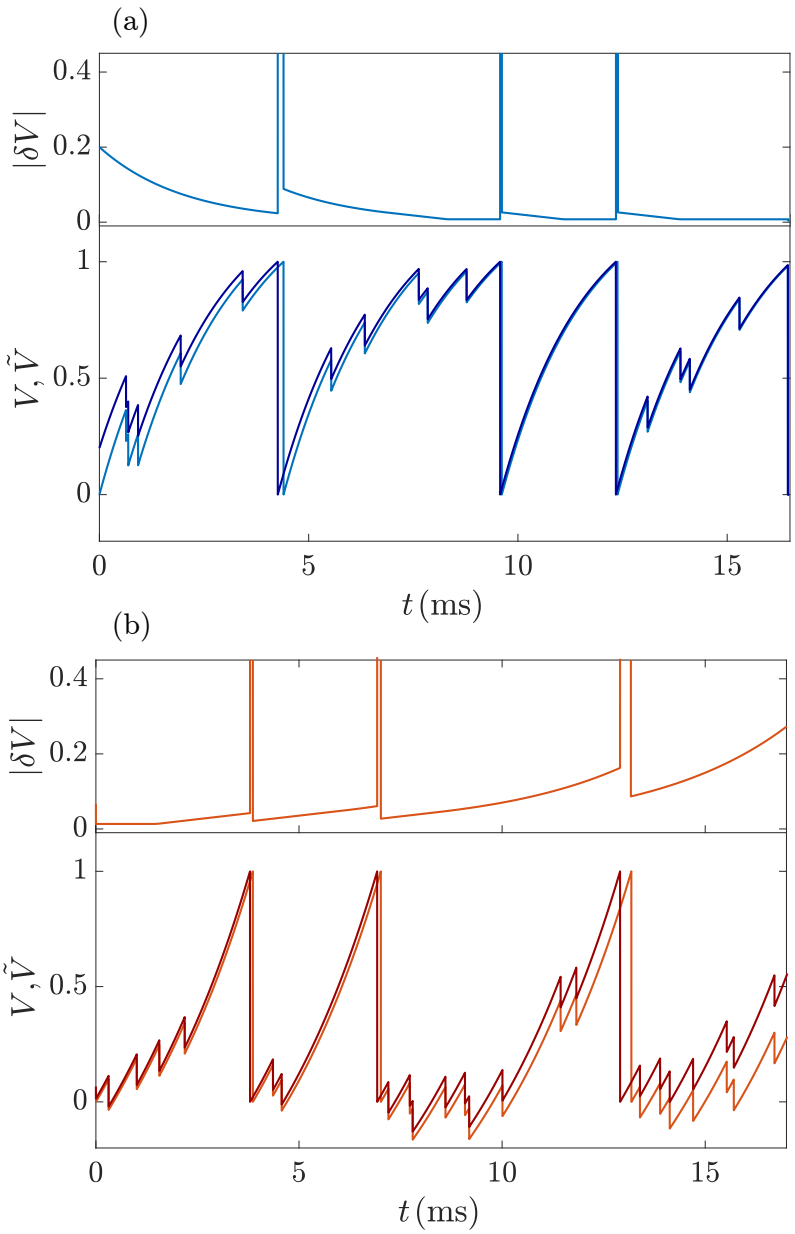

Figure 3.4: Evolution of perturbations during subthreshold evolution and spiking in (a) an LIF and (b) an XIF neuron. During subthreshold evolution the distance (perturbation) $|\delta V(t)|$ between two neighboring trajectories shrinks for LIF and grows for XIF neurons, while spike generation partially resets it. Due to the receiving of inhibitory spikes, the intervals between spike generations are generally longer than for freely oscillating neurons. The impact of the subthreshold dynamics therefore dominates and overall the perturbation in the LIF neuron decays while that in the XIF neuron grows. Here we use $V_{\infty,\mathrm{LIF}} = 1.33$, $V_{-\infty,\mathrm{XIF}} = -1$, $\gamma_{\mathrm{LIF}} = 0.5\,\mathrm{ms}^{-1}$, and $\gamma_{\mathrm{XIF}} = -0.3\,\mathrm{ms}^{-1}$ for better illustration of the mechanism.

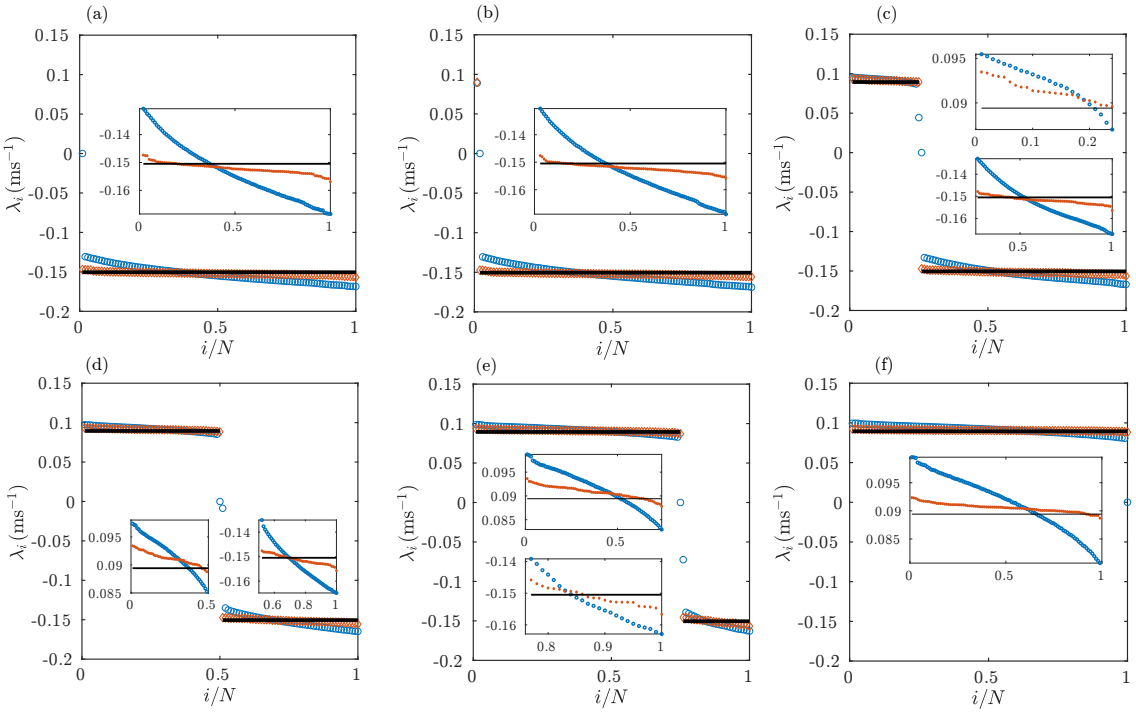

Figure 3.5: Lyapunov spectra of mixed networks. The numbers of LIF and XIF neurons in the networks are (a) $N_{\mathrm{LIF}} = 100$ and $N_{\mathrm{XIF}} = 0$, (b) $N_{\mathrm{LIF}} = 99$ and $N_{\mathrm{XIF}} = 1$, (c) $N_{\mathrm{LIF}} = 75$ and $N_{\mathrm{XIF}} = 25$, (d) $N_{\mathrm{LIF}} = 50$ and $N_{\mathrm{XIF}} = 50$, (e) $N_{\mathrm{LIF}} = 75$ and $N_{\mathrm{XIF}} = 25$, (f) $N_{\mathrm{LIF}} = 0$ and $N_{\mathrm{XIF}} = 100$. Blue circles display the numerically computed Lyapunov spectra using Eq. (3.22). Red diamonds display the mean-field result Eq. (3.21) with numerically measured neuron rates and black lines display the mean-field result Eq. (3.21) using the analytically obtained rates Eqs. (3.18) and (3.19). Insets show closeups of the positive and/or negative parts of the spectra.

XIF neurons in the network, while the number of stable directions equals the number of LIF neurons. This, however, does not give rise to a zero LE, which occurs in the full autonomous network due to time-translation symmetry. The mean-field spectrum and the rule for the number of stable and unstable directions can thus only be an approximation to the exact results.

Eq. (3.21) together with the analytical results Eqs. (3.18),(3.19) for ρ give a fully analytical estimate of the Lyapunov spectrum. Since all LIF or XIF neurons have the same analytical rate estimates and leak strengths, the spectrum consists of N_{LIF} identical negative and N_{XIF} identical positive exponents, see Fig. 3.5. Due to quenched noise from random coupling, the rates in the actual network are distributed. We can account for this by inserting the numerically measured rates into Eq. (3.21), see Fig. 3.5.

3.4.2 Network single spike Jacobian

To derive exact Lyapunov spectra we need to take into account the spreading of perturbations in the network. For this, we compute the full single spike Jacobian $J(k)$, which is a map from tangent vectors at the point $\boldsymbol{V}(t_k^+)$ in phase space to tangent vectors at $\boldsymbol{V}(t_{k+1}^+)$, where $\boldsymbol{V}(t) = \big(V_1(t), \ldots, V_N(t)\big)^T$ is the state of the system at time t. The resulting components of $J(k)$ read

$$J_{ij}(k) = \frac{\partial V_i(t_{k+1}^+)}{\partial V_j(t_k^+)} = \delta_{ij} \mathrm{e}^{-\gamma_i (t_{k+1} - t_k)} + \delta_{jl} \frac{\gamma_i}{\gamma_l} \frac{\delta_{il} V_{\mathrm{th}} - W_{il} h_i(V_i(t_{k+1}^-))}{\frac{I_l^{\mathrm{ext}}}{\gamma_l} - V_l(t_k^+)}, \tag{3.22}$$

for an LIF or an XIF neuron i, where l is the index of the neuron sending the $(k+1)$th spike, see Appendix. 3.8.C for details. We note that the mean-field theory accounts for the diagonal terms of this Jacobian.

3.4.3 Volume contraction

Owing to the simple form of the single spike Jacobians we can find an analytical expression for the full network dynamics' expansion rate of infinitesimal phase space volumes or, equivalently, for the sum of the LEs. The result in terms of the neuronal spike rates in the network is exact. It allows to analytically compute the Lyapunov spectra for two neuron systems and offers a test for the accuracy of their numerical estimates in larger networks.

The volume expansion and the sum of LEs are given by the time averaged logarithms of the determinants of the Jacobians [10]. We thus have

$$\sum_{i=1}^{N} \lambda_i = \lim_{L \to \infty} \frac{1}{t_L} \sum_{k=0}^{L-1} \ln \left| \det J(k) \right| \tag{3.23}$$

in terms of single spike Jacobians [69]. In Appendix 3.8.D we exploit the specific form of $J(k)$ to compute $\det J(k)$ with the matrix determinant lemma. The subsequent time averaging yields

$$\sum_{i=1}^{N} \lambda_i = -\sum_{j=1}^{N} \gamma_j \left(1 - \frac{\rho_j}{\rho_{\mathrm{free},j}} \right). \tag{3.24}$$

Notably, this shows that our mean-field theory yields an exact expression for the volume contraction rate and the sum of LEs: the estimate $\sum_{l=1}^{N} \lambda_{\mathrm{mf},l}$ with Eq. (3.21) agrees with the exact expression Eq. (3.24).

3.4.4 Numerical computation of the Lyapunov spectrum

The single spike Jacobians (3.22) allow us to iteratively compute the largest LE and the full Lyapunov spectrum [10, 70, 100], see also Appendix 3.8.F. In short, for the largest LE, one iterates an initial random perturbation vector by the single spike Jacobians, stores its growth every few steps and thereafter renormalizes it to its initial magnitude. The long-term average of the growth rate equals λ_1. For the full spectrum, one iterates a system of N orthogonal perturbation vectors with the single spike Jacobians. Every few steps, one records the growth of the different vectors. Thereafter one reorthogonalizes, always in the same order, and finally renormalizes the vectors. The long-term average growth rate of the first vector then equals λ_1, that of the second equals λ_2 etc. Ref. [100] suggested an efficient method to compute the Lyapunov spectrum and applied it to large networks; we use some of the ideas in our implementation.

For networks consisting only of LIF neurons we find in agreement with previous work [8, 9, 65, 66] and our mean-field theory that the largest nontrivial LE is negative, see Fig. 3.5a. However, we also

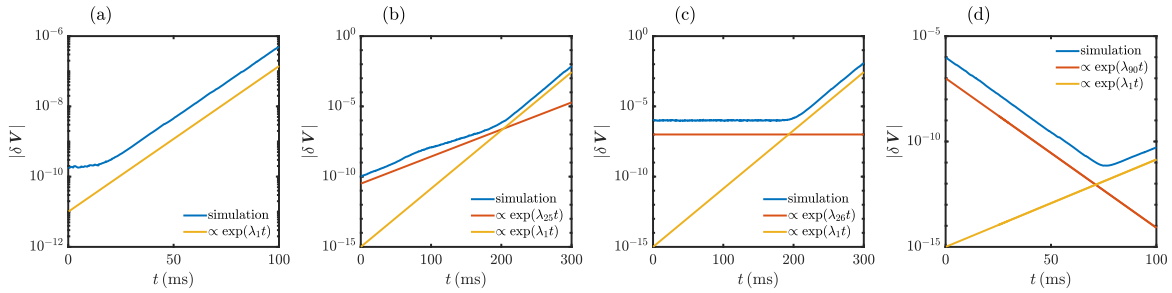

Figure 3.6: Tracking the growth or shrinkage of finite perturbations. We consider the same network as in Figs. 3.2 and 3.5c and explicitly perturb it in different directions. The semilogarithmic plots display the time evolution of distances between original and perturbed trajectories (blue) and compare them to exponential functions with growth rates equal to the relevant LEs (yellow, red). Transient large perturbations due to different event times in the two systems (cf. Fig. 3.4) are excluded. We perturb in (a) in a generic direction, in (b) in the direction of the unstable CLV $\boldsymbol{v}_{25}$ corresponding to LE λ_{25} (closest to the trivial one), in (c) in the direction of the trivial CLV $\boldsymbol{v}_{26}$ (in the direction of the trajectory) and in (d) in the direction of the stable CLV $\boldsymbol{v}_{90}$. In (a) after a short equilibration time we have growth with the largest LE. In (b),(c) and (d) the perturbation grows initially with the LE of the CLV. The small numerical error of the CLVs grows exponentially with the largest LE and eventually dominates the evolution.

find that already the presence of a single XIF neuron renders the largest LE positive, see Fig. 3.5b, indicating chaos in agreement with the mean-field theory. The computations also confirm that the destabilization of a network by a single XIF neuron is a special case of a general rule, namely that each XIF neuron introduces about one positive LE. This holds independently of N and K, see Fig. 3.5 and Appendix 3.8.E. The trivial (zero) exponent is an exception to the rule. Our numerical results indicate that it replaces a negative exponent if there are more LIF than XIF neurons in the network and a positive exponent otherwise. There is also good quantitative agreement with the mean-field spectrum, in particular the exponents are close to γ_{LIF} and γ_{XIF}. However, also when inserting the measured spike rates into Eq. (3.21) some discrepancy remains, showing that the spread of perturbations in the network and their transfer between neurons has a pronounced effect on their growth.

3.5 Stable and unstable directions

3.5.1 Lyapunov vectors and perturbation growth

To further elucidate the local phase space structure, we numerically investigate the characteristics of the perturbations that grow according to the individual LEs, i.e. how they are distributed across neurons and how they change during evolution. This will, in particular, allow us to understand why the mean-field theory works well. The directions of the perturbations are given by the CLVs or, in other words, by the stable and unstable manifolds along the trajectory [10, 63].

The ith CLV $\boldsymbol{v}_i(\boldsymbol{V}_0)$ at a point $\boldsymbol{V}_0$ in phase space is a normalized tangent vector that grows with long-term average rates λ_i and $-\lambda_i$ when evolved forward and backward in time. We call it a stable CLV if $\lambda_i < 0$ and an unstable one if $\lambda_i > 0$. We assume for simplicity that all LEs are different; the vector is then unique up to its orientation. Consider a trajectory $\boldsymbol{V}(t)$ that reaches shortly after the spike time t_0 the state $\boldsymbol{V}(t_0^+) = \boldsymbol{V}_0$. Using the single spike Jacobians $J(k)$, $\boldsymbol{v}_i(\boldsymbol{V}_0)$ may be defined as

the tangent vector satisfying

$$\left|\prod_{k=0}^{L} J(k)\, \boldsymbol{v}_i(\boldsymbol{V}_0)\right| \sim \mathrm{e}^{\lambda_i t_L}, \tag{3.25}$$

$$\left|\prod_{k=-M}^{-1} J^{-1}(k)\, \boldsymbol{v}_i(\boldsymbol{V}_0)\right| \sim \mathrm{e}^{\lambda_i t_{-M}}, \tag{3.26}$$

where L and M are chosen sufficiently large. The definition can be straightforwardly extended to states between spiking events. Both of its parts are important: The first part alone does not uniquely define the direction of $\boldsymbol{v}_i(\boldsymbol{V}_0)$, since adding any vector with growth rate less than λ_i yields the same asymptotics. The second part excludes such an addition, since its shrinkage rate is slower than $-\lambda_i$ and thus yields a different dominant asymptotics of backward evolution. As anticipated by the notation, the vector depends only on the state but not on the time when $\boldsymbol{V}(t)$ reaches it. Furthermore, the definition ensures covariance, that is the evolution of infinitesimal perturbations (the tangent flow) maps CLVs to CLVs. At subsequent spike times we thus have

$$J(k)\boldsymbol{v}_i(\boldsymbol{V}(t_k^+)) \propto \boldsymbol{v}_i(\boldsymbol{V}(t_{k+1}^+)). \tag{3.27}$$

The extension to states between spike times is again straightforward: the covariance implies that we can obtain CLVs $\boldsymbol{v}_i(\boldsymbol{V}(t))$ at a state between spike times t_k and t_{k+1} by propagating $\boldsymbol{v}_i(\boldsymbol{V}(t_k^+))$ forward with the Jacobian $\hat{J}_{ij}(t-t_k) = \delta_{ij}\exp(-\gamma_i\left(t-t_k\right))$ of subthreshold evolution.

We compute the CLVs in a dynamical manner by forward and restricted backward propagating sets of vectors, following refs. [10, 25]. Appendix 3.8.F provides a short description of the method. We note that the dynamics of our system are not invertible: given a state there is no unique way of propagating it back in time. This is because an ambiguity can arise at states where one neuron is at the reset potential; we generally cannot tell whether it was reset or crossed the reset potential from below (unless some postsynaptic neuron is too near to threshold to be able to have just received a spike). It is, however, still possible to compute the Lyapunov vectors by backward propagating along the trajectory that was previously taken for the forward propagation [25].

3.5.2 Stable and unstable directions in mixed networks

The CLVs yield the directions in which small but finite perturbations evolve according to the different LEs as shown in Fig. 3.6. We find that they generally contain perturbations to a variety of neurons and that they strongly change their direction during evolution. More specifically, we observe that the stable and unstable CLVs stay approximately confined to the subspaces of (strictly speaking: perturbations to) LIF and XIF neurons, respectively. Fig. 3.7a-d illustrates this by displaying the lengths $\sqrt{\sum_{j=1}^{N_{\mathrm{LIF}}} v_{i,j}^2(\boldsymbol{V}(t))}$ and $\sqrt{\sum_{j=N_{\mathrm{LIF}}+1}^{N} v_{i,j}^2(\boldsymbol{V}(t))}$ of the projections of different CLVs $\boldsymbol{v}_i$ onto the subspaces of LIF and XIF neurons. Here and in the following we assume that the LIF and XIF neurons have the indices $1, ..., N_{\mathrm{LIF}}$ and $N_{\mathrm{LIF}+1}, ..., N$, respectively. Fig. 3.7e,f further illustrates the confinement and shows the large temporal variability of single CLV components $v_{i,j}$ that are not close to zero. The confinement does not hold exactly since perturbations of LIF neurons usually also give rise to perturbations of XIF neurons and vice versa. In networks with inhomogeneous spike rates, we observe that single neurons that are strongly suppressed by inhibition have CLVs more aligned to

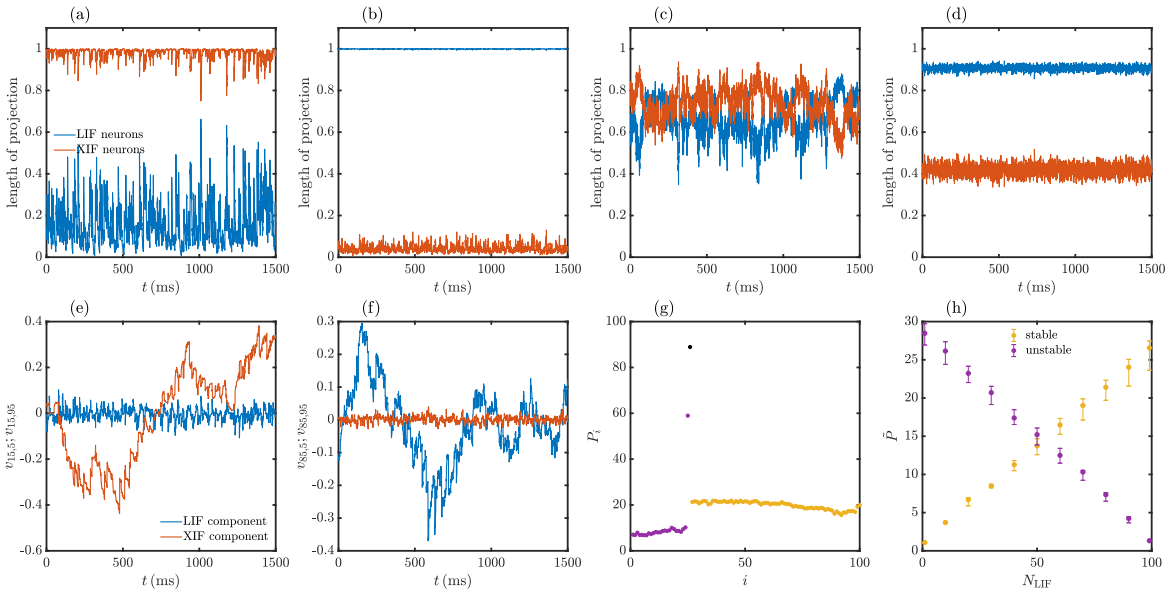

Figure 3.7: Stable and unstable CLVs are mostly confined to the subspaces of (perturbations to) LIF or XIF neurons, respectively. (a-d) show the time evolution of the lengths of the projections of different CLVs onto the subspaces of LIF (blue) and XIF (red) neurons: (a) unstable CLV $\boldsymbol{v}_6$, (b) stable CLV $\boldsymbol{v}_{88}$, (c) unstable CLV $\boldsymbol{v}_{25}$ (corresponding to the LE closest to the trivial one) and (d) trivial CLV $\boldsymbol{v}_{26}$. (e,f) show the time evolutions of the projections of different CLVs onto the subspace of a single LIF neuron (neuron index $n = 5$, blue) and a single XIF neuron ($n = 95$, red): (e) unstable vector ($\boldsymbol{v}_{15}$), (f) stable vector ($\boldsymbol{v}_{85}$). Values are plotted at event times $t = t_k$. (g) shows the participation ratios P_i of the unstable (purple), trivial (black) and stable (yellow) CLVs averaged over 10000 events. The network is the same as in Figs. 3.2, 3.5c and 3.6. (h) displays the median $\tilde{P}$ of the participation ratios of the stable (yellow) and unstable (purple) CLVs, in networks as in (a-g) with $N = 100$ neurons but with different fractions of LIF and XIF neurons. Bars indicate the first and third quartiles of the distribution.

them, because their perturbation spreads less in the network due to their lack of spiking.

We further quantify the localization of the CLVs using an inverse participation ratio (number) [25, 101], which we define for the ith CLV as

$$P_i^{-1} = \left\langle \sum_{j=1}^{N} v_{i,j}^4(\boldsymbol{V}(t_k)) \right\rangle_k . \tag{3.28}$$

Here, $\langle . \rangle_k$ is an average over sufficiently many events and we use that the CLVs are normalized, $\sum_{j=1}^{N} v_{i,j}^2(\boldsymbol{V}(t_k)) = 1$. The participation ratio P_i measures how many components contribute to a vector. If, for example, the vector $\boldsymbol{v}_i(\boldsymbol{V}(t_k))$ always has only one nonzero component, $P_i = 1$. If there are always m nonzero components of equal size, $P_i = m$. We observe that the participation ratio of unstable CLVs increases approximately linearly with the number of XIF neurons starting with $P_i \approx 1$ at $N_{\mathrm{XIF}} = 1$, consistent with a delocalization of these CLVs between the present XIF neurons, see Fig. 3.7g,h. P_i for stable CLVs increases likewise with the number of LIF neurons. The trivial CLV has a participation ratio close to N, because the components of the tangential vector $dV_j(t^+)/dt$ and thus the components of the CLV have roughly similar size.

Our mean-field approach uses the assumption that each LE is independently generated by the growth or shrinkage of a single neuron perturbation, with negligible influence of the perturbation's spread and backreaction in the network. Its suitability can now be understood as follows: The approximately N_{LIF} stable CLVs are confined to the N_{LIF}-dimensional subspace of perturbations to LIF neurons. The stable CLVs thus form a basis of the subspace of perturbations to LIF neurons. Likewise the unstable CLVs form a basis of the subspace of perturbations to XIF neurons. At each time point, a perturbation to a single LIF neuron can therefore be expressed as a linear combination of stable CLVs, while a perturbation to an XIF neuron can be expressed as a linear combination of unstable CLVs. The stable CLVs have similar decay rates (negative LEs) and the unstable CLVs have similar growth rates (positive LEs), see Fig. 3.5. Any linear combination of only stable or only unstable CLVs inherits this decay or growth rate. This holds in particular for the perturbation of a single neuron. At each time point the perturbation to a single LIF or XIF neuron thus grows according to the negative or according to the positive LEs, respectively. The mean-field approach therefore yields good results.

3.6 Computations with precisely timed spikes

3.6.1 Network architecture and task design

In the following, we employ our networks for computations. In particular, we investigate how their different phase space structures and CLVs may be exploited in specific tasks. This requires a computational scheme based on precise spiking, which is affected by the phase space structure. We design a setup where one of our recurrent neural networks acts as a kind of computational reservoir [69, 74, 75], in the sense that it randomly nonlinearly filters its inputs. An output neuron receives the generated spikes and learns to generate desired outputs, see Fig. 3.8a.

Inspired by experimental and computational neuroscience paradigms [102, 103], we assume that the networks receive inputs from context neurons, whose spiking defines the computation to be executed in the specific trial, and from input neurons. Their synaptic weights as well as the recurrent ones are static; only the output weights are learned. At the beginning of each trial, all membrane potentials are reset to zero. The recurrent network dynamics are therefore identical in trials with the same context and input neuron spikes. To keep the computational scheme consistent, we specify trains of precisely timed spikes as desired outputs.

The output neuron is an LIF neuron as used in the recurrent network. The subthreshold dynamics of its membrane potential $V_{\mathrm{out}}(t)$ are thus given by

$$\begin{aligned} V_{\mathrm{out}}(t) = &\sum_{j=1}^{N} w_j \sum_{k:t_{jk}<t} \mathrm{e}^{-\gamma_{\mathrm{LIF}}(t-t_{jk})} \\ &+ V_{\mathrm{th,out}} \left(- \sum_{t_{\mathrm{sp}}<t} \mathrm{e}^{-\gamma_{\mathrm{LIF}}(t-t_{\mathrm{sp}})} \right) \\ &+ V_{\infty,\mathrm{out}} \left(1 - \mathrm{e}^{-\gamma_{\mathrm{LIF}} t} \right), \end{aligned} \tag{3.29}$$

where w_j are the output weights, $V_{\mathrm{th,out}}$ is the threshold, t_{sp} the output spikes, $V_{\infty,\mathrm{out}}$ the asymptotic potential and N the number of spiking neurons in the recurrent network. Initially $V_{\mathrm{th,out}} = V_{\mathrm{th}}$,

$V_{\infty,\mathrm{out}} = V_{\infty,\mathrm{LIF}}$ and the w_j are are drawn randomly from the uniform distribution over $[2W, 0]$. We use Finite Precision Learning [61] to learn the input-output tasks. The shapes of the post-synaptic potentials in our single neuron dynamics are different from those in ref. [61] and there is an additional constant driving term. The learning rule can be readily adapted to this: We consider $V_{\mathrm{out}}(t) - V_{\mathrm{th,out}}$ and cast it into the form $V_{\mathrm{out}}(t) - V_{\mathrm{th,out}} = \sum_{k=1}^{N+2} w_k x_k(t)$. Spikes are generated when $V_{\mathrm{out}}(t) - V_{\mathrm{th,out}}$ reaches zero. At each time t, we thus have a kind of perceptron classification task, where w_j, $V_{\mathrm{th,out}}$ and $V_{\infty,\mathrm{out}}$ are the "weights" to be learned. The "inputs" belonging to these weights are

$$x_j(t) = \sum_{t_j<t} \mathrm{e}^{-\gamma_{\mathrm{LIF}}(t-t_j)}, \tag{3.30}$$

$$x_{N+1}(t) = -\sum_{t_{\mathrm{sp}}<t} \mathrm{e}^{-\gamma_{\mathrm{LIF}}(t-t_{\mathrm{sp}})} - 1, \tag{3.31}$$

$$x_{N+2}(t) = 1 - \mathrm{e}^{-\gamma_{\mathrm{LIF}}t}. \tag{3.32}$$

Following ref. [61], we assume a tolerance window of size ε around each desired spike (we use $\varepsilon = 1$ ms throughout). There are now two kinds of errors: (i) undesired spikes, i.e. spikes out of a tolerance window or second spikes within a tolerance window (Err = 1, the error time t_{Err} is the spike time) and (ii) missing spikes within a tolerance window (Err = -1, t_{Err} is the end of the tolerance window). The dynamics are stopped at the first error and w_j, $V_{\mathrm{th,out}}$ and $V_{\infty,\mathrm{out}}$ are corrected according to the perceptron rule,

$$\Delta w_j = -\eta \mathrm{Err} \sum_{k:t_{jk}<t_{\mathrm{Err}}} \mathrm{e}^{-\gamma_{\mathrm{LIF}}(t_{\mathrm{Err}}-t_{jk})}, \tag{3.33}$$

$$\Delta V_{\mathrm{th,out}} = \eta \mathrm{Err} \left(\sum_{t_{\mathrm{sp}}<t_{\mathrm{Err}}} \mathrm{e}^{-\gamma_{\mathrm{LIF}}(t_{\mathrm{Err}}-t_{\mathrm{sp}})} + 1 \right), \tag{3.34}$$

$$\Delta V_{\infty,\mathrm{out}} = -\eta \mathrm{Err} \left(1 - \mathrm{e}^{-\gamma_{\mathrm{LIF}}t_{\mathrm{Err}}} \right), \tag{3.35}$$

with learning rate η (we use $\eta = 0.01$). To focus on networks with inhibitory neurons throughout the article, we restrict the output weights to be inhibitory by clamping them at zero when they would become excitatory during learning. We note that a missed spike generates increases in w_j and $V_{\infty,\mathrm{out}}$ as well as a decrease in $V_{\mathrm{th,out}}$ to foster spiking. If an undesired spike occurs, the signs are reversed.

3.6.2 Switchable temporal XOR/AND

We exemplarily consider two tasks. In the first, the network of Fig. 3.8a learns to execute in context 1 a temporal XOR and in context 2 a temporal AND computation, see Fig. 3.8b-d. The weights from context and input neurons to the recurrent network are drawn randomly from the uniform distribution over $[2W, 0]$. At the beginning of a trial, at $t = 0$ ms, context neuron 1 or 2 sends a spike, specifying the context. Thereafter each input neuron sends a spike, either at time $t_{\mathrm{in,+}} = 5$ ms ("+"-input) or at $t_{\mathrm{in,-}} = 10$ ms ("−"-input). The desired output spike is at $t_{\mathrm{out,+}} = 15$ ms ("+"-output) or at $t_{\mathrm{out,-}} = 20$ ms ("−"-output), depending on the context and the input spike times.

The considered networks learn the task easily, whether the reservoir consists of LIF or XIF neurons

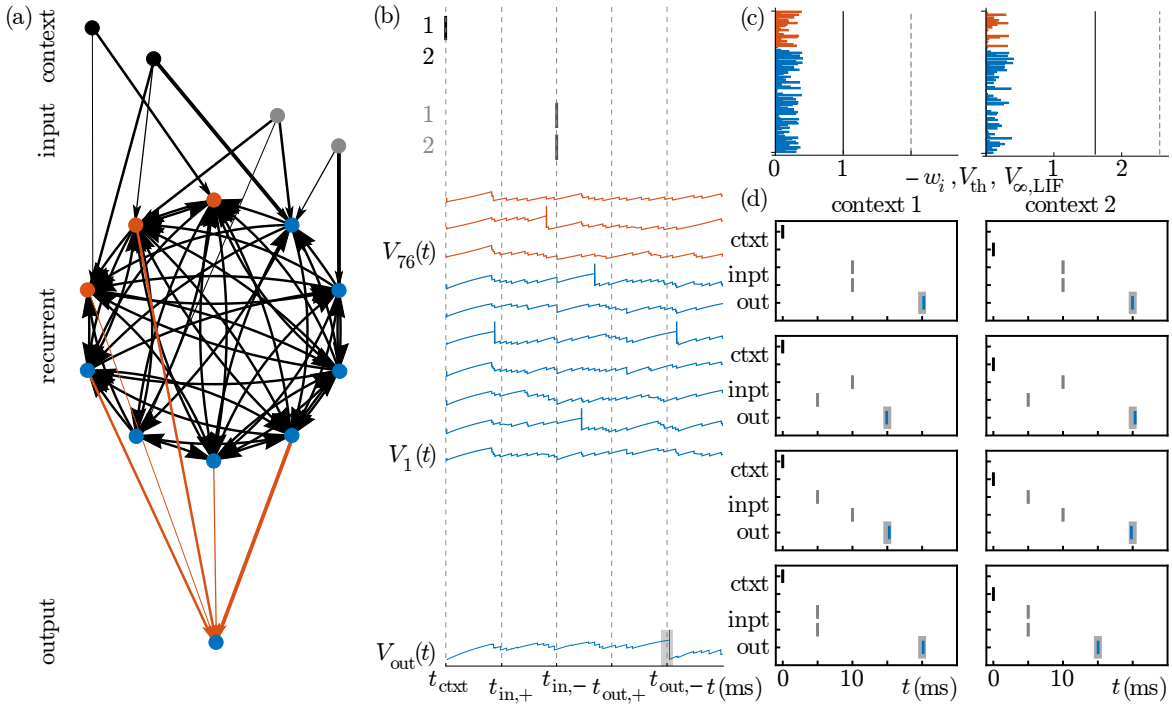

Figure 3.8: Network for precise spike-based computations and solution of the XOR/AND task. (a) Network architecture. The recurrent network (middle, blue and red for LIF and XIF neurons) receives input spikes from context (top, black) and from input neurons (gray). The output neuron (bottom) changes its plastic weights (red) according to Finite Precision Learning to learn the task. (b) Spiking of the context and input neurons (top), voltage traces of recurrent neurons (middle, LIF neurons 1-7, XIF neurons 76-78, spikes highlighted by vertical lines) and voltage traces of the output neuron (bottom) after learning the XOR/AND task. Dashed lines indicate the times of possible context (t_{ctxt}) and input neuron ($t_{in,+}$, $t_{in,-}$) spiking as well as the possible desired output spike times ($t_{out,+}$, $t_{out,-}$). The output neuron sends its spike in the tolerance window around the desired time of the specific trial (gray rectangle). (c) Output weights w_i from LIF (blue) and XIF neurons (red), V_{th} (black line) and $V_{\infty,LIF}$ (gray dashed line), before (left) and after (right) learning. The weights have overall decreased during learning, while V_{th} (black line) and $V_{\infty,LIF}$ have increased. The specific weight pattern after learning is crucial for executing the task, random weight shuffling leads to erroneous output spiking. (d) Overview of the eight spike patterns of the task after learning. In context 1, the system generates a temporal XOR computation, in context 2 a temporal AND computation. The output spikes are in the desired tolerance windows (gray rectangles) for all patterns.

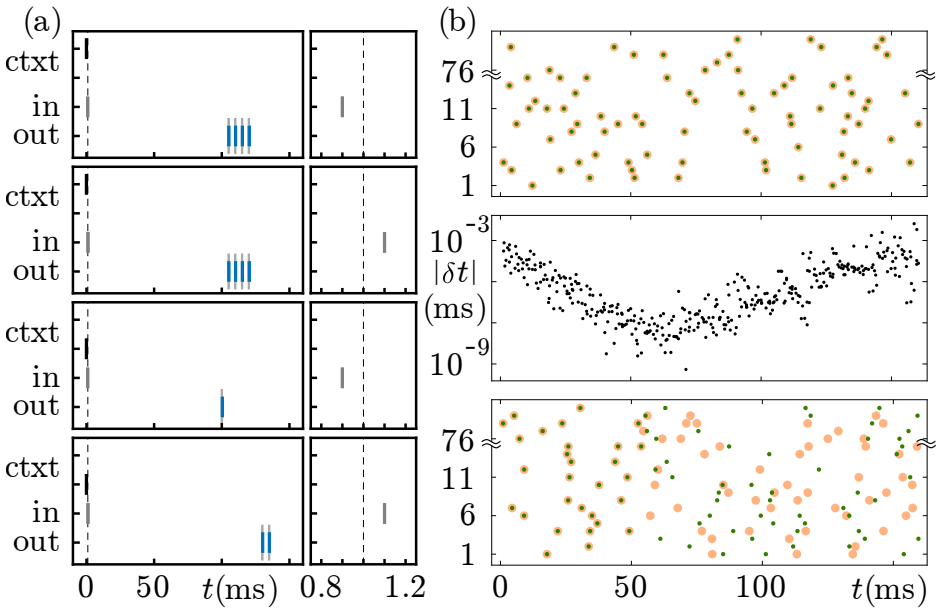

Figure 3.9: Network trained to detect or ignore input time differences, after learning. (a) Overview of the four spike patterns of the task. In context 1, the network ignores the small time difference in the input (right subpanels: closeups around input times), in context 2 it detects and highlights it by generating different numbers and timings of output spikes. After learning, the output spikes are in the desired tolerance windows (gray rectangles, appearing as lines at the displayed timescale) for all patterns. (b, top) The spiking dynamics in the recurrent reservoir (green and yellow dots) are in context 1 similar for both input times, due to the chosen context and input weights. This generally fosters and here enables learning of the same output. (middle) The temporal differences δt between reservoir spikes display the typical pattern of first shrinkage then growth, of perturbations along stable CLVs, cf. Fig. 3.6d. (bottom) The different spiking dynamics in the recurrent reservoir for different input times in context 2 allow the generation of different output.

or of a mixture of both. The example with a mixed network displayed in Fig. 3.8 (same network as in Figs. 3.2, 3.5c, 3.6 and 3.7) required 53 learning cycles, where in each cycle the four input-desired output patterns of both contexts were presented. The networks cannot learn the task, if the recurrent network dynamics at the desired output times are too similar for different contexts and input conditions. This happens for recurrent LIF networks, if the context or input neurons have coupling strengths that are so weak that the perturbations due to different input timing are small. The states are then within the same flux tube and the perturbation decays up to a time shift. In XIF and mixed networks, the recurrent dynamics are too similar if there is insufficient time for the perturbation to grow and spread before the first desired output.

3.6.3 Detect or ignore input time differences

In the second task, the system has to ignore a difference in input timing in context 1 and to detect it in context 2. The network setup is as in Fig. 3.8a, except that there is only one input neuron. This sends a spike at $t_{\text{in},+} = t_1 - \Delta t$ or at $t_{\text{in},-} = t_1 + \Delta t$ ($t_1 = 1\,\text{ms}$, $\Delta t = 0.1\,\text{ms}$). The output neuron shall generate in context 1 for both input conditions the same output, a burst of four spikes at $t = 105\,\text{ms}, 110\,\text{ms}, 115\,\text{ms}, 120\,\text{ms}$. In context 2 it shall detect the difference and highlight it by

sending one spike at $t = 100\,\mathrm{ms}$ (input at $t_{\mathrm{in},+}$) or two spikes at $t = 130\,\mathrm{ms}, 135\,\mathrm{ms}$ (input at $t_{\mathrm{in},-}$). For this task, for simplicity we assume that the impacts of input neurons do not depend on the membrane potential, i.e. for them $h_i(V_i^-) = 1$. Further, we allow context and input weights to be excitatory and inhibitory.

We find that networks with the previously chosen random parameters of external weights drawn from $[2W, 0]$ usually cannot solve the task (criterion: no convergence within 50000 cycles). The reason is different for pure LIF reservoirs and for reservoirs containing XIF neurons: In a pure LIF reservoir, the small difference in input times leads to state perturbations that are usually in the same flux tube. These decay to a temporal shift until the time of the desired outputs. The readout neuron thus cannot learn to generate two different output patterns as required in context 2. In presence of XIF neurons, the dynamics are locally unstable. The small input difference is amplified in both contexts and the reservoir spiking is different for all four patterns at the times of the desired output. The network therefore has to learn four input-output relations with eleven output spikes and silence periods in between, without being able to take advantage of the fact that two of the four output patterns are identical. This typically exceeds its learning capacity. We also observe for our parameters that the dynamics of the pure LIF reservoir can leave its flux tube due to the perturbation. If this happens only for context 2, the system can often learn the task.

To solve the problem, we design the network such that, reliably, in context 1 but not in context 2 the input differences leave the reservoir spiking at the time of the desired outputs largely unaffected. This can be achieved by choosing the context and input couplings such that the input difference generates a state perturbation along a stable CLV of the reservoir dynamics in context 1. In contrast, for context 2 the state perturbation should have a component in the direction of an unstable CLV such that it is quickly amplified. The setup requires mixed networks with both types of CLVs. We note that an alternative approach might exploit the dichotomy of large and small perturbations, which do and do not leave the flux tubes of pure LIF networks.

To derive appropriate weights, we compute the state perturbations in the reservoir assuming that in the "unperturbed" system the input arrives at t_1. We there have

$$V_j(t_1^+) = V_j(t_1^-) + W_j^{\mathrm{in}}, \tag{3.36}$$

where W_j^{in} is the coupling strength from the input neuron to neuron j. In the "perturbed" system, the input arrives shifted by δt (here: $\delta t = \pm\Delta t$), such that we have in linear approximation

$$\tilde{V}_j(t_1 + \delta t^+) = V_j(t_1^-) + \dot{V}_j(t_1^-)\delta t + W_j^{\mathrm{in}}. \tag{3.37}$$

To compute a perturbation in the $V_j(t_1^+)$ that corresponds to the perturbation due to the temporal shift of input, we propagate the perturbed potential in linear approximation from $t_1 + \delta t^+$ to t_1^+,

$$\begin{aligned}
\delta V_j(t^+) &\approx \tilde{V}_j(t_1 + \delta t^+) - \dot{\tilde{V}}_j(t_1 + \delta t^+)\delta t - V_j(t_1^+) \\
&\approx \dot{V}_j(t_1^-)\delta t - \left[(-\gamma_j)\left(V_j(t_1^-) + W_j^{\mathrm{in}}\right) + I_j^{\mathrm{ext}}\right]\delta t \\
&= \gamma_j W_j^{\mathrm{in}}\delta t.
\end{aligned} \tag{3.38}$$

A temporal input difference that should be ignored should be proportional to a stable CLV $\boldsymbol{v}_i$ at the

state $\boldsymbol{V}(t_1^+)$, i.e.

$$\gamma_j W_j^{\text{in}} \propto v_{i,j}(\boldsymbol{V}(t_1^+)). \tag{3.39}$$

We choose the same recurrent network as in Figs. 3.2, 3.5c, 3.6 and 3.7 and the same CLV as in Fig. 3.6d at $t = 0\,\text{ms}$, i.e. $i = 90$ and

$$\boldsymbol{V}(t_1^+) = \boldsymbol{V}_0, \tag{3.40}$$

where $\boldsymbol{V}_0$ is the state at which the vector was recorded. Context input 1 determines the state at t_1^- by fixing the initial conditions of the dynamics. We choose as context input weights

$$\begin{aligned} W_j^{\text{ctxt},1} &= V_j(0^+) \\ &= e^{\gamma_j t_1}\left(V_{0,j} - W_j^{\text{in}}\right) + V_\infty\left(1 - e^{\gamma_j t_1}\right), \end{aligned} \tag{3.41}$$

which lead to Eq. (3.40) after free propagation until t_1 and receiving of the input W_j^{in}. To ensure that the perturbation in context 2 has a component in the direction of an unstable CLV, it suffices to choose a random context weight vector, such that $\boldsymbol{V}(t_1^+) \neq \boldsymbol{V}_0$ and $\boldsymbol{v}_i(\boldsymbol{V}_0)$ is typically not a stable CLV or a linear combination of stable CLVs at the state $\boldsymbol{V}(t_1^+)$. We randomly permute the entries of $\boldsymbol{W}^{\text{ctxt},1}$ to obtain $\boldsymbol{W}^{\text{ctxt},2}$.

We find that the network constructed in this way can reliably learn the task. The example displayed in Fig. 3.9 uses a proportionality factor of 0.01 in Eq. (3.39); the output weights converged after 146 cycles.

3.7 Discussion

In the present article we investigate the spiking and membrane potential statistics, the stability properties and the phase space structure of mixed networks containing conventional LIF neurons and XIF neurons with convex rise function. The recurrent connections are inhibitory and the synaptic currents have infinitesimal temporal extent. We employ two analytical mean-field approaches, one for the statistics and one for the dynamical stability properties; numerical simulations yield additional features of the dynamics and a better understanding of the analytical approximations. Finally, we apply the networks for computation with spikes, exploiting our insights into the dynamics.

We investigate networks in the balanced state. To establish it in our networks, we introduce a voltage dependence in the XIF neuron inputs: below a certain potential, further input has no impact. This simple model of a conductance-based synapse prevents XIF neurons from switching off and provides a good-natured nonlinearity, which leaves the dynamics analytically tractable.

The balanced state is typically investigated using spiking network models with an excitatory and an inhibitory neuron population or with a single population of hybrid excitatory-inhibitory or inhibitory neurons [5, 48, 62, 70, 104]. While detailed models of small circuits with specific abilities such as central pattern generators commonly consider multiple neuron types [105], studies on the impact of mixed populations of multiple neuron types on the collective dynamics of larger networks are rare. Ref. [106] simulated networks with excitatory and inhibitory populations containing resonator and integrator type neurons. These mixed networks both persistently generated activity and quickly changed their overall rate in response to inputs, thereby combining abilities of their pure counterparts. Refs. [107] and [108] considered models for working memory and visual processing with different types of interneurons that were grouped into distinct populations with different connectivities.

We characterize the balanced dynamics of inhibitory mixed LIF and XIF networks first from a statistical perspective, adopting a shot noise approach, which accounts for the finite input rate and finite size of individual inputs [60, 86, 96, 97]. We extend this approach to XIF neurons and derive their steady-state firing rate and voltage probability density. In contrast to the case of LIF neurons, the final continuity equation needs to be integrated numerically, due to the nonlinearity in the XIF input. We apply the results to obtain neuron parameters that lead to homogeneous firing rates for our further considered networks. We insert these rates into the mean-field expressions of the Lyapunov exponents (LEs) and thus analytically determine the dynamical stability properties of the network.

While networks of LIF neurons have stable dynamics [8, 9, 65, 66, 69], we find that already one XIF neuron gives rise to a positive largest LE indicating chaos, in contrast to the robustness against introducing excitatory connections [9, 66]. We give an analytical argument for this and expand it to a mean-field estimate of the entire Lyapunov spectrum. Simply put, the destabilizing effect of excitatory inputs will be compensated by receiving inhibitory ones, if the latter dominate and the period of spiking is overall increased compared to the free neurons. If one introduces an XIF neuron there is nothing which could counteract the increase of its perturbation through inhibitory input other than an unlikely network backreaction triggered by its perturbed output spikes. We note that in the phase representation of LIF neurons used in ref. [9], in contrast to our voltage representation an excitatory input explicitly increases a perturbation, while an inhibitory input decreases it, unless the excitatory input is suprathreshold [109, 110].

While computing the largest LE is a standard procedure, few studies have so far obtained a large part or the entire spectrum of balanced spiking dynamics. They considered a single homogeneous or an excitatory and an inhibitory neuron population [69, 70, 100, 111–113]. We analytically and numerically obtain the full spectrum for mixed networks of inhibitory LIF and XIF neurons. Interestingly, we find that it separates into two parts, in contrast to the ones reported previously including those of networks with separate excitatory and inhibitory populations. Furthermore, we compute the covariant Lyapunov vectors (CLVs) of the dynamics [10, 63]. They provide us with further insight into the phase space structure and the approximations underlying the mean-field analysis of LEs. The stable (unstable) CLVs are approximately aligned to the subspace of perturbations to LIF (XIF) neurons.

Our mean-field analysis predicts that the number of negative (positive) LEs is equal to the number of LIF (XIF) neurons. Since the underlying arguments do not depend on the neurons' specifics, we expect this to hold for any types of neurons with purely concave and convex rise functions. The mean-field analysis further indicates that the size of the LEs is approximately given by the strength of the leak and the quotient of free and actual spike frequency. The LEs are thus largely independent of the collective dynamics but rather reflect properties of individual neurons. This implies in particular that the typical perturbation growth rate does not change with network size. It further implies that in the balanced state, where the ratio of actual and free spike rate is low, the LEs are mainly determined by the single neuron leak strengths, see ref. [69] for a similar finding in large networks of LIF neurons with high indegree. The result is a consequence of the linear subthreshold dynamics of the neurons, which imply that the increase or decrease of a perturbation is independent of the state of the neuron when receiving a spike. We note that ref. [114] defined the Lyapunov spectrum as consisting of mean-field LEs in a numerical study on LIF neuron networks.

Our numerical computations of the Lyapunov spectrum show that the mean-field result is a good approximation. We explain this by analyzing the CLVs. Furthermore, we derive an exact expression for the change of phase space volume, which agrees with the mean-field result.

The presence of discrete events and the possibly large impact of changing their order could in

principle render the transfer of insights on infinitesimal perturbations to finite ones difficult. Refs. [8, 9] studied the evolution of finite size perturbations in the pure LIF network model with stable dynamics and showed that finite size perturbations decay exponentially fast, while the minimal perturbation leading to a change of event order decreases only algebraically. Thus, for sufficiently small initial finite size perturbations the probability of a change of event order goes to zero and no difficulties occur. For unstable dynamics, we may expect generic interchanges of event order to be an additional source of deviations between trajectories so that small finite size perturbations grow as fast and larger ones at least as fast as their infinitesimal counterparts. We therefore focus mostly on linear stability analysis in the present article. Our numerical simulations employ finite size perturbations and confirm the results.

To illustrate the usefulness of our findings we apply the considered networks to neural computations. We propose a computing scheme based on precisely timed spikes where details of the phase space structure matter. In particular, our solution of the second task exploits details of the network's state space, the stability or instability of the spiking dynamics against perturbations in the direction of different CLVs. This may be especially relevant for neuromorphic computing, where precise spike-based schemes receive increasing interest [84, 85, 115–117]. In our setup, the inputs are fed into a random recurrent network, whose neurons generate precisely timed spike trains, which depend nonlinearly on the input. In this sense, the recurrent network acts like a random filter bank and computational reservoir. The spike trains are read out by a spiking neuron. In contrast to previous spiking reservoir computers [75, 118–121], we use trains of precisely timed spikes as targets. To train the readout neuron, we use Finite Precision Learning [61]. It was introduced for neurons with temporally extended input currents of either sign. In our study we adapt it to a neuron with inhibitory, infinitesimally short input currents and constant external drive. We note that the general phase space structure implies that the considered networks do not lend themselves to conventional reservoir computing: there is no global fixed point, which could be reached by the spiking dynamics such that sufficiently long past input is forgotten. In other words, our networks do not have the so-called echo state property [122]. We therefore introduce a forgetting mechanism by resetting the network at the beginning of a trial.

Our findings show that by choosing appropriate numbers of LIF and XIF neurons, one can straightforwardly construct spiking networks with a desired number of stable and unstable directions. The obtained CLVs allow to exploit them for computation: one can choose the input weights such that meaningless inputs and input perturbations happen along stable directions while meaningful ones have a component in an unstable direction; the former ones are suppressed while the latter ones are amplified. Our mixed networks thus combine the computational capabilities of purely stable and purely unstable networks. It is tempting to speculate that also in the brain the combination of different neuron types might globally change the phase space structure and lead to combinations of computational capabilities that can be selected with different input vectors. While we have chosen the input weights by hand, plasticity rules for spiking networks in the brain as well as future artificial ones may allow to find them by learning.

3.8 Appendices

3.8.A Voltage probability distribution of LIF neurons

In the following, we further discuss the discrepancy between the voltage probability density of an LIF neuron obtained by the shot noise approach and the one observed if the inputs are spike trains

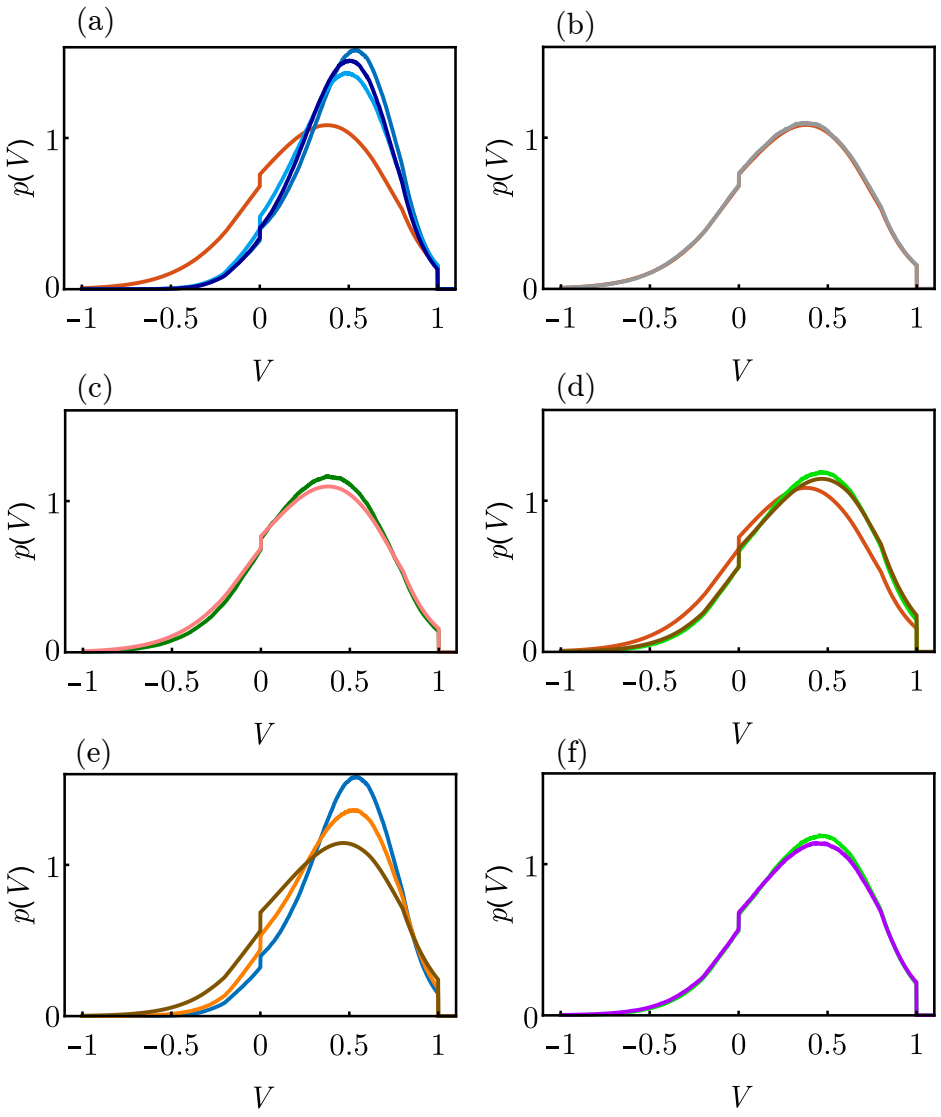

Figure 3.10: Comparison of different numerically sampled and analytical voltage probability densities $p(V)$ of LIF neurons. (a) Analytical density from Eqs. (3.9) and (3.18) (red) and numerically sampled $p(V)$ of three neurons receiving different sets of $K = 50$ spike trains from the network of Fig. 3.2 (different shades of blue). (b) Analytical estimate (red, mostly covered) and numerically sampled $p(V)$ of an LIF neuron receiving Poisson input with the same rate (gray). (c) Similar plot as in (a) with a single numerically sampled distribution of an LIF neuron receiving $K = 50$ spike trains from a larger network of $N = 10000$ neurons (dark green) and an additional analytically estimated $p(V)$ using Eqs. (3.9) and (3.15) with input rate r set to the rate of the superposed network spike trains (light red). (d) $p(V)$ of an LIF neuron receiving $K = 50$ time-shifted spike trains from the network of Fig. 3.2 (light green), analytical estimate with the same input rate (brown) and analytical estimate as in (a) (red). (e) $p(V)$ of LIF neurons receiving $K = 50$ spike trains from the network of Fig. 3.2 (blue), Gamma process input with the same rate and CV (orange) and Poisson input with the same rate (brown). (f) $p(V)$ of LIF neurons receiving shifted spike trains as in (d) (light green) and Gamma process input with the same rate and CV (purple).

recorded in the network of Fig. 3.2, cf. Figs. 3.3a and 3.10a. The analytical density obtained by the shot-noise approach, Eq. (3.9) with rate given by Eq. (3.18), matches that of an LIF neuron receiving a Poisson spike train with the same rate and spike impact strength, see Fig. 3.10b. Hence, we can attribute the observed discrepancy for LIF neurons with network spike train input to deviations of the spike trains' rate and the assumed Poisson statistics.

We expect that the discrepancy is mainly caused by spatial correlations that arise in a rather dense network of $N = 100$ neurons with an indegree of $K = 50$. To substantiate this we reduce the correlations in two ways: First, we use spike trains from a sparse network with $N = 10000$ and $K = 50$ to generate the neuron input. Secondly, we randomly shift the individual spike trains of the original $N = 100$ network in time before superposing them to generate the input; this eliminates spatial correlations while keeping the temporal correlations of the individual spike trains intact. Fig. 3.10c,d shows that both manipulations strongly reduce the discrepancy to the analytical density. Some of the remaining discrepancy is due to the difference between the network spike rate and the result of Eq. (3.18), see Fig. 3.10d.

Finally we explore the impact of the reduced variability of the inter-spike intervals. For this, we use Poisson and Gamma process input spike trains. The latter are completely characterized by their rate and the coefficient of variation of the inter-spike-interval distribution, which we match to those of the superposed spike trains of the network of Fig. 3.2. The quality of approximation increases when taking into account the reduced variability, see Fig. 3.10e. Also if the input is a superposition of shifted spike trains, accounting for it still slightly improves the similarity between the resulting $p(V)$, compare Fig. 3.10f and Fig. 3.10d (which matches the result for Poisson input).

3.8.B Mean-field Lyapunov exponents

In the following we compute the mean-field LEs. For this, we describe the dynamics by a sequence of discrete maps from the state at a time (strictly speaking: infinitesimally) shortly after generation of a spike to the state at a time shortly after generation of the next spike. We take a stroboscopic map approach, i.e. the times remain unchanged if a small perturbation is applied to the dynamics. The dynamics of small perturbations are encoded in the Jacobian matrices at each time point. Specifically, for our discrete description we need the single spike Jacobians [69, 70] $J(k)$. They generally describe the linear evolution of infinitesimal perturbations from time $t_k^+ = t_k + \varepsilon$ (with $\varepsilon > 0$ arbitrarily small) shortly after the kth spike event in a network to time t_{k+1}^+ shortly after the next spike. For our effective single neuron dynamics, they reduce to scalar factors

$$J(k) = \frac{\partial V(t_{k+1}^+)}{\partial V(t_k^+)}, \tag{3.42}$$

where the relevant events are the spike generations of the considered neuron. To compute $J(k)$, we first recall that the free evolution between spikes until time $t_{k+1}^- = t_{k+1} - \varepsilon$ yields

$$V(t_{k+1}^-) = V(t_k^+)\mathrm{e}^{-\gamma(t_{k+1}-t_k)} + \frac{I^{\mathrm{ext}}}{\gamma}\left[1 - \mathrm{e}^{-\gamma(t_{k+1}-t_k)}\right], \tag{3.43}$$

see Eq. (3.1). At t_{k+1} the neuron is reset,

$$V(t_{k+1}^+) = V(t_{k+1}^-) - V_{\mathrm{th}}, \tag{3.44}$$

and $V(t_{k+1}^-) = V_{\text{th}}$ implies an implicit dependence of t_{k+1} on $V(t_k^+)$ via

$$V_{\text{th}} = V(t_k^+)\mathrm{e}^{-\gamma(t_{k+1}-t_k)} + \frac{I^{\text{ext}}}{\gamma}\left[1 - \mathrm{e}^{-\gamma(t_{k+1}-t_k)}\right]. \tag{3.45}$$

We now consider the evolution of an infinitesimal perturbation of the membrane potential. According to Eqs. (3.43), (3.44), the perturbation $\delta V(t_k^+)$ changes until t_{k+1}^- by a factor $\mathrm{e}^{-\gamma(t_{k+1}-t_k)}$. Further, it generates a perturbation $\delta t_{k+1} = \left(\partial t_{k+1}/\partial V(t_k^+)\right)\delta V(t_k^+)$ of t_{k+1}. The different evaluation time before the spike event results in an additional membrane potential change $\dot{V}(t_{k+1}^-)\delta t_{k+1}$, which lets the neuron reach the threshold at $\left(t_{k+1} + \delta t_{k+1}\right)^-$. Since we have a stroboscopic description, we need to compensate the time shift to obtain the state at t_{k+1}^+. This is achieved by subtracting $-\dot{V}(t_{k+1}^+)\delta t_{k+1}$. A perturbation $\delta V(t_k^+)$ of the state at t_k^+ thus generates at time t_{k+1}^+ a perturbation

$$\delta V(t_{k+1}^+) = \mathrm{e}^{-\gamma(t_{k+1}-t_k)}\delta V(t_k^+) + \left[\dot{V}(t_{k+1}^-) - \dot{V}(t_{k+1}^+)\right] \times \frac{\partial t_{k+1}}{\partial V(t_k^+)}\delta V(t_k^+) \tag{3.46}$$

and the resulting mean-field Jacobian reads

$$J_{\text{mf}}(k) = \mathrm{e}^{-\gamma(t_{k+1}-t_k)} + \left[\dot{V}(t_{k+1}^-) - \dot{V}(t_{k+1}^+)\right] \frac{\partial t_{k+1}}{\partial V(t_k^+)}. \tag{3.47}$$

Application of the implicit function theorem,

$$\frac{\partial t_{k+1}}{\partial V(t_k^+)} = -\frac{1}{\dot{V}(t_{k+1}^-)}\frac{\partial V(t_{k+1}^-)}{\partial V(t_k^+)}, \tag{3.48}$$

and inserting Eqs. (3.1), (3.43), (3.44), (3.45), (3.4) or (3.5) results in

$$J_{\text{mf}}(k) = \frac{\dot{V}(t_{k+1}^+)}{\dot{V}(t_{k+1}^-)}\mathrm{e}^{-\gamma(t_{k+1}-t_k)} = \exp\left(\frac{\gamma}{\rho_{\text{free}}} - \gamma(t_{k+1} - t_k)\right). \tag{3.49}$$

We note that another, equivalent derivation of $J(k)$ first computes the voltages at a fixed time t' between t_{k+1} and t_{k+2} in terms of the voltages at another fixed time t between t_k and t_{k+1}. Taking derivatives leads to the Jacobian for the dynamical evolution from t to t'. The limits $t \searrow t_k$ and $t' \searrow t_{k+1}$ then yield $J(k)$.

The growth rate of perturbations and thus the mean-field LE are given by the long-term average of Eq. (3.49),

$$\begin{aligned}
\lambda_{\text{mf}} &= \lim_{L\to\infty}\frac{1}{t_L}\sum_{k=0}^{L-1}\ln\left|J_{\text{mf}}(k)\right| \\
&= \lim_{L\to\infty}\frac{1}{t_L}\sum_{k=0}^{L-1}\frac{\gamma}{\rho_{\text{free}}} - \gamma \\
&= -\gamma\left(1 - \frac{\rho}{\rho_{\text{free}}}\right).
\end{aligned} \tag{3.50}$$

3.8.C Network single spike Jacobian

The components of the single spike Jacobian $J(k)$ are given by

$$J_{ij}(k) = \frac{\partial V_i(t_{k+1}^+)}{\partial V_j(t_k^+)}. \tag{3.51}$$

To compute them, as in our mean-field approach we need to take into account the decay of perturbations between spikes as well as the reset of the neuron sending the $(k+1)$th spike, say neuron l. $V_l(t_{k+1}^-) = V_{\mathrm{th}}$ implies an implicit dependence of the spike time t_{k+1} on $V_l(t_k^+)$ as in Eq. (3.45). Additionally, we now have to include the jump-like potential change by $W_{il}h_i(V_i(t_{k+1}^-))$ that the spike induces in neuron i, such that

$$V_i(t_{k+1}^+) = V_i(t_k^+)\mathrm{e}^{-\gamma_i(t_{k+1}-t_k)} + \frac{I_i^{\mathrm{ext}}}{\gamma_i}\left[1-\mathrm{e}^{-\gamma_i(t_{k+1}-t_k)}\right] + W_{il}h_i(V_i(t_{k+1}^-)) - \delta_{il}V_{\mathrm{th}}. \tag{3.52}$$

The stroboscopic description yields a dependence

$$\delta V_i(t_{k+1}^+) = \mathrm{e}^{-\gamma_i(t_{k+1}-t_k)}\delta V_i(t_k^+) + \left[\dot V_i(t_{k+1}^-) - \dot V_i(t_{k+1}^+)\right]\frac{\partial t_{k+1}}{\partial V_l(t_k^+)}\delta V_l(t_k^+) \tag{3.53}$$

of the perturbation $\delta V_i(t_{k+1}^+)$ on the perturbations at the state at t_k^+. This is analogous to Eq. (3.46), with the difference that the neuron that sends the spike and determines the shift in t_{k+1} (neuron l) may be different from neuron i. The Jacobian thus reads

$$J_{ij}(k) = \delta_{ij}\mathrm{e}^{-\gamma_i(t_{k+1}-t_k)} + \delta_{jl}\left[\dot V_i(t_{k+1}^-) - \dot V_i(t_{k+1}^+)\right]\frac{\partial t_{k+1}}{\partial V_l(t_k^+)}. \tag{3.54}$$

Application of the implicit function theorem and inserting Eqs. (3.1), (3.52), (3.45) results in

$$J_{ij}(k) = \delta_{ij}\mathrm{e}^{-\gamma_i(t_{k+1}-t_k)} + \delta_{jl}\frac{\gamma_i}{\gamma_l}\frac{\delta_{il}V_{\mathrm{th}} - W_{il}h_i(V_i(t_{k+1}^-))}{\frac{I_l^{\mathrm{ext}}}{\gamma_l} - V_l(t_k^+)}, \tag{3.55}$$

for an LIF or an XIF neuron i.

3.8.D Volume contraction

The volume expansion and the sum of LEs are given by the time averaged logarithms of the determinants of the Jacobians [10]. We thus have

$$\sum_{i=1}^{N}\lambda_i = \lim_{L\to\infty}\frac{1}{t_L}\sum_{k=0}^{L-1}\ln\left|\det J(k)\right| \tag{3.56}$$

in terms of single spike Jacobians [69]. The specific form of our $J(k)$ allows to split it into a diagonal matrix $\hat J(k)$ covering the perturbation change during subthreshold evolution and a rank one correction,

$$J(k) = \hat J(k) + \boldsymbol{u}(k)\boldsymbol{v}(k)^{\mathrm{T}}, \tag{3.57}$$

where

$$\hat{J}_{ij}(k) = \delta_{ij} \mathrm{e}^{-\gamma_i (t_{k+1} - t_k)}, \tag{3.58}$$

$$v_j(k) = \delta_{jl}, \tag{3.59}$$

$$u_i(k) = \frac{\gamma_i}{\gamma_l} \left[\frac{\delta_{il} V_{\mathrm{th}} - W_{il} h_i(V_i(t_{k+1}^-))}{\frac{I_l^{\mathrm{ext}}}{\gamma_l} - V_l(t_k^+)} \right]. \tag{3.60}$$

The matrix determinant lemma now allows to compute $\det J(k)$ via

$$\det J(k) = \left[1 + \boldsymbol{v}(k)^{\mathrm{T}} \hat{J}(k)^{-1} \boldsymbol{u}(k) \right] \det \hat{J}(k). \tag{3.61}$$

Eq. (3.45) and the relation $1 + V_{\mathrm{th}}/(I_l^{\mathrm{ext}}/\gamma_l - V_{\mathrm{th}}) = \exp(\gamma_l/\rho_{\mathrm{free},l})$ for the free spike frequency $\rho_{\mathrm{free},l}$ of neuron l (see Eqs. (3.4), (3.5)) lead to

$$\det J(k) = \exp\left(\frac{\gamma_l}{\rho_{\mathrm{free},l}} - \left(\sum_{i=1}^{N} \gamma_i \right) (t_{k+1} - t_k) \right). \tag{3.62}$$

Time averaging yields

$$\begin{aligned} \sum_{i=1}^{N} \lambda_i &= \lim_{L\to\infty} \frac{1}{t_L} \sum_{k=0}^{L-1} \ln \det J_k \\ &= \lim_{L\to\infty} \frac{1}{t_L} \sum_{k=0}^{L-1} \frac{\gamma_{l(k)}}{\rho_{\mathrm{free},l(k)}} - \sum_{j=1}^{N} \gamma_j \\ &= -\sum_{j=1}^{N} \gamma_j \left(1 - \frac{\rho_j}{\rho_{\mathrm{free},j}} \right), \end{aligned} \tag{3.63}$$

where the index $l(k)$ denotes the neuron that spikes at time t_k and ρ_j is the spike rate of neuron j in the network.

3.8.E Dependence of the Lyapunov spectrum on indegree and network size

The rule that the number of negative (positive) LEs approximately equals the number of LIF (XIF) neurons holds independent of N and K, see Fig. 3.11. Fig. 3.11a-c indicates that for large N and sufficiently large fixed indegree the Lyapunov spectrum assumes a fixed shape, which differs from the result of our mean field approach. This is because the mean field approach neglects the nonzero off-diagonal entries of the single-spike Jacobians, whose strength and average number K do not depend on N, see Appendix 3.8.C. The shape of the Lyapunov spectrum varies with the indegree of the network. For larger ratios K/N the positive and negative parts of the Lyapunov spectrum become flatter, cf. Fig. 3.11d. We note, however, that also the spiking becomes more regular. We observe for very sparse but still strongly connected networks that the Lyapunov spectrum is no longer well approximated by our mean field theory, cf. Fig. 3.11d.

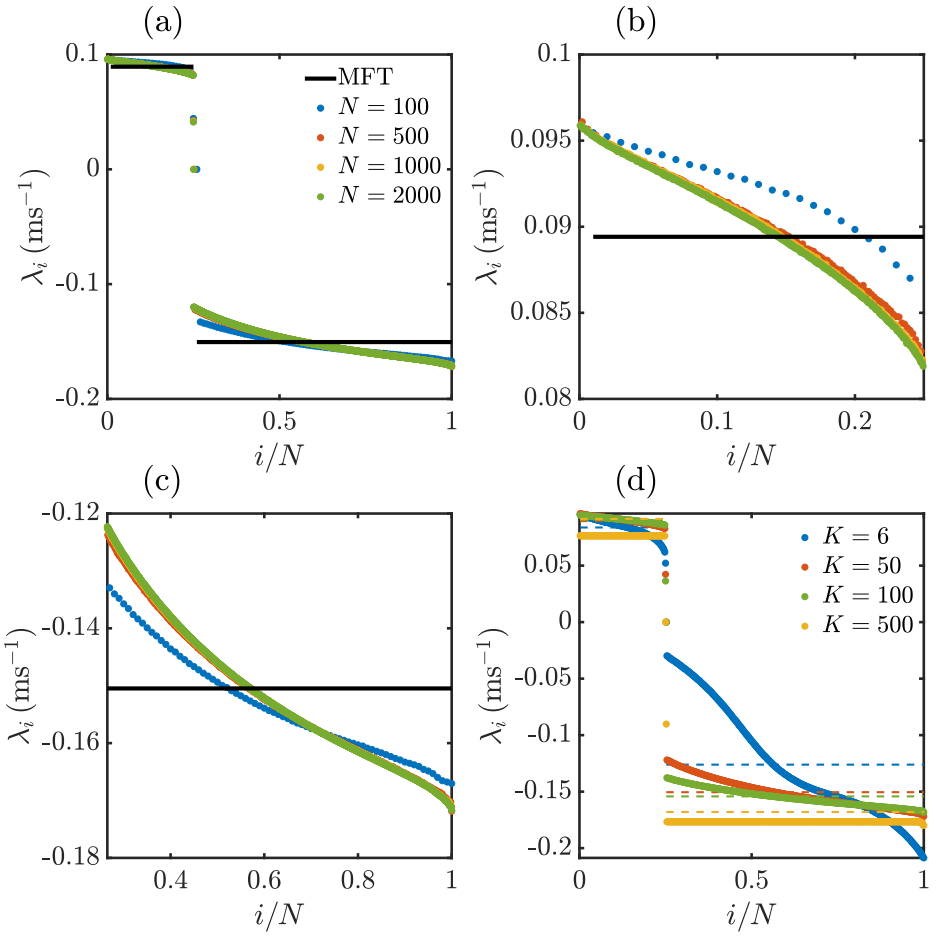

Figure 3.11: Lyapunov spectra of networks with different sizes and indegrees. (a) Lyapunov spectra for different network sizes N and constant indegree $K = 50$ (numerical results: color coded dots; result of the mean field theory: horizontal black lines). (b,c) Closeups of the positive and negative parts of the spectra in (a). (d) Lyapunov spectra for different indegrees K and size $N = 500$ (numerical results: color coded dots; results of the mean field theory: color coded dashed lines). Remaining network parameters are as in Fig. 3.2.

3.8.F Computing covariant Lyapunov vectors

We compute the CLVs in a dynamical manner [10, 25]. In short, if we want to compute them at $t = 0$, we start sufficiently long before with an arbitrary set of N orthonormal vectors $\boldsymbol{q}_i(t_0)$, which forms a basis of the tangent space. We evolve this basis forward until zero and further to a sufficiently long time t_f, using the single spike Jacobians. Every few steps, we reorthonormalize the basis. The orthogonalizations leave the first vector $\boldsymbol{q}_1(t)$ unchanged. It thus evolves freely (up to normalization) until it has aligned with the first covariant Lyapunov vector at $t = 0$ and thus also at $t = t_f$. The second vector, $\boldsymbol{q}_2(t)$, is kept orthogonal to $\boldsymbol{q}_1(t)$. Since it otherwise evolves freely, $\boldsymbol{q}_2(0)$ will lie in the subspace of the first and the second CLV at $t = 0$, which are in general not orthogonal; the same holds for $\boldsymbol{q}_2(t_f)$ at $t = t_f$. Analogously $\boldsymbol{q}_3(0)$ will be in the subspace of $\boldsymbol{v}_1(0)$, $\boldsymbol{v}_2(0)$, and $\boldsymbol{v}_3(0)$, and so on. As noted in Section 3.4.4, the growth rates of the vectors already yield the LEs. In order to find the CLVs one uses the time reversal property: we evolve the vectors $\boldsymbol{q}_i(t_f)$ along the previously taken forward trajectory back in time until $t = 0$. During this, we keep them restricted to their respective subspaces, which are known from the forward propagation. The vectors will then align with the least expanding directions of their subspaces, so the backpropagated $\boldsymbol{q}_2(t_f)$ will align with $\boldsymbol{v}_2$, the backpropagated $\boldsymbol{q}_3(t_f)$ with $\boldsymbol{v}_3$, and so on. We concretely implemented the simple and efficient algorithm derived in ref. [25], which performs the backpropagation by representing and mapping the vectors in terms of their components in the bases $\boldsymbol{q}_i(t)$. After obtaining the CLVs at $t = 0$, those in the not too distant future can be obtained using Eq. (3.27).

CHAPTER 4

Purely STDP-based assembly dynamics: stability, learning, overlaps, drift and aging

Contribution statement

This chapter is based on an early version of the following publication:

[2] **P. Manz** and R.-M. Memmesheimer
Purely STDP-based assembly dynamics: stability, learning, overlaps, drift and aging
PLOS CB **19** (2023) e1011006

In this chapter, we study the stability of long-term memory in the balanced state and how to achieve it in a biologically plausible manner. Memories may be encoded in the brain via strongly interconnected groups of neurons, called assemblies. The concept of Hebbian plasticity suggests that these assemblies are generated through synaptic plasticity, strengthening the recurrent connections within select groups of neurons that receive correlated stimulation. To remain stable in the absence of such stimulation, the assemblies need to be self-reinforcing under the plasticity rule. Previous models of such assembly maintenance require additional mechanisms of fast homeostatic plasticity that induce competition between synaptic strengths [12–14, 123–126]. However, these mechanisms often have biologically implausible timescales [15, 127–130].

Here we provide a model of neuronal assembly generation and maintenance purely based on spike-timing-dependent plasticity (STDP) between excitatory neurons. Our model, as described in Section 4.2 consists of populations of excitatory Hawkes model neurons whose connection strengths undergo spike-timing-dependent plasticity with a symmetric STDP window that crucially has a negative integral. In Section 4.3.1 we demonstrate that our model is capable of showing spontaneous formation of assemblies from random initial connectivity. More specifically, we show that with our model assembly structures form spontaneously from initially random connectivity with characteristic assembly sizes that match our theory. To explain the formation of stable finite size assemblies in the absence of homeostatic plasticity in Section 4.3.2 we analyze the time-averaged weight dynamics of homogeneously connected assemblies of varying sizes. Using previous results for the correlation structure of Hawkes networks, see Section 2.5, we obtain exact expressions for the time-averaged

weight changes, assuming a separation of synaptic and plasticity time scales. These results indicate that assemblies do not grow beyond a certain size because of negative contributions to weight change that increase with larger assembly size: these contributions are caused by the uncorrelated part of each neuron's firing rate, and by cascades of activity within larger assemblies, which spread out the distribution of time lags between spikes. For both reasons, temporally imprecise spike correlations dominate the plasticity in large assemblies and thus set a limit to assembly growth.

We then show via simulations further capabilities of networks with our plasticity rule: In Section 4.3.3 we show that our networks can store new assemblies in weakly connected networks through correlated feedforward input. We, therefore, demonstrate that learning new long-term memory items through experience is possible with our model. We then show in Section 4.3.4 that given appropriate background firing rates, assemblies in our model can establish overlaps, i.e., neurons can stably be part of more than one assembly. We demonstrate the stability of a network in which all neurons are part of two assemblies, a necessary step towards memory networks that store information compactly. In Section 4.3.5, we show that our networks can exhibit representational drift for appropriate learning rates, a result the authors of [13] previously showed with strong synaptic competition. We also develop and demonstrate a novel mechanism of drifting assemblies through changes in background firing rates, which involves insights gained from studying the spontaneous formation of overlaps in Section 4.3.4. Finally, in Section 4.3.6 we study how assemblies are affected by increased network sparsity which we use as a model for aging-related decreases in connectivity. Our results indicate that assemblies grow in the aging brain, in agreement with some experimental findings. I performed the largest parts of the simulations and analytical derivations, created the figures, and wrote large parts of the text for this paper. The article is presented with minor editorial changes and the formatting adapted to match this thesis.

Code availability

The code required to reproduce the results of this chapter is publicly available at PLOS Computational Biology, see [2].

4.1 Introduction

A widely used model of long-term memory posits that items are stored in the brain by strongly interconnected neuronal assemblies [23, 28, 131]. A memory item is represented by a group of neurons that coactivate upon memory recall. The assembly structure allows for associative recall from an incomplete input: due to the strong interconnections, activation of a part of the neurons in an assembly can trigger reactivation of the entire assembly and thus a recall of the full memory. The assemblies may each be created through experience-dependent plasticity or they may already form during development [132]. In the latter case, memories may form by connecting the pre-existing assemblies to appropriate input and output neurons [13]. The theory of the formation and maintenance of neuronal assemblies has been studied in much detail in previous works. The creation of new memories is commonly modeled using Hebbian plasticity [12, 14, 123–126, 133, 134]: if a set of neurons is co-activated, Hebbian plasticity increases the strength of their mutual connections leading to the formation of what is called a Hebbian assembly. However, memory networks in the brain also show ongoing spontaneous and irregular activity. If plasticity still takes place during this activity, it should not interfere with the

existing memory assemblies – otherwise memories would have implausibly short lifespans. Hebbian assemblies can be self-reinforcing under plasticity since their interconnectedness leads to higher correlations in the activities, which in turn leads to potentiation of the intra-assembly weights. Models of assembly maintenance, however, found that fast homeostatic plasticity was needed in addition to Hebbian learning. This introduces competition between synapses and prevents pathological growth of assemblies and exploding activity [12–14, 123–126]. Homeostatic plasticity has been observed in experiments, but it is much slower than Hebbian plasticity and does therefore not suffice to prevent runaway potentiation [15, 127–130] (see, however, [135] for a different view and [134, 136] for a small timescale implementation of homeostasis via inhibitory STDP).

Experiments indicate that distinct memory assemblies have a fraction of shared neurons, i.e. neurons that are part of both assemblies [16–18]; the size of these overlaps appears to correspond to the strength of the associations between the concepts encoded by the assemblies. Previous models of assemblies stabilized by recurrent synaptic plasticity and fast homeostatic normalization usually do not show prominent overlaps [12–14, 123, 124]. An example of a network with weight plasticity, structural plasticity, multiple synapses per connection and short-term depression that can store two strongly overlapping assemblies was given in [137]. We will explore whether our purely STDP-based model can maintain overlaps.

Another topic of interest in the study of memory networks is whether they can generate the representational drift observed in recent experiments [19]. Such drift has recently been modeled by drifting assemblies, which spontaneously exchange neurons with each other, leading to a gradual resorting of the whole network [13]. The network model incorporated fast homeostatic normalization to stabilize the assemblies. We will explore whether also our purely STDP-based model can exhibit drifting assemblies.

Finally, many conditions, such as aging are associated with a decrease in overall connectivity [20, 21]. We will therefore explore how the assemblies in our networks adapt to such a decrease.

The paper is structured as follows: We initially introduce the model of spiking neurons and STDP and describe existing analytical approximations for the time-averaged weight dynamics. As a first result, we show spontaneous assembly formation. We then obtain an analytical approximation of the weight growth in different assembly sizes to obtain an understanding of the numerically observed assembly formation. Next we study whether assemblies can be learned by correlated external input. The subsequent section shows that our networks can stably maintain overlapping assemblies. We then examine whether our model can be set up to exhibit representational drift. Finally, we investigate the dependence of assembly sizes on network sparsity and relate our results to effects of aging on the brain.

4.2 Materials and methods

4.2.1 Poisson neurons

Neural networks in the mammalian brain typically generate irregular and apparently largely stochastic spiking. To guarantee that our model networks generate similar spiking activity, we consider networks of stochastic linear Poisson (or “Hawkes”) model neurons [12, 58, 138, 139]. Such networks are additionally analytically well tractable. We explicitly model excitatory neurons. Since the irregular activity in biological neural networks is likely due to a balanced state, where excitatory and inhibitory

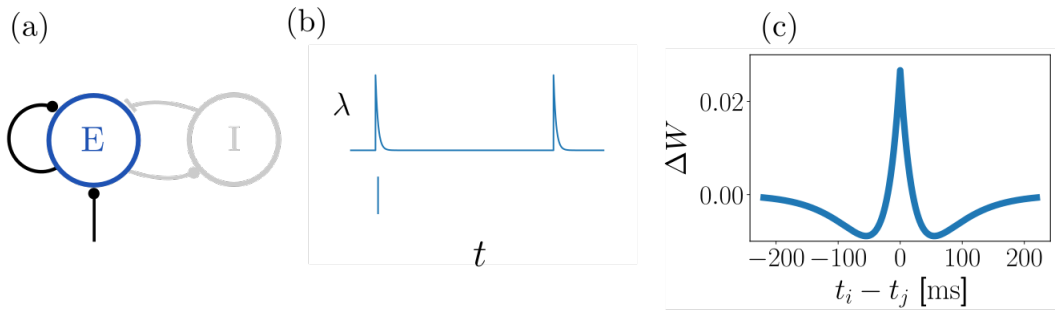

Figure 4.1: Network, neuron and plasticity model. (a): A network consists of recurrently connected excitatory neurons, which generate irregular, stochastic spiking. It may be interpreted as the excitatory population of a neural network where excitatory and inhibitory neurons generate a balanced state [5] of irregular spiking. Thus, an inhibitory neuron population is implicitly contained in our model. (b): Poisson model of stochastic single neuron spiking: each neuron is characterized by its instantaneous rate $\lambda(t)$ (upper subpanel), which depends on incoming spikes and determines the probability of emitting a spike (lower subpanel). (c): Strength of STDP updates as a function of the time difference between pre- and postsynaptic spikes for $\mu = 1$.

inputs to each neuron balance [1, 3–5, 8, 62], our model implicitly incorporates inhibitory neurons, see Fig. 4.1.

The spiking activity of each neuron i is an inhomogeneous Poisson process whose time-dependent instantaneous spike rate (intensity) $\lambda_i(t) = \langle S_i(t) \rangle$ given input spike trains $S_k(t') = \sum_m \delta(t' - t_k^m)$ up to time t and weights W_{ik} is

$$\lambda_i(t) = \lambda_i^0 + \sum_k W_{ik} \sum_m a(t - t_k^m). \tag{4.1}$$

The angular brackets here denote trial-averaging with fixed input spike trains and weights to neuron i. We use an exponentially decaying synaptic kernel

$$a(t) = \frac{\Theta(t)}{\tau_s} \mathrm{e}^{-\frac{t}{\tau_s}}, \tag{4.2}$$

where $\Theta(t)$ is the Heaviside distribution, and a constant external drive λ^0.

It is now useful to introduce quantities that are trial-averaged over the entire spiking network dynamics. The trial-averaged instantaneous rate (intensity) of neuron i is $r_i(t) = \langle \lambda_i(t) \rangle$, where angular brackets now denote the trial-averaging over the network dynamics. For static W with spectral radius less than 1, the trial-averaged rate dynamics relax to a fixed point where the vector $r(t)$ is constant,

$$r = (\mathbb{1} - W)^{-1} \lambda^0, \tag{4.3}$$

see [12, 53, 54, 138]. In addition, one can analytically compute the corresponding stationary cross-correlation functions $C_{ij}(\tau) = \langle S_i(t + \tau) S_j(t) \rangle$ of pairs of neurons i, j at arbitrary t; in the frequency domain they read

$$\tilde{C}(\omega) = 2\pi\delta(\omega) r r^T + \left(\mathbb{1} - \tilde{a}(\omega) W\right)^{-1} D \left(\mathbb{1} - \tilde{a}(-\omega) W^T\right)^{-1}, \tag{4.4}$$

where we used matrix notation and $D_{ij} = \delta_{ij} r_i$ [12, 53, 138]. The Fourier transform of a function $g(t)$ is defined as $\tilde{g}(\omega) = \int_{-\infty}^{\infty} \mathrm{d}t \mathrm{e}^{-i\omega t} g(t)$. If the scale of the STDP updates is sufficiently small, one can assume that the weight dynamics are quasistationary with respect to neuronal dynamics. Then Eqs. 4.3,4.4 still approximately hold true despite W changing over time due to plasticity.

4.2.2 Spike-timing-dependent plasticity

We consider networks of spiking neurons with pair-based STDP, i.e. the change of synaptic strength depends on the time lags of pairs of pre- and postsynaptic spikes. The characteristics of the plasticity function crucially determine how synaptic strengths evolve in a network. Networks with symmetric plasticity functions can establish a structure of neuronal assemblies whereas plasticity functions with a large antisymmetric part tend to establish feedforward chains of connectivity; see, however, [134, 136] for networks with asymmetric STDP maintaining assemblies and [124] for a triplet STDP rule that forms assemblies despite an asymmetric two-spike interaction part. In the present work, we consider symmetric plasticity functions because of their simplicity and analytical tractability, building on previous theoretical work that has employed them [12, 13]. Symmetric STDP has been found in the CA3 region of the rodent hippocampus [36], i.e. in a region that is assumed to serve as an associative memory network and to store assemblies [28]. Recently, near-symmetric STDP has also been observed in the primary motor cortex of primates [140].

The change induced by a spike pair with time lag t is given by the STDP function F,

$$F(t) = \mu \left(A_\mathrm{p} \mathrm{e}^{-\frac{|t|}{\tau_p}} + A_\mathrm{d} \mathrm{e}^{-\frac{|t|}{\tau_d}} \right), \tag{4.5}$$

where the scaling factor μ is the learning rate, $A_\mathrm{p} > 0$, $A_\mathrm{d} < 0$, $\left|A_\mathrm{p}\right| > \left|A_\mathrm{d}\right|$ and $\tau_\mathrm{p} < \tau_\mathrm{d}$, see Fig. 4.1c. For the analytical treatment of our plasticity rule we set $\mu = 1$. In all our simulations the parameters are chosen such that the integral $\int_{-\infty}^{\infty} \mathrm{d}t\, F(t)$ over F is negative.

At each spike time, plasticity acts additively on the pre- and postsynaptic weights of the spiking neuron, with amplitudes given by F. Specifically, at a spike time t_{k_i} of the postsynaptic neuron there is a jump-like change in W_{ij} of $\sum_{t_{k_j} \leq t_{k_i}} F(t_{k_i} - t_{k_j})$. At a spike time t_{j_k} of the presynaptic neuron W_{ij} jumps by $\sum_{t_{k_i} \leq t_{k_j}} F(t_{k_i} - t_{k_j})$. This can be compactly written as

$$\frac{\mathrm{d}}{\mathrm{d}t} W_{ij}(t) = \sum_{t_{k_i}, t_{k_j} \leq t} \left(F(t_{k_i} - t_{k_j}) \delta(t - t_{k_i}) + F(t_{k_i} - t_{k_j}) \delta(t - t_{k_j}) \right). \tag{4.6}$$

Here and henceforth we assume $i \neq j$; there is no self-interaction in our networks, $W_{ii} = 0$. We further stipulate that no weight can become negative or exceed a maximum value $\hat{w}$ due to STDP: if a synapse would be depressed below 0 it will be set to 0 instead and if a synapse would be potentiated to a value beyond $\hat{w}$ it will be set to $\hat{w}$.

In the regime of quasistationary weight dynamics the time-averaged drift in synaptic efficacy, $\overline{\Delta W_{ij}}$, can be approximated by

$$\overline{\Delta W_{ij}} := \frac{1}{T} \int_t^{t+T} \mathrm{d}t' \, \frac{\mathrm{d}}{\mathrm{d}t'} W_{ij}(t') = \int_{-\infty}^{\infty} \mathrm{d}\tau \, C_{ij}(\tau) F(\tau), \tag{4.7}$$

where $C_{ij}(\tau)$ are again the correlations of the spike trains of neurons i and j [58]. Plancherel's theorem then yields

$$\overline{\Delta W_{ij}} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\ \tilde{C}_{ij}(\omega)\tilde{F}(-\omega), \tag{4.8}$$

with the Fourier transforms of the correlation function and of the STDP window. Inserting the correlation function $\tilde{C}_{ij}(\omega)$ of Eq. (4.4) gives

$$\begin{aligned}\overline{\Delta W_{ij}} = {} & f_0 r_i r_j \\ & + \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\ \sum_k \left[(\mathbb{1} - \tilde{a}(\omega)W)^{-1}\right]_{ik} r_k \left[(\mathbb{1} - \tilde{a}(-\omega)W^T)^{-1}\right]_{kj} \\ & \times \tilde{F}(-\omega),\end{aligned} \tag{4.9}$$

see [12], where

$$f_0 = \tilde{F}(0) = \int_{-\infty}^{\infty} \mathrm{d}t\ F(t) = 2\left(A_\mathrm{p}\tau_\mathrm{p} + A_\mathrm{d}\tau_\mathrm{d}\right). \tag{4.10}$$

It is often useful to expand Eq. (4.4) into a power series with respect to W [138]. Inserting the series into the right hand side of Eq. (4.8) (or directly expanding Eq. (4.9)) results in a series expansion for $\overline{\Delta W_{ij}}$, [12]:

$$\overline{\Delta W_{ij}} = f_0 r_i r_j + \sum_{\alpha,\beta} f_{\alpha\beta} \sum_m r_m (W^\alpha)_{im} (W^\beta)_{jm}. \tag{4.11}$$

The terms of the sum encode contributions from motifs in which a source neuron affects post- and presynaptic neurons via a chain of α and β connections, respectively (the same connection may be counted more than once). If $\alpha = 0$ ($\beta = 0$) the post(pre)synaptic neuron itself is the source. The coefficients $f_{\alpha\beta}$ contain integrals over the Fourier transform of the STDP window and powers of the Fourier transform of the synaptic kernel function,

$$f_{\alpha\beta} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\ \tilde{F}(-\omega)\tilde{a}(\omega)^\alpha \tilde{a}(-\omega)^\beta. \tag{4.12}$$

4.3 Results

4.3.1 Spontaneous assembly formation

We first simulate the model described above with initially unstructured weight matrix and without structured external stimulation, in order to test its capability of spontaneous assembly formation and subsequent maintenance. We find that for appropriately chosen parameters in the plasticity function the network weights indeed converge towards a structure with segregated assemblies of a characteristic size.

The mechanism underlying the increase of weights between assembly neurons is well known: Initially basically randomly coincident spiking leads to strengthening of some weights. These weights induce more near-simultaneous spiking, which further strengthens them. The positive feedback loop leads to weights that increase until they reach $\hat{w}$ [22]. It has been shown that if the summed

weight strength to and from a neuron are additionally normalized by fast homeostasis, there can be spontaneous emergence of assemblies [12, 13, 124, 125]. The reason is that the homeostatic normalization lets synapses compete, such that more slowly growing ones are suppressed. Once an imbalance of connectivity and thus first assembly-like structures occur, the weight increase of intra-assembly synapses prior to normalization is stronger, due to the stronger co-spiking of the connected neurons. The normalization then suppresses the inter-assembly connections and the assembly structure consolidates.

In our networks, there is no fast homeostatic normalization that could introduce competition between synapses. Then, why do not all weights tend to $\hat{w}$? In other words: why does the network not turn into one big assembly?

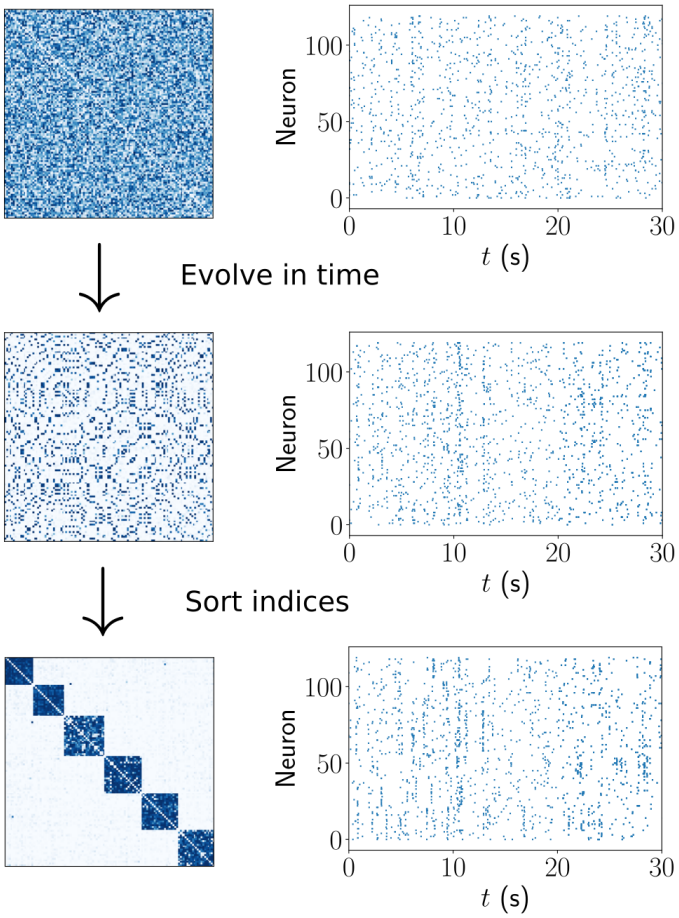

Figure 4.2: Spontaneous assembly formation. Several assemblies (strongly interconnected ensembles of neurons) emerge in a network with initially random connectivity, due to spontaneous activity.

4.3.2 Plasticity in homogeneously connected assemblies

In this section, we will argue that the assembly growth is in our model limited because of the depression dominance of the learning rule. For this we consider the special case of an isolated assembly of N

neurons that is homogeneously connected with synaptic strengths $\hat{w}$. We compute $\overline{\Delta W_{ij}}$ for weights within this type of assembly, disregarding the clipping at $\hat{w}$. This indicates whether and how vigorously weights that fall below $\hat{w}$ are restored towards it. Further, it indicates how successfully further neurons are recruited, as recruitment relies on increasing the weights to new neurons once they are randomly increased. We expect that the assembly size is limited by values of N above which $\overline{\Delta W_{ij}}(N)$ is negative: Larger assemblies will lose neurons due to the depression of weights, to other, smaller assemblies that tend to potentiate their connections with these neurons. This yields as an estimate for the size N of assemblies the solution of $\overline{\Delta W_{ij}}(N) = 0$. The assumption that an assembly loses neurons already if another assembly can potentiate its connections to them more yields as preferred assembly size the N that maximizes $\overline{\Delta W_{ij}}(N)$.

To analytically obtain $\overline{\Delta W_{ij}}$ as a function of the assembly size, we first compute from Eq. (4.3) the average rate of an assembly neuron

$$r = \lambda_0 \sum_{m=0}^{\infty} ((N-1)\hat{w})^m = \frac{\lambda_0}{1-(N-1)\hat{w}} = \frac{\lambda_0}{1-\rho}, \tag{4.13}$$

where $\rho = (N-1)\hat{w}$ is the branching parameter, which gives the average number of spikes that are directly induced by a single spike in the assembly. A stable network requires $\rho < 1$ [139, 141, 142]. Eqs. 4.4, 4.13 and 4.8 yield

$$\begin{aligned}\overline{\Delta W_{ij}}(N) = &\frac{2\lambda_0^2(A_\mathrm{p}\tau_\mathrm{p} + A_\mathrm{d}\tau_\mathrm{d})}{(1-(N-1)\hat{w})^2} + \\ &+ \frac{\lambda_0 A_\mathrm{p}\tau_\mathrm{p}\hat{w}\left[(2-(N-2)\hat{w})\tau_\mathrm{p} + (2-(N-2)\hat{w}-(N-1)\hat{w}^2)\tau_\mathrm{s}\right]}{(1+\hat{w})(1-(N-1)\hat{w})^2(\tau_\mathrm{s}+(1+\hat{w})\tau_\mathrm{p})(\tau_\mathrm{s}+(1-(N-1)\hat{w})\tau_\mathrm{p})} + \\ &+ \frac{\lambda_0 A_\mathrm{d}\tau_\mathrm{d}\hat{w}\left[(2-(N-2)\hat{w})\tau_\mathrm{d} + (2-(N-2)\hat{w}-(N-1)\hat{w}^2)\tau_\mathrm{s}\right]}{(1+\hat{w})(1-(N-1)\hat{w})^2(\tau_\mathrm{s}+(1+\hat{w})\tau_\mathrm{d})(\tau_\mathrm{s}+(1-(N-1)\hat{w})\tau_\mathrm{d})},\end{aligned} \tag{4.14}$$

see SI S.1 for details. Assuming that $N \gg 1$, we can neglect N's discrete nature and consider the limit where it approaches $1 + \frac{1}{\hat{w}}$. In this limit, the firing rates diverge and Eq. (4.14) gives

$$\begin{aligned}&\lim_{N\to 1+\frac{1}{\hat{w}}} \overline{\Delta W_{ij}}(N) \\ &= \lim_{N\to 1+\frac{1}{\hat{w}}} \left(\frac{2\lambda_0^2(A_\mathrm{p}\tau_\mathrm{p} + A_\mathrm{d}\tau_\mathrm{d})}{(1-(N-1)\hat{w})^2} + \frac{\lambda_0\hat{w}\left(A_\mathrm{p}\tau_\mathrm{p} + A_\mathrm{d}\tau_\mathrm{d}\right)}{\tau_\mathrm{s}(1+\hat{w})(1-(N-1)\hat{w})^2} \right) \qquad (4.15) \\ &= \lim_{N\to 1+\frac{1}{\hat{w}}} f_0 \left(\frac{\lambda_0^2}{(1-(N-1)\hat{w})^2} + \frac{\lambda_0\hat{w}}{2\tau_\mathrm{s}(1+\hat{w})(1-(N-1)\hat{w})^2} \right), \qquad (4.16)\end{aligned}$$

where we used Eq. (4.10) to obtain the last line. It implies

$$\lim_{N\to 1+\frac{1}{\hat{w}}} \overline{\Delta W_{ij}}(N) = -\infty \text{ if } f_0 < 0, \tag{4.17}$$

since the summands in the large bracket are positive. Thus, if $f_0 < 0$ and there is an N for which $\overline{\Delta W_{ij}}(N)$ is positive, $\overline{\Delta W_{ij}}(N)$ will have a maximum; we therefore expect limited assembly growth and assume $f_0 < 0$ throughout the article. Indeed, if $f_0 > 0$, $\overline{\Delta W_{ij}}(N)$ diverges to positive values for N where also the firing rate diverges, indicating that sufficiently large assemblies continue to grow until the network generates pathological activity.

The first term in the bracket of Eq. (4.16) covers the impact of uncorrelated pre- and postsynaptic spiking with rates r_j and r_i on the synaptic strengths ($C_{ij}(\tau) = r_i r_j$ for uncorrelated spiking). As the firing rates increase with N, the contribution of the STDP rule due to uncorrelated spiking tends to increasingly negative values, since $f_0 < 0$. The fact that the second term, which encodes the effects of connectivity motifs on STDP, see Eq. (4.11), also becomes negative for sufficiently large N is due to contributions from higher order motifs. We can see this by reconsidering Eq. (4.11): For homogeneously connected assemblies, it simplifies to

$$\overline{\Delta W_{ij}} = f_0 r^2 + \frac{r}{N} \sum_{\alpha+\beta>0} f_{\alpha\beta} \left((N-1)^{\alpha+\beta} - (-1)^{\alpha+\beta} \right) \hat{w}^{\alpha+\beta} \tag{4.18}$$

$$= f_0 r^2 + \frac{r}{N} \sum_{k=1}^{\infty} f_k \left((N-1)^k - (-1)^k \right) \hat{w}^k, \tag{4.19}$$

where

$$f_k := \sum_{\alpha+\beta=k} f_{\alpha\beta}, \tag{4.20}$$

see SI S.2. Since $f_{\alpha\beta}$ (cf. Eq. (4.12)) and thus f_k are independent of N, higher order terms in k grow in Eq. (4.19) faster with N than lower order ones. Therefore higher order motifs become more relevant for the weight change the larger N is. We find for high orders that

$$\lim_{k\to\infty} f_k = \frac{f_0}{2\tau_s}, \tag{4.21}$$

see SI S.2; therefore high order motifs induce synaptic depression.

Importantly, the relation Eq. (4.21) holds generally, for any plasticity window F, not only for the considered symmetric one. We explain it as follows: A term with specific exponents α and β in Eqs. 4.11, 4.18 covers the contribution of a specific connectivity motif to the spike correlation. This motif consists of a common presynaptic neuron that is separated from neurons j and i by a chain of β and α connections, respectively [12], see Fig. 4.3b for an illustration. Higher orders of $k = \alpha + \beta$ thus encode the effects of long cascades of spiking activity in an assembly. The longer these cascades get, the more spread-out and Gaussian the temporal distribution of their impacts becomes, due to the summation of inter-spike-intervals and their jitters. (We note that since the considered model is linear, one can indeed attribute the generation of each spike to a precursor spike or to the external drive.) Eq. (4.20) homogeneously sums over motifs with different connection chain lengths to the pre- and the postsynaptic neuron. For sufficiently large k the evoked spike time differences are thus equidistant, broad, overlapping Gaussians over the STDP window, which leads to the dependence on the integral of the STDP function f_0, as for uncorrelated pre- and postsynaptic spiking.

Fig. 4.3a shows $\overline{\Delta W_{ij}}(N)$ as in Eq. (4.14) for different values of $\hat{w}$: $\overline{\Delta W_{ij}}(N)$ is maximal for a particular assembly size and becomes negative for large assembly sizes. We confirmed these results using simulations of homogeneous assemblies, in which the weight updates that would occur due to

STDP without clipping are tracked but not applied and then averaged over time. We observe that the typical size N of spontaneously formed assemblies corresponds to the zero crossing $\overline{\Delta W_{ij}}(N) = 0$ if corrected for gaps in the assemblies.

We note that in our simulations we also find sparser assemblies where some internal connections are weak or basically missing. For these we expect the maximum of $\overline{\Delta W_{ij}}$ to be at larger sizes compared to our estimates with fully connected ones. Consistent with this, we observe that sparser assemblies in our simulations tend to be larger.

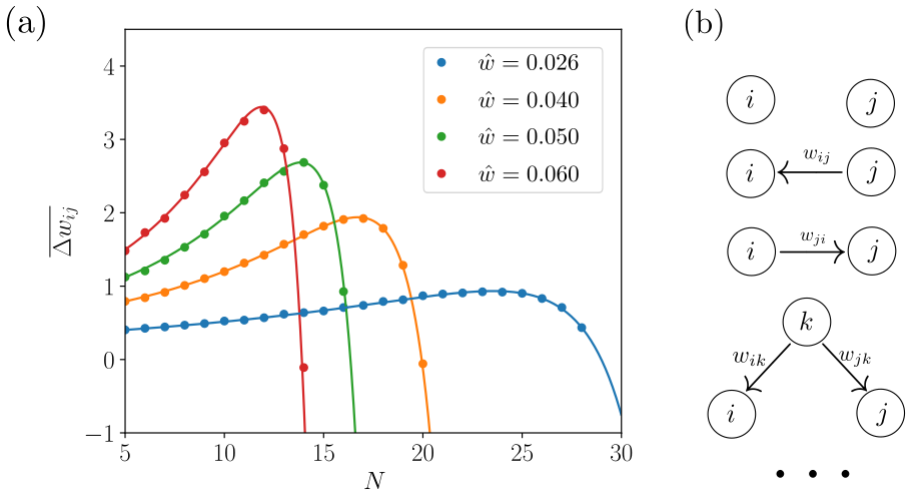

Figure 4.3: Explanation of assembly formation. (a): Time-averaged weight change (tracked, not applied; arbitrary units) for a synapse in an assembly (fully connected, no other connections, all weights fixed at $\hat{w}$) as a function of its size N, for different values of $\hat{w}$. Solid lines: theoretical prediction. (b): Illustration of zeroth, first and second order contributions to the time-averaged weight change for a given rate and weight configuration.

4.3.3 Storing new assemblies

Next we show how assemblies can be learned via correlated feedforward input. Neurons recruited for a new assembly may previously have only weak connections, i.e. they may belong to a background of neurons in front of which assemblies exist. Alternatively recruited neurons may already be part of other assemblies. In our simulations, during the learning phase, each neuron that is to be recruited receives Poisson spike input from the same source. This stimulates the neurons to spike near-simultaneously, such that the weights between them grow. Fig. 4.4 displays the resulting weight and spiking dynamics in a network which prior to stimulation hosts one assembly in front of a background of weakly coupled neurons. Before time $t = 0$s, the single assembly is stably stored in the network, despite the ongoing plasticity. At $t = 0$ s a group of background neurons receives correlated stimulation for $T_{\text{stim}} = 180$ s. This leads to the formation of a rudimentary assembly. After the stimulation has ended, its synapses grow over a longer period of time until a fully connected assembly is reached, where all interconnections have synaptic strength close to $\hat{w}$. The remaining background neurons, which do not receive correlated stimulation in the beginning, do not form assemblies. The stimulation furthermore does not affect the already existing assembly and neither does this assembly interfere with

the formation of the new assembly. Fig. 4.11 shows another example of assembly learning; here again mostly background neurons are recruited, but also one neuron that is already part of two pre-existing assemblies.

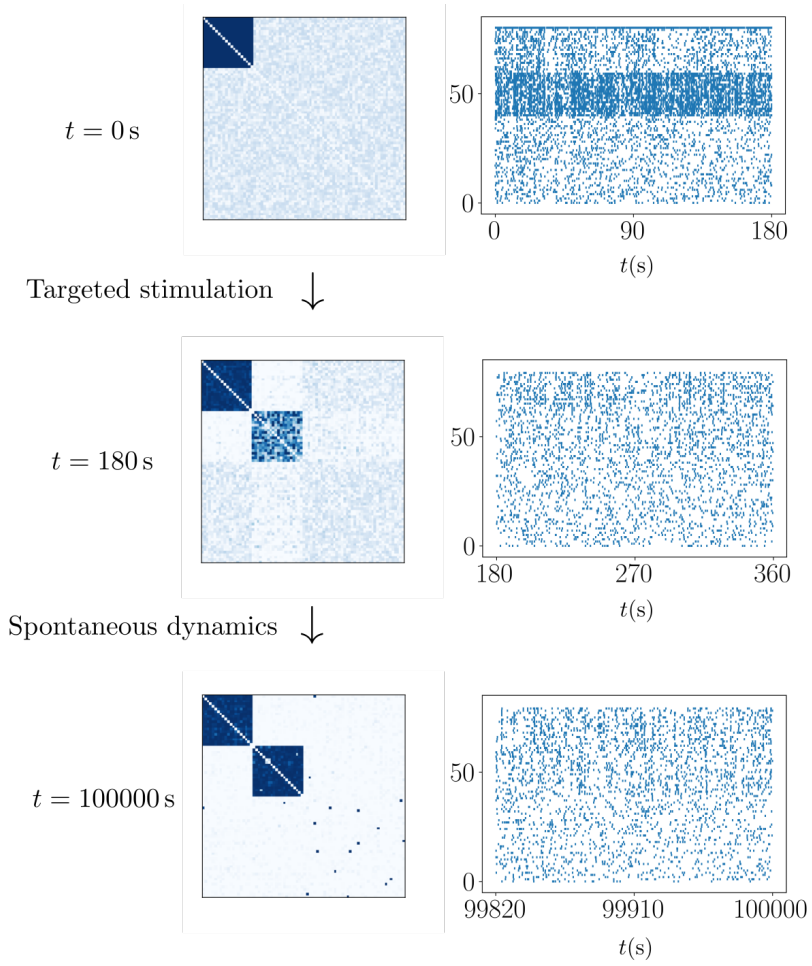

Figure 4.4: Learning a new assembly in a network starting with a single assembly and otherwise weak random connectivity. Initially, correlated external input stimulates the neurons of the new assembly, which leads to an increase of their interconnecting weights and thereby to the formation of a rudimentary assembly. The stimulation is then turned off and the network evolves on its own; the new assembly becomes fully connected.

4.3.4 Overlapping Assemblies

Experiments indicate that neurons can code for more than one memory item [16–18]. For assembly models of memory this implies that neurons can be part of more than one assembly, see Fig. 4.5a, i.e. assemblies have overlaps [143]. Such overlaps may encode associations between memories.

We find that for appropriate parameters, our networks can stably store assemblies with some overlaps, see Fig. 4.5b for an example weight matrix. However, we also observe that overlapping assemblies present a challenge for our models: overlapping assemblies either tend to merge or overlapping neurons

tend to disconnect from one of the assemblies they are a part of, dissolving the overlap. The first case occurred especially when the sizes of the overlapping assemblies were significantly smaller than the sizes that would maximize $\overline{\Delta W_{ij}}$. We hypothesize that in this case a feedback loop emerges: overlaps induce correlations between the assemblies, which facilitates the formation of additional overlaps until the assemblies have completely fused. The other case occurred when the sizes of the two assemblies were close to maximizing $\overline{\Delta W_{ij}}$. We hypothesize that here the higher firing rate of an overlap neuron and the resulting increased negative contribution (from, for example, the first term in Eq. (4.14)) imply that overlaps become more likely to disappear due to fluctuations in the weight dynamics.

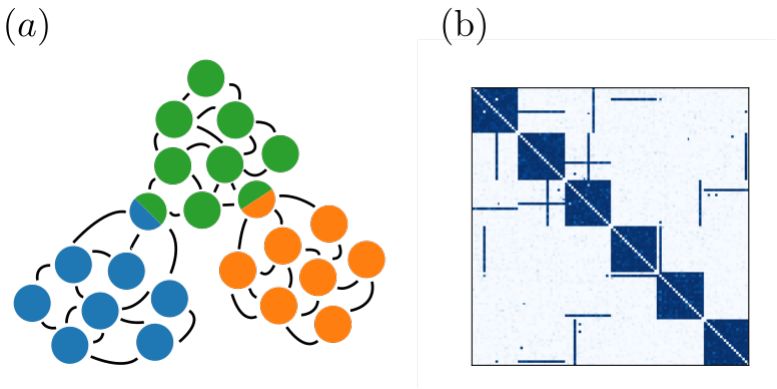

Figure 4.5: Overlapping assemblies. (a): Illustration of overlaps. (b): Weight matrix of a network with overlapping assemblies. The base assemblies (diagonal blocks) are each 20 neurons in size.

The hypotheses about the mechanism underlying the problems with overlap suggest two solutions how networks may solve them: First, if the synaptic long-term depression that occurs at high rates limits the connectivity of neurons, a neuron with a significantly lower spontaneous rate may be able to connect to multiple assemblies at the same time. In Fig. 4.6 we show how neurons with lower spontaneous firing rate join a second assembly. In Fig. 4.6a a neuron with lowered spontaneous rate is already partially connected to another assembly. This partial connection then causes enough correlation with the second assembly to induce the completion of the connections. In this case both the initial partial connection and the lower λ_0 were necessary to create this overlap. For an even lower λ_0, an overlap emerged spontaneously, with a randomly selected assembly, through the inherent stochasticity of the dynamics, as shown in Fig. 4.6b.

A second way of sustaining overlaps, is by having most, or all neurons be part of more than one assembly. In this case fusion of two assemblies appears less likely, because all assemblies and neurons are similarly saturated. Indeed, in the brain, we expect most or all neurons to be part of more than one assembly [16, 143]. To test whether our networks are capable of sustaining similarly prominent overlap structures, we consider a network of intertwined assemblies in which each neuron is part of exactly two assemblies. We choose equally sized assemblies of size n_A and the overlap structure such that for any given assembly each of its neurons is shared with a different assembly. In other words, the overlap between any two assemblies consists of one neuron or a fraction of $1/n_\mathrm{A}$ of the population. This implies that there are in total $n_\mathrm{A}+1$ assemblies and that the setup requires a network with $N = n_\mathrm{A}(n_\mathrm{A}+1)/2$ neurons. Notably, a network with the same number and size of assemblies, where each neuron is only part of one assembly (no overlap), would require twice as many, $(n_\mathrm{A}+1)n_\mathrm{A}$, neurons. Fig. 4.7 shows that this structure is stable under the STDP rule over long timescales. The

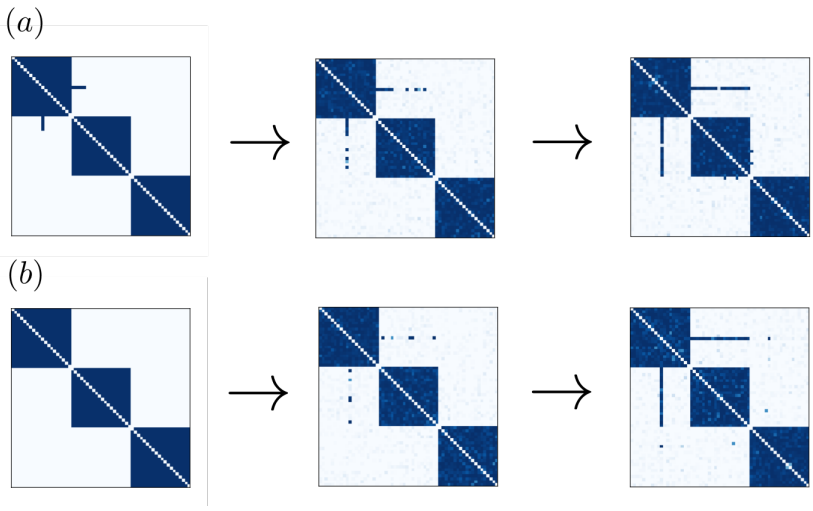

Figure 4.6: Creation of overlaps. (a): A partial overlap self-completes over time due to plasticity. (b): An overlap emerges for a neuron with lower spontaneous firing rate.

assemblies have size $n_A = 19$; the typical overlap between two assemblies is thus $1/n_A = 1/19 \approx 5.3\%$. Random correlations cause additional, extra-assembly connections to appear (more precisely: to strengthen) and intra-assembly connections to disappear (to weaken). To demonstrate that these deviations of individual weights are transient, we have tracked in Fig. 4.7b and c the strengthening of extra-assembly weights and the weakening of intra-assembly weights, respectively. The upper panel of Fig. 4.7b displays the sum of the extra-assembly weights over time, indicating that there is no positive drift, i.e. no overall increase of weights that do not belong to the stored assembly structure over time. The lower panel displays the Pearson correlations of the extra-assembly weights over time with the extra-assembly weights observed at five reference time points. With increasing distance to the reference time point, the correlations decay to chance level showing that no persistent pattern of strengthened extra-assembly weights emerges. Fig. 4.7c analogously displays the differences between the maximal (optimal) and the actual intra-assembly weights. The upper panel shows that there is no overall increase of this difference, i.e. no overall decay of the assembly structure. The lower panel shows that the patterns of weakened intra-assembly weights are transient.

4.3.5 Drifting assemblies

Experiments have shown that memory representations need not consist of the same neurons over time but can in fact exchange neurons without affecting behavior [19], a phenomenon called representational drift. It may occur because memory assemblies drift, by gradually exchanging neurons between each other [13]. The gradual exchange implies that at each point in time, each assembly is present and unambiguously identifiable by following the course of its evolution from the beginning. In the following, we show that our model networks can give rise to drifting assemblies. The drift happens due to two alternative mechanisms: (i) Neuron exchange between assemblies due to high weight plasticity noise, as in [13] and (ii) formation of temporal overlaps due to modulations in the spontaneous spike rate. Whether drift occurs due to mechanism (i) is chiefly determined by the learning rate μ: Fig. 4.8a displays the assembly dynamics in two networks with different values of μ while all other parameters

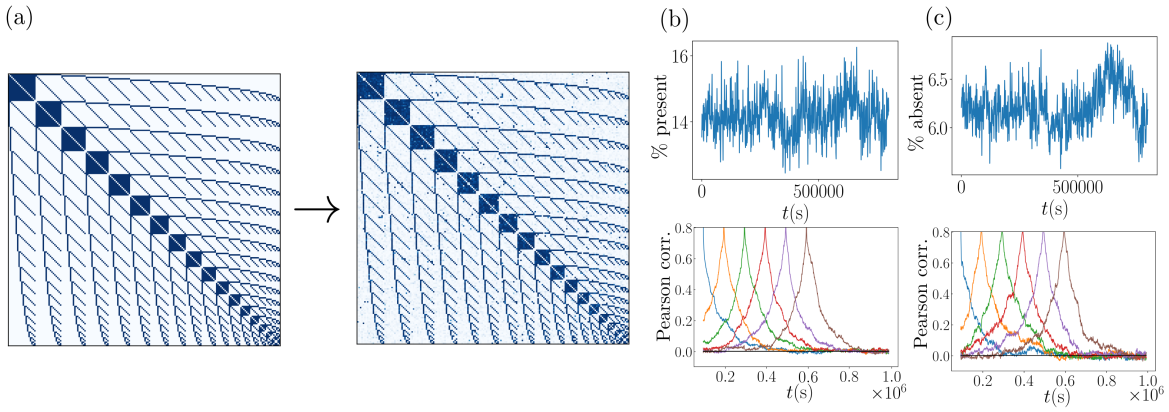

Figure 4.7: Intertwined assembly network. (a): Initial weight matrix (left hand side) and weight matrix after simulation over 10^6 s (right hand side). (b, upper): Sum of the extra-assembly weights, that is of those weights that are zero initially and when the assembly structure is optimally realized. The sum is given as a fraction of the total initial weights. (b, lower): Correlations of the extra-assembly weights with those at six different reference points in time. (c,upper): Sum of the absent intra-assembly weights, that is, the sum of distances of intra-assembly weights from their initial and optimal value $\hat{w}$. The sum is given as a fraction of the total initial weights. (c,lower): As in (b,lower) for absent intra-assembly weights.

are kept the same. The network with smaller μ has stationary assemblies. In contrast, the network with larger μ exhibits drifting assemblies. Specifically, in this network the ensemble of neurons forming an assembly completely changes over time (Fig. 4.12a). Simultaneously, one can track the identity of an assembly by comparing its constituent neurons over short time intervals; the neurons forming it at one time can be unambiguously matched to the neurons forming it shortly thereafter, since the difference between those two ensembles is still small (Fig. 4.12b). We can explain the occurrence of this type of drift as an effect of fluctuations in the weight dynamics. The time-averaged weight dynamics as described by $\overline{\Delta W_{ij}}$ reinforce connections between neurons in the same assembly while suppressing connections between neurons of different assemblies. However, large enough fluctuations in the weight dynamics may nevertheless cause neurons to lose connections within their assembly and form connections to other assemblies. The size of these fluctuations is governed by μ, and if they are sufficiently large there is a finite probability that a neuron switches assemblies, see [13] for a detailed discussion. If μ is too large, the strong fluctuations prevent assemblies from forming at all.

In the second mechanism switching of assemblies by neurons is a two-step process. In the first step a neuron with a sufficiently low λ_0 spontaneously forms a connection to an additional assembly, as in Fig. 4.6b. Then, when λ_0 increases again this neuron loses its connections to one of the assemblies. The neuron can thereby leave either of the assemblies it is connected to – if it loses its connections to the assembly that it was originally connected to, it has switched assemblies. If this happens sufficiently often for many neurons in the network, this also causes overall drift on a slow timescale, as shown in Fig. 4.6b.

4.3.6 Aging

We finally apply our assembly model to networks that undergo changes in their synaptic connectivity due to processes related to aging. Anatomical studies have shown that the aging cortex is characterized

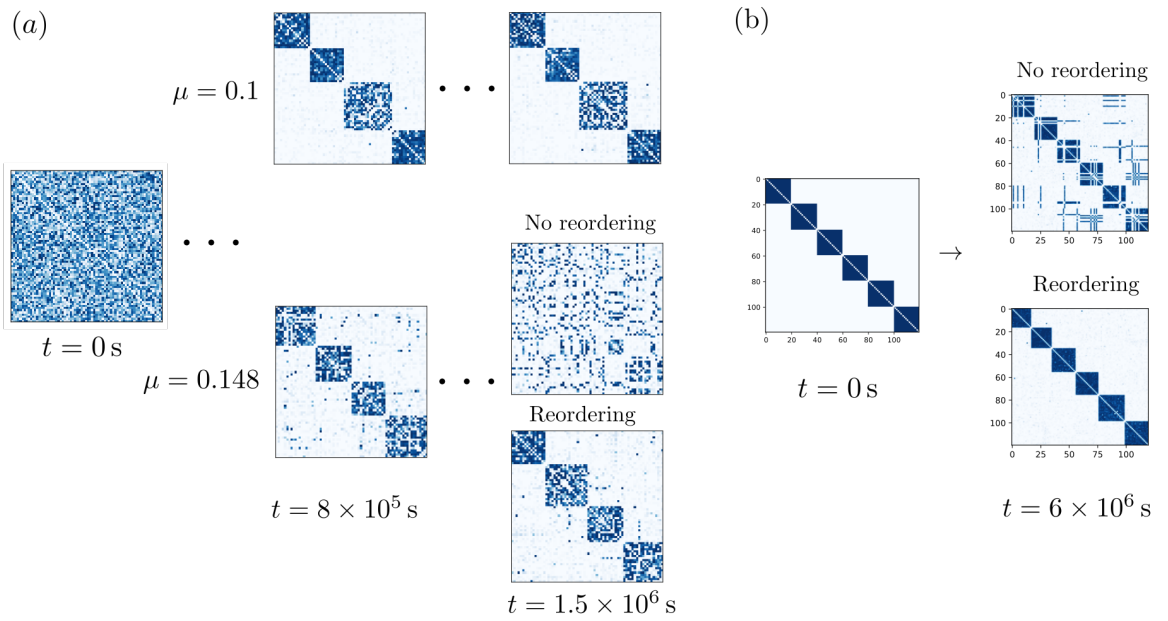

Figure 4.8: Drifting assemblies. (a): Drift through noisy spiking activity and resulting weight fluctuations. Assembly dynamics for lower ($\mu = 0.1$, upper part) and higher ($\mu = 0.145$, lower part) learning rate. The panel shows connection matrices W at different point in time. Initially, the weight matrices are random (left). After $t = 8 \times 10^5$ s assemblies have spontaneously emerged (middle). The neuron indices are sorted to reveal them (cf. Fig. 4.2). The simulation is thereafter continued, which shows that for lower learning rate the assemblies are static. In contrast, for higher learning rate the assemblies drift. Since the assemblies exchange neurons, the coupling matrix appears increasingly unstructured. Reordering of the indices, however, reveals that the assemblies are maintained at each point in time. (b): Drift through transient changes in intrinsic neuron properties and resulting transient overlaps. The spontaneous firing rates of neurons change on a slow timescale. This leads to the transient appearance of overlaps, cf. Fig. 4.6b. When an overlap vanishes, the neuron randomly decides for one of the assemblies, which leads to drifting assemblies.

by a decrease in the number of synaptic spines [20] and presynaptic terminals [21]. These changes may be interpreted as an increase in sparsity of connections between neurons and/or as an overall weakening of synaptic strengths (since connections between neurons often consist of multiple synapses). We can model the former by permanently setting a fraction of entries of the weight matrix, chosen at random, to zero. We can model the latter by decreasing $\hat{w}$. Fig. 4.3a shows the effect of decreasing $\hat{w}$: it shifts the potentiation maximum and thus the typical assembly size towards larger N. Increased sparsity similarly lowers the branching parameter for an assembly of a given size and thus also shifts the potentiation maximum and the characteristic assembly size towards larger N, see Fig. 4.9a. As a result, spontaneously forming assemblies are larger in networks with higher sparsity, see Fig. 4.9c. In addition Fig. 4.9c suggests that smaller learning rates further increase the tendency to form larger assemblies. We observe that if the assembly sizes in a network are significantly smaller than the characteristic size predicted by Figs. 4.3a and 4.9a, assemblies will merge to form larger ones, see Fig. 4.9b. This represents a loss in memory capacity. Assuming that in the brain assemblies representing closely related memories merge (due to existing overlaps), during this process the overall memory content becomes less detailed and differentiated. At the same time the neuronal activity during recall increases

due to larger assembly sizes indicating a less efficient use of neural resources.

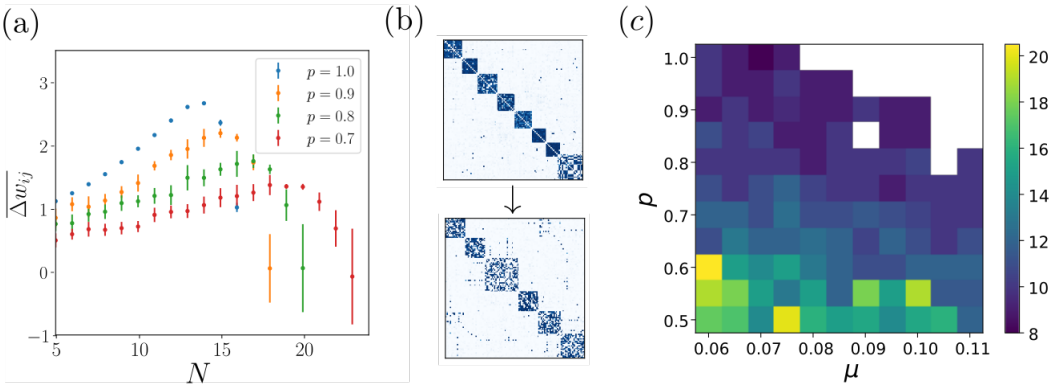

Figure 4.9: Effects of increasing network sparsity. (a): Time-averaged weight change like in Fig. 4.3a, of an assembly with intra-assembly connection probability p, as a function of its size for different values of p. Lower values of p lead to a shift of the optimal assembly size to larger N. Simulations with low $\overline{\Delta W_{ij}}$ converge slowly, which is reflected by large error bars. (b): An assembly structure becomes coarser (fewer, larger assemblies) as the connectivity sparsity increases. The panel shows the weight matrices of the same network when the probability p that a connection between neurons exists is decreased from $p = 1.0$ (upper subpanel, after spontaneous assembly emergence) to $p = 0.7$ (lower subpanel, after re-equilibration). (c): Dependence of the size of a typical assembly on the network connection probability p and the learning rate μ. The panel shows the median size of spontaneously forming assemblies in networks with random initial connectivity. White areas denote parameter regions for which activity becomes pathological, that is, firing rates diverge.

4.4 Discussion

We have studied assemblies in networks with plasticity that is purely spike-timing-dependent. We find that the networks stably store assemblies, which may spontaneously emerge or be learned. Further, the assemblies can have overlaps, spontaneously drift and adapt to changing network properties.

In biological spiking neural networks the change of synaptic efficacies depends to a large extent on STDP, where depression and potentiation are a function of the time lags between pre- and postsynaptic spikes [144–146]. Assemblies may be kept up by the co-spiking of their member neurons, which strengthens their interconnections. If the total input to a neuron is not constrained, large assemblies generate large inputs to each member neuron and thereby highly reliably co-activate them. Furthermore, larger assemblies generate stronger input to other neurons and thereby tend to recruit them, such that the assembly grows. This generates a positive feedback loop, which can without restraining mechanisms lead to excessive assembly growth. Previous work on assembly networks therefore usually added additional homeostatic plasticity that limits the total input strength to a neuron by fast normalization [12–14, 123–126]. This curtails the resulting excessive assembly growth. The homeostatic plasticity observed in neurobiological experiments is, however, much slower than that required to prevent runaway assembly growth [15, 127–130].

In the present work, we therefore studied assemblies in networks with purely STDP-based plasticity. We find that our depression-dominated STDP-rule (the integral of the STDP window is negative) restricts the growth of assemblies by two mechanisms: On the one hand, the spike rate of the neurons

in an assembly grows with the assembly size, which increases the depressing effect of the rate-based term of the time-averaged weight change. On the other hand, the contributions of the higher-order connectivity motifs to the time-averaged weight change become more significant the larger the assemblies are, because longer cascades of spikes become more likely. Further for sufficiently high order the appropriately summed contributions of the motifs are approximately proportional to the integral of the STDP window and thus negative. The contributions of the purely firing rate-based and the higher-order motifs therefore reduce or even invert the tendency of weights to grow in large assemblies. As a consequence, neurons are more prone to leave a larger assembly and for example join another, smaller one, which induces larger growth of the interconnecting weights. A mechanism similar to the purely firing rate-based effect that we have described has been shown to stabilize the output spike rate of a single neuron receiving feedforward input, since a strong dominance of the rate-based term ultimately leads to the reversal of the sign of plasticity [147]. Another related work, [134], demonstrated the learning of static, non-overlapping assemblies in networks with STDP in the recurrent excitatory and the inhibitory-to-excitatory connections. These assemblies are maintained at least over several hours. The excitatory STDP is thereby balanced (the integral over the STDP window is about zero) such that the inhibitory STDP is effectively much faster. This yields rate homeostasis in the excitatory neurons. In our models the integral over the excitatory STDP curve is negative. Like rate homeostasis, this restricts the maximal average excitatory spike rates. In contrast to rate homeostasis, it also allows for smaller weights, for example in our networks with assembly and background neurons. We note that the fast weight homeostasis in [12] is implemented such that it also only constrains the maximal summed input and output weight of a neuron. We further note that highly diverse average spike rates are observed in biology [148].

Our model networks generate irregular, probabilistic spiking activity. This is in agreement with experimentally observed irregular spiking in the cortex. In our models the irregularity of spiking is guaranteed by the usage of a Poisson spiking neuron model. In biological neural networks, it is usually assumed to result from the input fluctuation-driven spiking dynamics of individual neurons, which occur when there is an overall balance of excitatory and inhibitory input. When the excitatory synaptic weights are plastic, as in our model, this balance might be maintained by inhibitory plasticity [14, 123, 133, 134].

We observe that our networks are able to maintain prominent overlap structures, where each neuron belongs to more than one assembly. The assembly structure is saturated and remains stable. In particular, additionally increased connections are sparse and transient. Previous related works have not shown similarly prominent overlap structures [12–14, 123, 124, 134, 137], perhaps because the mostly assumed fast homeostatic normalization induces a stronger competition between the assemblies. This may force neurons to decide for one assembly. Networks with overlaps allow a more economic use of neural resources, in the sense that more assemblies of a specific size can be stored, in our example network twice as many as without overlaps. The overlap between assemblies in our simulations is about 5%. This agrees with the overlap estimated for assemblies representing associated concepts in the brain [143]. The overlap of randomly chosen assemblies is smaller, about 1%.

Our model networks can stably maintain assemblies in front of a background of neurons that are not part of any assembly. Such a scenario might be particularly relevant for early development when not many memories have been stored yet. Previous related works usually assume that the entire space is tiled by assemblies [12–14, 123, 124, 134]. The reason may be similar as for the prominent overlap structures, namely that fast homeostatic plasticity has a strong tendency to force neurons into assemblies. Ref. [137] shows assemblies in front of a background of weakly connected neurons in

networks with structural plasticity and multiple synapses per connection between neurons.

The assemblies in our networks can drift. Such assembly drift may explain the experimentally observed drift of memory representations in the brain [19]. In our model, assemblies drift by exchanging neurons. This exchange can on the one hand originate from sufficiently large synaptic weight fluctuations, as in [13]. These fluctuations occur in our models for sufficiently large learning rates due to the noisy spiking. On the other hand, our models also show a novel neuron exchange (and thus drift) mechanism: For high spontaneous rate, each neuron belongs to one assembly. If the spontaneous rate of a neuron in the network then transiently drops, the synaptic weights with another assembly increase, such that the neuron belongs to two assemblies. When the intrinsic rate recovers, the synaptic weights with one of the assemblies weaken. This may eliminate the strong weights with the assembly that the neuron originally belonged to and thus induce a switch to the other assembly. The observation suggests a new general mechanism for representational drift in the brain: Transient (or persistent) changes in single neuron properties may lead to changes in the synaptic weights. These in turn induce a change in the features represented by the neuron. The changes in the synaptic weights and the representation may thereby be much longer lasting than the changes in the intrinsic neuronal properties.

Since in the aging brain the overall connectivity decreases [20, 21], we have explored the impact of a such a decrease on the assemblies in our model networks. We observe that the size of assemblies is inversely related to the connection probability in the network. We expect that a similar relation can be observed rather generally, for example also in model networks where the assemblies are stabilized by fast homeostasis. In particular, sparser networks should, all other parameters being equal, lead to larger assemblies that are activated during recall. This is consistent with the observation that neural activity for the same task is stronger in aged individuals [149], where the neural networks are more sparsely connected. An increased assembly size and the resulting stronger activity during reactivation might also explain why episodic memories are experienced more vividly in elderly subjects [150]. In our model networks, we observe that assemblies merge to larger ones when networks become increasingly sparse. Such mergers might explain why episodic memories become less detailed in the aging brain [150, 151].

4.5 Appendices

4.5.A Time-averaged weight change in fully connected assemblies

The correlation function of our Poisson model neurons reads in frequency space

$$\tilde{C}(\omega) = \tilde{C}^0(\omega) + \tilde{C}^1(\omega) = 2\pi\delta(\omega) r r^T + \left(\mathbb{1} - \tilde{a}(\omega)W\right)^{-1} D \left(\mathbb{1} - \tilde{a}(-\omega)W^T\right)^{-1}, . \tag{4.22}$$

For a homogeneously coupled assembly of N neurons with recurrent weights $\hat{w}$ and identical spontaneous rates λ_0, all rates r_i are the same,

$$r_i =: r = \frac{\lambda_0}{1 - (N-1)\hat{w}}, . \tag{4.23}$$

This fixes the first right hand side term of Eq. (4.22) and $D = r\mathbb{1}$. W has zeros on the diagonal and otherwise entries $\hat{w}$. The second right hand side term of Eq. (4.22) can thus be written as

$$\begin{aligned}\tilde{C}^1(\omega) &= \left(\mathbb{1} - \tilde{a}(\omega)W\right)^{-1} D \left(\mathbb{1} - \tilde{a}(-\omega)W^T\right)^{-1} \\ &= r\left(\alpha_+\mathbb{1} + \beta_+ J_N\right)^{-1}\left(\alpha_-\mathbb{1} - \beta_+ J_N\right)^{-1},\end{aligned} \tag{4.24}$$

where J_N is the $N \times N$ matrix of ones, $\alpha_\pm = 1 + \hat{w}\tilde{a}(\pm\omega)$, and $\beta_\pm = -\hat{w}\tilde{a}(\pm\omega)$. We obtain the inverses of the matrices by assuming that they have the same general structure, that is using the ansatz

$$\left(\alpha_\pm\mathbb{1} + \beta_\pm J_N\right)^{-1} = \gamma_\pm\mathbb{1} + \delta_\pm J_N. \tag{4.25}$$

This yields

$$\gamma_\pm = \frac{1}{\alpha_\pm} = \frac{1}{1 + \hat{w}\tilde{a}(\pm\omega)}, \tag{4.26}$$

$$\delta_\pm = \frac{-\beta_\pm}{\alpha_\pm(\alpha_\pm + N\beta_\pm)} = \frac{\hat{w}\tilde{a}(\pm\omega)}{(1 + \hat{w}\tilde{a}(\pm\omega))(1 - \hat{w}\tilde{a}(\pm\omega)(N-1))}. \tag{4.27}$$

In terms of $\gamma_\pm$ and $\delta_\pm$, $C^1(\omega)$ reads

$$\begin{aligned}\tilde{C}^1(\omega) &= r\left(\gamma_+\mathbb{1} + \delta_+ J_N\right)\left(\gamma_-\mathbb{1} + \delta_- J_N\right) \\ &= r\left(\gamma_+\gamma_-\mathbb{1} + \gamma_+\delta_- J_N + \gamma_-\delta_+ J_N + N\delta_+\delta_- J_N\right).\end{aligned} \tag{4.28}$$

Since there are no autapses in the network, we are only interested in the off-diagonal elements,

$$\tilde{C}^1_{i\neq j}(\omega) = r\left(\gamma_+\delta_- + \gamma_-\delta_+ + N\delta_+\delta_-\right). \tag{4.29}$$

The Fourier transform of the synaptic kernel, Eq. (4.2), is

$$\tilde{a}(\omega) = \frac{1}{1 + i\omega\tau_\mathrm{s}}, \tag{4.30}$$

such that

$$\begin{aligned}\tilde{C}^1_{i\neq j}(\omega) = r\Bigg(&\frac{\hat{w}(1+i\omega\tau_\mathrm{s})(1-i\omega\tau_\mathrm{s})}{(1+i\omega\tau_\mathrm{s}+\hat{w})(1-i\omega\tau_\mathrm{s}+\hat{w})(1-i\omega\tau_\mathrm{s}-(N-1)\hat{w})} + \\ &+ \frac{\hat{w}(1-i\omega\tau_\mathrm{s})(1+i\omega\tau_\mathrm{s})}{(1-i\omega\tau_\mathrm{s}+\hat{w})(1+i\omega\tau_\mathrm{s}+\hat{w})(1+i\omega\tau_\mathrm{s}-(N-1)\hat{w})} + \\ &+ \frac{N\hat{w}^2(1+i\omega\tau_\mathrm{s})(1-i\omega\tau_\mathrm{s})}{(1+i\omega\tau_\mathrm{s}+\hat{w})(1+i\omega\tau_\mathrm{s}-(N-1)\hat{w})(1-i\omega\tau_\mathrm{s}+\hat{w})(1-i\omega\tau_\mathrm{s}-(N-1)\hat{w})}\Bigg).\end{aligned} \tag{4.31}$$

For the Fourier transform of the plasticity window, Eq. (4.5), we have

$$\tilde{F}(\omega) = \frac{2A_\mathrm{p}\tau_\mathrm{p}}{(1+i\tau_\mathrm{p}\omega)(1-i\tau_\mathrm{p}\omega)} + \frac{2A_\mathrm{d}\tau_\mathrm{d}}{(1+i\tau_\mathrm{d}\omega)(1-i\tau_\mathrm{d}\omega)} = \tilde{F}(-\omega). \tag{4.32}$$

Inserting Eqs. 4.22 and 4.23 into Eq. (4.8) gives

$$\begin{aligned}\overline{\Delta W_{ij}}(N) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\, \tilde{C}_{ij}(\omega)\tilde{F}(-\omega) \\ &= \frac{f_0\lambda_0^2}{(1-(N-1)\hat{w})^2} + \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\, \tilde{C}^1_{ij}(\omega)\tilde{F}(-\omega). \end{aligned} \tag{4.33}$$

Inserting Eqs. 4.23, 4.31 and 4.32 into Eq. (4.33) results in

$$\begin{aligned}\frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\, \tilde{C}^1_{ij}(\omega)\tilde{F}(-\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathrm{d}\omega\, \Bigg(&2\pi\frac{\hat{w}(1+i\omega\tau_\mathrm{s})(1-i\omega\tau_\mathrm{s})}{(1+i\omega\tau_\mathrm{s}+\hat{w})(1-i\omega\tau_\mathrm{s}+\hat{w})(1-i\omega\tau_\mathrm{s}-(N-1)\hat{w})} + \\ &+ \frac{\hat{w}(1-i\omega\tau_\mathrm{s})(1+i\omega\tau_\mathrm{s})}{(1-i\omega\tau_\mathrm{s}+\hat{w})(1+i\omega\tau_\mathrm{s}+\hat{w})(1+i\omega\tau_\mathrm{s}-(N-1)\hat{w})} + \\ &+ \frac{N\hat{w}^2(1+i\omega\tau_\mathrm{s})(1-i\omega\tau_\mathrm{s})}{(1+i\omega\tau_\mathrm{s}+\hat{w})(1+i\omega\tau_\mathrm{s}-(N-1)\hat{w})(1-i\omega\tau_\mathrm{s}+\hat{w})(1-i\omega\tau_\mathrm{s}-(N-1)\hat{w})}\Bigg)\times \\ &\times\left(\frac{2A_\mathrm{p}\tau_\mathrm{p}}{(1+i\tau_\mathrm{p}\omega)(1-i\tau_\mathrm{p}\omega)} + \frac{2A_\mathrm{d}\tau_\mathrm{d}}{(1+i\tau_\mathrm{d}\omega)(1-i\tau_\mathrm{d}\omega)}\right)\frac{\lambda_0}{1-(N-1)\hat{w}}. \end{aligned} \tag{4.34}$$

This integral can be straightforwardly computed using the residue theorem. Together with the zeroth-order term given by Eq. (4.33) we obtain the following closed-form expression for the time-averaged weight change:

$$\begin{aligned}\overline{\Delta W_{ij}}(N) = &\frac{2\lambda_0^2(A_\mathrm{p}\tau_\mathrm{p}+A_\mathrm{d}\tau_\mathrm{d})}{(1-(N-1)\hat{w})^2} + \\ &+ \frac{\lambda_0 A_\mathrm{p}\tau_\mathrm{p}\hat{w}\left[(2+(2-N)\hat{w})\tau_\mathrm{p} + (2-(N-2)\hat{w}-(N-1)\hat{w}^2)\tau_\mathrm{s}\right]}{(1+\hat{w})(1-(N-1)\hat{w})^2(\tau_\mathrm{s}+(1+\hat{w})\tau_\mathrm{p})(\tau_\mathrm{s}+(1-(N-1)\hat{w})\tau_\mathrm{p})} + \\ &+ \frac{\lambda_0 A_\mathrm{d}\tau_\mathrm{d}\hat{w}\left[(2+(2-N)\hat{w})\tau_\mathrm{d} + (2-(N-2)\hat{w}-(N-1)\hat{w}^2)\tau_\mathrm{s}\right]}{(1+\hat{w})(1-(N-1)\hat{w})^2(\tau_\mathrm{s}+(1+\hat{w})\tau_\mathrm{d})(\tau_\mathrm{s}+(1-(N-1)\hat{w})\tau_\mathrm{d})}. \end{aligned} \tag{4.35}$$

4.5.B Asymptotic behavior of motif contributions to time-averaged plasticity

We first show that the series expansion Eq. (4.11) of the time-averaged weight change in a homogeneously connected assembly simplifies to

$$\begin{aligned}\overline{\Delta W_{ij}} &= f_0 r_i r_j + \sum_{\alpha,\beta} f_{\alpha\beta} \sum_m r_m (W^\alpha)_{im} (W^\beta)_{jm} \\ &= f_0 r^2 + \frac{r}{N} \sum_{k=1}^{\infty} f_k \left((N-1)^k - (-1)^k \right) \hat{w}^k, \end{aligned} \tag{4.36}$$

where

$$f_k := \sum_{\alpha+\beta=k} f_{\alpha\beta}. \tag{4.37}$$

In homogeneous assemblies we have $r_i = r$ and $W = \hat{w}(J_N - \mathbb{1})$. Eq. (4.8) thus yields

$$\begin{aligned}\overline{\Delta W_{ij}} &= f_0 r^2 + r \sum_{\alpha,\beta} f_{\alpha\beta} \left(J_N - \mathbb{1}\right)^{\alpha+\beta}_{ij} \hat{w}^{\alpha+\beta} \\ &= f_0 r^2 + r \sum_{k=1}^{\infty} f_k \left(J_N - \mathbb{1}\right)^k_{ij} \hat{w}^k. \end{aligned} \tag{4.38}$$

We observe that $J_N^m = N^{m-1} J_N$ for $m \geq 1$ while $J_N^0 = \mathbb{1}$. With this the binomial in Eq. (4.38) can be expanded to

$$\begin{aligned}\left(J_N - \mathbb{1}\right)^k &= \sum_{l=0}^{k} \binom{k}{l} J_N^l \, (-\mathbb{1})^{k-l} \\ &= \sum_{l=0}^{k} \binom{k}{l} N^{l-1} J_N (-1)^{k-l} - \frac{(-1)^k}{N} J_N + (-\mathbb{1})^k \,. \end{aligned} \tag{4.39}$$

We are again interested only in off-diagonal elements:

$$\begin{aligned}\left(J_N - \mathbb{1}\right)^k_{i\neq j} &= \sum_{l=0}^{k} \binom{k}{l} N^{l-1} (-1)^{k-l} - \frac{(-1)^k}{N} \\ &= \frac{(N-1)^k - (-1)^k}{N}. \end{aligned} \tag{4.40}$$

Inserting this into Eq. (4.38) gives Eq. (4.36).

We now show that

$$\lim_{k\to\infty} \sum_{\alpha+\beta=k} f_{\alpha\beta} = \frac{f_0}{2\tau_s}. \tag{4.41}$$

We start by using Eq. (4.12), to write

$$\sum_{\alpha+\beta=k} f_{\alpha\beta} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathrm{d}\omega\ \tilde{F}(-\omega) \sum_{\alpha+\beta=k} \tilde{a}(\omega)^{\alpha} \tilde{a}(-\omega)^{\beta}. \tag{4.42}$$

The sum in the integrand can be rewritten using Eq. (4.30),

$$\begin{aligned}
\sum_{\alpha+\beta=k} \tilde{a}(\omega)^{\alpha} \tilde{a}(-\omega)^{\beta} &= \sum_{\alpha=0}^{k} \tilde{a}(\omega)^{\alpha} \tilde{a}(-\omega)^{k-\alpha} \\
&= \sum_{\alpha=0}^{k} \left(\frac{1}{1+i\omega\tau_s}\right)^{\alpha} \left(\frac{1}{1-i\omega\tau_s}\right)^{k-\alpha} \\
&= \sum_{\alpha=0}^{k} \left(\frac{1-i\omega\tau_s}{1+i\omega\tau_s}\right)^{\alpha} \left(\frac{1}{1-i\omega\tau_s}\right)^{k} \\
&= \frac{1-\left(\frac{1-i\omega\tau_s}{1+i\omega\tau_s}\right)^{k+1}}{1-\left(\frac{1-i\omega\tau_s}{1+i\omega\tau_s}\right)} \left(\frac{1}{1-i\omega\tau_s}\right)^{k} \\
&= \frac{\omega\tau_s - i}{2\omega\tau_s} \left(\frac{1}{1-i\omega\tau_s}\right)^{k} + \frac{\omega\tau_s + i}{2\omega\tau_s} \left(\frac{1}{1+i\omega\tau_s}\right)^{k}.
\end{aligned} \tag{4.43}$$

We substitute $\omega' := \omega\tau_s$ and insert Eq. (4.43) into Eq. (4.42):

$$\begin{aligned}
\sum_{\alpha+\beta=k} f_{\alpha\beta} &= \frac{1}{2\pi\tau_s} \int_{-\infty}^{\infty} \mathrm{d}\omega'\ \tilde{F}(-\omega'/\tau_s) \left(\frac{\omega'-i}{2\omega'} \left(\frac{1}{1-i\omega'}\right)^{k} + \frac{\omega'+i}{2\omega'} \left(\frac{1}{1+i\omega'}\right)^{k}\right) \\
&= \frac{1}{2\pi\tau_s} \int_{-\infty}^{\infty} \mathrm{d}\omega' \int_{-\infty}^{\infty} \mathrm{d}t\ \mathrm{e}^{\frac{i\omega' t}{\tau_s}} F(t) \left(\frac{\omega'-i}{2\omega'} \left(\frac{1}{1-i\omega'}\right)^{k} + \frac{\omega'+i}{2\omega'} \left(\frac{1}{1+i\omega'}\right)^{k}\right) \\
&= \frac{1}{2\pi\tau_s} \int_{-\infty}^{\infty} \mathrm{d}t\ F(t) \int_{-\infty}^{\infty} \mathrm{d}\omega'\ \mathrm{e}^{\frac{i\omega' t}{\tau_s}} \left(\frac{\omega'-i}{2\omega'} \left(\frac{1}{1-i\omega'}\right)^{k} + \frac{\omega'+i}{2\omega'} \left(\frac{1}{1+i\omega'}\right)^{k}\right).
\end{aligned} \tag{4.44}$$

In the second line we employed the definition of the Fourier transform to substitute $\tilde{F}(-\omega/\tau_s)$. (Alternatively, one could use Plancherel's theorem and the inverse Fourier transform of 4.43 to obtain the third line.) The integrand of the inner integral of Eq. (4.44) has poles at $\omega' = \pm i$ (the singularity at $\omega' = 0$ is removable). For $t > 0$ ($t < 0$) we can compute the integral using a contour over the upper

(lower) half complex plane. The residue theorem then yields

$$\begin{aligned}
&\frac{1}{2\pi\tau_\mathrm{s}}\int_{-\infty}^{\infty}\mathrm{d}t\ F(t)\int_{-\infty}^{\infty}\mathrm{d}\omega'\ \mathrm{e}^{\frac{i\omega' t}{\tau_\mathrm{s}}}\left(\frac{\omega'-i}{2\omega'}\left(\frac{1}{1-i\omega'}\right)^k+\frac{\omega'+i}{2\omega'}\left(\frac{1}{1+i\omega'}\right)^k\right)\\
&=-\frac{i}{\tau_\mathrm{s}}\int_{-\infty}^{0}\mathrm{d}t\ F(t)\,\mathrm{Res}\left(\mathrm{e}^{\frac{i\omega' t}{\tau_\mathrm{s}}}\ \frac{\omega'-i}{2\omega'}\left(\frac{1}{1-i\omega'}\right)^k,-i\right)+\frac{i}{\tau_\mathrm{s}}\int_{0}^{\infty}\mathrm{d}t\ F(t)\,\mathrm{Res}\left(\mathrm{e}^{\frac{i\omega' t}{\tau_\mathrm{s}}}\ \frac{\omega'+i}{2\omega'}\left(\frac{1}{1+i\omega'}\right)^k,i\right)
\end{aligned}\tag{4.45}$$

We use that

$$\mathrm{Res}(g,c)=\frac{1}{(k-1)!}\lim_{z\to c}\frac{\mathrm{d}^{k-1}}{\mathrm{d}z^{k-1}}\left((z-c)^k g(z)\right)\tag{4.46}$$

if $g(z)$ has a kth order pole at $z=c$. For $\alpha+\beta=k+1$ the residue at $\omega'=+i$ in Eq. (4.45) thus becomes

$$\begin{aligned}
\mathrm{Res}\left(\mathrm{e}^{\frac{i\omega' t}{\tau_\mathrm{s}}}\ \frac{\omega'+i}{2\omega'}\left(\frac{1}{1+i\omega'}\right)^{k+1},i\right)&=\frac{1}{k!}\lim_{\omega'\to i}\frac{\mathrm{d}^k}{\mathrm{d}\omega'^k}\left(\left(\omega'-i\right)^{k+1}\left(\frac{1}{1+i\omega'}\right)^{k+1}\left(\frac{1}{2}+\frac{i}{2\omega'}\right)\mathrm{e}^{\frac{i\omega' t}{\tau_\mathrm{s}}}\right)\\
&=\frac{1}{i^{k+1}k!}\lim_{\omega'\to i}\frac{\mathrm{d}^k}{\mathrm{d}\omega'^k}\left(\left(\frac{1}{2}+\frac{i}{2\omega'}\right)\mathrm{e}^{\frac{i\omega' t}{\tau_\mathrm{s}}}\right)\\
&=\frac{1}{i^{k+1}k!}\lim_{\omega'\to i}\left(\frac{1}{2}\left(\frac{it}{\tau_\mathrm{s}}\right)^k\mathrm{e}^{\frac{i\omega' t}{\tau_\mathrm{s}}}+\sum_{j=0}^{k}\binom{k}{j}\left(\frac{it}{\tau_\mathrm{s}}\right)^j\mathrm{e}^{\frac{i\omega' t}{\tau_\mathrm{s}}}\frac{(-1)^{k-j}(k-j)!\,i}{2\omega'^{k-j+1}}\right)\\
&=\frac{1}{i^{k+1}}\lim_{\omega'\to i}\left(\frac{1}{2k!}\left(\frac{it}{\tau_\mathrm{s}}\right)^k\mathrm{e}^{\frac{i\omega' t}{\tau_\mathrm{s}}}+\sum_{j=0}^{k}\left(\frac{it}{\tau_\mathrm{s}}\right)^j\mathrm{e}^{\frac{i\omega' t}{\tau_\mathrm{s}}}\frac{(-1)^{k-j}\,i}{2\,j!\,\omega'^{k-j+1}}\right)\\
&=\frac{\mathrm{e}^{-\frac{t}{\tau_\mathrm{s}}}}{2ik!}\left(\frac{t}{\tau_\mathrm{s}}\right)^k+\frac{1}{2i}\mathrm{e}^{-\frac{t}{\tau_\mathrm{s}}}\sum_{j=0}^{k}\frac{1}{j!}\left(\frac{t}{\tau_\mathrm{s}}\right)^j.
\end{aligned}\tag{4.47}$$

Similarly for the residue at $\omega'=-i$ we obtain

$$\mathrm{Res}\left(\mathrm{e}^{\frac{i\omega' t}{\tau_\mathrm{s}}}\ \frac{\omega'-i}{2\omega'}\left(\frac{1}{1-i\omega'}\right)^{k+1},-i\right)=\frac{i\,\mathrm{e}^{\frac{t}{\tau_\mathrm{s}}}}{2k!}\left(\frac{-t}{\tau_\mathrm{s}}\right)^k+\frac{i}{2}\mathrm{e}^{\frac{t}{\tau_\mathrm{s}}}\sum_{j=0}^{k}\frac{1}{j!}\left(\frac{-t}{\tau_\mathrm{s}}\right)^j.\tag{4.48}$$

Inserting Eq. (4.47) and Eq. (4.48) into Eq. (4.45) gives

$$\begin{aligned}
\sum_{\alpha+\beta=k}f_{\alpha\beta}&=\frac{1}{\tau_\mathrm{s}}\int_{-\infty}^{\infty}\mathrm{d}t\ F(t)\left(\frac{\mathrm{e}^{-\frac{|t|}{\tau_\mathrm{s}}}}{2(k-1)!}\left(\frac{|t|}{\tau_\mathrm{s}}\right)^{k-1}+\frac{1}{2}\mathrm{e}^{-\frac{|t|}{\tau_\mathrm{s}}}\sum_{j=0}^{k-1}\frac{1}{j!}\left(\frac{|t|}{\tau_\mathrm{s}}\right)^j\right)&(4.49)\\
&=:\frac{1}{\tau_\mathrm{s}}\int_{-\infty}^{\infty}\mathrm{d}t\ F(t)A_k(t).&(4.50)
\end{aligned}$$

We then take the limit $k \to \infty$:

$$\begin{aligned}\lim_{k\to\infty} \sum_{\alpha+\beta=k} f_{\alpha\beta} &= \lim_{k\to\infty} \frac{1}{\tau_s} \int_{-\infty}^{\infty} \mathrm{d}t\ F(t) \left(\frac{\mathrm{e}^{-\frac{|t|}{\tau_s}}}{2(k-1)!} \left(\frac{|t|}{\tau_s}\right)^{k-1} + \frac{1}{2} \mathrm{e}^{-\frac{|t|}{\tau_s}} \sum_{j=0}^{k-1} \frac{1}{j!} \left(\frac{|t|}{\tau_s}\right)^{j} \right) \\ &= \lim_{k\to\infty} \frac{1}{\tau_s} \int_{-\infty}^{\infty} \mathrm{d}t\ F(t) \frac{\mathrm{e}^{-\frac{|t|}{\tau_s}}}{2(k-1)!} \left(\frac{|t|}{\tau_s}\right)^{k-1} + \frac{1}{2\tau_s} \int_{-\infty}^{\infty} \mathrm{d}t\ F(t)\ \mathrm{e}^{-\frac{|t|}{\tau_s}} \lim_{k\to\infty} \sum_{j=0}^{k-1} \frac{1}{j!} \left(\frac{|t|}{\tau_s}\right)^{j}.\end{aligned} \tag{4.51}$$

The first limit vanishes as long as $F(t)$ goes to zero polynomially or faster for large $|t|$. This is guaranteed if $F(t)$ is integrable over $(-\infty, \infty)$, which we already implicitly assume for example to do the Fourier transform. The limit of the power series in the second term is the exponential function. (Limit and integral are there interchangeable due to the dominated convergence theorem.) From Eq. (4.51) we thus obtain the final result:

$$\begin{aligned}\lim_{k\to\infty} \sum_{\alpha+\beta=k} f_{\alpha\beta} &= \frac{1}{2\tau_s} \int_{-\infty}^{\infty} \mathrm{d}t\ F(t)\ \mathrm{e}^{-\frac{|t|}{\tau_s}} \mathrm{e}^{\frac{|t|}{\tau_s}} \\ &= \frac{f_0}{2\tau_s}.\end{aligned} \tag{4.52}$$

To develop an intuition about this result, we observe that the inner integral of Eq. (4.44) is up to a constant factor the inverse Fourier transformation of $\sum_{\alpha+\beta=k} \tilde{a}(\omega)^{\alpha} \tilde{a}(-\omega)^{\beta}$. Using its linearity we may apply the inverse Fourier transform to each summand individually. This yields $a(t)$ convolved α times with itself and with $a(-t)$ convolved β times with itself. Due to Eq. (4.2) this equals the probability distribution of a sum of α exponentially distributed independent stochastic variables with mean τ_s minus the sum of β stochastic variables of the same type. In other words, we have the probability distribution of a spike time occurring at the end of a cascade of α spikes with exponentially distributed inter-spike intervals minus the time of a spike occurring at the end of a similar cascade of β spikes. This reflects the fact that the term $f_{\alpha\beta}$ covers the impact of the motif where a neuron evokes a spike in the pre- and postsynaptic neurons after spike cascades of length β and α. The probability distribution has mean $(\alpha - \beta)\tau_s$ and standard deviation $\sqrt{k}\tau_s$. The standard deviations of all distributions are thus identical and neighboring distributions have distance $2\tau_s$. According to the central limit theorem, for large k the distributions approximate normal distributions. For α and β adding to the same k, these are equidistantly shifted but otherwise identical. With increasing k they broaden, such that their superposition forms a plateau. The increase in overlap thereby compensates the decrease in the distribution heights. The number of distributions increases with increasing k. The added distributions, however, do not lead to a non-compensatory increase of the superposition in the relevant center where it overlaps with F. This is because the added distributions initially do not reach the center, as their mean scales with k while their width scales only with $\sqrt{k}$.

4.5.C Parameters

	N	$\hat{w}$	τ_s	A_p	A_d	τ_p	τ_d	μ	λ_0
Fig. 4.2	120	0.04	0.01 s	0.08	-0.0533	0.025 s	0.05 s	0.07	0.15 Hz
Fig. 4.3a			0.01 s	0.08	-0.0533	0.025 s	0.05 s		0.15 Hz
Fig. 4.4	80	0.026	0.01 s	0.08	-0.066	0.035 s	0.05 s	0.07	0.15 Hz
Fig. 4.5	120	0.017	0.01 s	0.08	-0.044	0.022 s	0.055 s	0.02	0.15 Hz
Fig. 4.6a	60	0.0158	0.01 s	0.08	-0.044	0.022 s	0.055 s	0.021	0.15 Hz
Fig. 4.6b	60	0.024	0.01 s	0.08	-0.0533	0.025 s	0.05 s	0.045	0.15 Hz
Fig. 4.7	190	0.018	0.01 s	0.08	- 0.042	0.026 s	0.065 s	0.017	0.15 Hz
Fig. 4.8a	72	0.056	0.01 s	0.08	-0.0533	0.025 s	0.05 s		0.2 Hz
Fig. 4.8b	120	0.024	0.01 s	0.08	-0.0533	0.025 s	0.05 s	0.05	0.15 Hz
Fig. 4.9a			0.01 s	0.08	-0.0533	0.025 s	0.05 s		0.15 Hz
Fig. 4.9b	120	0.058	0.01 s	0.08	-0.0533	0.025 s	0.05 s	0.07	0.15 Hz
Fig. 4.9c	150		0.01 s	0.08	-0.0533	0.025 s	0.05 s		0.2 Hz
Fig. 4.10	80	0.024	0.01 s	0.08	-0.0533	0.025 s	0.05 s	0.055	0.15 Hz
Fig. 4.11				0.08	-0.0533	0.025 s	0.05 s		
Fig. 4.12	72	0.056	0.01 s	0.08	-0.0533	0.025 s	0.05 s		0.2 Hz

Fig. 4.2: The networks starts with an initially random connectivity where each weight is independently and uniformly drawn from $[0, 0.25\hat{w}]$.
Fig. 4.4: The initial weight matrix has an assembly of neurons 1-20 interconnected with $W_{ij} = \hat{w}$ and background connectivity as in Fig. 4.2. During the stimulation period neurons 21-40 are given input from a single source of Poisson spiking with $\lambda = 4.18$ Hz and input weights $w_{\text{in}} = 10\hat{w}$.
Fig. 4.6a: The neuron with index 10 has spontaneous rate $\lambda_0 = 0.08$ Hz.
Fig. 4.6b: The neuron with index 10 has spontaneous rate $\lambda_0 = 0.01$ Hz.
Fig. 4.8a: The initial weight matrix is as in Fig. 4.2 but with weights drawn from $[0, 0.15\hat{w}]$.
Fig. 4.8b: Every 3×10^5 s a neuron changes its spontaneous rates with probability $p = 0.03$ to $\lambda_0' = 0.03$ Hz for 3×10^5 s. For the last 3×10^5 s of the simulation all spontaneous rates are kept at $\lambda_0 = 0.15$ Hz.

4.5.D Additional figures

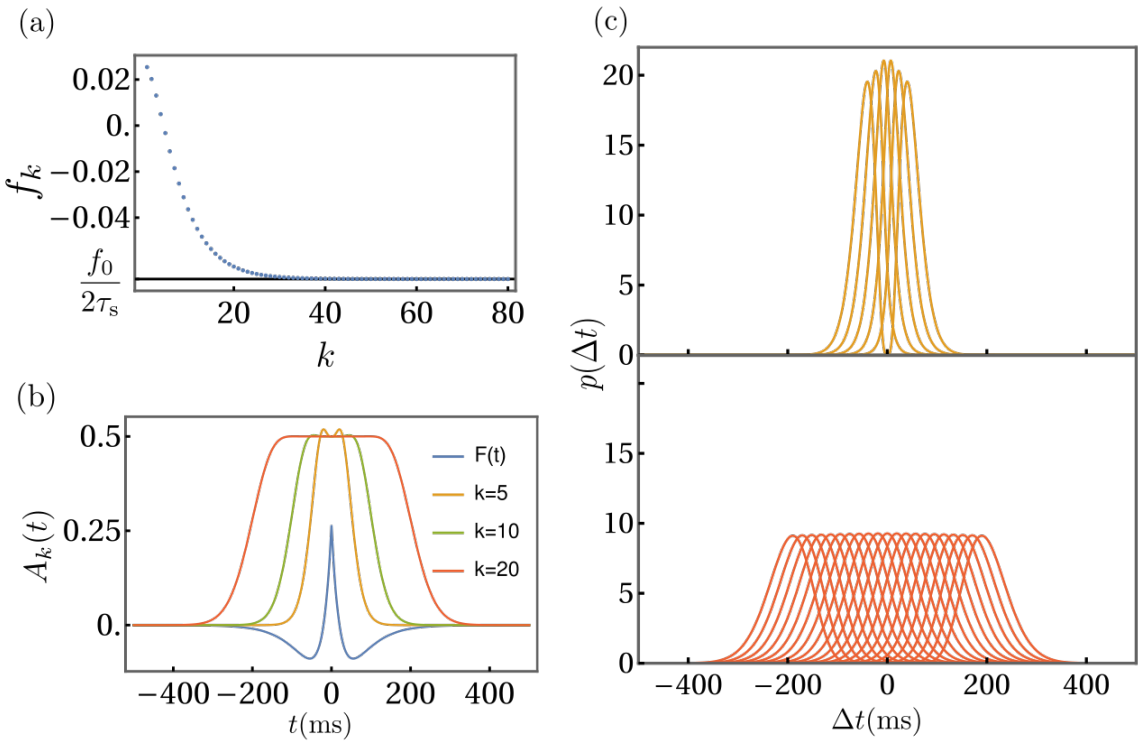

Figure 4.10: Asymptotic behavior of sums f_k of motif contributions. (a) Convergence of f_k to $\frac{f_0}{2\tau_s}$ for high orders of k. (b) Sums of convolutions of the synaptic current functions $A_k(t)$ (see Eq. (4.50)), for different values of k. As discussed in section S.2, these curves can be interpreted as resulting from the superposition of distributions of time-lags for spike cascades of total length k affecting pre- and postsynaptic neurons (see (c)). For high orders, $A_k(t)$ becomes a plateau of height $\frac{1}{2}$. The learning window (blue, a.u.), is shown to illustrate the time-scales. (c) Probability distributions of the time lags Δt between the last spikes of two spike-cascades of different lengths β and α. The total cascade length $k = \alpha + \beta$ equals 5 spikes (upper subpanel) or 20 spikes (lower subpanel). For larger k the distributions have larger variance and are more spread out. Their superpositions therefore assume the widening plateau shapes displayed in (b).

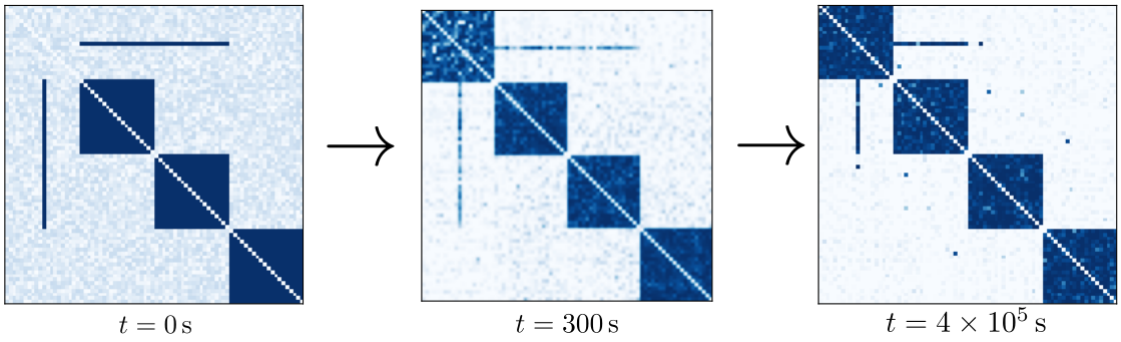

Figure 4.11: A network with preexisting assemblies, background neurons and a neuron that is part of two assemblies, learns a new assembly. The figure displays the weight matrices of the initial network configuration (left hand side), after stimulation (middle) and after a longer time (right hand side). The stimulation of neurons 1-20 lasts for 300 s with the same stimulation protocol as in Fig. 4. The initially strong connections of the "overlap neuron" (neuron 10) with two other assemblies experience depression after the neuron is recruited to the new assembly (middle). Subsequently it loses its connections to one of the assemblies it previously belonged to (right hand side).

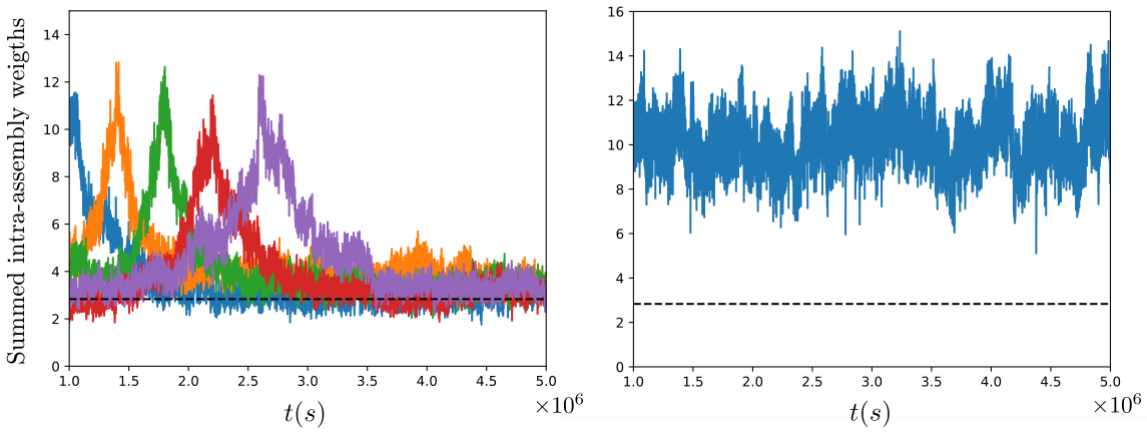

Figure 4.12: Drifting assembly dynamics. (a) The network with drifting assemblies of Fig. 4.8a is simulated over a longer time. We choose at five time points ($t = 1.0 \times 10^6$ s, 1.4×10^6 s, 1.8×10^6 s, 2.2×10^6 s, 2.6×10^6 s) an assembly as a reference and note the neurons forming it. Thereafter, we track the sum of weights between these neurons in the present, future and past weight matrices. For sufficiently long temporal distances this sum approximately reaches chance level (dashed black line, for an average size assembly), which we define as the average of the sum of interconnections in groups of randomly chosen neurons. We can therefore conclude that assemblies indeed completely drift, that is, there is no constant "stable core" set of neurons in an assembly. (b) Tracking of a single drifting assembly over time. At $t_0 = 1 \times 10^6$ s we choose assembly 1 (of the four assemblies) and set $t = t_0$. We then compute the sum of the weights at time $t + \Delta t$ ($\Delta t = 2$ s) between the neurons that formed assembly 1 at time t, to see if they are still strongly connected. We repeat the procedure using the actual assembly 1 at the new time $t = t_0 + \Delta t$ (which may have exchanged individual neurons compared to the one at t_0). In particular, we again compute the weights between its neurons at $t + \Delta t$ and so on. We find that the sum of weights stays at a consistently high level. In other words, neurons that form assembly 1 at a time t are still strongly connected at $t + \Delta t$. The change of the assembly is therefore gradual; most neurons that are part of the assembly at t are still part of it at $t + \Delta t$. This implies that the assembly can be tracked over time, despite the complete drift.

CHAPTER 5

Summary and outlook

In this thesis, we have dealt with different aspects of stability in balanced state networks in which neurons fire irregularly and asynchronously. In Chapter 3 we addressed the subject of dynamical stability, which studies the response of dynamical systems to small perturbations. Previous studies have proven, numerically and analytically, that networks of pulse-coupled inhibitory LIF neurons exhibit a negative leading Lyapunov exponent and that introducing recurrent excitatory interactions leads to a gradual shift to chaos [8, 9]. These studies further show that the stability of these networks is due to the concave shape of the rise function of LIF neurons. In biological neurons, however, the membrane potential accelerates toward the firing threshold at a certain level. We have studied the opposite case of completely convex rise functions, introducing the XIF neuron. These provide a mathematically tractable model of neurons whose membrane potential is balanced at the level where it accelerates toward spiking. We have achieved this by reversing the sign of the leak current of LIF neurons and by introducing a simple voltage dependence of the synapse. We have obtained the membrane potential distributions and firing rates of both LIF and XIF neurons by employing and adapting a mean-field theoretical approach that assumes Poissonian inputs and takes into account the finite size of the instantaneous postsynaptic potentials [60, 86, 96, 97]. We applied the results to obtain neuron parameters that lead to homogeneous firing rates; using these parameters, we created mixed networks of inhibitory LIF and XIF neurons with constant external drive that exhibit the characteristic irregular spiking of the balanced state. We have then developed another mean-field approach to estimate the Lyapunov exponents from first principles using only the firing rates and the neuron parameters. Our mean field estimate is in good agreement with the Lyapunov spectrum obtained from numerical simulation. We find that, setting aside the zero Lyapunov exponent, for each XIF neuron in the network there is one positive Lyapunov exponent. This means that already one XIF neuron is enough to render an otherwise stable network chaotic, in contrast to the introduction of recurrent excitatory connections, which only leads to a gradual transition to chaos. This sudden transition to chaos is a consequence of the fully convex rise function of the XIF neuron, for which nothing can realistically counteract the growth of perturbations. The mean field estimate of the Lyapunov exponents depends only on the additional time between a neuron's spikes caused by the recurrent inhibition and the signs and magnitudes of the leak currents. The LEs are, therefore, largely independent of the collective dynamics but rather reflect the properties of individual neurons. This implies, in particular, that the typical perturbation growth rate does not change with network size. Due to the linear subthreshold dynamics, the mean-field prediction of the Lyapunov exponents does not

depend on the statistics of the incoming spike trains. Therefore, the result should also apply to LIF/XIF networks that are not in the balanced state. Furthermore, we have established the leak time constants as the respective upper and lower bounds for the Lyapunov exponents in our model. While we have only studied linear subthreshold dynamics, we expect that for general networks with purely concave and convex rise functions, the rule of one positive Lyapunov exponent per convex neuron still holds. Our analysis of the Lyapunov exponents has proven that the neuron type composition of a network can have dramatic effects on its phase space structure. In addition, our model provides a simple recipe for a dynamical system of arbitrary dimension and with an arbitrary degree of hyperchaos.

We have, furthermore, computed and analyzed the covariant Lyapunov vectors of our networks, confirming our intuitions used in the mean-field approach. The CLVs additionally have allowed us to apply the properties of mixed networks to computations with precise spike times. For this we have used a mixed network as a dynamical reservoir and trained its outgoing weights using Finite Precision Learning [61] to produce desired spike patterns, depending on input and context. Notably, our approach does not treat the reservoir as monolithic, as reservoir computing usually does, but explicitly makes use of stable and unstable phase space directions, facilitated by the presence of the two different neuron types. This scheme allows for computations that ignore input variations in irrelevant directions while magnifying relevant ones. While stable and unstable directions, in general, depend on the state of the system, in our model, they are always mostly aligned to the corresponding neuron populations. Precise computation of the CLV, which requires knowledge of future and past trajectories, may therefore not be necessary to take advantage of the different sensitivities. It would be interesting to see if the brain, too, can distinguish between relevant and irrelevant input directions this way. In addition to a possible implementation in the brain, our computational scheme may also be interesting for computations with spikes on neuromorphic hardware.

In Chapter 4, we investigated the stability of memory assemblies in balanced state networks. In contrast to Chapter 3, here we take into account that network dynamics cause changes to the network's structure. We have introduced an STDP rule that supports the emergence, learning, maintenance, drift, and growth of assemblies without the need for implausible homeostatic mechanisms. It is characterized by a negative integral of the STDP window, which means that a synapse connecting two neurons with uncorrelated spiking will, on average, experience depression. We have analyzed the effects of this plasticity rule on networks of Hawkes neurons and identified why assembly growth ceases past a certain assembly size. This is due to two effects, both related to the negative integral of the STDP window: The first effect is increased firing rates in larger assemblies leading to an increase in uncorrelated spiking. The second effect is longer cascades of spikes within larger assemblies, spreading out the distribution of time gaps of spike pairs within these assemblies.

With our STDP model, networks can self-organize into a structure of assemblies from initially unstructured connectivity, which to our knowledge, has not been achieved before without fast homeostatic plasticity. The formation of assemblies not triggered by external input may occur early in brain development, as evidenced by observations in the optic tectum of larval zebrafish.

We have further demonstrated the creation of assemblies within networks of weakly connected neurons through correlated external input. In our model, weakly connected neurons are, thus, a suitable substrate for learning new memories. Notably, unaffected neurons stay weakly connected, which would not be possible if homeostasis caused summed weights of neurons to be normalized to a fixed value.

Our model can overcome the difficulty of creating and maintaining stable overlaps between assemblies. In doing so, we have contributed to the relatively unexplored relationship of STDP

with memory associations. Previous theoretical work has studied overlapping assemblies and their relationship with associative memory [143]; it has established the minimum overlap needed for an assembly activation to trigger activation of an associated assembly and the maximum overlap, that still allows for the distinction between two assemblies. We have demonstrated mechanisms that stabilize associated assemblies by preventing plasticity during background firing from causing overlaps to fade away or completely merge: something that seems to occur in models where assemblies are stabilized by strong homeostatic competition between weights.

We have further proven that the drifting of assemblies as proposed in [13] is not reliant on strict competition between weights. Representations in [13] drift because of synaptic changes unrelated to STDP [146] and due to weight fluctuations caused by the irregular dynamics at sufficiently high learning rates. We demonstrate drift with the latter mechanism and, using the insights gained in our study of overlapping assemblies, propose a novel drift mechanism that relies on changes in the intrinsic firing rates of neurons, causing the generation of temporary overlaps.

Furthermore, we have studied our model's dependence on the overall sparsity of connections. Our results suggest that when the number of synaptic connections in the brain decreases, as observed during aging [20, 21], the size of the neuron assemblies underlying memories increases. The larger assembly sizes may render memories in the aging brain more robust and prominent but also less specific.

For simplicity and following previous work, we have used a symmetric pairwise STDP rule. Symmetric STDP has been discovered in the CA3 region of the hippocampus [36] or the motor cortex [140]. Since the negative integral of our STDP rule is the primary reason for assembly stability, we are confident that one can obtain similar results with asymmetric or triplet STDP. Likewise, while we focused on excitatory neurons in this part, we expect our results to also hold if we explicitly include inhibitory neurons. In fact, inhibitory plasticity may be needed to some degree to maintain the balanced state. While we have used the linear Hawkes model, which is especially suited to study the effects of STDP analytically, we expect that our results should qualitatively also hold for general neuron models as long as the balanced state is maintained.

Our work in Chapter 4 makes several predictions for future experiments: First, we predict that homeostatic plasticity and heterosynaptic plasticity suitable to stabilize assemblies will not be experimentally found in networks generating assemblies. Second, impairment of the experimentally found slow homeostatic plasticity and fast heterosynaptic interactions need not abolish the ability of networks to store assemblies. Third, assemblies representing memories may, especially in ontogenesis, exist in front of a background of non-assembly neurons. Finally, we predict that assemblies in the brains of older individuals will be larger on average compared to younger ones.

In both parts of this thesis, we have studied neural networks in the balanced state with particular attention given to stability and the tradeoffs and limitations that come with it. While the brain needs to be able to stably represent information in order to process or store it, this stability may often come at the cost of sensitivity and plasticity. Maneuvering these dilemmas is a nontrivial task; indeed, stability-plasticity dilemmas are an ongoing challenge for the study of both biological [152, 153] and artificial [154–156] neural networks. The balanced state adds an additional layer of difficulty to this. On the one hand, it seems difficult to encode information via precise spike times if spiking activity is seemingly random, on the other hand, the irregular activity interacts with synaptic plasticity.

In Chapter 3 we examined the dilemma of the brain needing to be sensitive to variations in incoming information while at the same time being stable to random noise. For this, we have studied the dynamical stability and phase space structure of mixed networks with two dynamically distinct neurons.

We have shown a caveat to the previously made observation that inhibitory networks of LIF neurons tend to be stable, that is, nonchaotic: this stability can be broken by just one neuron with purely convex subthreshold dynamics. On the other hand, we found that a mixed neural network is able to spatially compartmentalize stability and sensitivity, such that it is sensitive to some inputs while generalizing others.

In contrast to the spiking patterns we studied in Chapter 3, in Chapter 4 we investigated the stability of synaptic weight configurations. Here we face a different sensitivity-stability dilemma: on the one hand, plasticity is needed to encode new memories, on the other hand, it needs to stably maintain old ones, even in the presence of irregular background activity. We have demonstrated a solution to this dilemma that does not rely on unrealistically fast homeostatic plasticity mechanisms. Furthermore, we have demonstrated the stability of overlaps between assemblies; here again, great care is needed, since the same forces that maintain stable overlaps between two assemblies can also cause these assemblies to merge completely.

With these works, we hope to have provided some valuable new perspectives on both dynamical and memory stability in networks of spiking neurons in the balanced state and stimulated future research.

Bibliography

[1] P. Manz, S. Goedeke, and R.-M. Memmesheimer, *Dynamics and computation in mixed networks containing neurons that accelerate towards spiking*, Physical Review E **100** (2019) 042404 (cit. on pp. vii, 19, 56).

[2] P. Manz and R.-M. Memmesheimer, *Purely STDP-based assembly dynamics: stability, learning, overlaps, drift and aging*, PLOS Computational Biology **19** (2023) e1011006 (cit. on pp. vii, 53, 54).

[3] G. L. Gerstein and B. Mandelbrot, *Random walk models for the spike activity of a single neuron*, Biophysical journal **4** (1964) 41 (cit. on pp. 1, 21, 56).

[4] M. N. Shadlen and W. T. Newsome, *Noise, neural codes and cortical organization*, Current opinion in neurobiology **4** (1994) 569 (cit. on pp. 1, 10, 21, 56).

[5] C. Van Vreeswijk and H. Sompolinsky, *Chaos in neuronal networks with balanced excitatory and inhibitory activity*, Science **274** (1996) 1724 (cit. on pp. 1, 10, 11, 21, 43, 56).

[6] S. H. Strogatz, *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*, CRC press, 2018 (cit. on p. 2).

[7] N. Bertschinger and T. Natschläger, *Real-time computation at the edge of chaos in recurrent neural networks*, Neural computation **16** (2004) 1413 (cit. on p. 2).

[8] S. Jahnke, R.-M. Memmesheimer, and M. Timme, *Stable irregular dynamics in complex neural networks*, Physical Review Letters **100** (2008) 048102 (cit. on pp. 2, 20–23, 34, 44, 45, 56, 81).

[9] S. Jahnke, R.-M. Memmesheimer, and M. Timme, *How chaotic is the balanced state?* Frontiers in computational neuroscience (2009) 13 (cit. on pp. 2, 21, 22, 34, 44, 45, 81).

[10] A. Pikovsky and A. Politi, *Lyapunov exponents: a tool to explore complex dynamics*, Cambridge University Press, 2016 (cit. on pp. 2, 5, 21, 34–36, 44, 49, 52).

[11] D. O. Hebb, *The organization of behavior: A neuropsychological theory*, Psychology Press, 2005 (cit. on pp. 3, 8).

[12] N. Ravid Tannenbaum and Y. Burak, *Shaping neural circuits by high order synaptic interactions*, PLoS Comput Biol **12** (2016) e1005056 (cit. on pp. 3, 53–59, 61, 68, 69).

[13] Y. F. Kalle Kossio, S. Goedeke, C. Klos, and R.-M. Memmesheimer, *Drifting assemblies for persistent memory: Neuron transitions and unsupervised compensation*, Proc Natl Adac Sci U S A **118** (2021) (cit. on pp. 3, 53–55, 57, 59, 65, 66, 68–70, 83).

[14] A. Litwin-Kumar and B. Doiron, *Formation and maintenance of neuronal assemblies through synaptic plasticity*, Nat Commun **5** (2014) 1 (cit. on pp. 3, 53–55, 68, 69).

[15] F. Zenke, W. Gerstner, and S. Ganguli, *The temporal paradox of Hebbian learning and homeostatic plasticity*, Curr Opin Neurobiol **43** (2017) 166 (cit. on pp. 3, 53, 55, 68).

[16] E. De Falco, M. J. Ison, I. Fried, and R. Q. Quiroga, *Long-term coding of personal and universal associations underlying the memory web in the human brain*, Nat Commun **7** (2016) 1 (cit. on pp. 3, 55, 63, 64).

[17] H. G. Rey et al., *Encoding of long-term associations through neural unitization in the human medial temporal lobe*, Nat Commun **9** (2018) 1 (cit. on pp. 3, 55, 63).

[18] H. G. Rey et al., *Single neuron coding of identity in the human hippocampal formation*, Curr Biol **30** (2020) 1152 (cit. on pp. 3, 55, 63).

[19] L. A. DeNardo et al., *Temporal evolution of cortical ensembles promoting remote memory retrieval*, Nat Neurosci **22** (2019) 460 (cit. on pp. 3, 55, 65, 70).

[20] M. L. Feldman and C. Dowd, *Loss of dendritic spines in aging cerebral cortex*, Anat Embryol (Berl) **148** (1975) 279 (cit. on pp. 3, 55, 67, 70, 83).

[21] E. Masliah, M. Mallory, L. Hansen, R. DeTeresa, and R. Terry, *Quantitative synaptic alterations in the human neocortex during normal aging*, Neurology **43** (1993) 192 (cit. on pp. 3, 55, 67, 70, 83).

[22] P. Dayan and L. F. Abbott, *Theoretical neuroscience: computational and mathematical modeling of neural systems*, MIT press, 2005 (cit. on pp. 5, 21, 58).

[23] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski, *Neuronal dynamics: From single neurons to networks and models of cognition*, Cambridge University Press, 2014 (cit. on pp. 5, 21, 54).

[24] G. Benettin, L. Galgani, A. Giorgilli, and J.-M. Strelcyn, *Lyapunov characteristic exponents for smooth dynamical systems and for Hamiltonian systems; a method for computing all of them. Part 1: Theory*, Meccanica **15** (1980) 9 (cit. on pp. 5, 14).

[25] F. Ginelli et al., *Characterizing dynamics with covariant Lyapunov vectors*, Physical review letters **99** (2007) 130601 (cit. on pp. 5, 17, 36, 37, 52).

[26] J. C. Eccles, P. Fatt, and K. Koketsu, *Cholinergic and inhibitory synapses in a pathway from motor-axon collaterals to motoneurones*, The Journal of physiology **126** (1954) 524 (cit. on p. 7).

[27] G. L. Gerstein and K. L. Kirkland,
Neural assemblies: technical issues, analysis, and modeling, Neural Networks **14** (2001) 589 (cit. on p. 8).

[28] G. Buzsáki, *Neural syntax: cell assemblies, synapsembles, and readers*,
Neuron **68** (2010) 362 (cit. on pp. 8, 54, 57).

[29] T. V. Bliss and G. L. Collingridge,
A synaptic model of memory: long-term potentiation in the hippocampus,
Nature **361** (1993) 31 (cit. on p. 8).

[30] G.-q. Bi and M.-m. Poo, *Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type*,
Journal of neuroscience **18** (1998) 10464 (cit. on p. 8).

[31] P. J. Sjöström, G. G. Turrigiano, and S. B. Nelson,
Rate, timing, and cooperativity jointly determine cortical synaptic plasticity,
Neuron **32** (2001) 1149 (cit. on p. 8).

[32] R. C. Froemke and Y. Dan,
Spike-timing-dependent synaptic modification induced by natural spike trains,
Nature **416** (2002) 433 (cit. on p. 8).

[33] G.-Q. Bi, *Spatiotemporal specificity of synaptic plasticity: cellular rules and mechanisms*,
Biological cybernetics **87** (2002) 319 (cit. on p. 8).

[34] H.-X. Wang, R. C. Gerkin, D. W. Nauen, and G.-Q. Bi,
Coactivation and timing-dependent integration of synaptic potentiation and depression,
Nature neuroscience **8** (2005) 187 (cit. on p. 8).

[35] J.-P. Pfister and W. Gerstner, *Triplets of spikes in a model of spike timing-dependent plasticity*,
Journal of Neuroscience **26** (2006) 9673 (cit. on p. 8).

[36] R. K. Mishra, S. Kim, S. J. Guzman, and P. Jonas, *Symmetric spike timing-dependent plasticity at CA3–CA3 synapses optimizes storage and recall in autoassociative networks*,
Nat Commun **7** (2016) 1 (cit. on pp. 9, 57, 83).

[37] J. Sjöström, W. Gerstner, et al., *Spike-timing dependent plasticity*,
Spike-timing dependent plasticity **35** (2010) 0 (cit. on p. 9).

[38] W. R. Softky and C. Koch, *The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs*, Journal of neuroscience **13** (1993) 334 (cit. on p. 10).

[39] A. Renart et al., *The asynchronous state in cortical circuits*, science **327** (2010) 587
(cit. on pp. 10, 11).

[40] T. Binzegger, R. J. Douglas, and K. A. Martin,
A quantitative map of the circuit of cat primary visual cortex,
Journal of Neuroscience **24** (2004) 8441 (cit. on p. 10).

[41] B. D. Burns and A. Webb, *The spontaneous activity of neurones in the cat's cerebral cortex*,
Proceedings of the Royal Society of London. Series B. Biological Sciences **194** (1976) 211
(cit. on p. 10).

[42] W. H. Calvin and C. F. Stevens,
Synaptic noise and other sources of randomness in motoneuron interspike intervals.,
Journal of neurophysiology **31** (1968) 574 (cit. on p. 10).

[43] H. L. Bryant and J. P. Segundo,
Spike initiation by transmembrane current: a white-noise analysis.,
The Journal of physiology **260** (1976) 279 (cit. on p. 10).

[44] Z. F. Mainen and T. J. Sejnowski, *Reliability of spike timing in neocortical neurons*,
Science **268** (1995) 1503 (cit. on p. 10).

[45] T. Tetzlaff, M. Helias, G. T. Einevoll, and M. Diesmann,
Decorrelation of neural-network activity by inhibitory feedback, (2012) (cit. on p. 11).

[46] M. Helias, T. Tetzlaff, and M. Diesmann, *The correlation structure of local neuronal networks intrinsically results from recurrent dynamics*,
PLoS computational biology **10** (2014) e1003428 (cit. on p. 11).

[47] N. Brunel and V. Hakim,
Fast global oscillations in networks of integrate-and-fire neurons with low firing rates,
Neural computation **11** (1999) 1621 (cit. on p. 11).

[48] N. Brunel,
Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons,
Journal of computational neuroscience **8** (2000) 183 (cit. on pp. 11, 25, 43).

[49] Y. Shu, A. Hasenstaub, and D. A. McCormick,
Turning on and off recurrent balanced cortical activity, Nature **423** (2003) 288 (cit. on p. 11).

[50] M. Wehr and A. M. Zador,
Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex,
Nature **426** (2003) 442 (cit. on p. 11).

[51] B. Haider, A. Duque, A. R. Hasenstaub, and D. A. McCormick, *Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition*,
Journal of Neuroscience **26** (2006) 4535 (cit. on p. 11).

[52] B. V. Atallah and M. Scanziani, *Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition*, Neuron **62** (2009) 566 (cit. on p. 11).

[53] A. G. Hawkes, *Point spectra of some mutually exciting point processes*,
J R Stat Soc Series B Stat Methodol **33** (1971) 438 (cit. on pp. 12, 13, 56, 57).

[54] A. G. Hawkes, *Spectra of some self-exciting and mutually exciting point processes*,
Biometrika **58** (1971) 83 (cit. on pp. 12, 56).

[55] Y. Ogata,
Statistical models for earthquake occurrences and residual analysis for point processes,
Journal of the American Statistical association **83** (1988) 9 (cit. on p. 12).

[56] A. G. Hawkes, *Hawkes processes and their applications to finance: a review*,
Quantitative Finance **18** (2018) 193 (cit. on p. 12).

[57] G. O. Mohler, M. B. Short, P. J. Brantingham, F. P. Schoenberg, and G. E. Tita,
Self-exciting point process modeling of crime,
Journal of the American Statistical Association **106** (2011) 100 (cit. on p. 12).

[58] R. Kempter, W. Gerstner, and J. L. Van Hemmen, *Hebbian learning and spiking neurons*, Phys Rev E **59** (1999) 4498 (cit. on pp. 13, 55, 58).

[59] V. I. Oseledets,
A multiplicative ergodic theorem. Characteristic Ljapunov, exponents of dynamical systems, Trudy Moskovskogo Matematicheskogo Obshchestva **19** (1968) 179 (cit. on p. 14).

[60] M. J. Richardson and R. Swarbrick,
Firing-rate response of a neuron receiving excitatory and inhibitory synaptic shot noise, Physical review letters **105** (2010) 178102 (cit. on pp. 20, 27, 44, 81).

[61] R.-M. Memmesheimer, R. Rubin, B. P. Ölveczky, and H. Sompolinsky,
Learning precisely timed spikes, Neuron **82** (2014) 925 (cit. on pp. 20, 22, 39, 45, 82).

[62] S. Denève and C. K. Machens, *Efficient codes and balanced networks*,
Nature neuroscience **19** (2016) 375 (cit. on pp. 21, 43, 56).

[63] P. V. Kuptsov and U. Parlitz, *Theory and computation of covariant Lyapunov vectors*,
Journal of nonlinear science **22** (2012) 727 (cit. on pp. 21, 35, 44).

[64] P. Frederickson, J. L. Kaplan, E. D. Yorke, and J. A. Yorke,
The Liapunov dimension of strange attractors, Journal of differential equations **49** (1983) 185 (cit. on p. 21).

[65] R. Zillmer, R. Livi, A. Politi, and A. Torcini, *Desynchronization in diluted neural networks*,
Physical Review E **74** (2006) 036203 (cit. on pp. 21, 34, 44).

[66] R. Zillmer, N. Brunel, and D. Hansel, *Very long transients, irregular firing, and chaotic dynamics in networks of randomly connected inhibitory integrate-and-fire neurons*,
Physical Review E **79** (2009) 031909 (cit. on pp. 21, 22, 34, 44).

[67] A. Politi, R. Livi, G.-L. Oppo, and R. Kapral, *Unpredictable behaviour in stable systems*,
EPL (Europhysics Letters) **22** (1993) 571 (cit. on p. 21).

[68] A. Politi and A. Torcini, *Stable chaos*,
Nonlinear dynamics and chaos: advances and perspectives (2010) 103 (cit. on p. 21).

[69] M. Monteforte and F. Wolf,
Dynamic flux tubes form reservoirs of stability in neuronal circuits,
Physical Review X **2** (2012) 041007 (cit. on pp. 21, 23, 31, 34, 38, 44, 47, 49).

[70] M. Monteforte and F. Wolf,
Dynamical entropy production in spiking neuron networks in the balanced state,
Physical review letters **105** (2010) 268104 (cit. on pp. 21, 22, 31, 34, 43, 44, 47).

[71] R. Tremblay, S. Lee, and B. Rudy,
GABAergic interneurons in the neocortex: from cellular properties to circuits,
Neuron **91** (2016) 260 (cit. on p. 22).

[72] D. V. Buonomano and M. M. Merzenich, *Temporal information transformed into a spatial code by a neural network with realistic properties*, Science **267** (1995) 1028 (cit. on p. 22).

[73] P. F. Dominey, *Complex sensory-motor sequence learning based on recurrent state representation and reinforcement learning*, Biological cybernetics **73** (1995) 265
(cit. on p. 22).

[74] H. Jaeger and H. Haas, *Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication*, science **304** (2004) 78 (cit. on pp. 22, 38).

[75] W. Maass, T. Natschläger, and H. Markram, *Real-time computing without stable states: A new framework for neural computation based on perturbations*, Neural computation **14** (2002) 2531 (cit. on pp. 22, 38, 45).

[76] M. B. Westover, C. Eliasmith, and C. H. Anderson, *Linearly decodable functions from neural population codes*, Neurocomputing **44** (2002) 691 (cit. on p. 22).

[77] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, nature **521** (2015) 436 (cit. on p. 22).

[78] D. Sussillo and L. F. Abbott, *Generating coherent patterns of activity from chaotic neural networks*, Neuron **63** (2009) 544 (cit. on p. 22).

[79] F. Ponulak and A. Kasiński, *Supervised learning in spiking neural networks with ReSuMe: sequence learning, classification, and spike shifting*, Neural computation **22** (2010) 467 (cit. on p. 22).

[80] R. V. Florian, *The chronotron: A neuron that learns to fire temporally precise spike patterns*, (2012) (cit. on p. 22).

[81] A. Mohemmed, S. Schliebs, S. Matsuda, and N. Kasabov, *Span: Spike pattern association neuron for learning spatio-temporal spike patterns*, International journal of neural systems **22** (2012) 1250012 (cit. on p. 22).

[82] Y. Xu, X. Zeng, and S. Zhong, *A new supervised learning algorithm for spiking neurons*, Neural computation **25** (2013) 1472 (cit. on p. 22).

[83] C. Albers, M. Westkott, and K. Pawelzik, *Learning of precise spike times with homeostatic membrane potential dependent synaptic plasticity*, PloS one **11** (2016) e0148948 (cit. on p. 22).

[84] F. Zenke and S. Ganguli, *Superspike: Supervised learning in multilayer spiking neural networks*, Neural computation **30** (2018) 1514 (cit. on pp. 22, 45).

[85] D. Huh and T. J. Sejnowski, *Gradient descent for spiking neural networks*, Advances in neural information processing systems **31** (2018) (cit. on pp. 22, 45).

[86] S. Olmi, D. Angulo-Garcia, A. Imparato, and A. Torcini, *Exact firing time statistics of neurons driven by discrete inhibitory noise*, Scientific reports **7** (2017) 1 (cit. on pp. 23, 27, 44, 81).

[87] R. E. Mirollo and S. H. Strogatz, *Synchronization of pulse-coupled biological oscillators*, SIAM Journal on Applied Mathematics **50** (1990) 1645 (cit. on p. 23).

[88] R.-M. Memmesheimer and M. Timme, *Designing complex networks*, Physica D: Nonlinear Phenomena **224** (2006) 182 (cit. on p. 23).

[89] P. Goel and B. Ermentrout, *Synchrony, stability, and firing patterns in pulse-coupled oscillators*, Physica D: Nonlinear Phenomena **163** (2002) 191 (cit. on p. 24).

[90] R. M. Smeal, G. B. Ermentrout, and J. A. White,
Phase-response curves and synchronized neural networks,
Philosophical Transactions of the Royal Society B: Biological Sciences **365** (2010) 2407 (cit. on p. 24).

[91] A. Viriyopase, R.-M. Memmesheimer, and S. Gielen, *Analyzing the competition of gamma rhythms with delayed pulse-coupled oscillators in phase representation*,
Physical Review E **98** (2018) 022217 (cit. on p. 24).

[92] C. van Vreeswijk and H. Sompolinsky, *Chaotic balanced state in a model of cortical circuits*,
Neural computation **10** (1998) 1321 (cit. on p. 25).

[93] L. P. Kadanoff, *More is the same; phase transitions and mean field theories*,
Journal of Statistical Physics **137** (2009) 777 (cit. on p. 25).

[94] D. Breuer, M. Timme, and R.-M. Memmesheimer,
Statistical physics of neural systems with nonadditive dendritic coupling,
Physical Review X **4** (2014) 011053 (cit. on p. 25).

[95] J. Schuecker, S. Goedeke, and M. Helias,
Optimal sequence memory in driven random networks, Physical Review X **8** (2018) 041029 (cit. on p. 25).

[96] H. C. Tuckwell, *Nonlinear and stochastic theories*,
Introduction to Theoretical Neurobiology. **2** (1988) (cit. on pp. 27, 44, 81).

[97] A. N. Burkitt,
A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input,
Biological cybernetics **95** (2006) 1 (cit. on pp. 27, 44, 81).

[98] D. Angulo-Garcia, S. Luccioli, S. Olmi, and A. Torcini,
Death and rebirth of neural activity in sparse inhibitory networks,
New Journal of Physics **19** (2017) 053011 (cit. on p. 27).

[99] J. K. Hale and S. M. V. Lunel, *Introduction to functional differential equations*, vol. 99,
Springer Science & Business Media, 2013 (cit. on p. 27).

[100] R. Engelken, *Chaotic neural circuit dynamics*,
PhD thesis: Georg-August-Universität Göttingen, 2017 (cit. on pp. 34, 44).

[101] B. Kramer and A. MacKinnon, *Localization: theory and experiment*,
Reports on Progress in Physics **56** (1993) 1469 (cit. on p. 37).

[102] S. Maren, K. L. Phan, and I. Liberzon,
The contextual brain: implications for fear conditioning, extinction and psychopathology,
Nature reviews neuroscience **14** (2013) 417 (cit. on p. 38).

[103] V. Mante, D. Sussillo, K. V. Shenoy, and W. T. Newsome,
Context-dependent computation by recurrent dynamics in prefrontal cortex,
nature **503** (2013) 78 (cit. on p. 38).

[104] B. Kriener, T. Tetzlaff, A. Aertsen, M. Diesmann, and S. Rotter,
Correlations and population dynamics in cortical networks,
Neural computation **20** (2008) 2185 (cit. on p. 43).

[105] A. A. Prinz, *Insights from models of rhythmic motor systems*, Current Opinion in Neurobiology **16** (2006) 615 (cit. on p. 43).

[106] C. Savin, I. Ignat, and R. C. Muresan, "Heterogeneous networks of spiking neurons: self-sustained activity and excitability," *Proceedings of 2nd IEEE Conference on Intelligent Computer Communication and Processing (ICCP)*, 2006 (cit. on p. 43).

[107] 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*, Proceedings of the National Academy of Sciences **101** (2004) 1368 (cit. on p. 43).

[108] A. Litwin-Kumar, R. Rosenbaum, and B. Doiron, *Inhibitory stabilization and visual coding in cortical circuits with multiple interneuron subtypes*, Journal of neurophysiology **115** (2016) 1399 (cit. on p. 43).

[109] R.-M. Memmesheimer and M. Timme, *Stable and unstable periodic orbits in complex networks of spiking neurons with delays*, Discrete & Continuous Dynamical Systems **28** (2010) 1555 (cit. on p. 44).

[110] Q.-L. L. Gu, Z.-Q. K. Tian, G. Kovačič, D. Zhou, and D. Cai, *The dynamics of balanced spiking neuronal networks under Poisson drive is not chaotic*, Frontiers in Computational Neuroscience **12** (2018) 47 (cit. on p. 44).

[111] S. Luccioli, S. Olmi, A. Politi, and A. Torcini, *Collective dynamics in sparse networks*, Physical review letters **109** (2012) 138103 (cit. on p. 44).

[112] G. Lajoie, J.-P. Thivierge, and E. Shea-Brown, *Structured chaos shapes spike-response noise entropy in balanced neural networks*, Frontiers in computational neuroscience **8** (2014) 123 (cit. on p. 44).

[113] E. Ullner and A. Politi, *Self-sustained irregular activity in an ensemble of neural oscillators*, Physical Review X **6** (2016) 011015 (cit. on p. 44).

[114] S. Coombes, "Chaos in integrate-and-fire dynamical systems," *AIP Conference Proceedings*, vol. 502, 1, American Institute of Physics, 2000 88 (cit. on p. 44).

[115] X. Lagorce and R. Benosman, *Stick: spike time interval computational kernel, a framework for general purpose computation using neurons, precise timing, delays, and synchrony*, Neural computation **27** (2015) 2261 (cit. on p. 45).

[116] S. J. Verzi et al., *Computing with spikes: The advantage of fine-grained timing*, Neural computation **30** (2018) 2660 (cit. on p. 45).

[117] M. Pfeiffer and T. Pfeil, *Deep learning with spiking neurons: opportunities and challenges*, Frontiers in neuroscience (2018) 774 (cit. on p. 45).

[118] D. Thalmeier, M. Uhlmann, H. J. Kappen, and R.-M. Memmesheimer, *Learning universal computations with spikes*, PLoS computational biology **12** (2016) e1004895 (cit. on p. 45).

[119] L. F. Abbott, B. DePasquale, and R.-M. Memmesheimer, *Building functional networks of spiking model neurons*, Nature neuroscience **19** (2016) 350 (cit. on p. 45).

[120] W. Nicola and C. Clopath,
Supervised learning in spiking neural networks with FORCE training,
Nature communications **8** (2017) 1 (cit. on p. 45).

[121] B. DePasquale, C. J. Cueva, K. Rajan, G. S. Escola, and L. Abbott,
full-FORCE: A target-based method for training recurrent networks,
PloS one **13** (2018) e0191527 (cit. on p. 45).

[122] H. Jaeger, *The "echo state" approach to analysing and training recurrent neural networks-with an erratum note*, Bonn, Germany: German National Research Center for Information Technology GMD Technical Report **148** (2001) 13 (cit. on p. 45).

[123] F. Zenke, E. J. Agnes, and W. Gerstner, *Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks*, Nat Commun **6** (2015) 1 (cit. on pp. 53–55, 68, 69).

[124] L. Montangie, C. Miehl, and J. Gjorgjieva,
Autonomous emergence of connectivity assemblies via spike triplet interactions,
PLoS Comput Biol **16** (2020) e1007835 (cit. on pp. 53–55, 57, 59, 68, 69).

[125] M. A. Triplett, L. Avitan, and G. J. Goodhill, *Emergence of spontaneous assembly activity in developing neural networks without afferent input*, PLoS Comput Biol **14** (2018) e1006421 (cit. on pp. 53–55, 59, 68).

[126] J. Herpich and C. Tetzlaff,
Principles underlying the input-dependent formation and organization of memories,
Netw Neurosci **3** (2019) 606 (cit. on pp. 53–55, 68).

[127] Y.-Y. Huang, A. Colino, D. K. Selig, and R. C. Malenka,
The influence of prior synaptic activity on the induction of long-term potentiation,
Science **255** (1992) 730 (cit. on pp. 53, 55, 68).

[128] 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 **391** (1998) 892 (cit. on pp. 53, 55, 68).

[129] K. Ibata, Q. Sun, and G. G. Turrigiano,
Rapid synaptic scaling induced by changes in postsynaptic firing, Neuron **57** (2008) 819 (cit. on pp. 53, 55, 68).

[130] F. Zenke and W. Gerstner,
Hebbian plasticity requires compensatory processes on multiple timescales,
Philos Trans R Soc Lond B Biol Sci **372** (2017) 20160259 (cit. on pp. 53, 55, 68).

[131] A. Holtmaat and P. Caroni,
Functional and structural underpinnings of neuronal assembly formation in learning,
Nat Neurosci **19** (2016) 1553 (cit. on p. 54).

[132] T. Pietri et al., *The emergence of the spatial structure of tectal spontaneous activity is independent of visual inputs*, Cell Rep **19** (2017) 939 (cit. on p. 54).

[133] 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 **334** (2011) 1569 (cit. on pp. 54, 69).

[134] G. K. Ocker and B. Doiron,
Training and spontaneous reinforcement of neuronal assemblies by spike timing plasticity,
Cereb Cortex **29** (2019) 937 (cit. on pp. 54, 55, 57, 69).

[135] C. Tetzlaff, C. Kolodziejski, M. Timme, and F. Wörgötter, *Synaptic scaling in combination with many generic plasticity mechanisms stabilizes circuit connectivity*,
Front Comput Neurosci **5** (2011) 47 (cit. on p. 55).

[136] 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)
(cit. on pp. 55, 57).

[137] M. J. Fauth and M. C. van Rossum,
Self-organized reactivation maintains and reinforces memories despite synaptic turnover,
Elife **8** (2019), URL: https://doi.org/10.7554/elife.43717 (cit. on pp. 55, 69).

[138] V. Pernice, B. Staude, S. Cardanobile, and S. Rotter,
How Structure Determines Correlations in Neuronal Networks,
PLoS Comput Biol **7** (2011) e1002059, ed. by O. Sporns,
URL: https://doi.org/10.1371/journal.pcbi.1002059 (cit. on pp. 55–58).

[139] Y. F. Kalle Kossio, S. Goedeke, B. van den Akker, B. Ibarz, and R.-M. Memmesheimer,
Growing Critical: Self-Organized Criticality in a Developing Neural System,
Phys Rev Lett **121** (5 2018) 058301,
URL: https://link.aps.org/doi/10.1103/PhysRevLett.121.058301
(cit. on pp. 55, 60).

[140] R. Yun, J. H. Mishler, S. I. Perlmutter, and E. E. Fetz, *Paired stimulation for spike-timing dependent plasticity quantified with single neuron responses in primate motor cortex*,
bioRxiv (2022) (cit. on pp. 57, 83).

[141] T. E. Harris et al., *The theory of branching processes*, vol. 6, Springer Berlin, 1963
(cit. on p. 60).

[142] J. M. Beggs and D. Plenz, *Neuronal avalanches in neocortical circuits*,
J Neurosci **23** (2003) 11167 (cit. on p. 60).

[143] C. Gastaldi, T. Schwalger, E. De Falco, R. Q. Quiroga, and W. Gerstner,
When shared concept cells support associations: Theory of overlapping memory engrams,
PLoS Comput Biol **17** (2021) e1009691 (cit. on pp. 63, 64, 69, 83).

[144] Y. Dan and M.-m. Poo, *Spike timing-dependent plasticity of neural circuits*,
Neuron **44** (2004) 23 (cit. on p. 68).

[145] N. Caporale and Y. Dan, *Spike timing–dependent plasticity: a Hebbian learning rule*,
Annu Rev Neurosci **31** (2008) 25 (cit. on p. 68).

[146] N. E. Ziv and N. Brenner,
Synaptic tenacity or lack thereof: spontaneous remodeling of synapses,
Trends Neurosci **41** (2018) 89 (cit. on pp. 68, 83).

[147] R. Kempter, W. Gerstner, and J. L. v. Hemmen,
Intrinsic stabilization of output rates by spike-based Hebbian learning,
Neural Comput **13** (2001) 2709 (cit. on p. 69).

[148] G. Buzsáki and K. Mizuseki,
The log-dynamic brain: how skewed distributions affect network operations,
Nat Rev Neurosci **15** (2014) 264, URL: https://doi.org/10.1038/nrn3687
(cit. on p. 69).

[149] J. Spaniol and C. Grady, *Aging and the neural correlates of source memory: over-recruitment and functional reorganization*, Neurobiol Aging **33** (2012) 425 (cit. on p. 70).

[150] A. Folville, J. S. Simons, A. D'Argembeau, and C. Bastin, *I remember it like it was yesterday: Age-related differences in the subjective experience of remembering*, Psychon Bull Rev (2021), URL: https://doi.org/10.3758/s13423-021-02048-y (cit. on p. 70).

[151] B. Levine, E. Svoboda, J. F. Hay, G. Winocur, and M. Moscovitch,
Aging and autobiographical memory: Dissociating episodic from semantic retrieval.,
Psychol Aging **17** (2002) 677, URL: https://doi.org/10.1037/0882-7974.17.4.677
(cit. on p. 70).

[152] W. C. Abraham and A. Robins,
Memory retention–the synaptic stability versus plasticity dilemma,
Trends in neurosciences **28** (2005) 73 (cit. on p. 83).

[153] M. Mermillod, A. Bugaiska, and P. Bonin, *The stability-plasticity dilemma: Investigating the continuum from catastrophic forgetting to age-limited learning effects*, 2013 (cit. on p. 83).

[154] S. Grossberg, *Competitive learning: From interactive activation to adaptive resonance*,
Cognitive science **11** (1987) 23 (cit. on p. 83).

[155] G. A. Carpenter and S. Grossberg,
ART 2: Self-organization of stable category recognition codes for analog input patterns,
Applied optics **26** (1987) 4919 (cit. on p. 83).

[156] R. M. French,
Pseudo-recurrent connectionist networks: An approach to the'sensitivity-stability'dilemma,
Connection Science **9** (1997) 353 (cit. on p. 83).