# Mean-field theory for complex neuronal networks

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Master of Science
Jannis Schücker
aus Paderborn

Berichter: Universitätsprofessor Dr. rer. nat. Carsten Honerkamp
Juniorprofessor Dr. rer. nat. Moritz Helias

Tag der mündlichen Prüfung: 14.06.2017

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

# Zusammenfassung


### Mean-field Theorie für komplexe neuronale Netzwerke

Ein wichtiges Ziel der Neurowissenschaft besteht darin, die Funktionsweise des Gehirns anhand der neuronalen Aktivität zu verstehen. Das Gehirn zeigt eine Dynamik auf verschiedenen räumlichen und zeitlichen Skalen. Diese dynamische Komplexität entsteht aus einer Wechselwirkung zwischen unzähligen Einzelkomponenten, den Neuronen. Daher ist es wichtig grundsätzlich zu verstehen, wie kollektive Dynamik aus der Dynamik einzelner Neurone sowie deren Verbindungen hervorgeht. Hierbei bezeichnet man die Verbindungen zwischen den Neuronen als strukturelle Konnektivität.

Um die Dynamik besser zu verstehen, werden innerhalb der theoretischen Neurowissenschaft analytische Methoden entwickelt, die die Komplexität der Dynamik durch Mittlung über räumliche oder zeitliche Skalen reduzieren. Diese so genannte Mean-field Theorie drückt die Netzwerk-Aktivität als mathematische Funktion der Eigenschaften einzelner Neuronen sowie der strukturellen Konnektivität aus. Diese mathematische Beziehung ermöglicht ein systematisches Verständnis der Dynamik. In dieser Arbeit entwickeln wir die Mean-field Theorie sowohl auf der Ebene der einzelnen Neurone als auch auf der Ebene der strukturellen Konnektivität weiter.

Eine zentrale mathematische Größe der Mean-field Theorie ist die Übertragungsfunktion, die charakterisiert, wie ein Neuron eine Modulation des Eingangssignals auf eine Modulation seines Ausgangssignals überträgt. Ein Neuron im Gehirn erhält Signale von einer großen Anzahl von präsynaptischen Neuronen. Diese große Anzahl verursacht effektiv ein verrauschtes Eingangssignal, welches durch die synaptische Übertragung zwischen den Neuronen tiefpassgefiltert wird. Diese Filterung führt zu einem farbigen Rauschen, welches die mathematische Bestimmung der Übertragungsfunktion schwierig macht. Wir entwickeln ein allgemeines Verfahren, um ein System mit farbigem Rauschen auf ein einfacheres System mit weißem Rauschen zu reduzieren, wobei die Farbe des Rauschens durch effektive Randbedingungen berücksichtigt wird. Wir wenden diesen Formalismus auf eines der am häufigsten verwendeten Modelle an und leiten einen neuartigen analytischen Ausdruck für dessen Übertragungsfunktion her. Wir zeigen weiterhin, wie dieses Ergebnis zum Verständnis dynamischer Fluktuationen in einem Netzwerk Modell beiträgt.

Auf der Ebene der strukturellen Konnektivität untersuchen wir den Einfluss von Verbindungen zwischen verschiedenen Teilen des Gehirns auf die Stabilität der Netzwerkaktivität. Dazu

entwickeln wir eine Mean-field basierte Methode, die es ermöglicht, den Phasenraum eines Netzwerk Modells durch systematische Verfeinerung seiner experimentell bestimmten Struktur zu manipulieren. Dabei stellen wir grundlegende Anforderungen an die Aktivität, die weder verschwindend gering noch unrealistisch hoch sein darf. Diese Anforderungen sind ausreichend, um eine realistische Aktivität im Modell zu erhalten. Somit trägt die Methode dazu bei, die experimentellen Daten in das Modell zu integrieren, indem sie die empirische Netzwerkstruktur auf eine Realisierung einschränkt, die mit physiologischen Experimenten kompatibel ist. Darüber hinaus stellt die dynamische Mean-field Theorie eine Beziehung zwischen der strukturellen Konnektivität und Fluktuationen in der Aktivität der Neuronen her. Wir erweitern die dynamische Mean-field Theorie auf stochastische Systeme, bei denen das Rauschen den Spike-Emissionsprozess von Neuronen repräsentiert. Dies ermöglicht es uns, die selbst-konsistente Statistik der Aktivität zu bestimmen. Darüber hinaus leiten wir den Transitions-Punkt her, an dem das neuronale Netzwerk einen Übergang zur chaotischen Dynamik zeigt. Wir finden heraus, dass das Rauschen die chaotische Fluktuationen durch einen dynamischen Mechanismus unterdrückt, der effektiv die Empfindlichkeit des Netzwerks auf kleine Störungen reduziert.

Die quantitative Validierung der verschiedenen Mean-field Ansätze erfordert Simulationen von zwei Modellklassen, den raten- und den spike-basierten Neuron-Modellen. Hierzu entwickeln wir eine vereinheitlichende Simulationsumgebung, die die Mean-field Validierung vereinfacht und deren Zuverlässigkeit erhöht.

# Summary


### Mean-field theory for complex neuronal networks

Understanding the working principles of the brain constitutes the major challenge in computational neuroscience. Reaching this goal is particularly difficult since the brain displays complex dynamics on various spatial and temporal scales, emerging from the interaction between a tremendous number of individual components, the neurons. Therefore a crucial step is to understand how collective dynamics in the brain arises from single neurons dynamics and the connectivity between the neurons. To better understand the collective dynamics in network of neurons, theoretical neuroscience has developed analytical tools, which reduce the complexity of the dynamics by averaging over spatial or temporal scales. In particular, this so called mean-field theory relates single neuron properties and connectivity to network activity, which allows for a systematic understanding of the emerging activity.

In this thesis we advance mean-field theory both on the level of single neuron properties and on the level of network structure.

One building block of mean-field theory is the transfer function characterizing how a modulation of the neuron's input is transmitted to a modulation of its output. A Neuron in the brain receives signals from a large number of presynaptic neurons, effectively causing a noisy input. Realistic activation kinetics of synaptic transmission amounts to a low-pass filtering of this input. Formally, the mathematical analysis of the transfer function therefore leads to a difficult colored-noise problem. We develop a general method to reduce this colored-noise to a white-noise system by capturing the color of the noise by effective boundary conditions. We apply this formalism to one of the most used neuron models, revealing a novel analytical expression for its transfer function. We further show how these results contribute to the understanding of fluctuations in a network of model neurons.

On the level of structural connectivity we investigate the impact of connections between different parts of the brain on the stability of the network activity. To this end, we extend mean-field theory to devise a method allowing us to shape the phase space of a large-scale model by systematically refining its experimentally obtained connectivity map. Fundamental constraints on the activity, i.e., prohibiting quiescence and requiring global stability, prove sufficient to obtain realistic activity. Thus, the method contributes to the data integration process by constraining the empirical connectivity map to a realization that is compatible with physiological experiments.

Moreover, dynamical mean-field theory relates connectivity to network generated fluctuations in the activity. We extend dynamical mean-field theory to stochastic systems where the noise mimics the spike emission process of neurons. This allows us to derive the self-consistent statistics of the activity as well as the point where the neural network displays a transition to chaos. We find that noise suppresses chaotic fluctuations by a dynamic mechanism effectively reducing the sensitivity of the network to small perturbations.

The quantitative validation of the various mean-field approaches requires simulations of two model classes, i.e., rate-based and spike-based model neurons. We develop a unified simulation framework supporting both models, which facilitates this validation and increases reliability.

# Contents

**Zusammenfassung** **i**

**Summary** **iii**

**I Introduction** **1**

1.1 Scope of the thesis . . . 3
1.2 Fundamentals . . . 3
1.3 Mean-field theory and structure of this thesis . . . 14

**II Mean-field theory for spiking dynamics** **19**

**2 Modulated escape from a metastable state driven by colored noise** **21**

2.1 Introduction . . . 22
2.2 Reduction from colored to white noise . . . 23
2.3 LIF-neuron . . . 32
2.4 Interim discussion . . . 39

**3 Stationary activity in spiking neuronal networks** **41**

3.1 Mean-field solution . . . 41
3.2 Application to balanced random network . . . 45
3.3 Application to the microcircuit model . . . 47

**4 Stationary activity and global stability in complex networks** **51**

4.1 Introduction . . . 52
4.2 Results . . . 54
4.3 Methods . . . 69
4.4 Interim discussion . . . 74

**5 Fluctuations around stationary states** **77**

5.1 Application to balanced random network . . . 78
5.2 Application to complex spiking networks . . . 80

**III Dynamic mean-field theory for stochastic rate dynamics 83**

**6 Breakdown of mean-field theory for spiking dynamics 85**

**7 Noise dynamically suppresses chaos in random neural networks 89**
7.1 Introduction . . . 90
7.2 Dynamic Mean-field equation . . . 90
7.3 Effective motion of particle in a potential . . . 92
7.4 Effect of noise on the transition to chaos . . . 95

**IV Simulation of rate-based neuron models 101**

**8 Integration of rate-models in a spiking neural network simulator 103**
8.1 Introduction . . . 104
8.2 Methods . . . 106
8.3 Results . . . 117
8.4 Interim discussion . . . 127

**V Discussion 131**

**9 Summary and conclusions 133**

**10 Outlook 141**

**VI Appendices 143**

**A Supplementary calculations for Fokker-Planck theory 145**
A.1 N-dimensional Fokker-Planck equation . . . 145
A.2 White-noise stationary rate . . . 148
A.3 White-noise transfer function . . . 149
A.4 Modulation of noise amplitude . . . 157

**B Supplementary calculations for dynamical mean-field theory 161**
B.1 Derivation of mean-field equation . . . 161
B.2 Stationary process . . . 165
B.3 Assessing chaos by a pair of identical systems . . . 165

**C Simulation of rate-based neuron models 171**
C.1 Numerical evaluation of the Siegert formula . . . 171
C.2 Usage of the NEST reference implementation . . . 172

**Bibliography 175**

**Author's list of publications** **i**

**Acknowledgment** **iii**

**Funding** **v**

# Part I

# Introduction

## 1.1 Scope of the thesis

The subject of computational neuroscience is to decipher the working principles of the brain. This goal is particularly difficult, since the brain constitutes one of the most complex systems we know: Its dynamics arises from an interaction between a tremendous number of individual components, the neurons.

One route towards understanding this system is to formulate models which describe the dynamics of the brain in terms of mathematical equations. Bottom-up approaches aim to model the individual neurons and investigate the emerging collective dynamics. However, as the brain displays dynamics on various spatial and temporal scales, also the emerging dynamics in the models is difficult to understand. To better assess the collective dynamics, theoretical neuroscience has developed analytical tools that reduce the complexity of neuronal networks. An important subgroup of these use mean-field theory, a well-known concept from statistical physics.

For networks that model a canonical building block of cortical networks, these existing tools can explain the dynamics. However, it is well known that brain function emerges from an interplay between different spatial scales in the brain. Therefore, large-scale models have been developed over the past years which begs the question whether these models can still be understood within the existing theoretical framework.

The scope of this study is to extend and adapt the current theoretical tools to apply them to the biologically complex and more realistic network models. In particular, we present a theoretical method which not only explains the network steady-state activity but moreover reduces uncertainties in the models' parameters, which are inevitable due to its complexity. Thus, we show how a symbiosis between modeling and analytical theory contributes to the development of large-scale models. In addition, we develop further the analytical tools which access the temporal fluctuations in the activity of the network, which are believed to be important for brain function. Together, this thesis helps to understand the dynamics in large-scale models which in turn can help to understand the working principles of the brain. Finally, we investigate in which regime the theoretical predictions break down due to a violation of the assumptions made. Subsequently, we develop the first step to overcome this mismatch and thereby lay the foundation for future theoretical developments.

## 1.2 Fundamentals

This section presents background information introducing fundamental terms which are used throughout the thesis. In particular, we explain modeling approaches to neurons, synapses and neural networks.

### 1.2.1 Neurons and synapses

*Neurons* constitute the most important cells in the brain, forming the substrate for information processing. The basic structures of each neuron are its cell body, the *soma*, as well as the *dendrites* by which they receive input from other neurons and the *axon* through which they send output to other neurons (Fig. 1.1).

The *membrane potential* is the voltage difference between the inside of the cell and the surrounding extracellular medium. In general, the value of the membrane potential changes due to currents which flow through membrane-spanning *ion channels*. The channels are specialized to certain ions among which the most prominent ones are sodium (Na), potassium (K), calcium (Ca) and chloride (Cl). An inward directed current depolarizes the neuron while an outward-directed current has a hyper-polarizing effect.

Under resting condition, i.e., without input from other neurons, the membrane potential takes a value of about $-70$ mV, which results from an equilibrium between electrical and chemical forces acting on the ions (Dayan & Abbott, 2001). The chemical force is caused by concentration gradients of the different ion species, which are maintained by so-called ion pumps.

When the neuron is depolarized above a certain *threshold* it generates an *action potential* or *spike*, a short ($\approx 1$ ms) departure in the membrane potential with an amplitude of around 100 mV. In the *refractory period*, the time-span directly after the spike, the neuron is not excitable and thus not able to generate another action potential. Typically, a neuron emits a low number of spikes (Sec. 1.2.3) and their frequency is called *firing rate.*

The spike propagates along the axon and eventually arrives at the *synapses*, contact points between the pre-synaptic axon and the post-synaptic dendrites with a synaptic cleft in between (Fig. 1.1). On the pre-synaptic side, the voltage transient of the action potential causes the release of neurotransmitters into the cleft which in turn bind to receptors on the post-synaptic side. These receptors open ion channels for a certain period changing the channel's *conductance* which results in the *post-synaptic current* (PSC). The corresponding change in the membrane potential of the post-synaptic neuron is termed *post-synaptic potential* (PSP). The time span between the emission of a spike and the evoked PSP is called the *transmission delay.*

Depending on the type of the released neurotransmitter, a spike of a presynaptic neuron can either excite or inhibit the post-synaptic neuron, i.e., hyperpolarize or depolarize its membrane potential. Dale's principle states that one neuron only releases one neurotransmitter type so that neurons can be classified into excitatory ($\mathcal{E}$) and inhibitory ($\mathcal{I}$) cells. Beside this high-level classification, neurons feature a large diversity with respect to their morphology as well as their dynamical properties (Markram et al., 2015). For instance, most of the excitatory neurons are *pyramidal cells,* where the cell body has a characteristic triangular shape. Typical inhibitory neurons are *spiny stellate cells*, where the name originates from their star-like shape, with dendrites extending in all directions.

While the synapses described above work on a chemical basis, there also exist electrical

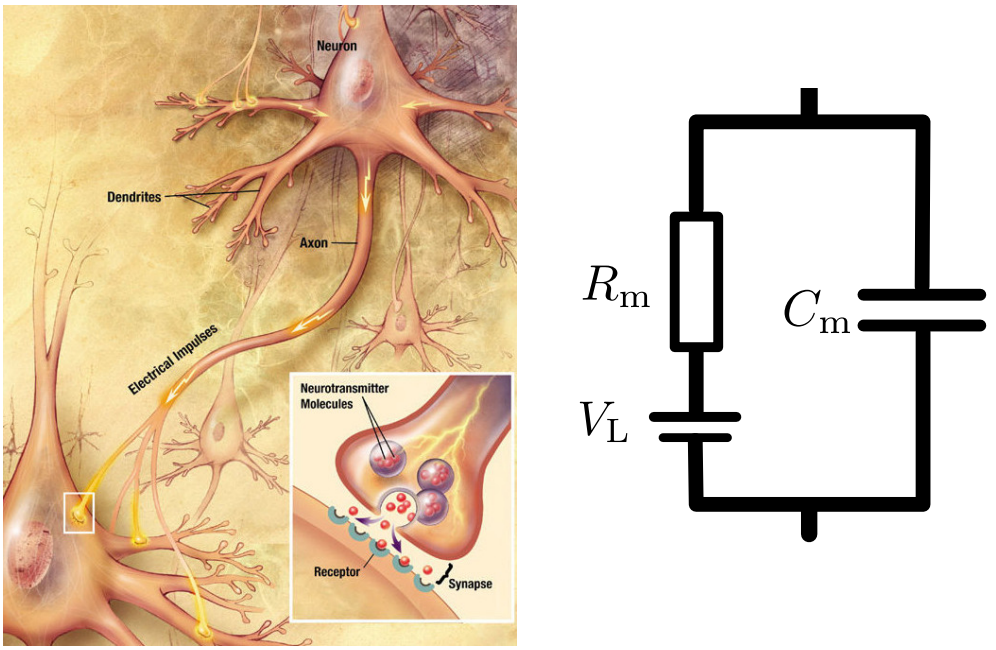



**Figure 1.1: Neurons and synapses.** **Left** Sketch of the most important structures of neurons and of chemical synapses (inset) (source: National Institute of Aging, National Institutes of Health, http://www.nia.nih.gov/alzheimers/publication/alzheimers-disease-unraveling-mystery/preface, in public domain.). **Right** Sketch of the $R-C$ circuit corresponding to the LIF neuron model

synapses, which form a direct contact between two cells continuously coupling the cells' membrane potentials. These synapses are called gap junctions and are found in various parts of the brain (Connors & Long, 2004).

### 1.2.2 Modeling of neurons and synapses

The basic idea of neuron models is to describe neurons and the communication between them with mathematical equations. They strongly vary in their complexity. Multi-compartment models take spatial properties of the cell into account and divide the morphology into small coupled compartments, for each of which a differential equation is solved. In contrast, single-compartment models collapse the morphology into a single point describing the membrane potential with a single variable. An overview over the different neuron models is for example given in Gerstner & Kistler (2002). Here we focus on point-neuron models, which are used throughout the thesis.

One of the most prominent examples from the point-neuron class is the leaky integrate-and-fire (LIF) neuron model, which dates back to the work by Lapicque (1907). As the name already indicates, the basic modeling idea is that the membrane potential integrates input and the neuron fires a spike when the potential hits a threshold. This is elaborated in detail in the following.

First, the neuron dynamics is reduced to an equivalent electrical circuit of a capacitance and a resistance in parallel (Fig. 1.1). The capacitor represents the ability of the neuron to store electrical charges on the two opposed sides of the isolating membrane, while the resistance represents the ion channels. The equation for the membrane potential $V$ reads

$$\tau_{\mathrm{m}}\dot{V} \quad = \quad -(V - E_{\mathrm{L}}) + R_{\mathrm{m}}I(t) + R_{\mathrm{m}}I_{\mathrm{const}}\,, \tag{1.1}$$

where the dot denotes the temporal derivative, $\frac{dV}{dt} =: \dot{V}$, and $E_{\mathrm{L}}$ represents the resting potential. The linear term in the brackets is called leak term and models the constant leakage of current through the ion channels. This term originates from the assumption that the conductances of the individual channels are constant over time and independent of the voltage and thus can be lumped together into an effective resistance $R_{\mathrm{m}}$ (Dayan & Abbott, 2001). $I_{\mathrm{const}}$ mimics an experimental condition where a constant current is injected into the cell. If it is the only input to the neuron ($I = 0$), the membrane potential follows the charging curve of an $R$-$C$ circuit. The membrane time constant $\tau_{\mathrm{m}}$ characterizes how fast the membrane is charged and the membrane resistance $R_{\mathrm{m}}$ together with the amplitude of the current determine the asymptotic value for the membrane potential.

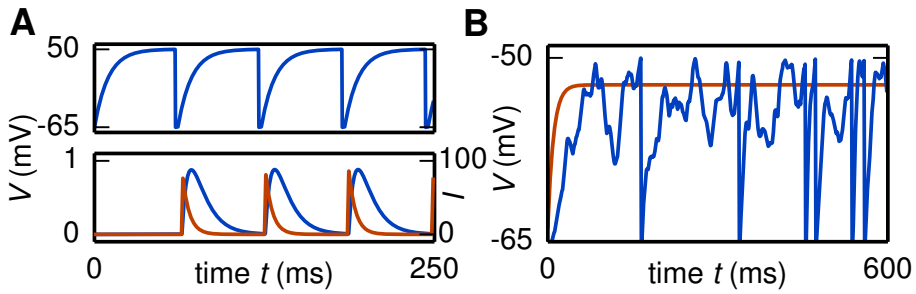



**Figure 1.2: Dynamics and interaction in the LIF neuron model.** **A** Upper panel: Voltage trace of a pre-synaptic LIF neuron model driven with super-threshold constant input current $I_{\mathrm{const}} = 378$ pA. The neuron is connected to a post-synaptic neuron model, from which the voltage- (blue) and current-trace (red) are shown in the lower panel. Neuron and synapse parameters: $\tau_{\mathrm{m}} = 10\,\mathrm{ms}$, $\tau_{\mathrm{s}} = 5\,\mathrm{ms}$, $\tau_{\mathrm{r}} = 2\,\mathrm{ms}$, $R_{\mathrm{m}} = 40 \cdot\,\mathrm{M\Omega}$, $V_\theta = -50\,\mathrm{mV}$, $V_{\mathrm{R}} = -65\,\mathrm{mV}$, $V_l = -65\,\mathrm{mV}$, $d = 5\,\mathrm{ms}$, $w = -87\,\mathrm{pA}$. **B** Voltage trace of a model neuron driven by sub-threshold constant input $I_{\mathrm{const}} = 320\,\mathrm{pA}$ (red) and of a model neuron (blue) driven by a spike train drawn from a Poisson point process with rate 729 spikes/s, chosen so that the mean of the input equals 320 pA as before. Other parameters as in A.

In biology, the time scale of the action potential is fast compared to the time scale of the sub-threshold dynamics. Therefore, the LIF model does not explicitly describe the time course of the action potential but rather replaces it by the fire-and-reset rule: At the time points $t^k$ at which the membrane potential hits the threshold $V_\theta$ the neuron emits a spike and $V$ is reset to the reset potential $V \leftarrow V_{\mathrm{R}}$ where it is clamped for the refractory period $\tau_{\mathrm{r}}$. The charging

curve as well as the threshold crossing are illustrated in (Fig. 1.2A, upper panel).

In the following we describe how the emitted spike affects the synaptic current $I$ in Eq. (1.1) for the post-synaptic neuron. In general, synaptic currents depend on the membrane potential and conductance-based neuron models, for instance the Hodgkin-Huxley model neuron (Hodgkin & Huxley, 1952), explicitly expresses this dependency by gating variables. In contrast, current-based models neglect the dependency on $V$ and instead assume a stereotypical time course for the PSC. We here loosely follow the notation introduced in Fourcaud & Brunel (2002).

The simplest form is a $\delta$-shaped PSC which follows from

$$R_{\mathrm{m}} I(t) = \tau_{\mathrm{m}} \tilde{J} \delta(t - t^k) \,,$$

where $\delta$ denotes the delta-distribution. It causes an instantaneous jump in the membrane potential of size $\tilde{J}$, which denotes the *synaptic efficacy* in volt, also termed *synaptic weight*. However, this time course is biologically unrealistic as it implies that the ion channels at the synapse open and close instantaneously. To model a more realistic time course, one can assume an exponentially decaying PSC, where the decay or synaptic time constant $\tau_{\mathrm{s}}$ characterizes the closing of the ion channels. More formally, the synaptic current evolves according to

$$\tau_{\mathrm{s}} \dot{I} \quad = \quad -I + \tau_{\mathrm{s}} w \delta(t - t^k) \,, \tag{1.2}$$

where $w$ is the synaptic efficacy in ampere. The time course of the PSC and the corresponding PSP are shown in (Fig. 1.2A, lower panel). A Fourier transformation of Eq. (1.2) shows that exponentially decaying PSC's lead to low-pass filtering of the input which we refer to as synaptic filtering. While we use this model of the PSC throughout this work, more realistic models exists, for example, also describing the time scale on which the ion channels open (see, e.g Fourcaud & Brunel (2002) and references therein).

In the situation shown in the upper panel Fig. 1.2A the constant current drives the membrane potential to threshold and therefore it is termed super-threshold current. As a consequence, the neuron emits a sequence of spikes (spike-train) which is regular. In contrast, if the constant current is sub-threshold (Fig. 1.2B, red curve) the neuron never hits the threshold. A neuron in the living brain, however, does not see a constant input current but rather receives a tremendous number of spikes from pre-synaptic neurons constituting a strongly fluctuating input $I(t)$ (Eq. (1.2)). This input is typically approximated as a Poisson point process (Dayan & Abbott, 2001) and can approximately be described by its mean value and the amplitude of the fluctuations (see Chap. 3 for a quantitative description). If the mean of the synaptic current, corresponding to $I_{\mathrm{const}}$ in the situation before, is much larger then the threshold, this would again result in a rather regular spike train constituting the so called *mean driven regime*. However, if the mean value is sub-threshold, the input fluctuations occasionally cause the membrane potential to hit the threshold (Fig. 1.2B, blue curve) resulting in an irregular spike train and this regime is termed *fluctuation driven.*

Two important quantities, namely the *gain function* and the *transfer function*, describe the behavior of the LIF neuron model in response to input. The gain function quantifies the constant firing rate of the neuron as a function of a constant input rate (rate of the pre-synaptic neuron). The transfer function characterizes how a modulation of the neuron's input rate is transmitted to a modulation of its output rate.

In conclusion, the LIF neuron model combines two important aspects of neuronal dynamics, the sub-threshold integration of inputs by the membrane potential as well as the interaction with other neurons via spikes. For our purposes, this model comes along with the right amount of detail: On the one hand it is amenable to mathematical treatment, because its sub-threshold dynamics is linear and on the other hand it can produce irregular activity as experimentally observed in cortex (Softky & Koch, 1993).

### 1.2.3 Cortical network structure and dynamics

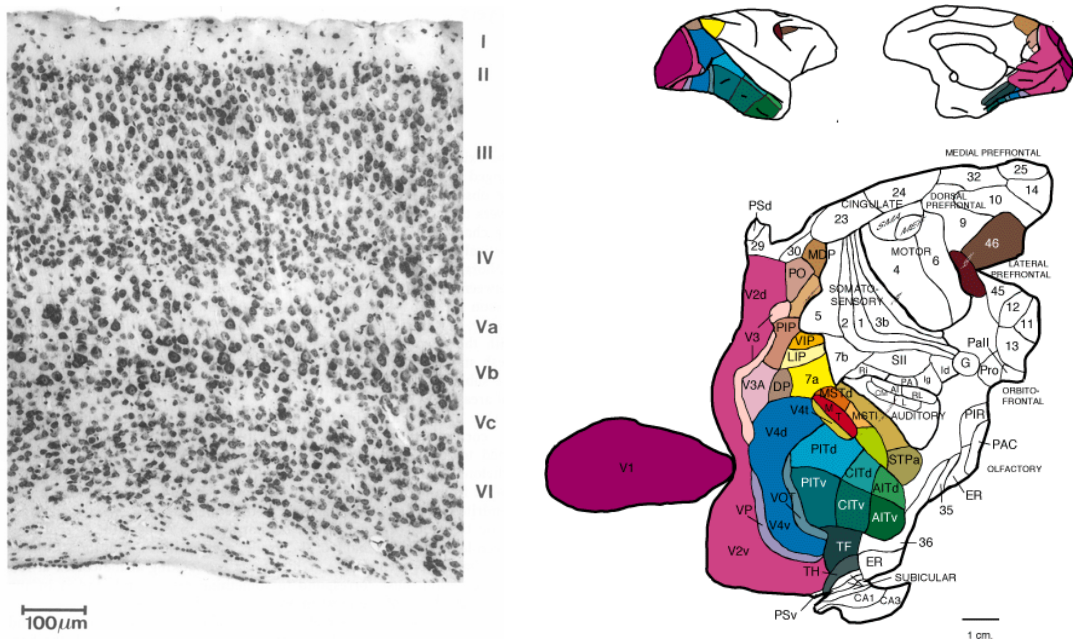



**Figure 1.3: Structure of cerebral cortex. Left** Layered structure of the visual cortex. Figure taken with permission from Braitenberg & Schüz (1991). **Right** Parcellation scheme of macaque visual cortex. Figure taken with permission from Felleman & Van Essen (1991).

The cerebral cortex is involved in many brain functions such as perception, speech, memory, learning and motor control. It constitutes the outermost part of the mammalian brain with a thickness on the order of a millimeter. The cortex is folded to maximize its surface area and consists mostly out of cell bodies and the wires (axons and dendrites) between them.

The cortex can be subdivided into areas according to structural, e.g. cytoarchitectonical features (Brodmann, 1909) and functional properties (Kandel et al., 2000). For example,

Felleman & Van Essen (1991) proposed a parcellation scheme of macaque visual cortex into 32 areas (Fig. 1.3, right) by considering cell densities, connectivity patterns and physiological properties of the neurons. Such a parcellation scheme can serve as a basis for constructing large-scale models (Sec. 1.2.4). While the connectivity inside the individual areas is dominating, also long-range connections between areas, termed *cortico-cortical*, are important (Markov et al. (2011, 2014a)).

The individual areas are further organized into cortical layers (Fig. 1.3, left), which differ in neuron density, the occurrence of certain cell types as well as their connectivity (Braitenberg & Schüz, 1991; Abeles, 1991a). As stated above cortical neurons are either excitatory or inhibitory. The number of excitatory neurons is around four times larger than the number of inhibitory neurons (Braitenberg & Schüz, 1991).

The neuronal density of the cortex is on the order of $4 \cdot 10^4/\text{mm}^3$ and each neuron receives input from around $2 \cdot 10^4$ synapses (Abeles, 1991b), where these numbers are obtained from an average over species. A local cortical circuit is located below a small patch of the cortical surface of around $1\text{mm}^2$ size. Within this scale, the connection probability $p$ between two neurons is approximately constant and around $p \approx 0.1$ (Braitenberg & Schüz, 1991). Such a local circuit is also called cortical column.

The activity of cortical neurons is measured on various scales using different techniques: One way to measure the spiking activity of single neurons is by inserting electrodes into cortical tissue. Complementary, the local field potential (LFP) is the low-pass filtered signal of the electrode and reflects the averaged synaptic activity of a population of neurons with a spatial reach up to a few mm (Lindén et al., 2011). On a macroscopic scale spanning the entire brain, the activity of whole areas is for example monitored with functional magnetic resonance imaging (fMRI) (Huettel et al., 2004). This is a non-invasive technique that measures the time course of the level of oxygen, which is related to the activity of a group of neurons. Various other measures exists and an overview is for example given in Brette & Destexhe (2012).

Cortical neurons emit roughly between 0.05 spikes/s and 30 spikes/s (Swadlow, 1988; de Kock & Sakmann, 2009). Furthermore, neurons fire irregularly (Softky & Koch, 1993) and correlations between neurons are typically low (Ecker et al., 2010; Cohen & Kohn, 2011) which can be beneficial since the presence of strong correlations could compromise information decoding (Zohary et al., 1994; Shadlen & Newsome, 1998; Tripp & Eliasmith, 2007; Tetzlaff et al., 2012). In addition, a dynamic balance between excitation and inhibition is observed (van Vreeswijk & Sompolinsky, 1996; Okun & Lampl, 2008; Renart et al., 2010; Dehghani et al., 2016). On the population level, oscillations are present on various frequency bands (see Buzsáki & Draguhn (2004) for a review), which are for example important for the communication between areas (Fries, 2005; Bastos et al., 2015; van Kerkoerle et al., 2014; Bastos et al., 2015b,a).

### 1.2.4 Modeling of cortical networks

To gain an understanding of the functional principles of the brain, experiments in computational neuroscience aim to relate neuronal activity to behavior and thus brain function. These relations have been discovered in various areas and are usually found in terms of correlations between observables and behavior (e.g, spike patterns in visual cortex (Berger et al., 2010) or motor cortex (Kilavik et al., 2009; Torre et al., 2016)), which do not necessarily imply a causal relationship (Brette, 2015). Therefore, it can be difficult to infer the underlying mechanism, hindering a true understanding of the working principles. A classical example is the concept of tuning curves (Hubel & Wiesel, 1959): Firing rates of recorded cells strongly vary with certain stimulus parameters. While these relations have been shown in many experiments, they could be only an epiphenomenon of the actual computation in the brain.

In general, models can help in overcoming these difficulties, as they can be controlled at will while one can access the full system at the same time. The first benchmark for a model is to reproduce experimental findings and then develop predictions or hypothesis, which in turn could be proved in an experiment. Thus, the experiment and the model ideally form a loop. In the following we briefly discuss important modeling strategies employed in computational neuroscience and explain which features of brain activity or brain function they can account for.

**Top-down models** Functionally inspired top-down approaches aim to understand the computation in neural networks. These models usually consider a defined function and focus on its implementation. They typically describe neurons or neuronal populations in terms of continuous variables, e.g. firing rates (Hertz et al., 1991; Schöner et al., 2015). Rate-based models originate from the seminal works by Wilson & Cowan (1972) and Amari (1977) and were introduced as a coarse-grained description of the overall activity of large-scale neuronal networks. Being amenable to mathematical analyses and exhibiting rich dynamics such as multistability, oscillations, traveling waves, and spatial patterns (Coombes, 2005; Roxin et al., 2005; Bressloff, 2012), rate-based models have fostered progress in the understanding of memory, sensory and motor processes (Kilpatrick, 2014). Particular examples include visuospatial working memory (Camperi & Wang, 1998; Laing et al., 2002), decision making (Usher & McClelland, 2001; Bogacz et al., 2006, 2007), spatial navigation (Samsonovich & McNaughton, 1997), and movement preparation (Erlhagen & Schöner, 2002). Ideas from functional network models have further inspired the field of artificial neuronal networks in the domain of engineering. While the majority of top-down models use a rate-based description, recent developments model functional networks also with spiking neuron models (for an overview see Abbott et al. (2016)).

**Bottom-up models** Bottom-up models typically have a different focus: How does the macroscopic activity in the network emerge from the interplay between the individual com-

ponents on the microscopic scale, i.e., on the scale of individual neurons. This is particularly hard to answer from the experimental side since current experimental technologies, e.g. the 100-electrode Utah array (Blackrock Microsystems), can simultaneously only record from hundreds of individual neuron in a small patch leading to a tremendous undersampling problem compared to approximately $10^6$ neurons within this patch.

Within the bottom-up approach, individual neurons of the network are typically modeled with biophysically inspired spiking neuron models. These network models can explain a variety of salient features of microscopic neural activity observed *in vivo*, such as spike-train irregularity (Softky & Koch, 1993; van Vreeswijk & Sompolinsky, 1996; Amit & Brunel, 1997b; Shadlen & Newsome, 1998), membrane-potential fluctuations (Destexhe & Paré, 1999), self-sustained activity (Ohbayashi et al., 2003; Kriener et al., 2014a), rate distributions across neurons (Griffith & Horn, 1966; Koch & Fuster, 1989; Roxin et al., 2011), as well as resting-state activity (Deco & Jirsa, 2012). Bottom-up model are not exclusively spike-based. For example, on the brain scale, rate models have been used to study resting-state activity (Deco et al., 2011) and hierarchies of time scales (Chaudhuri et al., 2015).

Spiking bottom-up models of cortical networks express the connectivity with random networks. These models are based on the fact that experimentally measured connectivity between neurons is typically described on a statistical level as an average over individuals and experiments. Then, the connectivity in the network is realized according to a statistical rule. Therefore, to learn about general principles, an implicit assumption of the random network approach is that a similar macroscopic activity emerges for different realization of the connectivity. This assumption is justified for various macroscopic observables such as time-averaged firing rate or oscillations in the population activity.

More concretely, in such a network model the individual neurons are randomly connected forming for example an Erdős–Rényi graph, where each possible connection is realized with $p$, or a graph with fixed in- (out-) degree, i.e., a fixed number of incoming (outgoing) connections per neuron. For example, Brunel (2000) combined this approach with the LIF neuron to model a local cortical circuit with randomly connected excitatory and inhibitory neuron populations. From now on we refer to this model as the $\mathcal{E} - \mathcal{I}$ network model. Using this model, Amit & Brunel (1997b) and Brunel (2000) could show that inhibitory feedback leads to the experimentally observed asynchronous and irregular firing, illustrating that bottom-up models provide a useful bridge between structure and activity.

The $\mathcal{E} - \mathcal{I}$ network models a small patch ($\sim$ mm$^2$) of a cortical layer compared to the full surface area which is on the order of square meters. In this model the neurons receive input from around $10^3$ synapses. Comparing this number to the actual number of synapses in the cortex ($\sim 10^4$), around 90% of the synapses and the corresponding pre-synaptic neurons are not represented in the model. These neurons may be located in other layers inside the same column, in the same area but outside the column or in other areas. In the $\mathcal{E}-\mathcal{I}$ network model, input of these neurons is replaced by external drive with spike trains drawn from Poissonian statistics (as in Fig. 1.2B). This naturally precludes an interaction between neurons in the

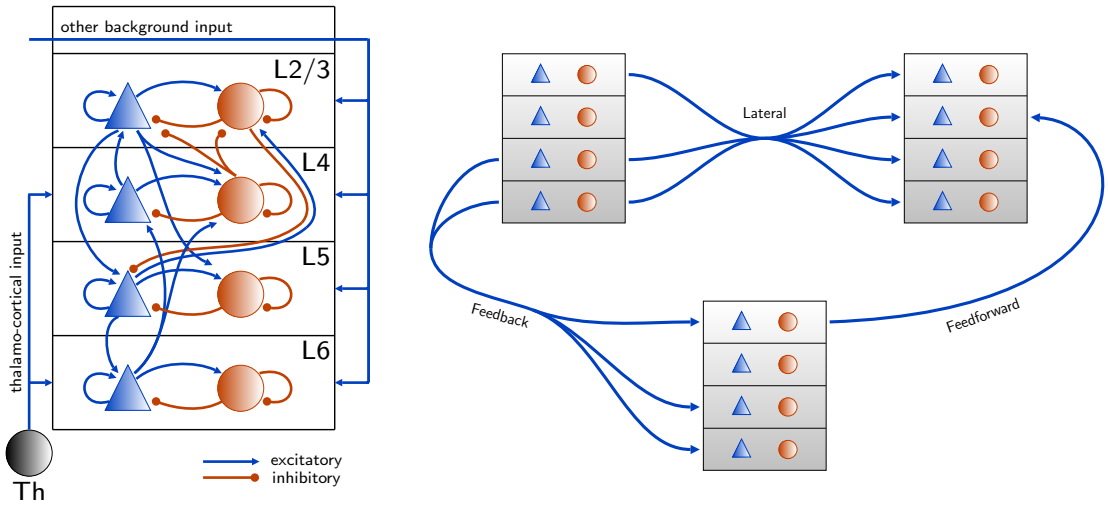



**Figure 1.4: Complex network models of cortex. Left** Sketch of microcircuit model with four layers and one excitatory (blue triangles) and one inhibitory population (red circles) each. The model represents one cortical column (compare to Fig. 1.3 left). **Right** Sketch of multi-area model (Schmidt et al., 2016) showing three example areas with indicated inter-area connections. Figure adapted from Maximilian Schmidt, personal communication.

$\mathcal{E}-\mathcal{I}$ network model and the rest of the cortex. However, it is well known that communication between areas is involved in cortical function, for instance in visual perception (Fries, 2005; van Kerkoerle et al., 2014; Bastos et al., 2015,b). Moreover, the $\mathcal{E}-\mathcal{I}$ network model does not feature experimentally observed slow oscillations (Steriade et al., 1993) and we believe that the long transmission delays between remote areas are important for this dynamical feature.

For these reasons there is a need for large scale models and two recent advances enable the development of such models. On the one hand progress in simulation (Hines et al., 2008; Kumar et al., 2010; Hines et al., 2011; Helias et al., 2012; Kunkel et al., 2014) and computing technology (Jülich Supercomputing Centre, 2015) allow scaling up the number of neurons and synapses in the models. On the other hand data on large-scale connectivity is gradually being filled through classical techniques combined with innovative quantitative analyses (Markov et al., 2014a,b) and new technologies (Wedeen et al., 2008; Axer et al., 2011).

Potjans & Diesmann (2014) has developed a model of a cortical column (Fig. 1.4, left), also called microcircuit model, including four layers with an excitatory and an inhibitory population each and realistic connection probabilities between the populations. This model shows that already a realistic connectivity pattern can explain the experimentally observed distribution of firing rates across its laminar populations. In an ongoing study, Senk et al. (2015) further scale up the model from a square millimeter patch to four times four millimeters. Moreover, the microcircuit model serves as the basis for the model of vision-related areas of macaque cortex developed by Schmidt et al. (2016) (Fig. 1.4, right). In this model, individual areas are represented by adapted microcircuit models which are then connected to each other with

a realistic cortico-cortical connectivity map (for more details see Sec. 4.3).

The large-scale models come along with a higher number of populations increasing the network complexity. In particular, the interplay between the populations can lead to dynamic phenomena which are difficult to understand from simulations alone, which begs for a theoretical understanding of the dynamics in these complex network models.

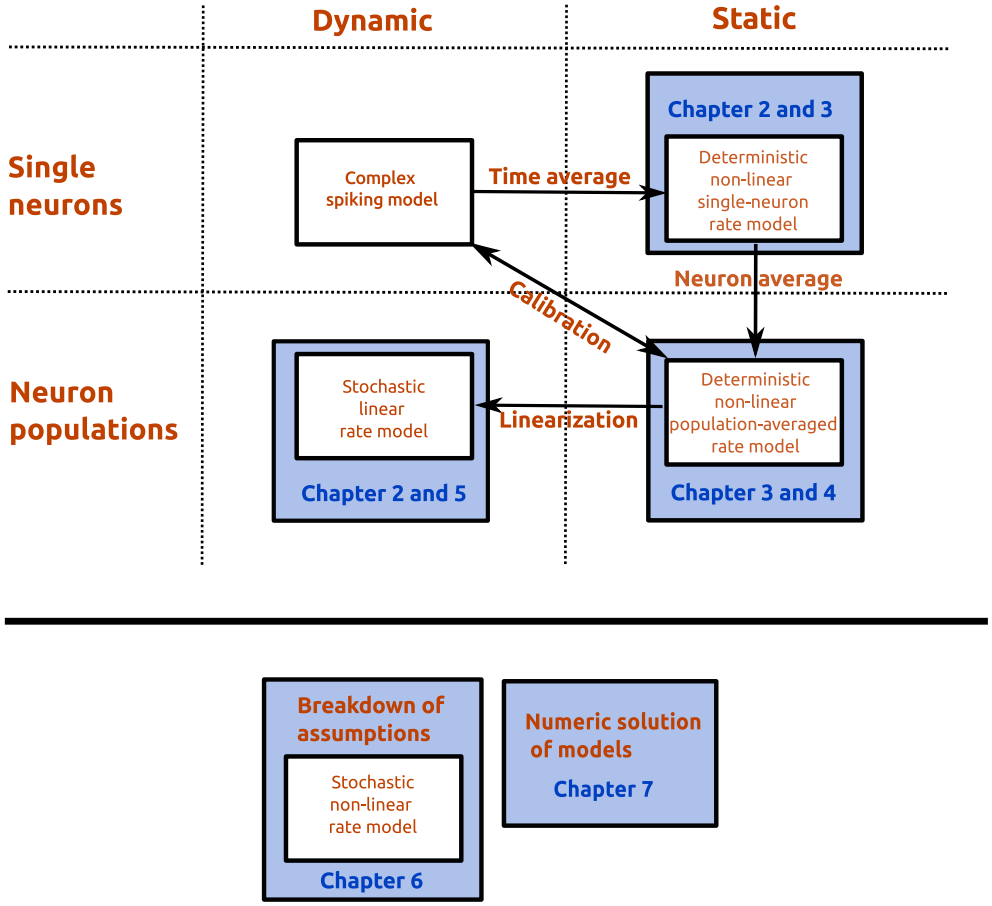



**Figure 1.5: Structure of this thesis. Upper part** The table shows the different description levels of our theoretical framework for the analysis of the dynamics in spiking network models (upper left box). First, we consider time-averaged activity of the individual neurons (upper right box). We subsequently describe the population-averaged activity (lower right box), which helps us to calibrate the spiking network models. Moreover, a linearization around the stationary states allows for the investigation of small temporal fluctuations in the activity of the spiking model (lower left box). **Lower part** We discuss the breakdown of the theoretical framework together with a first step to a solution (left box). Finally, the numerical solution of the various models appearing in the theoretical framework is described (right box).

## 1.3 Mean-field theory and structure of this thesis

This thesis addresses the question, **how we can understand the dynamics of large-scale network models**. Answering this question is particularly relevant in constructing these models, exploring and predicting their behavior and finding mechanisms for the emergence of certain phenomena from the underlying connectivity structure and neuron dynamics.

To investigate the dynamics of the network models, one first needs to choose a proper level of description. In this thesis, we do not consider the exact realization of spike times in the network, inspired by the observation that the activity in the cortex as well as in bottom-up models is typically irregular. Within a common approach (Brunel & Hakim, 1999; Trousdale et al., 2012; Helias et al., 2013; Grytskyy et al., 2013) this irregular spiking activity can be approximated by Poisson processes, whereby the realization of the spikes amounts to a source of noise. In the neuroscientific community it is controversially debated whether the brain uses a rate-based or a spike-time based coding scheme. However, it has been pointed out that these are not exclusive (Brette, 2015). In particular the rate-based description could serve as a base-line in a spike-based view. Thus, we do not refuse the potential importance of individual spike times but rather chose firing rates as a pragmatic level of description to understand the principle dynamic properties.

Fig. 1.5 illustrates the different scales on which we investigate the dynamics of complex networks. In the following we outline the content of this thesis along this sketch and set it into context with existing works.

We start with a complex spiking network model (Fig. 1.5, upper left corner). We assume that the firing rates of the neurons only weakly fluctuates in time around a stationary state, where the latter is our first quantity of interest. The stationary state is described by the time-averaged activity of the neurons in the network (Fig. 1.5, Time average).
To determine the latter analytically one first needs to know the non-linear gain function or input-output relationship of an individual neuron. Here, the input to the neuron is described in a stochastic manner to account for fluctuations in the spiking input. Deriving the gain function can be formulated as an escape problem of a particle in a potential which is driven by noise. Exploiting the stochastic nature of the problem one moves from a description of the state of the neuron to a description of the density of states, where the evolution of the density is governed by the Fokker-Planck equation (Risken, 1996). The gain function follows from the stationary solution of this equation and is well-known from the literature (Siegert, 1951; Brunel, 2000; Fourcaud & Brunel, 2002; Risken, 1996).
In Chap. 2 we present a new method which simplifies and generalizes the derivation of the known results. The advantage of this method is that it can be extended to the dynamic case as explained further below.

Besides the gain function, the second ingredient to arrive at the stationary firing rates in the network are closure assumptions. In the network situation the output of the neurons con-

stitutes the input to the neurons, i.e., the system needs to be solved self-consistently. This is achieved using standard arguments from mean-field theory (e.g. used in van Vreeswijk & Sompolinsky, 1996; Amit & Brunel, 1997b; Brunel, 2000), effectively replacing the input to the neuron from the network by Gaussian noise.
In Chap. 3 we discuss the closure assumptions in detail and further show that these assumptions combined with the gain function lead to a reduction of the spiking network to a deterministic, non-linear rate model. Its solution quantifies the stationary state of the spiking network on the level of single neurons, as for example used in Sadeh & Rotter (2015). For an illustrative purpose we show this approach for a simple network example.

To understand the network activity, we further simplify the description by reducing the dimensionality (Fig. 1.5, Neuron average): Assuming single neurons in one population to be homogeneous with respect to their parameters and input statistics, the dimension of the non-linear rate model reduces to the number of populations. This reduction has been successfully applied to analyze stationary activities of spiking populations in small-scale network models (Brunel, 2000; Wong & Wang, 2006), i.e, networks with a small number of populations. However, the low dimensionality of these model renders the phase space and attractor landscape simple.
In Chap. 4 we extend this approach to analyze the phase space of the multi-area model (Schmidt et al., 2016), comprising 254 populations. Due to the high dimensionality qualitatively new features emerge as for example the coexistence of multiple attractors. To deal with these complications we develop a novel technique which allows us to analyze and control the separatix and thus the global stability in the network. The developed method not only helps us to understand the models' activity but more importantly allows the inclusion of activity constraints into the process of model construction. This is in particular important as this high dimensional model with a huge parameter set is strongly underconstrained. Finally, the method enables the selection of model parameters from within the range of experimental uncertainty so that the model reaches a realistic set point of stationary activity.

So far we have considered the stationary state firing rate. Various approaches address the dynamics of rate fluctuations around this state. For example the dynamics of oscillatory modes of the population activity can be obtained using a mode decompositions of the Fokker-Planck operator (Mattia & Del Guidice, 2002, 2004). Another Ansatz determines the autocorrelation function of the population averaged activity performing a linearization (Brunel, 2000) or a weakly non-linear analysis (Brunel & Hakim, 1999) of the Fokker-Planck equation. The autocorrelation of the population rate is an interesting quantity, as it includes the cross-correlations between neurons and there is experimental evidence that these are functionally relevant in the brain (Kilavik et al., 2009; Okun & Lampl, 2008). The cross-correlations can also be derived using the related framework of linear response theory (Lindner et al., 2005; Ostojic & Brunel, 2011; Trousdale et al., 2012; Grytskyy et al., 2013) (Fig. 1.5: Linearization), which reduces the network of spiking neurons to a network of stochastic linear rate models. This reduction could identify inhibitory feedback as the mechanism leading to low correlations

in spiking networks (Tetzlaff et al., 2012; Helias et al., 2013, 2014). Furthermore, the dynamical linear-rate model can reveal the origin of collective oscillations (Bos et al., 2016).
The central building block of the linear-response theory is the transfer function of individual neurons. An analytic expression for the transfer function of the LIF neuron model has been derived with the Fokker-Planck formalism (Brunel & Hakim, 1999; Lindner & Schimansky-Geier, 2001), but realistic synaptic filtering has been neglected in the derivations so far. The reason is that it complicates the mathematical treatment substantially constituting a colored-noise problem. The models considered in this work employ synaptic filtering to prevent the network state from entering an unrealistic degree of synchronization between neuron (Helias et al., 2010; Schultze-Kraft et al., 2013). Thus, to successfully apply the linear-response framework to the considered networks, the analytic difficulty of synaptic filtering needs to be overcome.
In Chap. 2 we solve this problem by developing a general framework that reduces a first order stochastic differential equation driven by fast colored noise to an effective system driven by white noise capturing the color of the noise by effective boundary conditions. The second step to the solution is a novel mathematical treatment of the LIF neuron model with an operator notation exploiting the model's analogy to the quantum harmonic oscillator. The combination of the reduction step with the operator notation finally allows us to derive a novel analytic expression for the transfer function including synaptic filtering.
In Chap. 5 we show how the linear-response framework including the new result for the transfer function is applied to spiking models to analyze rate fluctuations around the stationary states. For an illustrative purpose we first calculate correlations in the $\mathcal{E}-\mathcal{I}$ network. Subsequently, we extend the framework to the multi-area model laying the foundation for future studies on the communication between different areas by investigating the coherence between them.

The validity of the theory describing the stationary and dynamic properties in a spiking network is limited to certain regimes. For example, Ostojic (2014) discovered that mean-field theory breaks down (Fig. 1.5: Breakdown) for increased but still realistic coupling weights between the neurons. In this regime, even the explanation of the population-averaged stationary rates is beyond the analytical framework introduced above. The discrepancies originate from the emergence of large fluctuations in the firing rate of single neurons. Recent works (Mastroguiseppe & Ostojic, 2016; Ostojic, 2014; Kadmon & Sompolinsky, 2015a) aim to explain and handle these fluctuations using dynamical mean-field theory, which provides a self-consistent treatment of fluctuations. However, this theory has been developed for rate-based neuron models (Sompolinsky et al., 1988) and the transfer of the results to spiking neuron models is not known, leaving the breakdown of the theory an open problem. The application of dynamical mean-field theory to spiking models is partly hindered by the fact that the original deterministic formulation (Sompolinsky et al., 1988) does not include noise. Such a noise term naturally appears in the mapping of spiking models to a rate-based description (Grytskyy et al., 2013).
In Chap. 7 we develop the dynamical mean-field theory for a stochastic rate model, a pre-

requisite to apply the framework of dynamical mean-field theory to spiking neural networks. In particular, we investigate the generic effect of noise on the transition to chaos, which is believed to be related to the emergence of fluctuations in the activity of spiking networks (Ostojic, 2014; Mastroguiseppe & Ostojic, 2016; Harish & Hansel, 2015).

As indicated in Fig. 1.5 various rate-based models appear, which have to be solved to make predictions for the activity in spiking network models. A unifiying simulation framework for spiking and rate-based models would simplify the cross-check of these predictions. However, in the beginning of this work, such a framework does not exists due to the fundamental differences between the two models.
In Chap. 8 we overcome these difficulties and develop a unifying simulation framework by integrating rate-based models into a spiking neural network simulator (Gewaltig & Diesmann, 2007; Bos et al., 2015).

Each chapter starts with a short outline of the context. In addition, the major chapters close with "Interim discussion", where we discuss issues very specific to the corresponding topic. A broader discussion of all chapters is given in Chap. 9. In Chap. 10 we further give an outlook and mention potential new research avenues which are opened up by this thesis.

# Part II

# Mean-field theory for spiking dynamics

# Chapter 2

# Modulated escape from a metastable state driven by colored noise

---

The following chapter is based on the following publications:

Schuecker, J., Diesmann, M., & Helias, M. (2015). Modulated escape from a metastable state driven by colored noise. Phys. Rev. E 92, 052119.
Schuecker, J., Diesmann, M., & Helias, M. (2014). Reduction of colored noise in excitable systems to white noise and dynamic boundary conditions. arXiv preprint arXiv:1410.8799.
**Author contributions:**
Under the supervision of M Helias, the author performed all parts of the above publications. All authors contributed to the writing of the manuscripts.

In this chapter we derive the gain function as well as a novel analytical expression for the transfer function of the LIF neuron model including synaptic filtering, both quantities needed to analyze the dynamics of complex spiking networks. We consider a general excitable systems driven by colored noise, while the LIF neuron model including synaptic filtering is a special case of this general formulation. The temporal correlations in the fluctuations hinder an analytical treatment. We here present a method of reduction to a white-noise system, capturing the color of the noise by effective boundary conditions. We subsequently apply this reduction method to the LIF neuron model, where we in addition exploit the models analogy to the quantum harmonic oscillator which allows a straight forward mathematical treatment of the model.

## 2.1 Introduction

In his pioneering work Kramers (1940) investigated chemical reaction rates by considering the noise-activated escape from a metastable state as a problem of Brownian motion over a barrier. In the overdamped case, this generic setting applies to various phenomena ranging from ion channel gating (Goychuk & Hänggi, 2002), Josephson junctions (Malakhov & Pankratov, 1996; Pankratov & Spagnolo, 2004; Mantegna & Spagnolo, 2000), tunnel diodes (Mantegna & Spagnolo, 1996), semiconductor lasers (Hales et al., 2000) to abstract models of cancer growth (Spagnolo et al., 2007). These studies assume the noise statistics to have a white spectrum. This idealization simplifies the analytical treatment and is the limit of non-white processes for vanishing correlation time (Gardiner, 1983). White noise cannot exist in real systems, which is obvious considering, for example, the voltage fluctuations generated by thermal agitation in a resistor (Johnson, 1928; Nyquist, 1928): if fluctuations had a flat spectrum, power dissipation would be infinite. In a non-equilibrium setting a system is driven by an external fluctuating force and the simplest model for a non-white driving noise is characterized by a single time constant, representing a more realistic colored-noise model.

The effect of colored noise is relevant in modeling the phase difference in lasers (Vogel et al., 1987), the velocity in turbulence models (Frisch & Morf, 1981), genetic selection (Horsthemke, 1984) and chemical reactions (Lindenberg & West, 1983) (for further examples see (Moss & McClintock, 1989a,b)). Its influence on the escape from a metastable state has been studied theoretically as the stationary mean first passage time (MFPT) of a particle in the Landau potential (Sancho et al., 1982; Lindenberg & West, 1983; Hanggi et al., 1985; Fox, 1986; Grigolini, 1986; Doering et al., 1987). Treating colored noise analytically comes along with considerable difficulties, since it adds a dimension to the governing Fokker-Planck equation (FPE) and the common strategy is to reduce the colored-noise problem to an effective white-noise system. Doering et al. (1987) and Kłosek & Hagan (1998) developed these approaches further by singular perturbation methods and boundary-layer theory, showing that the leading order correction to the static MFPT stems from an appropriate treatment of the boundary conditions of the effective system.

The organization of this chapter is as follows. In Sec. 2.2 we extend the works by Doering

et al. (1987) and by Kłosek & Hagan (1998) to the time-dependent case by a perturbation expansion of the flux operator appearing in the FPE itself. This approach leads to a general method that reduces a first order differential equation driven by additive and fast colored noise to an effective one-dimensional system, allowing the study of time-dependent phenomena and thus revealing the spectral properties of the system. The effective formulation implicitly contains the matching between outer and boundary-layer solutions. The latter appear close to absorbing boundaries and are obtained by a half-range expansion. Our main result is that colored-noise approximations for stationary but more importantly also for dynamic quantities are directly obtained by shifting the location of the boundary conditions in the solutions for the corresponding white-noise system.

In Sec. 2.3 we apply this general result to the particular problem of modeling biological membranes: the leaky integrate-and-fire (LIF) neuron model (Lapicque, 1907; Stein, 1967) with exponentially decaying post-synaptic currents (Sec. 1.2.2) can equivalently be described as a first-order differential equation driven by colored noise. Based on the works by Doering et al. (1987) and by Kłosek & Hagan (1998), Brunel et al. and Fourcaud & Brunel (2002) calculated the high-frequency limit of its transfer function. Here we complement these works deriving a novel analytical expression for the transfer function valid up to moderate frequencies, which we confirm by direct simulations. While for slow noise, an adiabatic approximation for the transfer function is known (Moreno-Bote & Parga, 2006), this first order correction in the time scale of the noise is a qualitatively new result. In the previous work by Fourcaud & Brunel (2002) it was even claimed that the first order correction to the transfer function vanishes in this biologically relevant regime. However, this was only shown in case of the integrate-and-fire neuron without a leak term, i.e., for the perfect integrator (PIF).

In conclusion, previous to this thesis the correction to the transfer function of the LIF neuron model was believed to vanish, while we here prove its existence and calculate its exact form.

## 2.2 Reduction from colored to white noise

Consider a pair of coupled stochastic differential equations (SDE) with a slow component $y$ with time scale $\tau_{\mathrm{m}}$, driven by a fast Ornstein-Uhlenbeck process $z$ with time scale $\tau_{\mathrm{s}}$. In dimensionless time $s = t/\tau_{\mathrm{m}}$ and with $k = \sqrt{\tau_{\mathrm{s}}/\tau_{\mathrm{m}}}$ relating the two time constants we have

$$\begin{aligned} \frac{dy}{ds} &= f(y,s) + \frac{z}{k} \\ k\frac{dz}{ds} &= -\frac{z}{k} + \xi, \end{aligned} \tag{2.1}$$

with a unit variance white noise $\langle \xi(s+u)\,\xi(s)\rangle = \delta(u)$. The setting Eq. (2.1) describes an overdamped Brownian motion of a particle in a possibly time dependent potential $F(y,s) = -\int^y f(y',s)\,dy'$, driven by colored noise. The generic problem of escape from such a potential is illustrated in Fig. 2.1A where the particle has to overcome a smooth barrier. The latter

can be simplified, assuming an absorbing boundary at the right end $y = \theta$ of the domain, illustrated in Fig. 2.1B for the quadratic potential $F(y) = y^2$. The potential $F(y) = -y^2 + y^4$ (Fig. 2.1C) is regarded as the archetypal form giving rise to a bistable system and is reviewed in (Hänggi et al., 1990) and (Hänggi & Jung, 1995). The MFTP from one well to the peak of the potential can be calculated by assuming an absorbing boundary at the maximum, formally reducing this case to the one shown in Fig. 2.1B. Note that the MFTP from one well to the other is not directly related to the MFTP to the peak, but rather is twice the MFTP to an absorbing boundary on the separatrix in the two-dimensional domain (Hänggi et al., 1988) spanned by $y$ and $z$. However, we are here interested in the case where the component $y$ crosses a constant threshold value (Doering et al., 1988).

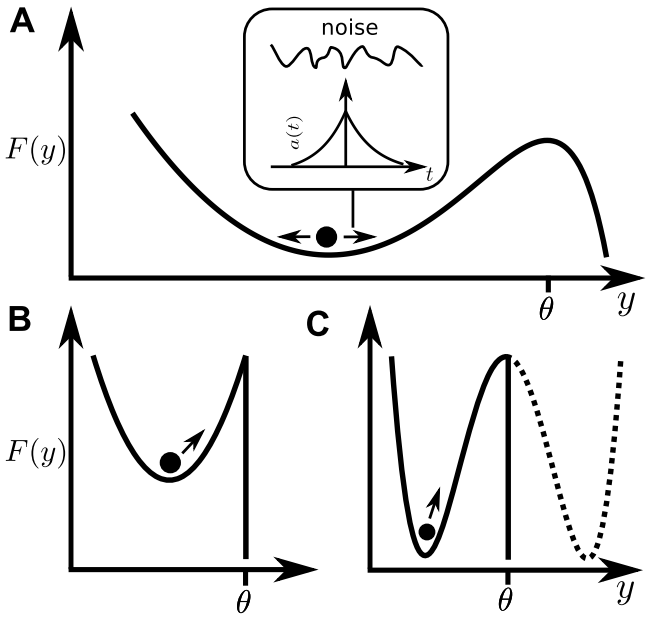



**Figure 2.1: Escape from a metastable state**. **A** Overdamped motion of a particle driven by colored noise (with exponentially decaying autocovariance $a(t)$, inset) in a quadratic potential $F(y) = y^2$ with smooth boundary near $\theta$. **B** Quadratic potential with hard absorbing boundary. **C** Bistable potential $F(y) = -y^2 + y^4$ with absorbing boundary at the maximum.

### 2.2.1 Effective diffusion

Initially we revisit the intuitive argument that in the case of fast noise $\tau_\mathrm{s} \ll \tau_\mathrm{m}$ the slow component $y$ approximately obeys the one-dimensional SDE $\frac{dy}{ds} = f(y, s) + \xi(s)$. Heuristically this can be viewed as $y$ integrating $z$ on a time scale $\tau_\mathrm{m}$. For finite $\tau_\mathrm{s}$ the integral of the autocorrelation of $z$ is $k^2$, identical to the limit of vanishing $\tau_\mathrm{s}$, where $z(s) = k\xi(s)$ becomes a white noise. We here set out to formally derive an effective diffusion equation for $y$ to obtain a formulation in which we can include the absorbing boundaries that are essential to study an

escape problem. We consider the FPE Risken (1996) corresponding to the two-dimensional system Eq. (2.1)

$$k^2 \partial_s P = \partial_z \left( \frac{1}{2} \partial_z + z \right) P - k^2 \partial_y S_y P, \tag{2.2}$$

where $P(y, z, s)$ denotes the probability density and we introduce the probability flux operator in $y$-direction as

$$S_y = f(y, s) + z/k \,. \tag{2.3}$$

Factoring-off the stationary solution of the fast part of the Fokker-Planck operator, $P = Q \frac{e^{-z^2}}{\sqrt{\pi}}$, we get with the chain rule

$$\begin{aligned} \partial_z e^{-z^2} \circ &= e^{-z^2} (\partial_z - 2z) \circ \\ \text{and} & \\ \left( \frac{1}{2} \partial_z + z \right) e^{-z^2} \circ &= e^{-z^2} \left( \frac{1}{2} \partial_z - z + z \right) \circ = e^{-z^2} \frac{1}{2} \partial_z \circ \\ \text{so compactly} & \\ \partial_z \left( \frac{1}{2} \partial_z + z \right) e^{-z^2} \circ &= e^{-z^2} \left( \frac{1}{2} \partial_z - z \right) \partial_z \circ \equiv e^{-z^2} L \circ, \end{aligned} \tag{2.4}$$

where we defined the operator $L \equiv \left( \frac{1}{2} \partial_z - z \right) \partial_z$. This transforms Eq. (2.2) to

$$k^2 \partial_s Q = LQ - kz \partial_y Q - k^2 \partial_y \, f(y, s) \, Q. \tag{2.5}$$

We refer to $Q$ as the outer solution, since initially we do not consider the absorbing boundary condition. The strategy is as follows: we show that the terms of first and second order in the small parameter $k$ of the perturbation ansatz

$$Q = \sum_{n=0}^{2} k^n \, Q^{(n)} + O(k^3) \tag{2.6}$$

cause an effective flux acting on the $z$-marginalized solution $\tilde{P}(y, s) = \int dz \, \frac{e^{-z^2}}{\sqrt{\pi}} Q(y, z, s)$ that can be expressed as a one-dimensional FPE which is correct up to linear order in $k$. To this end we need to know the first order correction to the marginalized probability flux

$$\nu_y(y, s) \equiv \int dz \, \frac{e^{-z^2}}{\sqrt{\pi}} S_y Q(y, z, s) = \sum_{n=0}^{1} k^n \, \nu_y^{(n)}(y, s) + O(k^2) \,.$$

Inserting the perturbation ansatz Eq. (2.6) into Eq. (2.5) we have $LQ^{(0)} = 0$. Noting the property $Lz^n = \frac{1}{2} n(n-1) z^{n-2} - n z^n$, which implies $Lz^0 = 0$, we see that the lowest order does not imply any further constraints on the $z$-independent solution $Q^{(0)}(y, s)$, which must be consistent with the solution to the one-dimensional Fokker-Planck equation corresponding

to the limit $k \to 0$ of Eq. (2.1). The first and second orders are

$$\begin{aligned} LQ^{(1)} &= z\partial_y Q^{(0)} \\ LQ^{(2)} &= \partial_s Q^{(0)} + z\partial_y Q^{(1)} + \partial_y f(y,s) Q^{(0)}. \end{aligned} \tag{2.7}$$

With $Lz = -z$, the general solution for the first order

$$Q^{(1)}(y,z,s) = Q_0^{(1)}(y,s) - z\partial_y Q^{(0)}(y,s) \tag{2.8}$$

leaves the freedom to choose a homogeneous ($z$-independent) solution $Q_0^{(1)}(y,s)$ of $L$. To generate the term linear in $z$ on the right hand side of the second order in Eq. (2.7), we need a term $-z\partial_y Q^{(1)}$. The terms constant in $z$ require contributions proportional to $z^2$, because $Lz^2 = -2z^2 + 1$. However, they can be dropped right away, because terms $\propto k^2 z^2$ only contribute to the correction of the flux in order of $k^2$, while their contribution to $\nu_y^{(1)}$ resulting from the application of the $\propto k^{-1}$ term in Eq. (2.3) vanishes after marginalization. For the same reason the homogeneous solution $Q_0^{(2)}(y,s)$ can be dropped. Collecting all terms which contribute to $\nu_y$ in orders $k^2$ and higher in $\iota$, the only relevant part of the second order solution is $-z\partial_y Q_0^{(1)}(y,s)$, which leaves us with

$$\begin{aligned} Q(y,z,s) = Q^{(0)}(y,s) &+ kQ_0^{(1)}(y,s) \\ &- kz\partial_y Q^{(0)}(y,s) - k^2 z\partial_y Q_0^{(1)}(y,s) + \iota\,. \end{aligned} \tag{2.9}$$

We obtain the marginalized flux as

$$\begin{aligned} \nu_y(y,s) &= \int dz\, \frac{e^{-z^2}}{\sqrt{\pi}} S_y Q(y,z,s) \\ &= \left(f(y,s) - \frac{1}{2}\partial_y\right)\left(Q^{(0)}(y,s) + kQ_0^{(1)}(y,s)\right) + O(k^2), \end{aligned}$$

where we used $\int dz\, \frac{e^{-z^2}}{\sqrt{\pi}} = 1$ and $\int dz\, \frac{z^2 e^{-z^2}}{\sqrt{\pi}} = \frac{1}{2}$. We observe that $f(y,s) - \frac{1}{2}\partial_y$ is the flux operator of a one-dimensional system driven by a unit variance white noise and $\tilde{P}(y,s) \equiv Q^{(0)}(y,s) + kQ_0^{(1)}(y,s)$ is the marginalization of Eq. (2.9) over $z$ leading to

$$\nu_y(y,s) = \left(f(y,s) - \frac{1}{2}\partial_y\right)\tilde{P}(y,s) + O(k^2). \tag{2.10}$$

Note that in Eq. (2.9) the higher order terms in $k$ appear due to the operator $kz\partial_y$ in Eq. (2.5) that couples the $z$ and $y$ coordinate. Eq. (2.10) shows that these terms cause an effective flux that only depends on the $z$-marginalized solution $\tilde{P}(y,s)$. This allows us to obtain the time evolution by applying the continuity equation to the effective flux (Eq. (2.10) yielding the effective FPE

$$\partial_s \tilde{P} = -\partial_y \nu_y(y,s) = \partial_y\left(-f(y,s) + \frac{1}{2}\partial_y\right)\tilde{P}. \tag{2.11}$$

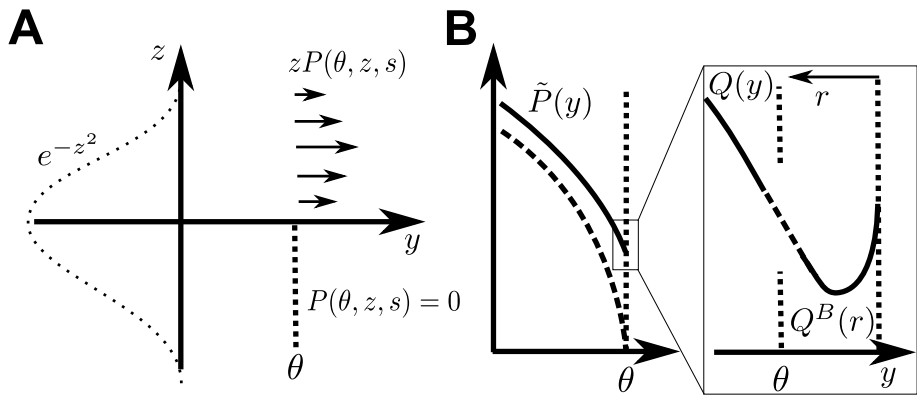



**Figure 2.2: Boundary conditions. A** Flux and boundary condition in the two-dimensional colored-noise system. On the negative half plane $z < 0$ the density must vanish at threshold $\theta$. Dashed curve: density marginalized over $y$ and neglecting boundary conditions. **B** Density of the white-noise system (dashed curve) vanishes at threshold. The density of the effective system (solid curve) has a finite value at threshold which is determined by the matching between the outer solution $Q$ and the boundary layer solution $Q^B$ (shown in enlargement).

### 2.2.2 Effective boundary condition

Effective FPEs, to first order identical to (2.11), but also including higher order terms, have been derived earlier (Sancho et al., 1982; Lindenberg & West, 1983; Hanggi et al., 1985; Fox, 1986; Grigolini, 1986). These approaches have been criticized for an untended treatment of the boundary conditions of the effective system (Kłosek & Hagan, 1998). Doering et al. (1987) and Kłosek & Hagan (1998) used singular perturbation methods and boundary-layer theory to show that the $O(k)$ correction to the static MFPT stems from colored-noise boundary conditions for the marginalized density of the effective system. Extending this approach to the transient case requires a time-dependent boundary condition for $\tilde{P}$ or, equivalently, for $Q_0^{(1)}$, because we assume the one-dimensional problem to be exactly solvable and hence the boundary value of $Q^{(0)}$ to be known. Without loss of generality, we assume an absorbing boundary at the right end $y = \theta$ of the domain. Thus, trajectories must not enter the domain from above threshold, implying the flux to vanish at $\theta$ for all points with negative velocity in $y$ given by $f(\theta, s) + \frac{z}{k} < 0$. A change of coordinates

$$z + kf(\theta, s) \to z \tag{2.12}$$

simplifies the condition to

$$\frac{z}{k} Q(\theta, z, s) = 0 \quad \forall z < 0$$

and it follows that

$$Q(\theta, z, s) = 0 \quad \forall z < 0. \tag{2.13}$$

The flux and this half range boundary condition are shown in Fig. 2.2A.

Due to the time dependence of the coordinate $z$ (Eq. (2.12)) the Fokker-Planck equation Eq. (2.2) with the time derivative of the density $\partial_s P(y, z(s), s) = \partial_s P - k(\partial_s f(\theta, s))\,\partial_z P$ transforms to

$$k^2\partial_s P = \partial_z(\frac{1}{2}\partial_z + z - kf(\theta, s))\,P - k^2\partial_y(f(y, s) - f(\theta, s) + \frac{z}{k})P - k^3\,\partial_s f(\theta, s)\partial_z P. \tag{2.14}$$

With $P = Q\,\frac{e^{-z^2}}{\sqrt{\pi}}$ we obtain

$$\begin{aligned} k^2\partial_s Q &= LQ - kz\partial_y Q \\ &\quad +k\left[f(\theta, s)(2z - \partial_z) - k\partial_y\left(f(y, s) - f(\theta, s)\right)\right]Q \\ &\quad +k^3\left(\partial_s f(\theta, s)\right)(2z - \partial_z)Q \end{aligned} \tag{2.15}$$

and we define the term in the angular brackets as $G(\theta, r, s, z)$. The last term originating from the time dependence of $f$ is of third order in $k$ and will therefore be neglected in the following.

To derive the boundary condition for the effective diffusion, we need to describe the behavior of the original system near these boundaries by transforming to the shifted and scaled coordinate $r = \frac{y-\theta}{k}$. The coordinate $r$ zooms into the region near the boundary, also termed boundary layer, and changes the order in $k$ of the interaction term $-kz\partial_y Q$ between the $y$ and the $z$ component from first to zeroth order, namely

$$\begin{aligned} k^2\partial_s Q^B = LQ^B - z\partial_r Q^B \\ + k\,G(\theta, r, s, z)\,Q^B + O(k^3), \end{aligned} \tag{2.16}$$

with $Q^B(r, z, s) \equiv Q(y(r), z, s)$. The boundary condition then takes the form

$$Q^B(0, z, s) = 0 \quad \forall z < 0. \tag{2.17}$$

With the perturbation ansatz $Q^B = \sum_{n=0}^{1} k^n\,Q^{B(n)} + O(k^2)$ we obtain

$$LQ^{B(0)} - z\partial_r Q^{B(0)} = 0 \tag{2.18}$$

$$LQ^{B(1)} - z\partial_r Q^{B(1)} = G(\theta, r, s, z)\,Q^{B(0)}, \tag{2.19}$$

the solution of which must match the outer solution. The latter varies only weakly on the scale of $r$ and therefore a first order Taylor expansion at the boundary yields the matching condition

$$Q^B(r, z, s) = Q(\theta, z, s) + kr\,\partial_y Q(\theta, z, s), \tag{2.20}$$

illustrated in Fig. 2.2B.

To zeroth order in $k$ we hence have

$$Q^{B(0)}(0,z,s) \quad = \quad 0, \tag{2.21}$$

because the white-noise system with $k=0$ has a vanishing density at threshold. Together with the homogeneous partial differential equation Eq. (2.18) this implies $Q^{B(0)}=0$ everywhere.
To perform the matching of the first order we insert Eq. (2.8) into Eq. (2.20) to obtain

$$Q^{B(1)}(r,z,s) \quad = \quad Q_0^{(1)}(\theta,s) - z\partial_y Q^{(0)}(\theta,s) + r\partial_y Q^{(0)}(\theta,s). \tag{2.22}$$

This can be simplified to

$$Q^{B(1)}(r,z,s) \quad = \quad Q_0^{(1)}(\theta,s) + 2\nu_y^{(0)}(s)\,(z-r), \tag{2.23}$$

where we again exploit that $Q^{(0)}(\theta,s)=0$ in the white-noise system and therefore

$$\partial_y Q^{(0)}(\theta,s) \quad = \quad -2(-\frac{1}{2}\partial_y + f(\theta,s))\,Q^{(0)}(\theta,s) = -2\nu_y^{(0)}(\theta,s) \tag{2.24}$$

with $\nu_y^{(0)}(\theta,s)\equiv\nu_y^{(0)}(s)$ the instantaneous flux in the white-noise system.
Using Eq. (2.21) the first order solution Eq. (2.19) must satisfy

$$LQ^{B(1)} - z\partial_r Q^{B(1)} \quad = \quad 0. \tag{2.25}$$

With the definitions $v=\sqrt{2}r$, $w=\sqrt{2}z$, and $g(v,w,s)=Q^{B(1)}(r,z,s)$, equation Eq. (2.25) takes the form

$$(\partial_w^2 - w\partial_w)\,g(v,w,s) \quad = \quad w\partial_v g(v,w,s).$$

Note that the time argument plays the role of a parameter here, since the time derivatives in Eq. (2.15) and (2.16) are of higher order in $k$. With the absorbing boundary condition $g(0,w,s)=0$ for $w<0$, following from Eq. (2.13), the solution not growing faster than linear in $v\to-\infty$ is given by Kłosek & Hagan (1998, B.11)

$$g(v,w,s) \quad = \quad \frac{C(s)}{\sqrt{2}}\left(\frac{\alpha}{\sqrt{2}} + w - v + \sqrt{2}\sum_{n=1}^{\infty} b_n(w/\sqrt{2})\,e^{\sqrt{n}v}\right) \tag{2.26}$$

with $\alpha=\sqrt{2}|\zeta(\frac{1}{2})|$ given by Riemann's $\zeta$-function and $b_n$ proportional to the $n$-th Hermitian polynomial. The constant $\alpha$ defined here follows the notation used in Fourcaud & Brunel (2002) and differs by a factor of $\sqrt{2}$ from the notation in Kłosek & Hagan (1998, B.11). At threshold and in the original coordinates we equate Eq. (2.26) to Eq. (2.23) which reads

$$Q_0^{(1)}(y,s) + 2\nu_y^{(0)}(s)\,(z-r) \quad = \quad C(s)\left(\frac{\alpha}{2} + z - r + \sum_{n=1}^{\infty} b_n(z)\,e^{\sqrt{2n}r}\right).$$

The term proportional to $(z - r)$ fixes the time dependent function $C(s) = 2\nu_y^{(0)}(s)$. The exponential term on the right hand side has no equivalent term on the left hand side. It varies on a small length scale inside the boundary layer, while the terms on the left hand side originate from the outer solution, varying on a larger length scale. Therefore the exponential term can not be taken into account and the term proportional to $\alpha$ fixes the boundary value

$$Q_0^{(1)}(\theta, s) = \alpha\nu_y^{(0)}(s). \tag{2.27}$$

This yields the central result of our theory: We have reduced the colored-noise problem Eq. (2.1) to the solution of a one-dimensional FPE (Eq. (2.11)) with the time-dependent boundary condition

$$\tilde{P}(\theta, s) = k\alpha\nu_y^{(0)}(s), \tag{2.28}$$

where the time-dependent flux $\nu_y^{(0)}(s)$ is obtained from the solution of the corresponding white-noise problem. The boundary condition can be understood intuitively: the filtered noise slows down the diffusion at the absorbing boundary with increasing $k$; the noise spectrum and the velocity of $y$ are bounded, so in contrast to the white-noise case, there can be a finite density at the boundary. Its magnitude results from a momentary equilibrium of the escape crossing the boundary and the flow towards it, where the latter is approximated by the flow $\nu_y^{(0)}$ at threshold in the white-noise system. In principle one can obtain colored-noise solutions combining the effective system Eq. (2.11) with the time-dependent colored-noise boundary condition Eq. (2.28), which is explicitly shown in (Schuecker et al., 2014). Here we aim for an even further reduction, recasting this time-dependent boundary condition into a static condition by considering a shift in the threshold

$$\tilde{\theta} = \theta + k\frac{\alpha}{2}. \tag{2.29}$$

We perform a Taylor expansion of the effective density

$$\tilde{P}(\tilde{\theta}, s) = \tilde{P}(\theta, s) + k\frac{\alpha}{2}\partial_y\tilde{P}(\theta, s) + O(k^2) = O(k^2) \tag{2.30}$$

using Eq. (2.28) and $\partial_y\tilde{P}(\theta, s) = -2\nu_y^{(0)}(s) + O(k)$ to show that the density vanishes to order $k^2$. As a result, to first order in $k$ the dynamic boundary condition Eq. (2.28) can be rewritten as a perfectly absorbing (white-noise) boundary at shifted $\tilde{\theta}$ Eq. (2.29). For the particular problem of the stationary MFPT this was already found by Kłosek & Hagan (1998) and Fourcaud & Brunel (2002) as a corollary deduced from the steady state density rather than as the result of a generic reduction of a colored- to a white-noise problem. It remained unclear from these earlier works, whether and how it is possible to apply the effective FPE to time-dependent problems. We show above that the effective FPE Eq. (2.11) that directly follows from a perturbation expansion of the flux operator can be re-summed to an effective operator acting on a one-dimensional density. This, together with the effective boundary condition

Eq. (2.28) or Eq. (2.29), is the crucial step to reduce the time-dependent problem driven by colored noise to an effective time-dependent white-noise problem, rendering the transient properties of the system accessible to an analytical treatment.

### 2.2.3 Boundary condition at reset

Certain physical systems, such as biological membranes, exhibit a reset to a smaller value by assigning $y \leftarrow R$ after $y$ has escaped over the threshold, while leaving the noise variable $z$ unchanged. Note that the noise density at reset therefore does not follow the marginalized density shown in Fig. 2.2. Instead, due to the fire and reset rule, it is biased to strictly positive noise values and in the limit of weak noise its approximate form was calculated for the PIF neuron (Lindner, 2004). Our reset condition corresponds to the flux in $y$-direction escaping at threshold $\theta$ being re-inserted at reset $R$. The corresponding boundary condition is

$$\begin{aligned} \nu_y(\theta, z, s) &= \frac{z}{k} Q(\theta, z, s) \\ &= S_y \left( Q(R+, z, s) - Q(R-, z, s) \right). \end{aligned}$$

With the half-range boundary condition $\frac{z}{k}\, Q(\theta, z, s) = 0 \; \forall z < 0$ we get

$$\left( f(R, z, s) - f(\theta, z, s) + \frac{z}{k} \right) \left( Q(R+, z, s) - Q(R-, z, s) \right) = 0 \quad \forall z < 0,$$

from which we conclude

$$0 = Q(R+, z, s) - Q(R-, z, s) \quad \forall z < 0. \tag{2.31}$$

We now consider the boundary layer at reset described with the rescaled coordinate $r = \frac{y-R}{k}$. We define the boundary layer solution

$$\begin{aligned} Q^B(r, z, s) &= Q^+(y(r), z, s) - Q^-(y(r), z, s) \\ &\equiv \Delta Q(y(r), z, s), \end{aligned}$$

where we introduced two auxiliary functions $Q^+$ and $Q^-$: here $Q^+$ is a continuous solution of Eq. (2.5) on the whole domain and, above reset, agrees to the solution that obeys the boundary condition at reset. Correspondingly, the continuous solution $Q^-$ agrees to the searched-for solution below reset. With Eq. (2.31) it follows $Q^B(0, z, s) = 0 \quad \forall z < 0$. Thus we have the same boundary condition as in the boundary layer at threshold Eq. (2.17). Therefore the calculations Eq. (2.19) - Eq. (2.28) can be performed analogously, where one has to use the continuity of the white-noise solution at reset, i.e. $\Delta Q^{(0)}(R, s) = 0$. The result is

$$\Delta Q_0^{(1)}(R, s) = \tilde{P}(R+, s) - \tilde{P}(R-, s) = \alpha \nu_y^{(0)}(s), \tag{2.32}$$

a time-dependent boundary condition for the jump of the first-order correction of the outer solution at reset proportional to the probability flux $\nu_y^{(0)}$ of the white-noise system. Considering a shifted reset $\tilde{R} = R + k\frac{\alpha}{2}$, a Taylor expansion again reveals that to first order in $k$ Eq. (2.32) is equivalent to a white-noise boundary condition

$$\tilde{P}(\tilde{R}+, s) - \tilde{P}(\tilde{R}-, s) = O(k^2)$$

where we used that the flux at threshold $\nu_y^{(0)}(\theta, s) = -\frac{1}{2}\partial_y Q^{(0)}(\theta, s)$ is reinserted at reset, e.g $\partial_y Q^{(0)}(\theta, s) = \partial_y Q^{(0)}(y, s)\Big|_{R-}^{R+}$. The extension to the reset boundary condition is a further generalization of the theory making it applicable to systems that physically exhibit a reset, such as excitable biological membranes. Besides, it is a standard computational procedure to re-insert the trajectories directly after escape at a reset value in order to derive the stationary escape rate ((Farkas, 1927), reviewed in (Hänggi et al., 1990)).

## 2.3 LIF-neuron

We now apply the theory to the LIF model exposed to filtered synaptic noise. The corresponding system of coupled differential equations (Fourcaud & Brunel, 2002)

$$\begin{aligned} \tau_{\mathrm{m}}\dot{V} &= -V + I + \mu \\ \tau_{\mathrm{s}}\dot{I} &= -I + \sigma\sqrt{\tau_{\mathrm{m}}}\,\xi \end{aligned} \tag{2.33}$$

describes the evolution of the membrane potential $V$ and the synaptic current $I$ driven by a Gaussian white noise with mean $\mu$ and variance $\propto \sigma^2$. The equations are identical to Eq. (1.1) and Eq. (1.2) except that the membrane resistance has been absorbed into the current and the resting potential into the membrane potential (for details see Sec. 3.1). Moreover, the synaptic spike events have been replaced by white noise as a result of a diffusion approximation (Ricciardi et al., 1999). This will be discussed in detail in Chap. 3. Note that the system Eq. (2.33) can be obtained from Eq. (2.1) by introducing the coordinates

$$y = \frac{V-\mu}{\sigma};\ z = \frac{k}{\sigma}I;\ s = t/\tau_{\mathrm{m}} \tag{2.34}$$

and the choice of $f(y, s) = -y$ as a linear function. The two-dimensional Fokker-Planck equation reads

$$\frac{\partial P(V, I, t)}{\partial t} = \left(-\frac{1}{\tau_{\mathrm{m}}}\partial_V(-V + I + \mu) + \frac{1}{\tau_{\mathrm{s}}}\partial_I I + \frac{\sigma^2\tau_{\mathrm{m}}}{2\tau_{\mathrm{s}}^2}\partial_I^2\right) P(V, I, t) \tag{2.35}$$

corresponding to Eq. (2.2) and explicitly derived in Sec. A.1 for completeness.

In the first subsection we consider the static case showing how our central result simplifies the derivation of the earlier obtained stationary firing rate. Hereby we introduce an operator

notation exploiting the analogy of the LIF neuron model to the quantum harmonic oscillator. In the second subsection we combine this operator notation and our general result Eq. (2.30) to study the effect of synaptic filtering on the transfer of a time-dependent modulation of the input to the neuron to the resulting time-dependent modulation of its firing rate, characterized to linear order by the transfer function.

### 2.3.1 The LIF model: the harmonic oscillator of neuroscience

For the LIF neuron model the white-noise Fokker-Planck equation (2.11) is

$$\begin{aligned} \partial_s \rho(x,s) &= -\partial_x \Phi(x,s) \equiv \mathcal{L}_0\, \rho(x,s) \\ \Phi(x,s) &= -(x+\partial_x)\, \rho(x,s), \end{aligned} \tag{2.36}$$

where we used $x = \sqrt{2}y$, the transformations Eq. (2.34), and the density $\rho(x,s) \equiv \frac{1}{\sqrt{2}}\tilde{P}(x/\sqrt{2},s)$. The Fokker-Planck operator $\mathcal{L}_0$ is not Hermitian. However, we can transform the operator to a Hermitian form, for which standard solutions are available. We therefore follow Risken (1996, p. 134, eq. 6.9) and apply a transformation that is possible whenever the Fokker-Planck equation possesses a stationary solution $\bar{\rho}_0$ (here $\bar{\rho}_0 = e^{-\frac{x^2}{2}}$ is the stationary solution of Eq. (2.36) if threshold and reset are absent). The square root of the stationary solution $u(x) = e^{-\frac{1}{4}x^2}$ fulfills the following relations

$$\begin{aligned} \partial_x u(x) \circ &= u(x) \underbrace{(-\frac{1}{2}x + \partial_x)}_{\equiv -a^\dagger} \circ \\ (x+\partial_x) u(x) \circ &= u(x) \underbrace{(\frac{1}{2}x + \partial_x)}_{\equiv a} \circ, \end{aligned} \tag{2.37}$$

where we defined the operators $a$ and $a^\dagger$, which fulfill the commutation relation $[a, a^\dagger] = 1$. Writing $\rho(x,s) = u(x)\, q(x,s)$, the flux operator $\Phi$ and the Fokker-Planck operator $\mathcal{L}_0$ from Eq. (2.36) transform to

$$\begin{aligned} \Phi\, u \circ = (-x - \partial_x)\, u \circ &= -ua \circ \\ \mathcal{L}_0\, u \circ = \partial_x (x+\partial_x)\, u \circ &= -ua^\dagger a \circ \end{aligned} \tag{2.38}$$

which yields for the Fokker Planck equation

$$\begin{aligned} \partial_s q(x,s) &= -a^\dagger a\, q(x,s) \\ &= (\partial_x^2 - \frac{1}{4}x^2 + \frac{1}{2})\, q(x,s). \end{aligned} \tag{2.39}$$

The right hand side is the Hamiltonian of the quantum harmonic oscillator. Note, however, that the $i\hbar$ is missing on the left hand side. The operator now is Hermitian and the eigenfunctions of $a^\dagger a$ form a complete orthogonal set.

We denote the stationary solution of (2.39) as $q_0$ obeying $a^\dagger a q_0 = 0$. Then, higher eigenfunctions are obtained by repeated application of $a^\dagger$, as the commutation relation $[a, a^\dagger] = 1$ holds and hence

$$\begin{aligned} a^\dagger a (a^\dagger)^n q_0 &= a^\dagger (a^\dagger a + [a, a^\dagger])(a^\dagger)^{n-1} q_0 \\ &= a^\dagger (a^\dagger a + 1)(a^\dagger)^{n-1} q_0 \\ &= \dots \\ &= (a^\dagger)^n (a^\dagger a + n) q_0 = n (a^\dagger)^n q_0. \end{aligned} \tag{2.40}$$

So $a^\dagger$ is the ascending, $a$ the descending operator and $(a^\dagger)^n q_0$ is the $n$-th eigenstate of the system (2.39) as in the quantum harmonic oscillator.

### 2.3.2 Stationary solution

In the stationary case, the probability flux between reset $x_R$ and threshold $x_\theta$ (with $x_{\{\theta,R\}} = \sqrt{2}\frac{\{\theta,R\}-\mu}{\sigma}$) is constant ($\tau_\mathrm{m}\nu_0$), whereas it vanishes below $x_R$ and above $x_\theta$. With (2.38) the flux takes the form

$$-ua\, q_0 = \tau_\mathrm{m}\nu_0\, H(x - x_R) H(x_\theta - x). \tag{2.41}$$

The latter equation can be solved by standard methods given the white-noise boundary conditions $0 = uq_0(x_\theta) = uq_0(x_{R+}) - uq_0(x_{R-})$, The normalization of the resulting stationary density then already leads to the known result for the firing rate (Siegert, 1951; Ricciardi, 1977), explicitly derived in Sec. A.2. It yields

$$(\tau_\mathrm{m}\nu_0)^{-1} = \int_{x_R}^{x_\theta} u^{-2}(x) F(x)\, dx, \tag{2.42}$$

with $F = \sqrt{\frac{\pi}{2}}(1 + \mathrm{erf}(\frac{x}{\sqrt{2}}))$.
In order to obtain the correction of the firing rate due to colored noise we only have to shift the boundaries in this expression according to our general result Eq. (2.29), i.e.

$$\{\theta, R\} \to \{\theta, R\} + \sqrt{\tau_\mathrm{s}/\tau_\mathrm{m}}\frac{\alpha}{2}, \tag{2.43}$$

which yields

$$(\tau_\mathrm{m}\nu)^{-1} = \int_{x_R+\frac{\alpha k}{\sqrt{2}}}^{x_\theta+\frac{\alpha k}{\sqrt{2}}} u^{-2}(x) F(x)\, dx, \tag{2.44}$$

identical to the earlier found solution (Brunel et al., 2001; Fourcaud & Brunel, 2002). In contrast to these works, we here obtain the expression without any explicit calculation, since the shift of the boundaries emerges from the generic reduction from colored to white noise (Sec. 2.2).
The analytical expression Eq. (2.44) is in agreement with direct simulations up to $k = \sqrt{0.1}$

(Fig. 2.3A). All simulations were carried out with NEST (Eppler et al., 2015). The firing rate shows a dependence on $\sqrt{\tau_\mathrm{s}}$ , which becomes obvious from a Taylor expansion of Eq. (2.44) in $k$. The figure shows additional simulations in which the threshold and the reset parameter in the neuron model are shifted according to

$$\{\theta, R\} \to \{\theta, R\} - \sqrt{\tau_\mathrm{s}/\tau_\mathrm{m}}\frac{\alpha}{2}, \tag{2.45}$$

counteracting the shift Eq. (2.29) caused by the synaptic filtering. As predicted by the theory the firing rate approximately stays constant as the synaptic time constant is increased.

In conclusion the leading order correction to the stationary rate or equivalently the MFPT stems only from the correct treatment of the boundary conditions. In order to show that this effect is not due to the simplified assumption of a perfectly absorbing boundary, we next consider the case of a continuous boundary of the potential in Fig. 2.3B. For concreteness we choose the exponential integrate-and-fire neuron model. For a steep falloff of the potential we still observe a $\sqrt{\tau_\mathrm{s}}$ dependence of the firing rate and more importantly are able to keep the firing rate constant as we shift the location of the boundaries, indicating that our central result holds true if the condition of a perfectly absorbing threshold is relaxed. It is therefore adequate to approximate the escape process by the system reaching a fixed threshold. Close to this point, the density resembles the boundary layer solution for the hard potential. For smoother potentials, the boundary layer will successively be softened and its correction to order $\sqrt{\tau_\mathrm{s}/\tau_\mathrm{m}}$ diminishes (Doering et al., 1988). This is in line with the recent observation that the dominant correction to the transfer function of the exponential integrate-and-fire model with a smooth falloff is of order $\tau_\mathrm{s}/\tau_\mathrm{m}$ (Alijani & Richardson, 2011).

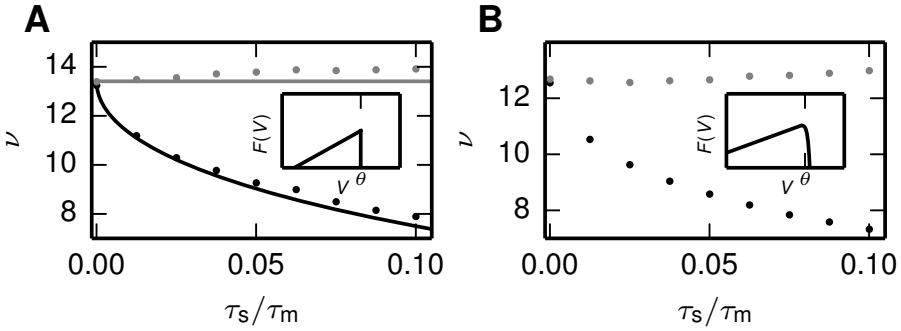



**Figure 2.3: Dependence of stationary firing rate of LIF model on synaptic filtering.** **A** Analytical prediction Eq. (2.44) of stationary firing rate $\nu$ (solid) in comparison to direct simulation (dots) for $V_\theta = 20\,\mathrm{mV}$, $V_r = 15\,\mathrm{mV}$, $\mu = 16.42\,\mathrm{mV}$, $\sigma = 4\,\mathrm{mV}$, $\tau_\mathrm{m} = 20\,\mathrm{ms}$ (black) and shifted threshold and reset $\{V_\theta, V_R\} \to \{V_\theta, V_R\} - \sigma\sqrt{\tau_\mathrm{s}/\tau_\mathrm{m}}\frac{\alpha}{2}$ (gray). Enlargement of potential $F(V) = \frac{1}{2}V^2$ close to perfectly absorbing boundary at $\theta$ shown in inset. **B** Stationary firing rate $\nu$ for same parameters but with smooth boundary shown in inset with $F(V) = \frac{1}{2}V^2 - \Delta_T^2 e^{(V-V_\theta)/\Delta_T}$ and $\Delta_T = 0.01\,\mathrm{mV}$. Spike time $t_\mathrm{s}$ defined as $V(t_\mathrm{s}) = 30\,\mathrm{mV}$.

### 2.3.3 Transfer function

We now derive a novel first order correction for the input-output transfer function valid up to moderate frequencies, which in the case of the PIF model neuron has been shown to vanish (Fourcaud & Brunel, 2002). This complements the earlier obtained limit for the modulation at high frequencies (Fourcaud & Brunel, 2002; Brunel et al., 2001).

We first simplify the derivation for white noise (Brunel & Hakim, 1999; Lindner & Schimansky-Geier, 2001) by exploiting the analogies to the quantum harmonic oscillator introduced in Sec. 2.3.1 and then study the effect of colored noise. In the following we summarize the derivation including the main intermediate results. A detailed derivation is given in A.3.

To linear order a periodic input with $\mu(t) = \mu + \epsilon\mu\, e^{i\omega t}$ and $\sigma^2(t) = \sigma^2 + H\sigma^2 e^{i\omega t}$ modulates the firing rate $\nu_0(t)/\nu_0 = 1 + n(\omega)e^{i\omega t}$, proportional to the transfer function $n(\omega)$ to be determined. With a perturbation ansatz for the modulated density $\rho(x,s) = \rho_0(x) + \rho_1(x,s)$ and the separation of the time dependent part, $\rho_1(x)\, e^{i\omega\tau s}$ follows a second order ordinary linear differential equation $i\omega\tau_{\rm m}\,\rho_1 = \mathcal{L}_0\rho_1 + \mathcal{L}_1\rho_0$ (see Eq. (A.15)). Here $\mathcal{L}_1 = -G\,\partial_x + H\partial_x^2$ is the perturbation operator, the first term of which originates from the periodic modulation of the mean input with $G = \sqrt{2}\epsilon\mu/\sigma$, the second terms stems from the modulation of the variance. In operator notation the perturbed FPE transforms to (see Eq. (A.17))

$$(i\omega\tau_{\rm m} + a^\dagger a)\, q_1 = (G\, a^\dagger + H\,(a^\dagger)^2)\, q_0$$

with the particular solution (see Eq. (A.18))

$$q_p = \frac{G}{1 + i\omega\tau_{\rm m}}\, a^\dagger q_0 + \frac{H}{2 + i\omega\tau_{\rm m}}\,(a^\dagger)^2 q_0,$$

where we used the properties of the ladder operators $a^\dagger a(a^\dagger)q_0 = (a^\dagger)q_0$ and $a^\dagger a(a^\dagger)^2 q_0 = 2(a^\dagger)^2 q_0$. We observe that the variation of $\mu$ contributes the first excited state, the modulation of $\sigma^2$ the second. The homogeneous solution $q_h$ can be expressed as a linear combination of parabolic cylinder functions $U(i\omega\tau_{\rm m} - \frac{1}{2}, x)$, $V(i\omega\tau_{\rm m} - \frac{1}{2}, x)$ (Abramowitz & Stegun, 1974; Lindner & Schimansky-Geier, 2001).

For white noise, the boundary condition on $q_1 = q_h + q_p$ is

$$\begin{aligned} 0 &= q_1(x_\theta) = q_1(x_{R+}) - q_1(x_{R-}) \\ &= q_1(x)|_{\{x_R, x_\theta\}}\,, \end{aligned} \tag{2.46}$$

where we used the shorthand notation introduced in Eq. (A.23). Moreover, a boundary conditions for the derivative follows from considering the probability flux due to the perturbation. It can be expressed with the transfer function as a sum of two contributions, the first resulting from the unperturbed flux operator acting on the perturbed density, the second from the perturbed operator acting on the unperturbed density and consistently neglecting the term of

second order, i.e.

$$\tau_{\mathrm{m}}\nu_0\, n(\omega) = u\left(-aq_1 + (G + H\, a^\dagger)q_0\right)\Big|_{\{x_R,x_\theta\}}. \tag{2.47}$$

Knowing the particular solution yields four conditions, for the homogeneous solution $q_h|_{\{x_R,x_\theta\}}$ (Sec. A.3.3) and its derivative $\partial_x q_h|_{\{x_R,x_\theta\}}$ (Sec. A.3.4) that determine the homogeneous solution on the whole domain as well as the transfer function arising from the solvability condition as (Sec. A.3.5)

$$n(\omega) = \underbrace{\frac{G}{1+i\omega\tau_{\mathrm{m}}}\,\frac{\Phi'_\omega|_{x_\theta}^{x_R}}{\Phi_\omega|_{x_\theta}^{x_R}}}_{\equiv n_G(\omega)} + \underbrace{\frac{H}{2+i\omega\tau_{\mathrm{m}}}\,\frac{\Phi''_\omega|_{x_\theta}^{x_R}}{\Phi_\omega|_{x_\theta}^{x_R}}}_{\equiv n_H(\omega)}, \tag{2.48}$$

where $x_{\{R,\theta\}} = \sqrt{2}\frac{\{V_R,V_\theta\}-\mu}{\sigma}$ and we introduced $\Phi_\omega(x) = u^{-1}(x)\,U(i\omega\tau_{\mathrm{m}} - \frac{1}{2}, x)$ as well as $\Phi'_\omega = \partial_x\Phi_\omega$ to obtain the known result (Brunel & Hakim, 1999; Brunel et al., 2001; Lindner & Schimansky-Geier, 2001).

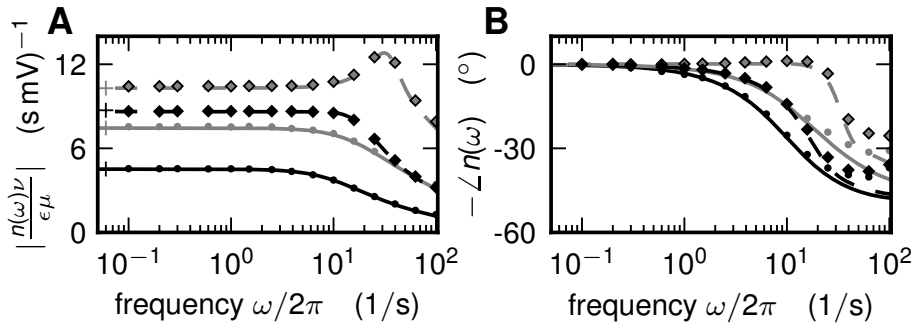



**Figure 2.4: Colored-noise transfer function of LIF model in different regimes.** Absolute value **A** and phase **B** of the transfer function (vertical) for $V_\theta = 20\,\mathrm{mV}$, $V_r = 15\,\mathrm{mV}$, $\tau_m = 20\,\mathrm{ms}$, $\tau_{\mathrm{s}} = 0.5\,\mathrm{ms}$, $\sigma = 4\,\mathrm{mV}$ (solid), $\sigma = 1.5\,\mathrm{mV}$ (dashed) as a function of frequency (log-scaled horizontal axis). The mean input $\mu$ was adapted to obtain different firing rates $\nu = 10\,\mathrm{Hz}$ (black) and $\nu = 30\,\mathrm{Hz}$ (gray). Analytical prediction $\tilde{n}_G$ (solid curves, (2.49)), direct simulations (dots, diamonds), and zero frequency limit $\frac{d\nu}{d\mu}$ (crosses, obtained from (2.44)).

With the general theory developed above we directly obtain an approximation for the colored-noise transfer function by replacing $x_{\{R,\theta\}} \to x_{\{\tilde{R},\tilde{\theta}\}}$ in the white-noise solution Eq. (2.48), denoted by

$$\tilde{n}_G(\omega) = \frac{G}{1+i\omega\tau_{\mathrm{m}}}\,\frac{\Phi'_\omega|_{x_{\tilde{\theta}}}^{x_{\tilde{R}}}}{\Phi_\omega|_{x_{\tilde{\theta}}}^{x_{\tilde{R}}}}. \tag{2.49}$$

Here we only consider a modulation of the mean $\mu$, because it dominates the response properties. We discuss a modulation of the variance $\sigma^2$ in A.4. Note that here the modulation enters the equation for $V$. If one is interested in the linear response of the system with respect to a perturbation of the input to $I$, as it appears in the neural context due to synaptic input, one needs to take into account the additional low pass filtering $\propto (1+i\omega\tau_{\mathrm{s}})^{-1}$ (see also Chap. 5),

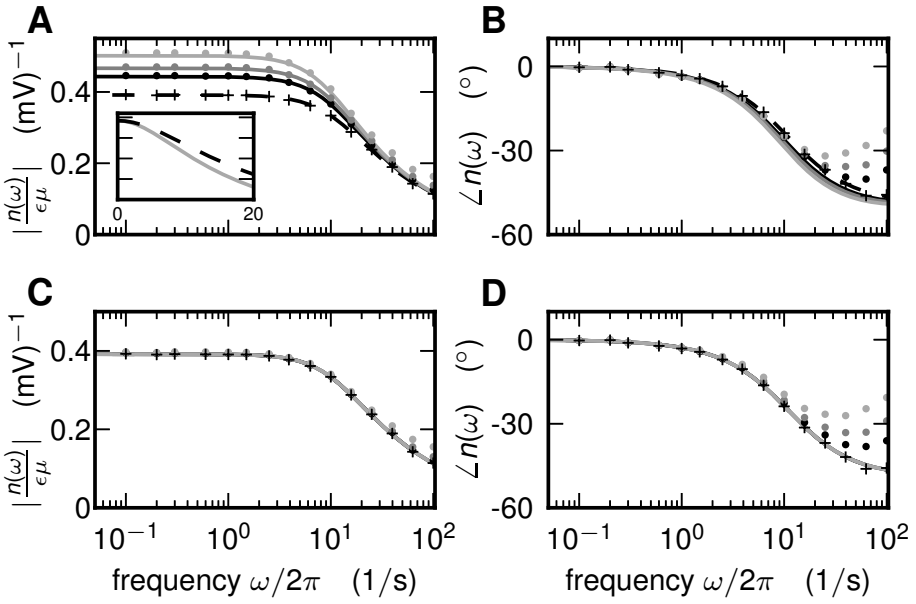



**Figure 2.5: Comparison between colored- and white-noise transfer function.** Upper row: Absolute value **A** and phase shift **B** for $\sigma = 4\,\mathrm{mV}$, white noise (dashed), colored noise $\tau_\mathrm{s} \in [0.5, 1, 2]$ ms (from black to gray) and $\tau_\mathrm{s} = 2$ ms normalized to zero frequency limit of white noise (gray, inset). Lower row **C**,**D**: same as **A**,**B** but threshold and reset shifted $\{\theta, R\} \to \{\theta, R\} - \sqrt{\tau_\mathrm{s}/\tau_\mathrm{m}}\frac{\alpha}{2}$ to maintain constant firing rate. Display and other parameters as in Fig. 2.4.

which is trivial.

A Taylor expansion of the resulting function $\tilde{n}_G$ around the original boundaries $x_{\{R,\theta\}}$ reveals the first order correction in $k$

$$\begin{aligned} n_{\mathrm{cn,G}}(\omega) =& n_G(\omega) \\ &+ \sqrt{\frac{\tau_\mathrm{s}}{\tau_\mathrm{m}}}\frac{\alpha}{\sqrt{2}}\,\frac{G}{1+i\omega\tau_\mathrm{m}}\left(\frac{\Phi''_\omega|^{x_R}_{x_\theta}}{\Phi_\omega|^{x_R}_{x_\theta}} - \left(\frac{\Phi'_\omega|^{x_R}_{x_\theta}}{\Phi_\omega|^{x_R}_{x_\theta}}\right)^2\right), \end{aligned} \tag{2.50}$$

valid for arbitrary noise intensity $\sigma$ entering the expression via the boundaries $x_{\{R,\theta\}}$. The first correction term is similar to the $H$-term in Eq. (2.48) indicating that colored noise has a similar effect on the transfer function as a modulation of the variance (Lindner & Schimansky-Geier, 2001) in the white-noise case. In the high-frequency limit this similarity was already found: modulation of the variance leads to finite transmission at infinite frequencies in the white-noise system (Lindner & Schimansky-Geier, 2001). The same is true for modulation of the mean in the presence of filtered noise (Brunel et al., 2001). However, for infinite frequencies our analytical expression behaves differently. The two correction terms in the second line of Eq. (2.50) cancel each other, since $\Phi''_\omega \to (i\omega\tau_\mathrm{m})^2\,\Phi_\omega$ and $\Phi'^2_\omega \to (i\omega\tau_\mathrm{m})^2\,\Phi^2_\omega$. Thus the transfer function decays to zero as in the white-noise case. This deviation originates from neglecting the time derivative on the left hand side of Eq. (2.16) that is of order $k^2$, which is true only up

to moderate frequencies $\omega\tau_{\mathrm{m}}k \ll 1$. The high frequency limit $\omega\tau_{\mathrm{m}}k \gg 1$ is known explicitly Brunel et al. (2001), but it was shown that the finite high-frequency transmission is due to the artificial hard threshold in the LIF model (Naundorf et al., 2003; Fourcaud-Trocmé et al., 2003).

Fig. 2.4 shows a comparison of the analytical prediction Eq. (2.50) to direct simulations for different noise levels $\sigma$ and firing rates controlled by the mean $\mu$. The absolute value of the analytical result is in agreement with simulations for the displayed range of frequencies. Above 100Hz deviations occur as expected. However, these are less important, since our theory predicts the response properties well in the frequency range where transmission is high. The deviations are more pronounced and already observed at lower frequencies in the phase shift. At high firing rate and low noise, the neuron is mean driven and exhibits a resonance at its firing frequency, again well predicted by the analytical result. The upper row in Fig. 2.5 shows a comparison to the white-noise case. A qualitative change can be observed: synaptic filtering on the one hand increases the dc-susceptibility in contrast to the decreased firing rate (Fig. 2.3A). For $\tau_{\mathrm{s}} = 2\,\mathrm{ms}$ this increase is already 25%, implying that this correction is not negligible even for short time constants $\tau_{\mathrm{s}}$. On the other hand the cutoff frequency (inset in panel A) is reduced, a behavior which is not expected from the existing knowledge that colored noise enhances the transmission in the high frequency limit. Our theory fully explains these qualitative changes by a shift in the reset and the threshold analogously to the stationary case (Fig. 2.3). This is shown explicitly in panels C and D: for different synaptic time constants the response in the low frequency regime is not altered in comparison to the white-noise case if reset and threshold are adapted according to $\{\theta, R\} \to \{\theta, R\} - \sqrt{\tau_{\mathrm{s}}/\tau_{\mathrm{m}}}\frac{\alpha}{2}$ to compensate for the effect of colored noise.

## 2.4 Interim discussion

In the following we discuss specific issues, regarding this chapter. A general summary and a broader discussion of the results follow in Chap. 9.

As discussed in Sec. 2.3.3 the general reduction presented here is valid up to moderate frequencies, while for high frequencies the time derivative on the left hand side of Eq. (2.16) cannot be treated as a second order perturbation in $k$. The presented theory could be extended by absorbing the time derivative into the operator $\tilde{L} = L - i\omega\tau_{\mathrm{m}}$. However, this significantly complicates the form of the outer as well as the boundary-layer solution and the matching between the two. Additionally, the theory could be generalized to multiplicative noise, i.e. an arbitrary factor $g(y)$ in front of the noise $\xi$ in (2.1).

We have further seen that the transfer function consists of two components, due to a modulation of the mean input and due to a modulation of the incoming fluctuations. Their contributions to signal transmission have been studied theoretically (Lindner & Schimansky-Geier, 2001) and experimentally (Silberberg et al., 2004). The formal analogy between the

LIF neuron model and the quantum harmonic oscillator (Sec. 2.3.1) shows that the two contributions are related to the first and second excited state, which casts a new light on this known distinct contributions.

The novel analytical expression Eq. (2.49) of the transfer function is used in this thesis to study fluctuations in spiking networks (Chap. 5). Furthermore, the transmission of correlated activity by pairs of neurons exposed to common input (Shea-Brown et al., 2008; De la Rocha et al., 2007), can now be studied in a time-resolved manner. In connection with the recently proposed method of linear stability analysis of pulse-coupled oscillators on the basis of perturbations of the inter-event density (Farkhooi & van Vreeswijk, 2015), our modified boundary condition can replace the white-noise boundary condition (Farkhooi & van Vreeswijk, 2015, their eq. (9)), to obtain an approximation for the case of colored noise avoiding the need to extend their method to a two-dimensional Fokker-Planck equation.

Our results are important also for areas outside the neuroscientific scope of this thesis. Excitable systems are ubiquitous in nature and a large body of literature discusses their properties. In particular the escape from a meta-stable state activated by colored noise appears in diverse areas of the natural sciences, including semiconductor physics, physical chemistry, genetics, hydrodynamics and laser physics (Moss & McClintock, 1989a,b). The simplicity of the presented method, reducing this general problem to the well understood escape from a potential driven by white-noise, but with displaced boundaries, opens a wide field of applications, where the effect of realistic colored noise can now be accessed in a straight-forward manner.

# Chapter 3

# Stationary activity in spiking neuronal networks

In this section we set up mean-field framework, which allows the prediction of stationary-state firing rates in a network of spiking neurons. We introduce the mathematical notation and subsequently apply the framework first to the $\mathcal{E}-\mathcal{I}$ network and then to the microcircuit model. The framework combines the gain function derived in Chap. 2 with closure assumption and is based on earlier works for simple networks (Amit & Brunel, 1997b; Brunel & Hakim, 1999; Brunel, 2000; Wong & Wang, 2006). At the end of this chapter the mean-field approach is applied to more complex spiking networks which forms the basis to analyze the stationary states and global stability of the multi-area model in Chap. 4. Moreover, the limits of the framework are tested in Chap. 6.

## 3.1 Mean-field solution

We consider a recurrent network of $N$ LIF neuron models. Each neuron model $i$ obeys the coupled set of differential equations

$$\begin{aligned} \tau_{\mathrm{m}}\frac{dV_i}{dt} &= -(V_i - E_{\mathrm{L}}) + R_{\mathrm{m}}I_i(t) \\ \tau_s\frac{dI_i}{dt} &= -I_i + \tau_{\mathrm{s}}\sum_{j=1}^{N} w_{ij}\, s_j(t-d) \end{aligned} \tag{3.1}$$

with the presynaptic spike trains $s_j(t) = \sum_k \delta(t - t_j^k)$, where the $t_j^k$ mark the time points at which neuron $j$ emits an action potential, i.e., where $V_j$ exceeds the threshold and $d$ denotes the synaptic delay. The synaptic efficacy is denoted by $w_{ij}$, where $i$ indicates the post- and $j$ the pre-synaptic neuron. If a connection between $j$ and $i$ is absent, $w_{ij}$ is zero.

We absorb the leak term in the membrane potential, i.e., $V \to V + E_{\mathrm{L}}$. Moreover, we use $\tau_{\mathrm{m}} = R_{\mathrm{m}}C_{\mathrm{m}}$, where $C_{\mathrm{m}}$ denotes the membrane capacitance and multiply the second line in Eq. (3.1) with $\tau_{\mathrm{m}}/C_{\mathrm{m}}$. At the same time we absorb the membrane resistance into the definition

of the current ($I \to \frac{C_\mathrm{m}}{\tau_\mathrm{m}} I$ ). Together this yields

$$\begin{aligned} \tau_\mathrm{m} \frac{dV_i}{dt} &= -V_i + I_i(t) \\ \tau_s \frac{dI_i}{dt} &= -I_i + \tau_\mathrm{m} \sum_{j=1}^{N} J_{ij}\, s_j(t-d)\,, \end{aligned} \tag{3.2}$$

where we defined the rescaled synaptic efficacy $J = \frac{\tau_s w}{C_\mathrm{m}}$ in volt. Eq. (3.2) corresponds to Eq. (2.33), but with input provided by spike trains from the other neurons in the network.

In the previous Chap. 2 we have derived the gain function, i.e., the stationary output firing rate of a single neuron when it is driven by Gaussian noise with mean $\mu$ and variance $\propto \sigma^2$. The starting point has been the Fokker-Planck equation (Eq. (A.7)), which includes a drift term ($\sim \partial_I \mu P$) due to the mean input and a diffusion term ($\sim \partial_I^2 \sigma^2 P$) due to the variance of the input. The drift term originates from the first order and the diffusion term from the second order of the so called Kramers-Moyal expansion of the joint probability distribution (Sec. A.1). The Fokker-Planck equation considers only these two terms, while higher orders in the Kramers-Moyal expansion are neglected. In the Gaussian case the Fokker-Planck equation is exact, since higher orders vanish (Sec. A.1.1).

However, in the network situation, the neuron is not driven by Gaussian noise but rather by spike trains emitted by other neurons. Moreover, due to the recurrent connections in the network, the output of the neurons likewise constitutes the input to the neurons. Therefore, the system has to be solved self-consistently. We approximate the spike times to follow Poisson statistics (Amit & Brunel, 1997b; Brunel & Hakim, 1999; Brunel, 2000), because for networks operating in the fluctuation driven regime (Sec. 1.2.2) the spike emission is irregular. In the following we show how this approximation on the one hand allows a mapping to the Gaussian case and on the other hand closes the system.

We first consider the situation of one neuron receiving Poisson input with rate $\nu_\mathrm{ext}$ and synaptic efficacy $J$, which constitutes a shot-noise process. The derivation of the Fokker-Planck equation (A.1) reveals that the drift and diffusion term can be mapped to the Gaussian case by identifying

$$\begin{aligned} \mu &= \tau_\mathrm{m} J \nu_\mathrm{ext} \\ \sigma^2 &= \tau_\mathrm{m} J^2 \nu_\mathrm{ext}\,, \end{aligned} \tag{3.3}$$

resulting in a Fokker-Planck identical to the Gaussian case. Here, the diffusion term accounts for fluctuations in the input due to the Poisson process. As previously mentioned, the Fokker-Planck equation neglects higher order terms in the Kramers-Moyal expansion. In contrast to the Gaussian case, these terms do not vanish for Poisson input with finite efficacy $J$ and thus the Fokker-Planck equation is an approximation, also called diffusion approximation. This approximation is justified, if the synaptic weight is small compared to the length scale on which the probability distribution varies, thus $J \ll V_\theta - V_\mathrm{R}$.

In summary, Gaussian noise and Poisson shot noise lead to the same Fokker-Planck equation, which is an approximation for the latter case. This correspondence is often used in an ad-hoc manner by directly replacing the spike input $\tau_\mathrm{m} \sum_{j=1}^{N} J_{ij}\, s_j(t-d)$ in Eq. (3.2) by Gaussian white noise $\mu + \sigma\sqrt{\tau_\mathrm{m}}\,\xi$.
The term 'mean-field' originates from the similarity to the mean-field theory employed for example in spin-glasses (Kirkpatrick & Sherrington, 1978): In a field-theoretical framework the replacement of the fluctuating input with Gaussian noise corresponds to a saddle-node approximation for an auxiliary field which replaces the fluctuations in the field by their mean. This replacement, in turn leads to a Gaussian noise on the level of the dynamic variables (see Chap. 6). However, for spiking dynamics a field-theoretical framework does not exist and the term 'mean-field' is used in an ad-hoc manner.

**Table 3.1:** Specification of the neuron parameters. If not stated otherwise, these are the default parameters for all networks of LIF model neurons considered in this thesis. In contrast to Fig. 1.2, the resting potential is zero and the threshold is positive here, because these are displaced according to a shift in the membrane potential $V \to V + E_\mathrm{L}$ (Sec. 3.1).

| **Neuron model parameters** | | |
|---|---|---|
| **Name** | **Value** | **Description** |
| $\tau_\mathrm{m}$ | 10 ms | membrane time constant |
| $\tau_\mathrm{r}$ | 2 ms | absolute refractory period |
| $\tau_\mathrm{s}$ | 0.5 ms | postsynaptic current time constant |
| $C_\mathrm{m}$ | 250 pF | membrane capacity |
| $V_\mathrm{R}$ | 0 mV | reset potential |
| $V_\theta$ | 15 mV | firing threshold |

### 3.1.1 Single-neuron level

We now consider a neuron embedded in a network, where its input is given by the output of the other neurons. Since we assume the output spike trains to follow Poisson statistics, we can describe the input due to one pre-synaptic neuron with Eq. (3.3) where we replace $\nu_\mathrm{ext}$ by the rate $\nu_j$ of the pre-synaptic neuron $j$. We further neglect cross-correlations between neurons so that the input statistics from different pre-synaptic neurons is given by a sum resulting in the mean and the variance of each neuron $i$

$$\begin{aligned} \mu_i &= \tau_\mathrm{m} \sum_j J_{ij}\nu_j + \mu_{i,\mathrm{ext}} \\ \sigma_i^2 &= \tau_\mathrm{m} \sum_j J_{ij}^2\nu_j + \sigma_{i,\mathrm{ext}}^2, \end{aligned} \tag{3.4}$$

where $\mu_{i,\mathrm{ext}}$ and $\sigma_{i,\mathrm{ext}}$ are due to external drive later to be specified.

Given the input parameters Eq. (3.4), the non-linear gain function (Sec. 2.3.2) provides the output rate $\nu_i^\mathrm{out}$ of neuron $i$. Due to recurrent connections the neurons project back into the

network and therefore the system needs to be solved self-consistently, i.e., $\nu_i^{\text{out}} = \nu_i$. The self-consistent rates are given by

$$\begin{aligned} \frac{1}{\nu_i} &= \tau_{\text{r}} + \tau_{\text{m}}\sqrt{\pi} \int_{\frac{V_r-\mu_i}{\sigma_i}+\gamma\sqrt{\frac{\tau_{\text{s}}}{\tau_{\text{m}}}}}^{\frac{V_\theta-\mu_i}{\sigma_i}+\gamma\sqrt{\frac{\tau_{\text{s}}}{\tau_{\text{m}}}}} e^{x^2}\,(1+\text{erf}(x))\,\text{d}x \\ &=: 1/\Phi_i(\boldsymbol{\nu})\,, \end{aligned} \tag{3.5}$$

which for one component $i$ corresponds to Eq. (2.44) in original units and in addition includes the refractory period $\tau_{\text{r}}$.

If the input statistics $\mu_i, \sigma_i$ is heterogeneous across neurons, each neuron possesses its individual firing rate $\nu_i$. Therefore we need to solve Eq. (3.5) on the level of single neurons, which is numerically difficult due to the large dimension $N \gtrsim O(10^3)$ of the problem. However, the fixed point of the dynamical system

$$\dot{\boldsymbol{\nu}} := \frac{\text{d}\boldsymbol{\nu}}{\text{d}s} = \boldsymbol{\Phi}(\boldsymbol{\nu}) - \boldsymbol{\nu} \tag{3.6}$$

is the solution to Eq. (3.5). Here, $s$ is a pseudo-time and the bold symbols denote vectors. We find the fixed point by integrating Eq. (3.6) until it converges to a stationary state. Note, that the trajectories of Eq. (3.6) do not reflect the real time evolution of the rates in the spiking network; considering this dynamical system is rather a numerical trick to find the self-consistent solution for the stationary rates. On the technical side, we implement Eq. (3.6) in NEST (Chap. 8) as a network of rate units.

### 3.1.2 Population level

We can further structure the neurons in the network into $N_{\text{pop}}$ interconnected populations. Then, we define the connectivity on a population level: a neuron $i$ in population $\alpha$ receives $K_{\alpha\beta}$ incoming connections from neurons in population $\beta$, each with synaptic efficacy $J_{\alpha\beta}$. This assumes that the neurons belonging to $\alpha$ receive exactly the same number of inputs. We further assume that the external drive is the same for all neurons in $\alpha$ rendering them homogeneous with respect to their total input.

In this situation it becomes obvious from Eq. (3.4) and Eq. (3.5) that the solution in which all neurons in one population have the same rate $\nu_\alpha$ is self-consistent: As the input parameters Eq. (3.4) follow from a sum over the columns of the connectivity we have $\mu_i = \mu_\alpha$ and $\sigma_i^2 = \sigma_\alpha^2\ \forall i \in \alpha$ leading to $\nu_i = \nu_\alpha\ \forall i \in \alpha$. Therefore, the self-consistent rates follow from Eq. (3.5) with the neuron index $i$ replaced by the population index $\alpha$ and the population

specific input parameters as

$$\begin{aligned} \nu_\alpha &= \Phi_\alpha(\boldsymbol{\nu}) \\ \mu_\alpha &= \tau_\mathrm{m} \sum_\beta K_{\alpha\beta} J_{\alpha\beta} \nu_\beta + \mu_{\alpha,\mathrm{ext}} \\ \sigma_\alpha^2 &= \tau_\mathrm{m} \sum_\beta K_{\alpha\beta} J_{\alpha\beta}^2 \nu_\beta + \sigma_{\alpha,\mathrm{ext}}^2 \,. \end{aligned} \tag{3.7}$$

To find its solution we solve Eq. (3.6) on the level of populations $N_\mathrm{pop}$. Note, that this reduction is only exact for fixed indegree $K$. If one instead realized each single connection with probability $p_{\alpha\beta} = K_{\alpha\beta}/N_\beta$ or with fixed outdegree, Eq. (3.7) can be regarded as an additional average over the connectivity.

## 3.2 Application to balanced random network

We consider the $\mathcal{E}-\mathcal{I}$ network model consisting of $N_\mathcal{E} = N$ excitatory and $N_\mathcal{I} = \gamma N$ inhibitory model neurons. The network is similar to the one studied in (Brunel, 2000) with the extension to synaptic filtering. We can define the connectivity on the population level $\alpha \in \{\mathcal{E}, \mathcal{I}\}$ as

$$\boldsymbol{K} = K_\mathcal{E} \begin{pmatrix} 1 & \gamma \\ 1 & \gamma \end{pmatrix} \; ; \boldsymbol{J} = J \begin{pmatrix} 1 & -g \\ 1 & -g \end{pmatrix} \tag{3.8}$$

with $K_\mathcal{E} = pN_\mathcal{E}$ denoting the indegree from excitatory neurons, $p$ being the connection probability. Here, $g$ denotes the ratio of the inhibitory to excitatory synaptic weight. Additionally, each neuron is driven by $K_{i,\mathrm{ext}}$ Poisson sources with rate $\nu_\mathrm{ext}$ and synaptic efficacy $J_\mathrm{ext}$. We choose the neuron parameters to be homogeneous across neurons, specified in Tab. 3.1. Due to its generic nature this architecture is studied in many occasions in this thesis under various aspects (next subsection, Sec. 5.1, Chap. 6). The various cases only differ in the definition of the external drive, which is defined in the corresponding sections.

### 3.2.1 Heterogeneous state

We introduce heterogeneity by drawing the number of external inputs from a normal distribution according to $K_{i,\mathrm{ext}} \propto \mathcal{N}(\bar{K}_\mathrm{ext}, \delta K_\mathrm{ext}^2)$ which leads to neuron specific input paramters $\mu_{i,\mathrm{ext}} = \tau_\mathrm{m} K_{i,\mathrm{ext}} J_\mathrm{ext} \nu_\mathrm{ext}$ and $\sigma_{i,\mathrm{ext}} = \tau_\mathrm{m} K_{i,\mathrm{ext}} J_\mathrm{ext}^2 \nu_\mathrm{ext}$ in Eq. (3.4). The raster plot shows heterogeneous frequencies of spike emission across neurons (Fig. 3.1A). For the mean-field prediction we integrate Eq. (3.6) on the level of the $N$ neurons which eventually converges to a fixed point. The same strategy has been recently employed by Sadeh & Rotter (2015) for a different network type.

The obtained fixed-point rates are in good agreement with the time-averaged firing rates of the corresponding simulation of the LIF-network (Fig. 3.1B). Only minor deviations occur in

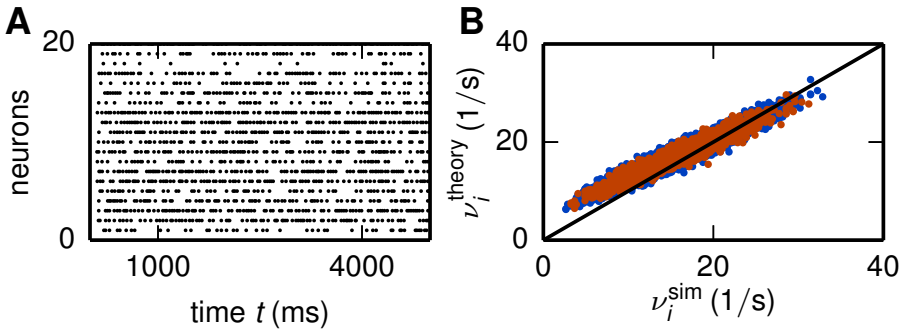



**Figure 3.1: Stationary firing rates in the excitatory-inhibitory network with heterogeneous drive. A** Raster plot (neuron id vs. spike times) for 20 excitatory neurons. **B** Scatter plot of analytic prediction of firing rates $\nu_i$ versus rates $\nu_i^{\text{sim}}$ obtained from a direct simulation of the LIF-neuron network. Excitatory neurons shown in blue, inhibitory ones in red. Network parameters are: $N_{\mathcal{E}} = 10,000$, $p = 0.1$, $\gamma = 0.25$, $g = 4.5$, $J = 0.1756\,\text{mV}$, $d = 1.5\,\text{ms}$, $J_{\text{ext}} = J$, $\bar{K}_{\text{ext}} = 1000$, $\delta K_{\text{ext}} = 50$, $\nu_{\text{ext}} = 8\,\text{Hz}$. Simulation parameters for LIF network: Simulation time $T = 5000\,\text{ms}$. Initial transient (100 ms) discarded for the calculation of the time-averaged rates.

terms of an overestimation of the smaller rates and a slight underestimation of the larger rates, resulting in an underestimation of the width of the rate distribution. This could be partly explained by the finite measurement time in the LIF-network which results in a positive bias to the measured width. The deviations can further originate from the neglected cross-correlations in the mean-field framework, as a very similar deviation was found for binary neurons, which could be explained and cured by considering cross-correlations in the input to the neurons (Dahmen et al., 2016). In principle it is also possible to consider cross-correlation in the input to the LIF-neurons (Helias et al., 2008), which could possibly be generalized to the network situation using the methods presented in Chap. 5 combined with an iterative method. However, the deviations in (Fig. 6.2B) are rather small and do not concern us at this stage.

### 3.2.2 Homogeneous state

We now choose the external inputs to be homogeneous across neurons of one population, i.e., $\mu_{\alpha,\text{ext}} = \tau_{\text{m}} K_{\alpha,\text{ext}} J_{\text{ext}} \nu_{\text{ext}}$ and $\sigma^2_{\alpha,\text{ext}} = \tau_{\text{m}} K_{\alpha,\text{ext}} J^2_{\text{ext}} \nu_{\text{ext}}$ for $\alpha \in \{\mathcal{E}, \mathcal{I}\}$. Therefore, Eq. (3.6) reduces to a two-dimensional problem. Note that only for such low-dimensional cases one could directly solve Eq. (3.7) using standard numerical methods like a simple bi-sectioning. However, we here use the integration of Eq. (3.6) as this can be easily extended to a higher number of populations, as seen in the next sub-section and in Chap. 4.

The mean-field predictions are in excellent agreement with the rates obtained from the spiking network simulations, where we varied the two main parameters of the model, namely the external drive $\nu_{\text{ext}}$ and the weight ratio $g$. Only the excitatory rates are shown. The inhibitory and excitatory rates are the same, as they receive the same input, since the rows in Eq. (3.8) are

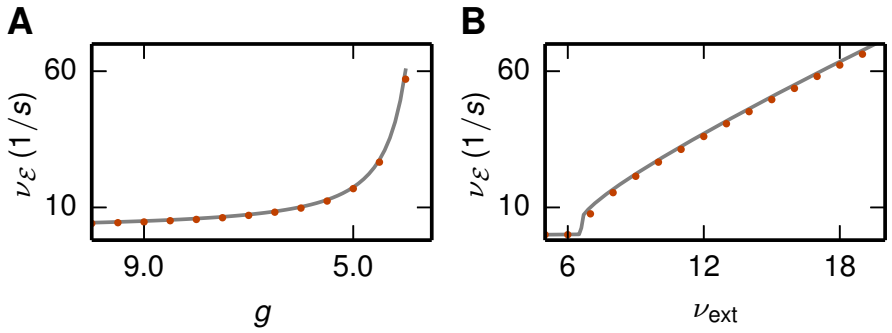



**Figure 3.2: Population-averaged firing rates in the $\mathcal{E}-\mathcal{I}$ network model. A** Analytic prediction for firing rates obtained from Eq. (3.6) with Eq. (3.7) (solid curves) and from direct simulation of the LIF-network (dots) for the excitatory population as a function of the inhibitory coupling ratio $g$. **B** Same as A for a variation of the external drive $\nu_{\text{ext}}$. Network and simulation parameters as in Fig. 3.1, except the external drive which is homogeneous here, i.e., $K_{\alpha,\text{ext}} = 1000$ for $\alpha \in \{\mathcal{E}, \mathcal{I}\}$.

equal. Therefore, one could even reduce the system to one dimension. However, we here keep the separation between excitatory and inhibitory populations due to conceptual differences between these two neurons types.

## 3.3 Application to the microcircuit model

Next we apply the framework to a more complex network. The cortical microcircuit model (Potjans & Diesmann, 2014) constitutes roughly 80000 spiking neurons structured into 8 populations across 4 layers $[L23, L4, L5, L6]$, with one excitatory and one inhibitory cell type each (Fig. 3.3A).

The model exhibits irregular and asynchronous spiking activity (Fig. 3.3C). Replacing each population by a single rate neuron (3.3B) results in an eight-dimensional rate network (Eq. (3.6)) which converges to a fixed point corresponding to the population-averaged firing rates obtained from direct simulation of the spiking model (Fig. 3.3D).

Having only eight coupled equations, the rate network can be solved considerable faster compared to the spiking network, which has $10^5$ times as many neurons. While the solution for the rates is obtained within a second on a single process, simulating the full spiking networks takes several minutes on multiple processes ($\approx$ 200s on 24 processes). This advantage allows for an efficient parameter exploration using the rate network.

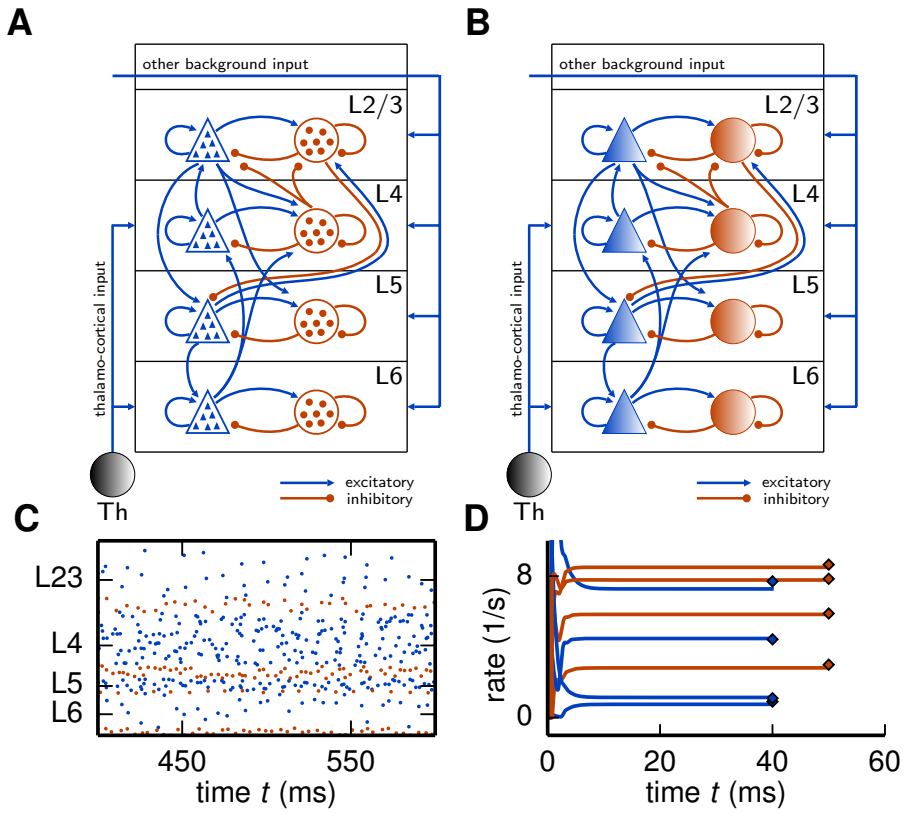



**Figure 3.3: Reduction of spiking microcircuit model to rate dynamics**. **A** Sketch of the microcircuit model (Potjans & Diesmann, 2014) with excitatory (blue triangles) and inhibitory (red circles) neuron populations, each consisting of a large number of neurons indicated by the small triangles and disks respectively. Arrows between populations indicate the in-degree $K$. **B** Sketch of the corresponding reduced model where each population is replaced by a single rate unit. **C** Spiking activity in the different layers. Only 0.1% of the neurons shown. **D** Dynamics of the eight neurons of the rate network (Eq. (3.6)) (curves) and comparison to population- and time-averaged firing rates obtained from direct simulations of the spiking network (diamonds).

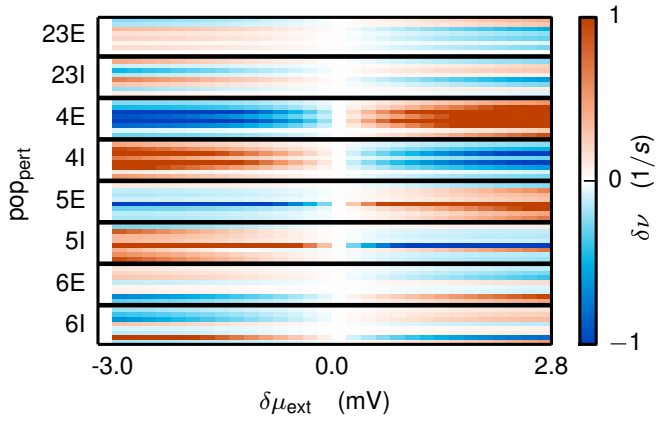



**Figure 3.4: Perturbation of the input to the microcircuit**. One block of eight rows shows the change in the stationary rate $\delta\nu$ of the populations in the microcircuit in response to a perturbation $\delta\mu_{\text{ext}}$ to population $\text{pop}_{\text{pert}}$, as specified on the y-axis. Each individual block is ordered from top to bottom according to [2/3E,2/3I,4E,4I,5E,5I,6E,6I] with black rows separating the blocks. For example, the second line in the top row shows the change in rates of population 2/3I after perturbing the external input to population 2/3E. The range of $\delta\mu_{\text{ext}}$ corresponds roughly to a change of the external rate ($r_{\text{ext}} = 8\,\text{Hz}$) of about $\pm 1\,\text{Hz}$.

A further application is to investigate the behavior of the stationary rates in the microcircuit with respect to a perturbation in the external input. We consider a perturbation of the mean external input $\mu_{\alpha,\text{ext}} \to \mu_{\alpha,\text{ext}} + \delta\mu_{\text{ext}}$ to one population and are interested in the effect of this perturbation on the firing rates of all populations in the network. For different values of $\delta\mu_{\text{ext}}$ we solve Eq. (3.6) and defined $\delta\nu$ as the induced change in the rates (Fig. 3.4). A perturbation to one population has an effect on all eight populations, where by trend a positive perturbation to an excitatory population causes an increase in the rates while the opposite is true for a perturbation of an inhibitory population (compare $\text{pop}_{\text{pert}} = 4\text{E}/4I$). However, increased input can also yield higher rates in some populations and lower ones in others (see $\text{pop}_{\text{pert}} = 6E$ or $\text{pop}_{\text{pert}} = 6I$). Another tendency is that excitatory populations are more strongly affected by perturbations than inhibitory ones. In particular, population 5E is susceptible to perturbations of populations in L4 or L5, while populations in L4 are very sensitive to perturbations in L4.

In this thesis the considered perturbation experiment serves as a show-case illustrating how our framework helps to understand the behavior of a spiking-network model. For a detailed interpretation of the results in the neuroscientific context we refer to Bos et al. (2016).

# Chapter 4

# Stationary activity and global stability in complex networks

---

The following chapter is based on the following publication:
Schuecker, J., Schmidt, M., van Albada, S., Diesmann, M., & Helias, M. (2016). Fundamental activity constraints lead to specific interpretations of the connectome. PloS Comput Biol (in press).

**Author contributions:**
Maximilian Schmidt and the author performed the work concertedly with equal shares on all parts of the study. Sacha J. van Albada, Markus Diesmann and Moritz Helias supervised the work and contributed to the writing of the manuscript.

In the previous chapter we have explained how the mean-field solution can predict the stationary state activity in spiking networks with a low number of populations. In this chapter we investigate the multi-area model, a complex network with 254 populations. The large number of populations causes an intricate phase space in which we analyze the global stability of the stationary states. In particular, we develop a mean-field based method which includes stability constraints to improve the data integration of experimentally determined structural connectivity into the model. In principle, this is possible because Eq. (3.7) provides a link between the structure $\boldsymbol{K}$ and the activity $\boldsymbol{\nu}$.

## 4.1 Introduction

The neural wiring diagram, the connectome, is gradually being filled through classical techniques combined with innovative quantitative analyses (Markov et al., 2014a,b) and new technologies (Wedeen et al., 2008; Axer et al., 2011). The connectivity between neurons is considered to shape resting-state and task-related collective activity (Cole et al., 2014; Shen et al., 2015). For simple networks, clear relationships with activity are known analytically, e.g., a dynamic balance between excitatory and inhibitory inputs in inhibition-dominated random networks leads to an asynchronous and irregular state (van Vreeswijk & Sompolinsky, 1996; Amit & Brunel, 1997a; Tetzlaff et al., 2012). Structures and mechanisms underlying large-scale interactions have been identified by means of dynamical models (Deco & Jirsa, 2012; Cabral et al., 2014) and graph theory (Markov et al., 2013; Goulas et al., 2014). Furthermore, the impact of local network structure on spike-time correlations is known in some detail (Ostojic et al., 2009; Pernice et al., 2011; Trousdale et al., 2012). Under certain conditions, there is a one-to-one mapping between correlations in neuronal network activity and effective connectivity, a measure that depends on the network structure and on its activity (van Albada et al., 2015). In conclusion, anatomy does not uniquely determine dynamics, but dynamical observations help constrain the underlying anatomy. Despite advances in understanding simple networks, a complete picture of the relationship between structure and dynamics remains elusive.

Fig. 4.1 visualizes the integrative loop between experiment, model, and theory that guides the investigation of structure and dynamics. In the traditional forward modeling approach, structural data from experimental studies determine the connectivity between model neurons. Combined with the specification of the single-neuron dynamics and synaptic interactions, simulations yield a certain network dynamics. There is a fundamental problem with this approach.

Despite the impressive experimental progress in determining network parameters, any neural network model is underdetermined, because of the sheer complexity of brain tissue and inevitable uncertainties in the data. For instance, counting synapses on a large scale presently takes a prohibitive amount of time, and no available technique allows determining precise synaptic weights for entire neural populations. Although it may be possible to quantify the

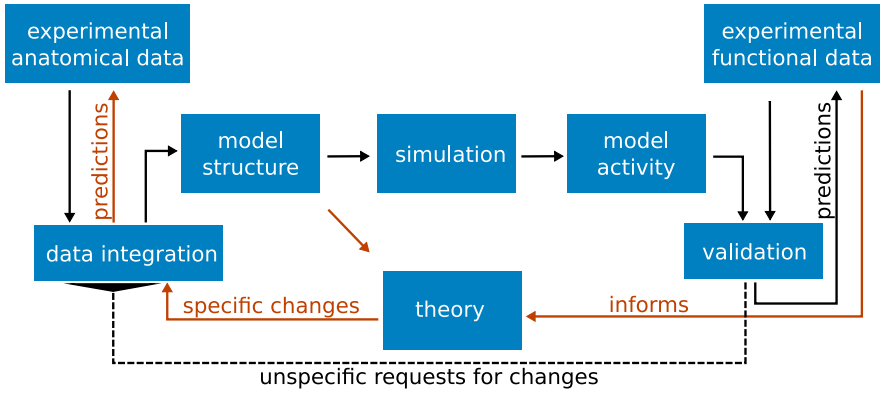



**Figure 4.1: The integrative loop between experiment, model, and theory.** Black arrows represent the classical forward modeling approach: Experimental anatomical data are integrated into a model structure, which gives rise to the activity via simulation. The model activity is compared with experimental functional data. The usual case of disagreement leads to the need to change the model definition. By experience and intuition the researcher identifies the parameters to be changed, proceeding in an iterative fashion. Once the model agrees well with a number of experimental findings, it may also generate predictions on the outcome of new experiments. Red arrows symbolize our new approach: informed by functional data, an analytical theory of the model identifies critical connections to be modified, refining the integration of data into a consistent model structure and generating predictions for anatomical studies.

full connectome of an individual in future, inter-individual variability will require modelers to express connectivity as connection probabilities or rules of self-organization if they want to learn about general principles of the brain. In modeling studies, parameters are usually tuned manually to achieve a satisfactory state of activity, which becomes unfeasible for high-dimensional models due to the size of the parameter space. In particular, it is a priori unclear how parameters of the model influence its activity. In consequence, modifications cannot be performed in a targeted fashion, and it is difficult to find a minimal set of modifications necessary for aligning a model with experimental activity data.

Overcoming this problem requires a shift of perspective. Instead of regarding the model exclusively in a forward manner, generating predicted activity from structure, we in addition consider the system in a reverse manner, predicting the structure necessary to explain the observed activity. Our theory, starting from the mean-field description (Sec. 3.1), provides a direct link between structure and activity. In contrast to simulations, the theory is invertible, which we exploit to identify connections critical for the dynamics and to find a minimal set of modifications to the structure yielding a realistic set point of activity. The predicted alterations generate hypotheses on brain structure, thus feeding back to anatomical experiments.

Applying the method to the multi-area model, we derive targeted modifications of a set of critical connections, bringing the model closer to experimental observations of cortical activity. Based on the global properties of the bistable phase space of the model, the method refines the model's construction principles within experimental uncertainties and identifies the

connections that decisively shape the dynamics.

## 4.2 Results

### Global stability in a simple network

We consider networks with neurons structured into interconnected populations, where the neurons inside one population are homogeneous with respect to their parameters and their input (Sec. 3.1.2).

The possible stationary states of these networks are characterized by the firing rates that are equilibria of (see Sec. 4.3)

$$\dot{\boldsymbol{\nu}} = \frac{d\boldsymbol{\nu}}{ds} = \boldsymbol{\Phi}(\boldsymbol{\nu}, \boldsymbol{A}) - \boldsymbol{\nu}, \tag{4.1}$$

where $s$ is a pseudo-time. The gain function $\boldsymbol{\Phi}$ (Eq. (3.5)) is known analytically and $\boldsymbol{A}$ indicates its dependence on the model parameters $\{\boldsymbol{K}, \boldsymbol{J}, \nu_{\text{ext}}, \dots\}$ (see Sec. 4.3).

The input-output relationship $\boldsymbol{\Phi}$ typically features a non-linearity which, in combination with feedback connections, can cause multi-stability in the network. In particular excitatory-excitatory loops cause the system defined by Eq. (4.1) to exhibit multi-stable behavior in the stationary firing rates. A necessary condition for the bistability is that the gain function has an inflection point. The LIF neuron model can have such an inflection point, originating from the interplay of its leak term and the threshold. Less realistic neuron models, such as the perfect integrate-and-fire model, do not have such an inflection point. To illustrate its origin, we first consider the noiseless case (Dayan & Abbott, 2001) without absolute refractoriness ($\tau_{\text{r}} = 0$). The gain function initially grows from zero with infinite slope due to the threshold and crosses the identity line (Fig. 4.2C). For larger input the leak term can be neglected and the gain function approaches a linear function with finite slope $\frac{1}{\tau_{\text{m}}} \frac{R_{\text{m}}}{V_\theta - V_{\text{R}}}$ (see, e.g., (Kriener et al., 2014b), eq. 11), equivalent to a perfect integrator. This is only possible with a negative curvature at intermediate rates, i.e., a reduction in the slope, which makes the gain function cross the identity line another time, causing the bistability. Network-generated noise only affects the low-rate regime where it smears out the kink causing the gain function to grow from zero with positive curvature (see, e.g. (Brunel, 2000), eq. 22). Importantly, the qualitative picture, i.e., the bistable behavior, is not altered. A finite refractory period only has an effect for very high input values where the gain function saturates at $1/\tau_{\text{r}} = 500$ spikes/s for the given parameters.

The basic problem is that there is a trade-off between excitation at the fixed points and their stability. In particular, exciting the model to bring a fixed point closer to experimental observations requires a method to preserve its stability. We achieve this by controlling the influence of the excitatory-excitatory loops on the phase space of the network.

As an illustration we first study the mechanism using the simple network architecture depicted

in Fig. 4.2A: a population of excitatory neurons is coupled to itself with indegree $K$ and is driven by external Poisson sources with the same indegree $K_{\text{ext}} = K$ and rate $\nu_{\text{ext}}$. All connections have identical synaptic weights $J$. An increase in the external drive shifts the input-output relationship $\Phi(\nu, \nu_{\text{ext}})$ of this one-dimensional system (Fig. 4.2C) to the left. The bifurcation diagram is shown in Fig. 4.2E: for low $\nu_{\text{ext}}$ there is only one fixed point with low activity (LA). When increasing $\nu_{\text{ext}}$, an additional pair of fixed points of which one is stable and the other is unstable emerges via a saddle-node bifurcation, leading to a bistable system. The second stable fixed point exhibits high firing rates, denoted as the high-activity (HA) state. For even higher values of $\nu_{\text{ext}}$, the LA state loses stability.

The equilibria, given by the zeros of the velocity $\dot{\nu}$ in the bistable case, are shown in Fig. 4.2F. An increase of the drive on the one hand increases the firing rate of the LA fixed point (inset) but on the other hand reduces its basin of attraction, indicated by the colored bars in the top left corner. For illustrative purposes, we extend the problem to two dimensions by splitting the excitatory population into two subpopulations of equal size (Fig. 4.2B), mimicking a loop between excitatory populations in the model of the multi-area model. The corresponding (symmetric) two-dimensional phase space is shown in Fig. 4.2D. The basin of attraction for the LA fixed point, limited by the separatrix (Strogatz, 1994), is reduced with increasing external drive.

Since we have a bistable system, there must be at least one unstable fixed point on the separatrix at the intersection of the nullclines, i.e., the subspace for which the velocity $\dot{\nu}_\alpha$ in direction $\alpha$ vanishes. We use the unstable fixed point to preserve the basin of attraction when the external drive $\nu_{\text{ext}}$ is increased. For this purpose, we modify the connectivity $K \to K'$ to reverse the shift of the unstable fixed point due to the parameter change $\nu_{\text{ext}} \to \nu'_{\text{ext}}$ (see Sec. 4.3 for a detailed derivation). Since the separatrix follows the unstable fixed point, this approximately restores the original basin of attraction.

The resulting velocity of the system $\Phi(\nu'_{\text{ext}}, K')$ ($\Phi$ defines the system Eq. (4.1)) is shown in Fig. 4.2F. The firing rate in the LA fixed point is increased as desired (inset), and the unstable fixed point coincides with that obtained in the original system $\Phi(\nu_{\text{ext}}, K)$. This pattern of fixed points is also indicated by the zero vectors of the velocity field $\dot{\boldsymbol{\nu}}$ (Fig. 4.2D). The separatrix follows the unstable fixed point, and the basin of attraction in the system $\boldsymbol{\Phi}(\nu'_{\text{ext}}, K')$ is restored to that in the original system $\boldsymbol{\Phi}(\nu_{\text{ext}}, K)$. Fig. 4.2G shows the behavior of the LA fixed point in more detail. The modification of $K$ does not noticeably change the location of the LA fixed point. In conclusion, the method allows us to increase the firing rates in the LA fixed point without modifying its basin of attraction.

The purely excitatory network is the simplest model to explain a phase space configuration with a LA and a HA fixed point. Inhibitory feedback is not necessary for this bistability, but it would certainly alter the input-output relationship. For example, in the balanced regime, the classical $\mathcal{E} - \mathcal{I}$ network (Brunel, 2000), investigated in Sec. 3.2.2, has an input-output relationship with a negative slope and thus only one fixed point exists. However, if a pair of such balanced E-I networks is coupled by sufficiently strong mutually excitatory connections,

these connections cause a bistability in a similar manner. Thus the mechanism shown in the purely excitatory network can also lead to the emergence of a HA attractor in more complex networks with inhibition.

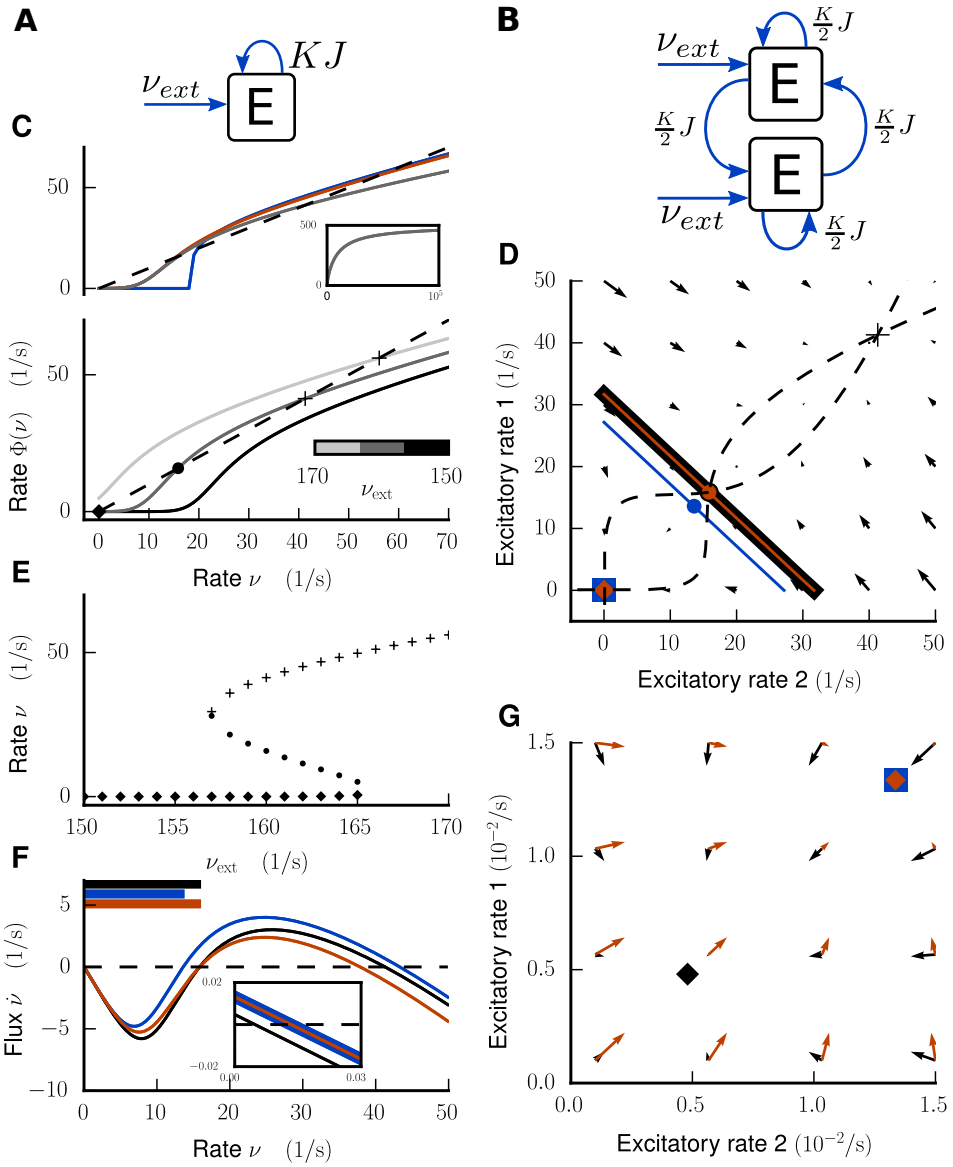



**Figure 4.2: Activity flow in an illustrative network example. Left column:** Global stability analysis in the single-population network. **A** Illustration of network architecture. **C** Upper panel: Input-output relationship $\Phi(\nu, \nu_{\text{ext}})$ for external Poisson drive $\nu_{\text{ext}} = 160\,\frac{1}{\text{s}}$ shown in gray. In addition $\Phi(\nu, \nu_{\text{ext}})$ for $\tau_{\text{r}} = 0$ for the noiseless case (blue) and the noisy case (red). Inset: gray curve over a larger input range. Lower panel: $\Phi(\nu, \nu_{\text{ext}})$ for different rates of the external Poisson drive $\nu_{\text{ext}} = [150, 160, 170]\,\frac{1}{\text{s}}$ from black to light gray. Intersections with the identity line (dashed) mark fixed points of the system, which are shown in **E** as a function of $\nu_{\text{ext}}$. **F** Flux $\dot{\nu}$ in the bistable case for $\Phi(\nu_{\text{ext}} = 160\frac{1}{\text{s}}, K)$ in black, $\Phi(\nu'_{\text{ext}} = 161\frac{1}{\text{s}}, K)$ in blue, and modified system $\Phi(\nu'_{\text{ext}} = 161\frac{1}{\text{s}}, K')$ in red. Intersections with zero (dashed line) mark fixed points. Inset: enlargement close to the LA fixed point. Horizontal bars at top of figure denote the size of the basin of attraction for each of the three settings. **Right column:** Global stability analysis in the network of two mutually coupled excitatory populations. **B** Illustration of network architecture. **D** Flow field and nullclines (dashed curves) for $\mathbf{\Phi}(\nu_{\text{ext}} = 160\frac{1}{\text{s}}, K)$ and separatrices (solid lines), LA fixed point (rectangle), HA fixed point (cross) and unstable fixed points (circles) for $\mathbf{\Phi}(\nu_{\text{ext}} = 160\frac{1}{\text{s}}, K)$ in black, $\mathbf{\Phi}(\nu_{\text{ext}} = 161\frac{1}{s}, K)$ in blue, and $\mathbf{\Phi}(\nu_{\text{ext}} = 161\frac{1}{s}, K')$ in red. **G** Enlargement of D close to the LA fixed points. Flow field of original (black) and modified system (red).

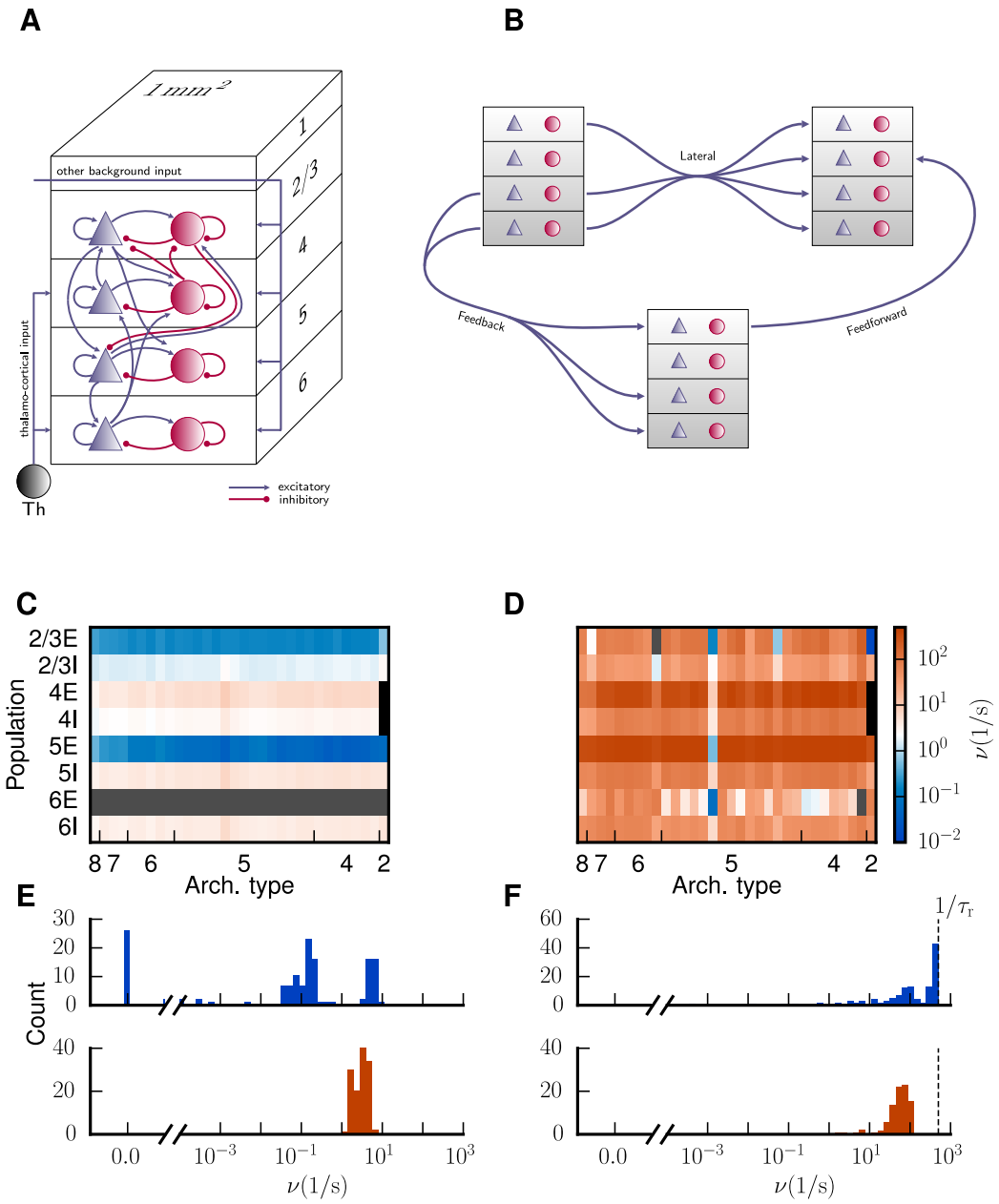



**Figure 4.3: Bistability of the model.** **A** Sketch of the microcircuit model serving as a prototype for the areas of the multi-area model (figure and legend adapted from figure 1 of Potjans and Diesmann (Potjans & Diesmann, 2014), with permission). **B** Sketch of the most common laminar patterns of cortico-cortical connectivity of the multi-area model. **C** Population-averaged firing rates encoded in color for a spiking network simulation of the multi-area model with low external drive ($\kappa = 1.0$). **D** As C but for increased external drive ($\kappa = 1.125$). The color bar refers to both panels. Areas are ordered according to their architectural type along the horizontal axis from V1 (type 8) to TH (type 2) and populations are stacked along the vertical axis. The two missing populations 4E and 4I of area TH are marked in black and firing rates $< 10^{-2}$ Hz in gray. **E** Histogram of population-averaged firing rates shown in **C** for excitatory (blue) and inhibitory (red) populations. The horizontal axis is split into linear- (left) and log-scaled (right) ranges. **F** as E corresponding to state shown in D.

### Bistability in the multi-area model

We investigate the multi-area model (Sec. 4.3) which is a multi-scale network model of cortical areas to understand the structural features essential for a realistic state of baseline activity. The model extends and adapts the microcircuit model presented in (Potjans & Diesmann, 2014), which covers 1 $\mathrm{mm}^2$ surface area of early sensory cortex (Fig. 4.3A), to all vision-related areas of macaque cortex (Fig. 4.3B). Based on the microcircuit model, an area is composed of 4 layers (2/3, 4, 5, and 6) each having an excitatory (E) and an inhibitory (I) neural population, except parahippocampal area TH, which consists of only 3 layers (2/3, 5, and 6). A detailed description of the data integration is given in (Schmidt et al., 2016).

Simulations of the model (Fig. 4.3C) reveal that, though realistic levels of activity can be achieved for populations in layers 2/3 and 4, populations 5E and 6E of the majority of areas show vanishingly low or zero activity in contrast to empirical data (Swadlow, 1988; de Kock & Sakmann, 2009). Inputs from subcortical and non-visual cortical areas are modeled as Poissonian spike trains, whose rate $\nu_{\mathrm{ext}}$ is a free, global parameter. To elevate the firing rates in the excitatory populations in layers 5 and 6, we enhance the external Poisson drive onto these populations parametrized by $\kappa$ (see Sec. 4.3). However, already a perturbation of a few percent leads to a state with unrealistically high rates (Fig. 4.3D), caused by the reduced basin of attraction of the low-activity state similar to Fig. 4.2D. Our aim is to improve the working point of the model such that all populations exhibit spiking activity $\gtrsim 0.05$ spikes/s while preventing the model from entering a state with unrealistically high rates of $\gtrsim 30$ spikes/s (figure 13 of (Swadlow, 1988), (de Kock & Sakmann, 2009)). The employed technique exposes the mechanism giving rise to the observed instability and identifies the circuitry responsible for this dynamical feature.

### Targeted modifications preserve global stability

We apply the procedure derived in Methods and find targeted modifications to the connectivity $\boldsymbol{K}$ that preserve the global stability of the low-activity fixed point for increased values of the external drive $\kappa$.

In the following we choose the inactive state $\boldsymbol{\nu}(0) = (0, \ldots, 0)^{\mathrm{T}}$ as the initial condition. The exact choice is not essential since we are only interested in the fixed points of the system. Fig. 4.4B shows the integration of Eq. (4.1) over pseudo-time $s$ for different levels of the external drive to populations 5E and 6E parametrized by $\kappa$. For low values of $\kappa$, the integration converges to the LA fixed point shown in Fig. 4.4D, and is in agreement with the activity emerging in the simulation (Fig. 4.3C). For increased values of $\kappa$, the system settles in the HA fixed point (Fig. 4.4E), again in agreement with the simulation. The population-specific firing rates in the HA state found in the mean-field predictions (Fig. 4.4E) are also close to those in the simulation (Fig. 4.3D), but minor deviations occur due to the violation of the assumptions made in the diffusion approximation (Sec. 3.1). In particular, at these pathologically high rates, the neurons fire regularly and close to the reciprocal of their refractory period, while in

the mean-field theory we assume Poisson spike statistics. Still, the mean-field theory predicts the bistability found in the simulation. Since the theory yields reliable predictions in both stable fixed points, we assume that also the location of the unstable fixed point in between these two extremes is accurately predicted by the theory.

To control the separatrix we need to find the unstable fixed point of the system. This is nontrivial since the numerical integration of Eq. (4.1) for finding equilibria by construction only converges to stable fixed points. If the unstable fixed point has only one repelling direction (Fig. 4.5A), it constitutes a stable attractor on the $N_{\mathrm{pop}} - 1$ dimensional separatrix. The separatrix is a stable manifold (Strogatz, 1994), and therefore a trajectory originating in its vicinity but not near an unstable fixed point initially stays in the neighborhood. If an initial condition just outside the separatrix is close to the basin of attraction of a particular unstable fixed point, the trajectory initially approaches the latter. Close to the fixed point the velocity is small. Ultimately trajectories diverge from the separatrix in the fixed point's unstable direction, as illustrated in Fig. 4.4A. In conclusion, we expect a local minimum in the velocity along the trajectories close to the unstable fixed point. To estimate the location of the unstable fixed point in this manner, we need to find initial conditions close to the separatrix. Naively, we would just fix the value of $\kappa$ and vary the initial condition. However, due to the high dimensionality of our system this is not feasible in practice. Instead, we vary $\kappa$ for a fixed initial condition. Fig. 4.4B shows the firing rate averaged across populations for two trajectories starting close to the separatrix, where the first one converges to the LA fixed point and the second one to the HA state. The trajectories diverge near the unstable fixed point and thus we define the last local minimum of the Euclidean norm of the velocity vector as the critical time $s_{\mathrm{c}}$ at which we assume the system to be close to the unstable fixed point (Fig. 4.4C). We find four relevant and distinct unstable fixed points, of which two are shown in Fig. 4.6.

To counteract the shift of the separatrix caused by the increase in $\kappa$, we follow the procedure described in Sec. 4.3. We subject the modifications of connectivity to the additional following constraints. In line with the anatomical literature, we do not allow for changes of the connectivity that would lead to cortico-cortical connections originating in the granular layer 4 (Felleman & Van Essen, 1991), and we also disallow inhibitory cortico-cortical connections, as the vast majority of long-range connections are known to be excitatory (Salin & Bullier, 1995; Tomioka & Rockland, 2007). In addition, we naturally restrict indegrees to positive values. We find that four iterations (numbered by index $j$) corresponding to the four distinct unstable fixed points suffice to preserve the basin of attraction of the LA state with respect to an increase of the external drive up to $\kappa = 1.15$. In the following we concentrate on iterations 1 and 2, where the second one is also representative for iterations 3 and 4, which are qualitatively alike. To derive the required modifications of the indegree matrix, we decompose $\boldsymbol{K}$ into its $N_{\mathrm{pop}}$ eigenmodes and quantify the contribution of each eigenmode to the shift of the unstable fixed point (see Sec. 4.3). This allows us to identify the most effective eigendirection: in each iteration $j$ there is exactly one unstable eigendirection with an eigenvalue $\mathrm{Re}\left(\lambda_{\mathrm{c}}^{(j)}\right) > 1$

(Fig. 4.5A). The associated critical eigenvector is approximately anti-parallel to the shift of the fixed point, $\boldsymbol{\delta\nu}^*$ (inset of Fig. 4.5B), and of similar length. The critical eigendirection (red dot in Fig. 4.5B) constitutes the most effective modification, giving the largest contribution to the desired shift while requiring only a small change of 2.3 % in average total indegrees. In the chosen space of eigenmodes, the modifications are minimal in the sense that only this most effective eigenmode is changed.

The associated eigenvector $u_c^{(1)}$ predominately points into the direction of populations 4E and 5E of areas FEF and 46 (Fig. 4.5C), while $u_c^{(2)}$ has large entries in the 5E populations of two areas (Fig. 4.5D). The high rates of these populations at the unstable fixed points (cf. Fig. 4.6A,B with Fig. 4.5C,D) reflect that the instability is caused by increased rates in excitatory populations, particularly in population 5E. Each iteration shifts the transition to the HA state (the value of $\kappa$ for which the separatrix crosses the initial condition) to higher values of $\kappa$ and increases the attainable rates of populations 5E and 6E in the LA state (Fig. 4.7A). After all four iterations, the average total indegrees (summed over source populations) of the system are changed by 11.3%. The first iteration mainly affects connections within and between areas 46 and FEF (Fig. 4.7B). In particular, the excitatory loops between the two areas are reduced in strength, especially those involving layer 5 (Fig. 4.7C). We thus identify two areas forming a critical loop. In the remaining iterations, the changes are spread across areas and especially connections originating in layer 5 are weakened (Fig. 4.7D). In conclusion, the method identifies critical structures in the model both on the level of areas and on the level of layers and populations, and leads to a small but specific structural change of the model.

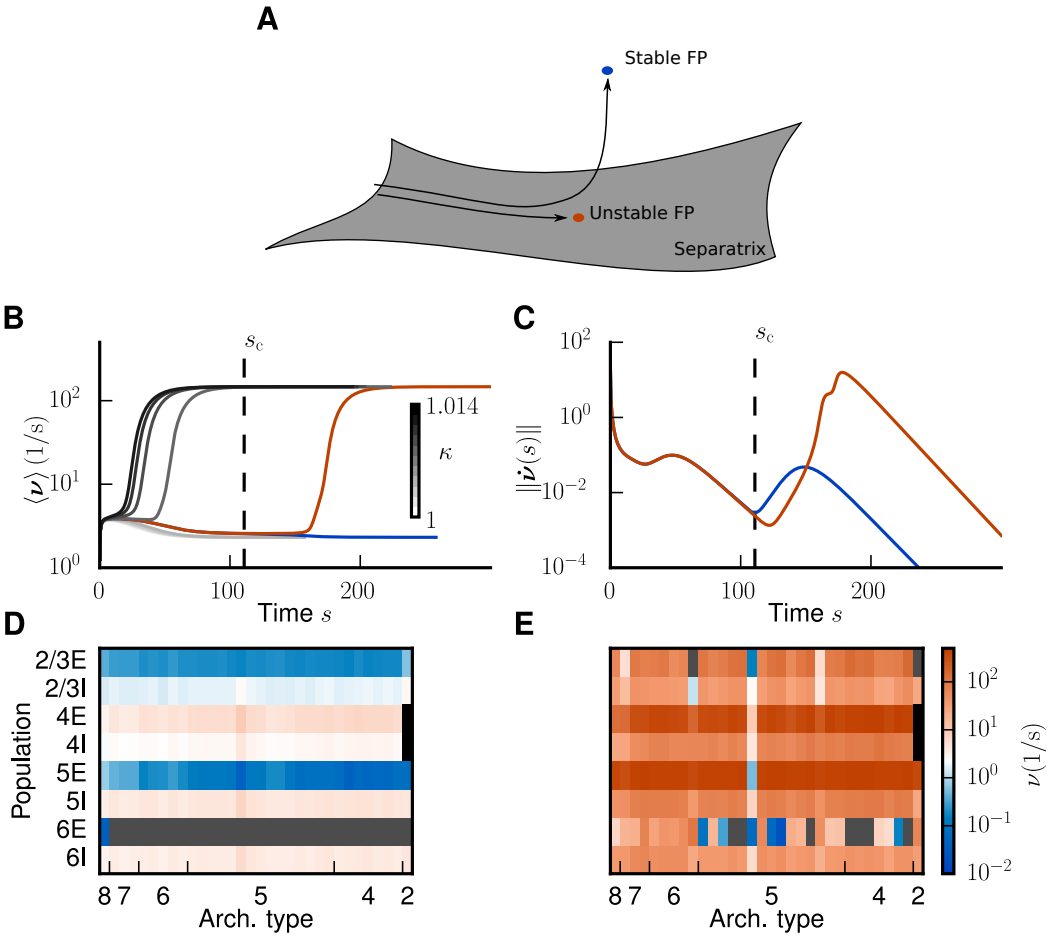



**Figure 4.4: Application of the mean-field theory to the multi-area model. A** Trajectories of Eq. (4.1) starting inside the separatrix converge to an unstable fixed point. Trajectories starting close to the separatrix are initially attracted by the unstable fixed point but then repelled in the fixed point's unstable direction and finally converge to a stable fixed point. **B** Firing rate averaged across populations over time. Integration of Eq. (4.1) leads to convergence to either the low-activity (LA) or the high-activity (HA) attractor for different choices of the external input factor $\kappa$, with $\kappa = 1$ the original level of external drive. We show eight curves with $\kappa$ varying from 1.0 to 1.014 in steps of 0.002 and two additional curves for $\kappa = 1.007662217$, $1.007662218$. The curves for the largest factor ($\kappa = 1.007662217$) that still leads to the LA state and for the smallest factor ($\kappa = 1.007662218$) that leads to the HA state are marked in blue and red, respectively. The four curves with $\kappa \leq 1.006$ coincide with the blue curve. **C** Euclidean norm of the velocity vector in the integration of Eq. (4.1) for the different choices of $\kappa$. The vertical dashed line indicates the time $s_c$ of the last local minimum in the blue curve. **D** Stationary firing rate in the different areas and layers of the model in a low-activity state for $\kappa = 1.0$ as predicted by the mean-field theory (same display as in Fig. 4.3). **E** As D , but showing the high-activity state for $\kappa = 1.125$.

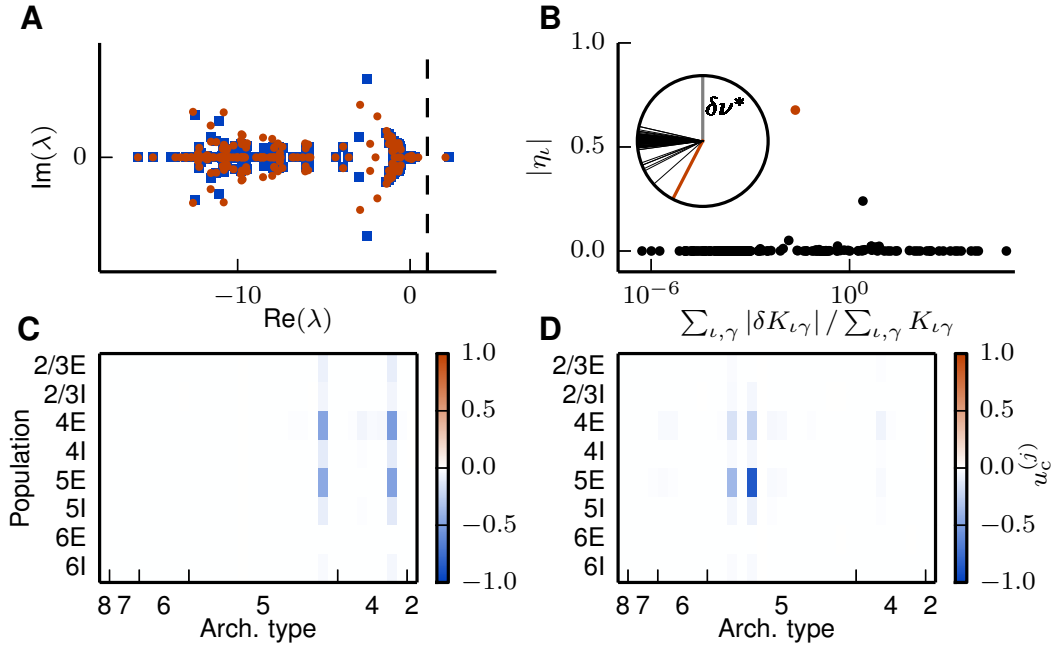



**Figure 4.5: Eigenspectrum analysis of network stability.** **A** Eigenvalue spectrum of the effective connectivity matrix $\boldsymbol{M}$ for the first (blue squares) and second (red dots) iteration. The dashed vertical line marks the edge of stability at a real part of 1. **B** Contribution $\eta_\iota$ (Eq. (4.12)) of an individual eigenprojection $\iota$ to the shift of the unstable fixed point versus the relative change in indegrees associated with $\iota$ for the first iteration. The data point corresponding to $\lambda_c^{(1)}$ is marked in red. The inset shows the relative angles between $\boldsymbol{\delta\nu}^*$ and the eigenvectors $\boldsymbol{u}^\iota$. The red line corresponds to the critical eigendirection. **C** Entries of the eigenvector $u_c^{(1)}$ associated with $\lambda_c^{(1)}$ in the populations of the model. The affected areas are 46 and FEF. **D** Same as C for the second iteration.

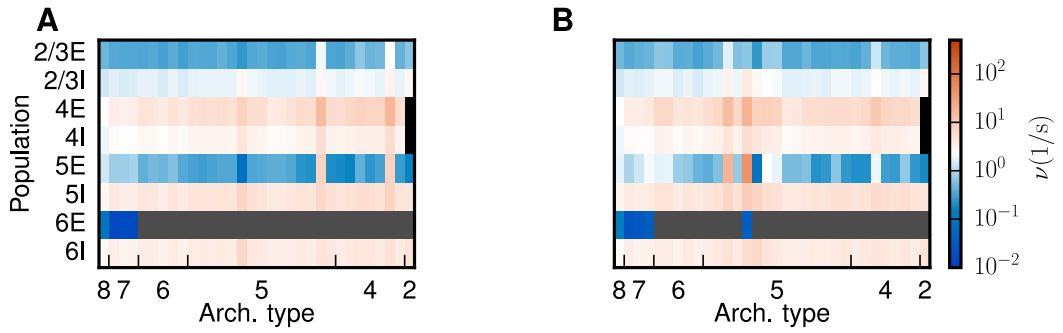



**Figure 4.6: Unstable fixed points in subsequent iterations**. Population firing rates at the unstable fixed point as predicted by the mean-field theory encoded in color for iterations 1 (**A**) and 2 (**B**). Same display as in Fig. 4.3.

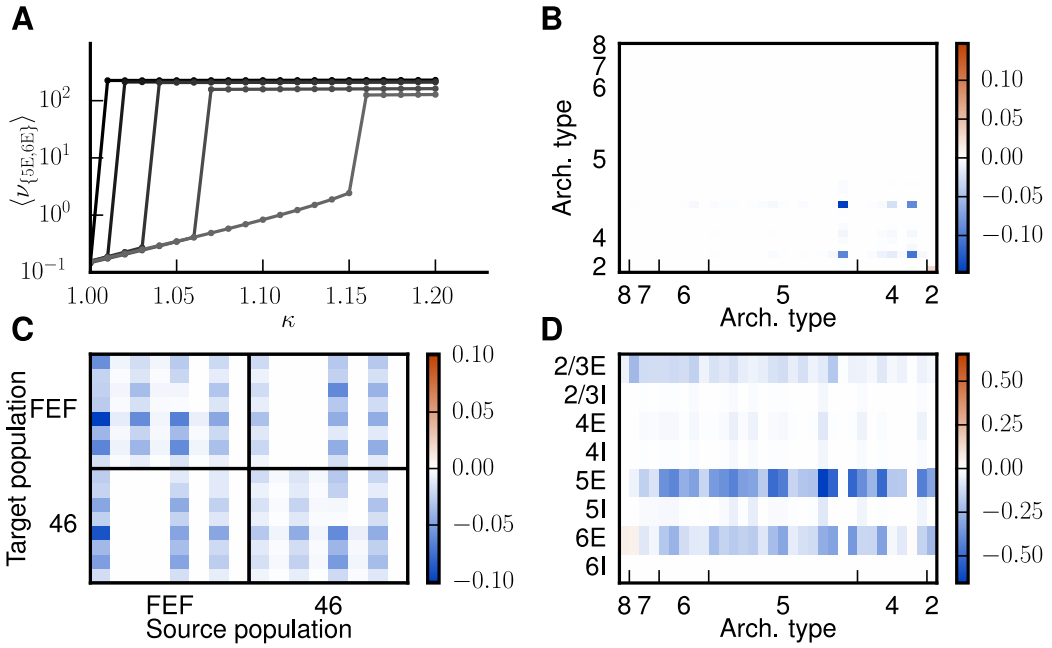



**Figure 4.7: Altered phase space and modified connections.** **A** Firing rates averaged across populations 5E and 6E and across areas for different stages from the original model (black) to iteration 4 (light gray) as a function of $\kappa$, predicted by the mean-field theory. **B** Relative changes in the indegree $\delta K_{AB} / \sum_B K_{AB}$ between areas $A, B$ in the first iteration. **C** Layer-specific relative changes $\delta K_{\iota\gamma} / \sum_\gamma K_{\iota\gamma}$ in the connections within and between areas FEF and 46, for the first iteration. Populations are ordered from 2/3E (left) to 6I (right) on the horizontal axis and from 6I (bottom) to 2/3E (top) on the vertical axis as in panel D. **D** Relative changes in population-specific indegrees summed over target populations, $\sum_\iota \delta K_{\iota\gamma} / \sum_\iota K_{\iota\gamma}$, combined for iterations two, three and four.

### Analysis of the modifications

In the following we analyze the modifications of the connectivity with respect to the internal and inter-area connections in detail. The intrinsic circuits of the areas are modified in different directions, as shown for two exemplary areas V4 and CITv in Fig. 4.8A. Despite this heterogeneity, significant changes affect mostly excitatory-excitatory connections (Fig. 4.8A, bottom panel) with connections from population 5E experiencing the most significant changes (top panel of Fig. 4.8A). In fact, the anatomical data (Binzegger et al., 2004) underlying the microcircuit model (Potjans & Diesmann, 2014) contain only two reconstructed excitatory cells from layer 5, but considerably more for other cell types, indicating a higher uncertainty for layer 5 connections. Fig. 4.8B shows the correlation between intrinsic connectivity changes for all pairs of areas, with areas ordered according to a hierarchical clustering using a farthest point algorithm (Voorhees, 1986) on the correlation matrix. We find four clusters each indicating a group of areas which undergo changes with similar patterns. The groups are displayed in different colors in the histogram in Fig. 4.8B. The areas of the model are categorized into architectural types based on cell densities and laminar thicknesses (see Sec. 4.3). Areas with architectural type 4, 5 and 6 are distributed over several clusters. We can interpret this as a differentiation of these types into further subtypes. The resulting changes of the intra-areal connectivity are small (Fig. 4.8A), but still significant for network stability.

Connections between areas can be characterized by their $FLN$ and $SLN$ (see Sec. 4.3). The $FLN$ reflects the overall strength of an inter-areal connection and is only weakly affected across connections (Fig. 4.8C), with a correlation between original and modified logarithms of $FLN$ of $r_{\text{Pearson}} = 0.79$. Significant variations in the $FLN$ occur mostly for very weak connections that are likely to have substantial relative uncertainties in the experimental data. The two overlapping red dots in Fig. 4.8C represent the connections between areas 46 and FEF, which are modified in the first iteration (Fig. 4.7C). The $SLN$ determines the laminar pattern of the location of source neurons for cortico-cortical connections. Overall, data are available for 24% of the inter-areal connections in the parcellation of Felleman & van Essen (Felleman & Van Essen, 1991), while the $SLN$ for the rest are derived from the sigmoidal law. The majority of connections undergo small changes in their laminar source pattern (Fig. 4.8D) and connections with large modifications ($|\delta SLN| > 0.5$) are weak (average $\overline{FLN} = 6 \cdot 10^{-4}$ compared to $\overline{FLN} = 10^{-2}$ in the model as a whole). Because weak connections are represented by low counts of labeled neurons, they have a relatively large uncertainty in their laminar patterns, justifying larger adjustments. Spearman's rank correlation between the $SLN$ of the original model that were directly taken from experiments and the logarithmic ratios of cell densities is $\rho = -0.63$ ($p = 3 \cdot 10^{-11}$, p-value of a two-sided test for uncorrelated data). For the modified model, we take the $SLN$ of all connections into account and obtain $\rho = -0.40$ ($p = 6 \cdot 10^{-20}$), indicating a reduced, but still significant, monotonic dependence between $SLN$ and the logarithmic ratios of cell densities.

To judge the size of the modification to the connectivity, we compare it to the variability

of measured cortico-cortical connection densities (Markov et al., 2014a). We quantify the latter as the average inter-individual standard deviation of the logarithmic $FLN$, i.e., $\sigma = \left\langle \sqrt{\overline{\left(\log FLN - \overline{\log FLN}\right)^2}} \right\rangle$, where the overbar $\bar{\cdot}$ denotes the average over injections and $\langle\cdot\rangle$ the average over connections. This variability equals 2.17 while the average modification of the logarithmic $FLN$ is 1.34. The main experimental connection probabilities used to construct the intra-areal connectivity of the model have an average relative standard deviation of 30% across electrophysiological experiments (cf. Table 1 of (Potjans & Diesmann, 2014)) while the intra-areal connection probabilities of the model are modified by 9% on average. The authors of (Scannell et al., 2000) report even greater variability in their review on cortico-cortical and thalamocortical connectivity. These considerations show that on average, the changes applied to the connectivity are well within the uncertainties of the data. Overall, 35 out of 603 connections were removed from the network. In the CoCoMac database, 83 % of these are indicated by only a single tracer injection, while the overall proportion of connections measured by a single injection is 59 %.

For the modified connectivity and $\kappa = 1.125$, which we choose to avoid being too close to the transition (Fig. 4.7A), the theory predicts average rates in populations 5E and 6E of 1.3 and 0.18 spikes/s, which is closer to experimentally observed rates compared to the original model. Furthermore we find that the modified connectivity allows us to decrease the inhibition in the network to $g = -11$. Simulating the full spiking network model then results in reasonable rates across populations and areas (Fig. 4.9B, D). The average rates in populations 5E and 6E are increased compared to a simulation of the original model from 0.09 and $2 \cdot 10^{-5}$ spikes/s to 3.0 and 0.4 spikes/s, respectively. All populations exhibit firing rates within a reasonable range of 0.05 to 30 spikes/s (Fig. 4.9D), as opposed to the original state in which a considerable fraction of excitatory neurons is silent (Fig. 4.3E). The theoretical prediction is in excellent agreement with the rates obtained in the simulation (Fig. 4.9A, C). Small discrepancies are caused by violations of the employed assumptions, i.e., Poissonian spiking statistics (Sec. 3.1). Differences between theory and simulation are small, and negligible for the central aim of the study: the integration of activity constraints into the data-driven construction of multi-scale neuronal networks.

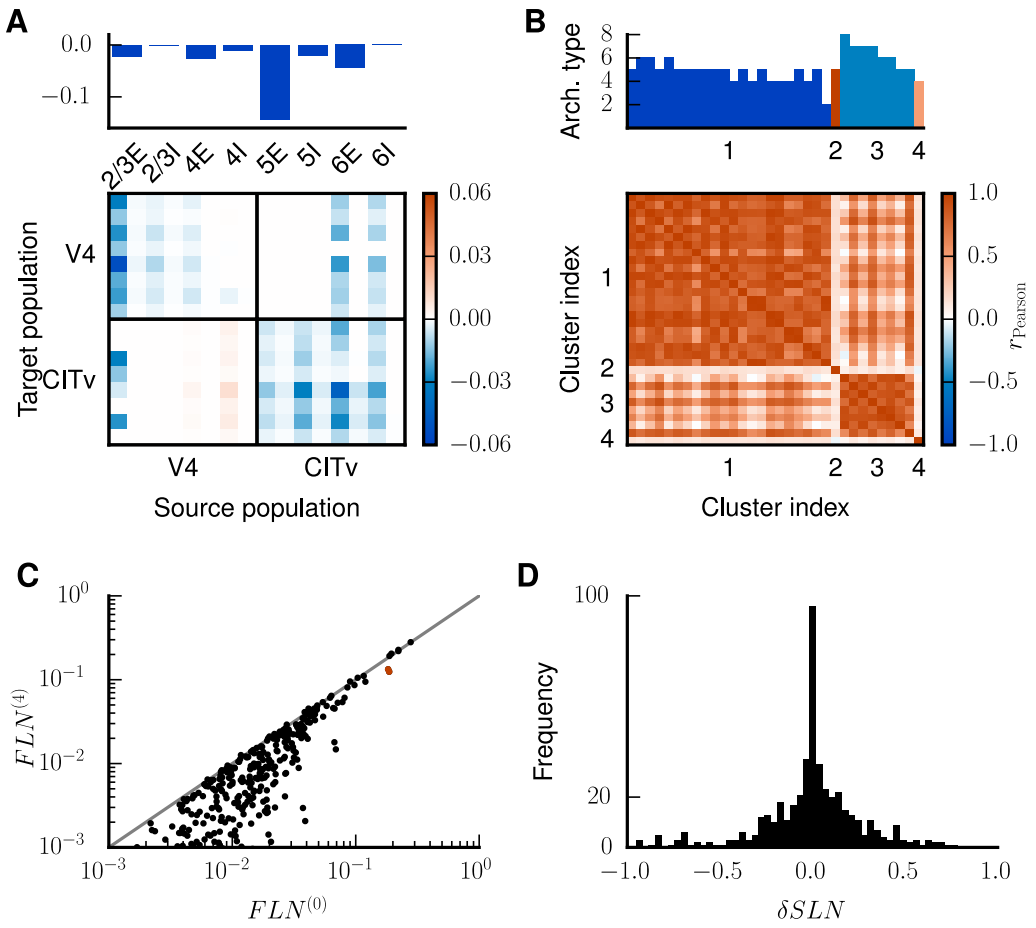



**Figure 4.8: Analysis of changes in connectivity.** **A** Top panel: relative changes in population-specific intrinsic indegrees summed over target populations and averaged across areas, $\langle \sum_\alpha \delta K_{AA}^{\alpha\beta} / \sum_\alpha K_{AA}^{\alpha\beta} \rangle_A$. Bottom panel: changes in the indegrees within and between exemplary areas V4 and CITv relative to the total indegrees of the target populations, i.e., $\delta K_{AA}^{\alpha\beta} / \sum_\beta K_{AA}^{\alpha\beta}$. Populations ordered as in Fig. 4.7C. **B** Pearson correlation coefficient of the changes of the internal indegrees $\delta K_{AA}^{\alpha\beta}$ between all pairs of the 32 areas. Areas ordered according to hierarchical clustering using a farthest point algorithm (Voorhees, 1986). The heights of the bars on top of the matrix indicate the architectural types of the areas (types 1 and 3 do not appear in the model) with color representing the respective clusters. **C** $FLN$ of the modified connectivity after 4 iterations versus the original $FLN$ of the model. Only $FLN > 10^{-3}$ are shown for a better overview. The overlapping red dots represent the connections between areas 46 and FEF. Unity line shown in gray. **D** Histogram of the cumulative changes in $SLN$ over all four iterations ($\delta SLN = SLN^{(4)} - SLN^{(0)}$).

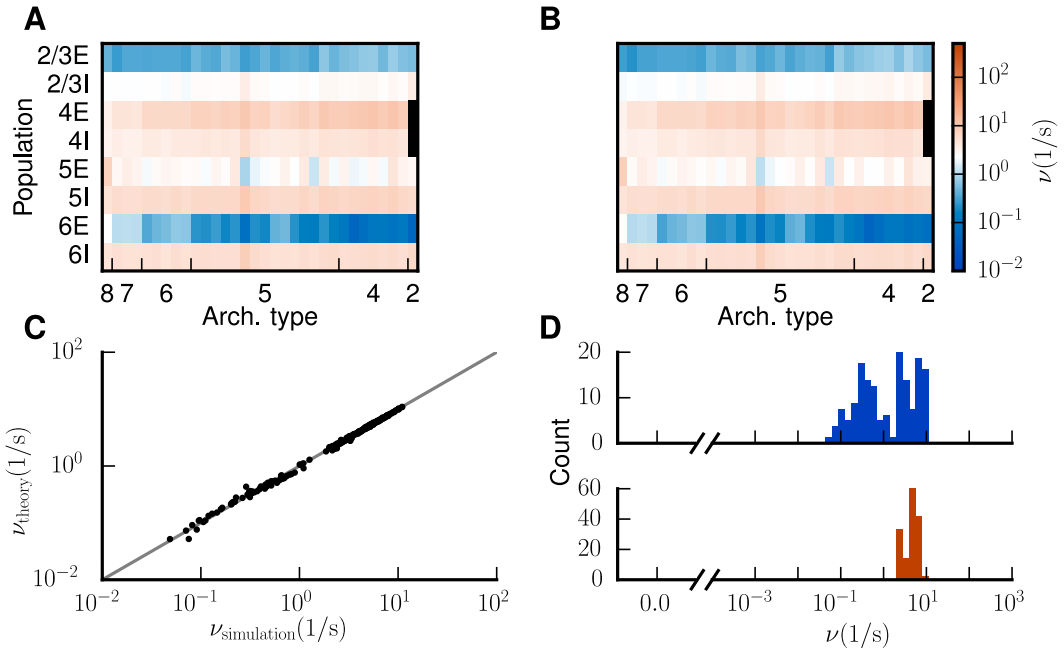



**Figure 4.9: Improved low-activity fixed point of the model.** Population-averaged firing rates for $\kappa = 1.125$ encoded in color (**A**) predicted by the analytical theory and (**B**) obtained from the full simulation of the spiking network. Same display as in Fig. 4.3. **C** Analytical versus simulated firing rates (black dots) and identity line (gray). **D** Histogram of population-averaged simulated firing rates. Same display as in Fig. 4.3.

## 4.3 Methods

The stationary firing rates for each population $\alpha$ are given by Eq. (3.7). We here need to extent the notation and introduce a parameter $\boldsymbol{A}$ in the equation, which reads

$$\begin{aligned}
\frac{1}{\nu_\alpha} &= \tau_\mathrm{r} + \tau_\mathrm{m}\sqrt{\pi}\int_{\frac{V_r-\mu_\alpha(\boldsymbol{A})}{\sigma_\alpha(\boldsymbol{A})}+\gamma\sqrt{\frac{\tau_\mathrm{s}}{\tau_\mathrm{m}}}}^{\frac{V_\theta-\mu_\alpha(\boldsymbol{A})}{\sigma_\alpha(\boldsymbol{A})}+\gamma\sqrt{\frac{\tau_\mathrm{s}}{\tau_\mathrm{m}}}} e^{x^2}\,(1+\mathrm{erf}(x))\,\mathrm{d}x \\
&= 1/\Phi_\alpha(\boldsymbol{\nu},\boldsymbol{A}) \qquad (4.2)\\
\mu_\alpha(\boldsymbol{A}) &= \tau_\mathrm{m}\sum_\beta K_{\alpha\beta}J_{\alpha\beta}\nu_\beta + \tau_\mathrm{m}K_\mathrm{ext}J_\mathrm{ext}\nu_\mathrm{ext} \qquad (4.3)\\
\sigma^2_\alpha(\boldsymbol{A}) &= \tau_\mathrm{m}\sum_\beta K_{\alpha\beta}J^2_{\alpha\beta}\nu_\beta + \tau_\mathrm{m}K_\mathrm{ext}J^2_\mathrm{ext}\nu_\mathrm{ext}. \qquad (4.4)
\end{aligned}$$

Here, $\boldsymbol{A}$ is chosen from the set of model parameters $\{\boldsymbol{K},\boldsymbol{J},\nu_\mathrm{ext},\dots\}$. If $\boldsymbol{A}$ is a matrix, we vectorize it by concatenating its rows and indicate this by lower case, i.e., $\boldsymbol{a} = (a_{00}, a_{01},\dots,a_{0N_\mathrm{pop}}, a_{10},\dots,a_{1N_\mathrm{pop}}, a_{N_\mathrm{pop}0}\dots,a_{N_\mathrm{pop}N_\mathrm{pop}}) = \mathrm{vec}(\boldsymbol{A}^\mathrm{T})^\mathrm{T}$ following (Magnus & Neudecker, 1995). If the chosen parameter is a scalar we denote it with $a$.

As in Chap. 3, we find the fixed points of Eq. (4.2) by solving the first-order differential equation (Wong & Wang, 2006). We rewrite Eq. (3.6) with the new parameter $\boldsymbol{A}$, i.e., (Eq. (4.1))

$$\dot{\boldsymbol{\nu}} = \frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}s} = \boldsymbol{\Phi}(\boldsymbol{\nu},\boldsymbol{A}) - \boldsymbol{\nu}.$$

The results shown in this chapter were obtained with PYTHON by integrating Eq. (4.1) using the classical fourth-order Runge-Kutta method (RK4) with step size $h = 0.01$, where $s$ denotes a dimensionless pseudo-time. The implementation of the rate dynamics in NEST, used in Chap. 3 and described in Chap. 8, has been done after the results from this chapter were obtained. Of course, the results shown in this chapter can also be obtained with the NEST implementation. Note that Eq. (4.1) does not reflect the real time evolution of the population rates, but rather is a mathematical method to obtain the system's fixed points. In contrast to (Wong & Wang, 2006) we do not only search for stable fixed points, but also use Eq. (4.1) to obtain unstable attractors (cf. Results), an idea originating from the study of simple attractor networks (Amit & Brunel, 1997b, esp. their Fig 2 and Eq 12).

In a bistable situation, the initial condition of Eq. (4.1) determines which fixed point the system settles in. However, studying the behavior for a particular initial condition is of minor interest, since the actual spiking network is a stochastic system which fluctuates around the fixed points of the deterministic system defined by Eq. (4.1). Even if we knew that Eq. (4.1) would relax to the LA fixed point for one particular initial condition, this would not necessarily imply that this state is indefinitely stable. Global stability is determined by the size of the basin of attraction of the LA fixed point.

In the following, we derive the equations leading us to targeted modifications of a parameter $\boldsymbol{b}$ necessary to compensate for the changes in the global stability induced by the change of

another parameter $\boldsymbol{a}$. To this end, we study the behavior of the fixed points with respect to an infinitesimal change $\boldsymbol{\delta a} = \boldsymbol{a'} - \boldsymbol{a}$ in the chosen model parameter. Let $\boldsymbol{\nu}^*(\boldsymbol{a})$ and $\boldsymbol{\nu}^*(\boldsymbol{a'})$ be the corresponding locations of the fixed points and $\boldsymbol{\delta\nu}^* = \boldsymbol{\nu}^*(\boldsymbol{a'}) - \boldsymbol{\nu}^*(\boldsymbol{a})$ their separation. We can then expand $\boldsymbol{\nu}^*(\boldsymbol{a'})$ into a Taylor series up to first order in $\boldsymbol{\delta a}$ and obtain

$$\begin{aligned} \boldsymbol{\nu}^*(\boldsymbol{a'}) &= \boldsymbol{\nu}^*(\boldsymbol{a}) + \boldsymbol{\Delta_a}\boldsymbol{\delta a} \\ \Leftrightarrow \boldsymbol{\delta\nu}^* &= \boldsymbol{\Delta_a}\boldsymbol{\delta a}\,, \end{aligned} \tag{4.5}$$

with

$$\begin{aligned} \Delta_{a,\alpha\beta} &= \frac{\mathrm{d}\nu_\alpha\left(\mu_\alpha,\sigma_\alpha\right)}{\mathrm{d}a_\beta} \\ &= \frac{\mathrm{d}\Phi_\alpha\left(\mu_\alpha,\sigma_\alpha\right)}{\mathrm{d}a_\beta} \\ &= \underbrace{\frac{\partial\Phi_\alpha}{\partial\mu_\alpha}}_{S_\alpha}\frac{\mathrm{d}\mu_\alpha}{\mathrm{d}a_\beta} + \underbrace{\frac{\partial\Phi_\alpha}{\partial\sigma_\alpha}\frac{1}{2\sigma_\alpha}}_{T_\alpha}\frac{\mathrm{d}\sigma_\alpha^2}{\mathrm{d}a_\beta}, \end{aligned} \tag{4.6}$$

where we notice that $S_\alpha$ and $T_\alpha$ only depend on the target population $\alpha$. We accordingly define two diagonal matrices $\boldsymbol{S}$ and $\boldsymbol{T}$ with $S_{\alpha\alpha} = S_\alpha$ and $T_{\alpha\alpha} = T_\alpha$. We further define the connectivity matrix $\mathbf{W} = \boldsymbol{K} \circledast \mathbf{J}$, where $\circledast$ denotes element-wise multiplication, also called the Hadamard product (see Cichocki et al., 2009, for a consistent set of symbols for operations on matrices). The derivatives with respect to $a_\beta$ have the compact expressions

$$\begin{aligned} \frac{\mathrm{d}\mu_\alpha}{\mathrm{d}a_\beta} &= \frac{\partial\mu_\alpha}{\partial a_\beta} + \sum_\gamma \frac{\partial\mu_\alpha}{\partial\nu_\gamma}\frac{\mathrm{d}\nu_\gamma}{\mathrm{d}a_\beta} \\ &= \left(\mathcal{D}_{\boldsymbol{a}}\boldsymbol{\mu}\right)_{\alpha\beta} + \tau_\mathrm{m}\sum_\gamma \underbrace{(K_{\alpha\gamma}J_{\alpha\gamma})}_{=W_{\alpha\gamma}}\Delta_{a,\gamma\beta}\,, \end{aligned}$$

with the Jacobian $\left(\mathcal{D}_{\boldsymbol{a}}\mathbf{f}\right)_{\alpha\beta} \coloneqq \frac{\partial f_\alpha}{\partial a_\beta}$ of some vector-valued function $\mathbf{f}$ and

$$\frac{\mathrm{d}\sigma_\alpha^2}{\mathrm{d}a_\beta} = \left(\mathcal{D}_{\boldsymbol{a}}\boldsymbol{\sigma}^2\right)_{\alpha\beta} + \tau_\mathrm{m}\sum_\gamma \underbrace{(K_{\alpha\gamma}J_{\alpha\gamma}^2)}_{=W_{2,\alpha\gamma}}\Delta_{a,\gamma\beta}\,,$$

where we use the Hadamard product again to define the matrix $\mathbf{W_2} \coloneqq \mathbf{K} \circledast \mathbf{J} \circledast \mathbf{J}$. Inserting the total derivatives into Eq. (4.6), we derive the final expression for $\boldsymbol{\Delta_a}$, reading

$$\begin{aligned} \boldsymbol{\Delta_a} &= \boldsymbol{S}\left[\mathcal{D}_{\boldsymbol{a}}\boldsymbol{\mu} + \tau_\mathrm{m}\boldsymbol{W}\boldsymbol{\Delta_a}\right] + \boldsymbol{T}\left[\mathcal{D}_{\boldsymbol{a}}\boldsymbol{\sigma}^2 + \tau_\mathrm{m}\boldsymbol{W_2}\boldsymbol{\Delta_a}\right] \\ \Leftrightarrow \boldsymbol{\Delta_a} &= \left[\mathbb{1} - \underbrace{\tau_\mathrm{m}(\boldsymbol{SW} + \boldsymbol{TW_2})}_{=:\boldsymbol{M}}\right]^{-1} \underbrace{\left[\boldsymbol{S}\mathcal{D}_{\boldsymbol{a}}\boldsymbol{\mu} + \boldsymbol{T}\mathcal{D}_{\boldsymbol{a}}\boldsymbol{\sigma}^2\right]}_{=:\bar{\boldsymbol{\Delta}}_{\boldsymbol{a}}}, \end{aligned} \tag{4.7}$$

where we use $\mathbb{1}$ for the identity matrix and define the effective connectivity matrix $\boldsymbol{M}$ and

the matrix $\bar{\boldsymbol{\Delta}}_{\boldsymbol{a}}$. The latter has dimensionality $N_{\text{pop}} \times P$, where $P$ is the dimension of $\boldsymbol{a}$ (for example, $P = N_{\text{pop}}^2$ for $\boldsymbol{a} = \boldsymbol{k}$ the vector of indegrees). With the aid of Eq. (4.7), evaluating Eq. (4.5) at the unstable fixed point predicts the shift of the separatrix (Fig. 4.2D) to linear order. We now consider an additional parameter $\boldsymbol{b}$ which is modified to counteract the shift of the unstable fixed point caused by the change in parameter $\boldsymbol{a}$, i.e.,

$$\begin{aligned} \boldsymbol{\Delta}_{\boldsymbol{a}}\boldsymbol{\delta a} &\stackrel{!}{=} -\boldsymbol{\Delta}_{\boldsymbol{b}}\boldsymbol{\delta b} \\ \Leftrightarrow \bar{\boldsymbol{\Delta}}_{\boldsymbol{a}}\boldsymbol{\delta a} &= -\bar{\boldsymbol{\Delta}}_{\boldsymbol{b}}\boldsymbol{\delta b}, \end{aligned} \tag{4.8}$$

where we used that the inverse of $\mathbb{1} - \boldsymbol{M}$ appears on both sides of the equation and hence drops out. Note that the tuple $(\boldsymbol{a}, \boldsymbol{b})$ may represent any combination of model parameters, for example external input, indegrees, synaptic weights, etc. For a particular choice of $(\boldsymbol{a}, \boldsymbol{b})$ we solve Eq. (4.8) for $\boldsymbol{\delta b}$. For the illustrative example shown in Fig. 4.2, where $a = \nu_{\text{ext}}$ and $b = K$, Eq. (4.8) simplifies to

$$\delta K = \frac{\bar{\Delta}_{\nu_{\text{ext}}}\delta\nu_{\text{ext}}}{\bar{\Delta}_K} = \frac{K_{\text{ext}}\delta\nu_{\text{ext}}}{\nu},$$

since $S$ and $T$ appearing in the respective $\bar{\Delta}'$s cancel each other.

To determine critical connections in a more complex model, we choose $\boldsymbol{b} = \boldsymbol{k}$, i.e. the vector of indegrees, and solve Eq. (4.8) with a decomposition into eigenmodes. We can write the right-hand side as

$$-\left(\bar{\boldsymbol{\Delta}}_{\boldsymbol{k}}\boldsymbol{\delta k}\right)_\alpha = -\sum_\beta \tau_{\text{m}}\left(S_\alpha J_{\alpha\beta} + T_\alpha J_{\alpha\beta}^2\right)\nu_\beta \delta K_{\alpha\beta}\,. \tag{4.9}$$

The equation holds because $\mu_\alpha, \sigma_\alpha$ are only affected by connections to population $\alpha$, and therefore their derivatives $\partial\mu_\alpha/\partial k_\iota, \partial\sigma_\alpha/\partial k_\iota$ and hence $\bar{\Delta}_{k,\alpha\iota}$, vanish for $\iota \notin [(\alpha-1)N_{\text{pop}} + 1, \alpha N_{\text{pop}}]$. We now make the ansatz

$$\delta K_{\alpha\beta} = \sum_\iota \frac{\epsilon_\iota}{\tau_{\text{m}}} \frac{\left(\boldsymbol{u}^\iota \boldsymbol{v}^{\iota\text{T}}\right)_{\alpha\beta}}{\left(\boldsymbol{SJ} + \boldsymbol{T}(\boldsymbol{J} \circledast \boldsymbol{J})\right)_{\alpha\beta}}\,, \tag{4.10}$$

which decomposes the changes $\boldsymbol{\delta K}$ into eigenmodes of the effective connectivity. The $\boldsymbol{u}^\iota$ and $\boldsymbol{v}^\iota$ are the $\iota$-th right and left eigenvectors of $\boldsymbol{M}$ as defined in Eq. (4.7), fulfilling the bi-orthogonality $\boldsymbol{v}^{\iota\text{T}}\boldsymbol{u}^\gamma = \delta_{\iota\gamma}$ and the completeness relation $\sum_\iota \boldsymbol{u}^\iota \boldsymbol{v}^{\iota\text{T}} = \mathbb{1}$. Inserting Eq. (4.9) with Eq. (4.10) into Eq. (4.8) yields

$$\bar{\boldsymbol{\Delta}}_{\boldsymbol{a}}\boldsymbol{\delta a} = -\sum_\iota \epsilon_\iota \boldsymbol{u}^\iota \boldsymbol{v}^{\iota\text{T}}\boldsymbol{\nu}. \tag{4.11}$$

Thus we can solve for $\epsilon_\iota$ by multiplying from the left with $\boldsymbol{v}^{\gamma\mathrm{T}}$

$$\begin{aligned}\underbrace{\boldsymbol{v}^{\gamma\mathrm{T}}\bar{\boldsymbol{\Delta}}_a\boldsymbol{\delta a}}_{=:\hat{a}^\gamma} &= -\sum_\iota \epsilon_\iota \underbrace{\boldsymbol{v}^{\gamma\mathrm{T}}\boldsymbol{u}^\iota}_{\delta_{\gamma\iota}}\underbrace{\boldsymbol{v}^{\iota\mathrm{T}}\boldsymbol{\nu}}_{=:\hat{\nu}^\iota} \\ \epsilon_\gamma &= -\frac{\hat{a}^\gamma}{\hat{\nu}^\gamma}\,.\end{aligned}$$

Our goal is to find a set of connections which dominate the size of the basin of attraction of the LA fixed point. Inserting Eq. (4.11) into Eq. (4.5) leads to

$$\begin{aligned}\boldsymbol{\delta\nu}^* &= \sum_\iota (\mathbb{1}-\boldsymbol{M})^{-1}\epsilon_\iota \boldsymbol{u}^\iota \boldsymbol{v}^{\iota\mathrm{T}}\boldsymbol{\nu} \\ &= \sum_\iota \frac{\epsilon_\iota}{1-\lambda_\iota}\boldsymbol{u}^\iota\hat{\nu}^\iota \\ &= \sum_\iota \frac{-\hat{a}^\iota}{1-\lambda_\iota}\boldsymbol{u}^\iota,\end{aligned}$$

where $\lambda_\iota$ are the eigenvalues of $\boldsymbol{M}$, which are either real or complex conjugate pairs since $\boldsymbol{M} \in \mathbb{R}^{N_{\mathrm{pop}}\times N_{\mathrm{pop}}}$. To determine the influence of each eigenmode on the shift of the fixed point, we project the eigenvectors $\boldsymbol{u}^\iota$ onto the fixed-point shift $\boldsymbol{\delta\nu}^*$ by multiplying each side with $\boldsymbol{\delta\nu}^*\boldsymbol{\delta\nu}^{*\mathrm{T}}$ and solve again for $\boldsymbol{\delta\nu}^*$ to obtain

$$\boldsymbol{\delta\nu}^* = \sum_\iota \underbrace{\frac{-\hat{a}^\iota}{1-\lambda_\iota}\frac{\boldsymbol{\delta\nu}^{*\mathrm{T}}\boldsymbol{u}^\iota}{\boldsymbol{\delta\nu}^{*\mathrm{T}}\boldsymbol{\delta\nu}^*}}_{=:\tilde{\eta}_\iota}\boldsymbol{\delta\nu}^*, \tag{4.12}$$

where we define the (possibly complex-valued) coefficients $\tilde{\eta}_\iota$. We aim at a decomposition of $\boldsymbol{\delta\nu}^*$ into real components. If $\mathrm{Im}(\lambda_\iota) = 0$, $\tilde{\eta}_\iota$ is real, so we can work directly with $\eta_\iota \coloneqq \tilde{\eta}_\iota$. Complex eigenvalues $\mathrm{Im}(\lambda_\iota) \neq 0$ and corresponding eigenvectors come in conjugate pairs, so in this case we combine the corresponding coefficients $\eta_\iota \coloneqq \tilde{\eta}_\iota + \tilde{\eta}_\iota^*$, to have all contributions $\eta_\iota \in \mathbb{R}$ and $\sum_\iota \eta_\iota = 1$ by construction (Eq. (4.12)). Each $\eta_\iota$ quantifies how much of the total fixed-point shift can be attributed to the $\iota$-th eigenmode, which allows identification of the most effective eigendirection (see Results), where we apply the ansatz Eq. (4.10) to the multi-area, multi-layer model of the vision-related areas of macaque visual cortex.

The spiking simulations of the network model were carried out on the JUQUEEN supercomputer (Jülich Supercomputing Centre, 2015). All simulations were performed with NEST version 2.8.0 (Eppler et al., 2015) with optimizations for the use on the supercomputer which will be included in a future release. The simulations use a time step of 0.1 ms and exact integration for the subthreshold dynamics of the leaky integrate-and-fire neuron model (reviewed in Plesser & Diesmann, 2009).

### Multi-area model

The multi-area model of the vision-related areas of macaque visual cortex uses the microcircuit model of (Potjans & Diesmann, 2014) as a prototype for all 32 areas in the FV91 parcellation (Felleman & Van Essen, 1991) and customizes it based on experimental findings on cortical structure. From anatomical studies, it is known that cortical areas in the macaque monkey are heterogeneous in their laminar structure and can be roughly categorized into 8 different architectural types based on cell densities and laminar thicknesses. This distinction was originally developed for prefrontal areas (Barbas & Rempel-Clower, 1997), and then extended to the entire cortex (Hilgetag et al., 2016). The visual cortex, and thus the model, comprises areas of categories 2, 4, 5, 6, 7 and 8. Precise layer-specific neuron densities are available for a number of areas, while for other areas, the neuron density is estimated based on their architectural type (see Schmidt et al. (2016) for details).

The inter-areal connectivity is based on binary data from the CoCoMac database (Stephan et al., 2001; Bakker et al., 2012; Felleman & Van Essen, 1991; Rockland & Pandya, 1979; Barnes & Pandya, 1992) indicating the existence of connections, and quantitative data from (Markov et al., 2014a). The latter are retrograde tracing data where connection strengths are quantified by the fraction of labeled neurons ($FLN$) in each source area. The original analysis of the experimental data was performed in the M132 atlas (Markov et al., 2014a). Both the FV91 and the M132 parcellations have been registered to F99 space (Van Essen, 2002), a standard macaque cortical surface included with the software tool Caret (Van Essen et al., 2001). This enables mapping between the two parcellations.

On the target side, we use the exact coordinates of the injections to identify the equivalent area in the FV91 parcellation. To map the data on the source side from the M132 atlas to the FV91 parcellation, we count the number of overlapping triangles on the F99 surface between any given pair of regions and distribute data proportionally to the amount of overlap using the F99 region overlap tool on http://cocomac.g-node.org. In the model, this $FLN$ is mapped to the indegree $K_{AB}$ the target area $A$ receives from source area $B$ divided by its total indegree, i.e., $FLN_{AB} = K_{AB} / \sum_{B'} K_{AB'}$. Here, $K_{AB}$ is defined as the total number of synapses between $A$ and $B$ divided by the total number of neurons in $A$. On the source side, laminar connection patterns are based on CoCoMac (Felleman & Van Essen, 1991; Barnes & Pandya, 1992; Suzuki & Amaral, 1994; Morel & Bullier, 1990; Perkel et al., 1986; Seltzer & Pandya, 1994) and on fractions of labeled neurons in the supragranular layers ($SLN$, Markov et al., 2014b). Gaps in these data are bridged exploiting a sigmoidal relation between $SLN$ and the logarithmized ratio of overall cell densities of the two areas, similar to (Beul et al., 2015). We map the $SLN$ to the ratio between the number of synapses originating in layer 2/3 and the total number of synapses between the two areas, assuming that only excitatory populations send inter-area connections, i.e., $SLN_{AB} = \sum_{\alpha} K_{AB}^{\alpha,2/3\mathrm{E}} N_A^{\alpha} / \sum_{\alpha,\beta} K_{AB}^{\alpha\beta} N_A^{\alpha}$, where the indices $\alpha$ and $\beta$ go over the different populations within area $A$ and $B$, respectively. In the context of the model, we use the terms $FLN$ and $SLN$ to refer to the relevant relative indegrees given here. On the

target side, the CoCoMac database provides data from anterograde tracing studies (Felleman & Van Essen, 1991; Rockland & Pandya, 1979; Jones et al., 1978; Seltzer & Pandya, 1991).

Missing inputs in the model, i.e., from subcortical and non-visual cortical areas, are replaced by Poissonian spike trains, whose rate $\nu_{\text{ext}}$ is a free, global parameter. In the original model all populations of a particular area receive the same indegree of external inputs $K_{\text{ext}}$. The only exception to this rule is area TH where the absence of granular layer 4 is compensated by an increase of the external input to populations 2/3E and 5E by 20 %. To elevate the firing rates in the excitatory populations in layers 5 and 6, we increase the external drive onto these populations. The possibility of a higher drive onto these populations is left open by the sparseness of the corresponding experimental data. We enhance the external Poisson drive of the 5E population, parametrized by the $K_{5E,\text{ext}}$ incoming connections per target neuron (indegree), in all areas by increasing $\kappa = K_{5\text{E,ext}}/K_{\text{ext}}$. The simultaneous increase in the drive of 6E needs to be stronger, since the firing rates in population 6E of the original model (Fig. 4.3C) are even lower than the rates of 5E (averaged across areas: 0.09 spikes/s for 5E compared to $2 \cdot 10^{-5}$ spikes/s for 6E). We thus scale up $K_{6\text{E,ext}}$ linearly with $\kappa$ such that $\kappa = 1.15$ results in $K_{6\text{E,ext}}/K_{\text{ext}} = 1.5$.

The synapse parameters of the multi-area model are defined in Tab. 4.1. The synaptic weights are normally distributed. Note, that the standard deviation $\delta w$ in the synaptic weights is neglected in the mean-field framework, where we assume the neurons to be homogeneous with respect to the input parameters. Taking this distribution into account would yield a contribution to the variance $\propto \delta w^2$ which is however negligible, i.e., it has almost no effect on the predicted rates (data not shown).

**Table 4.1:** Specification of the synapse parameters of the multi-area model.

| **Synapse parameters** | | |
|---|---|---|
| **Name** | **Value** | **Description** |
| $w \pm \delta w$ | $87.8 \pm 8.8$ pA | excitatory synaptic strength |
| $g$ | $-16$ (Fig. 4.3)<br>$-11$ (Fig. 4.9<br>Fig. 5.3) | relative inhibitory synaptic strength |
| $d_{\text{e}} \pm \delta d_{\text{e}}$ | $1.5 \pm 0.75$ ms | local excitatory transmission delay |
| $d_{\text{i}} \pm \delta d_{\text{i}}$ | $0.75 \pm 0.375$ ms | local inhibitory transmission delay |
| $d \pm \delta d$ | $d = s/v_{\text{t}} \pm \frac{1}{2} s/v_{\text{t}}$ | inter-areal transmission delay, with $s$ the distance between areas |
| $v_{\text{t}}$ | 3.5 m/s | transmission speed |

## 4.4 Interim discussion

In the following we discuss specific issues, regarding this chapter. A general summary and a broader discussion of the results follow in Chap. 9.

The cortico-cortical connectivity of the model is compiled from the extensive dataset of (Markov et al., 2014a) combined with the CoCoMac database (Stephan et al., 2001; Bakker et al., 2012), which collects data from hundreds of tracing studies. One could consider alternative methods for combining this information into one connectivity graph, for instance taking into account how consistently a given connection is reported across studies (Schmitt et al., 2014), and compare different methods by analyzing the resulting network dynamics. The presented mean-field theory could then be used to estimate the firing rates of each network instance without performing time-consuming simulations.

We restrict this study to networks of leaky integrate-and-fire model neurons, consistent with the key concept of the models we consider: individual cells are modeled in a simple manner to expose the impact of structural connectivity on the network dynamics. Moreover, the current-based leaky integrate-and-fire neuron can reproduce in-vivo like activity (Rauch et al., 2003; Jolivet et al., 2006) and is analytically tractable, which enables the identification of mechanisms underlying specific network effects. More complex neuron models can be incorporated into the method by replacing the gain function of each neuronal population with an analytical expression or an interpolated function obtained from spiking single-neuron simulations. For example, one could use a conductance-based point-neuron model for which the network dynamics can be described by population rate models (Shriki et al., 2003), featuring a non-monotonic gain function: the gain is reduced if excitatory and inhibitory inputs are increased in a balanced manner (Kuhn et al., 2004). Generally, this renders a system more stable. However, the bistability considered in our work is caused by excitatory inputs. Since conductance-based models also have a monotonically increasing gain function in dependence of the excitatory conductance alone, we expect the bistability to occur for such models as well.

Biological networks have various stabilization mechanisms not considered here, which render them less critical. For instance, during growth, homeostatic mechanisms guide the system toward the right structure. Furthermore, short-term synaptic plasticity (reviewed in (Morrison et al., 2008)), homeostatic synaptic scaling (Turrigiano et al., 1998) and spike-frequency adaptation (e.g. summarized in (Benda & Herz, 2003)) may prevent the system from entering the high-activity state. However, introducing these self-organizing mechanisms increases model complexity, causing a more intricate relation between structure and activity. Therefore, we start from a mean-field description on the level of neuronal populations, ignoring details of synaptic dynamics. Mild constraints on the activity lead to a network structure within the anatomical range of parameters. This network yields globally stable activity, suggesting that additional stabilization mechanisms are not required to achieve this. Nonetheless, they can potentially render the network more robust against external stimulation.

# Chapter 5

# Fluctuations around stationary states

---

The following chapter is based on the following publication:

Schuecker, J., Diesmann, M., & Helias, M. (2015). Modulated escape from a metastable state driven by colored noise. Phys. Rev. E 92, 052119.

**Author contributions:**

Under the supervision of M Helias, the author performed all parts of the above publication. All authors contributed to the writing of the manuscript.

So far we have investigated stationary properties in spiking neural networks by considering time-averages (right column in Fig. 1.5). In this chapter we address temporal fluctuations around the stationary state. To this end we use standard linear-response theory (Fig. 1.5: linearization) (Lindner et al., 2005; Ostojic & Brunel, 2011; Trousdale et al., 2012; Grytskyy et al., 2013) combined with the novel expression for the transfer function for the biologically relevant case of synaptic filtering derived in Chap. 2. Again, for illustrative purpose we first consider the simple $\mathcal{E} - \mathcal{I}$ network before investigating fluctuations in complex spiking networks.

## 5.1 Application to balanced random network

We here consider the $\mathcal{E} - \mathcal{I}$ network defined in Sec. 3.2. In contrast to the earlier chapters, the neurons here receive excitatory and inhibitory external Poisson drive with indegree $K_{\text{ext}} = 1$, $J_{\text{ext}} = J$, and rates $\nu_{\text{ext}}^{\mathcal{E},\mathcal{I}}$ calculated to fix the set point defined by the mean $\mu_{\mathcal{E}/\mathcal{I}} = \tau K_{\mathcal{E}} J(1 - \gamma g)\nu + \mu_{\text{ext}}$ and the standard deviation $\sigma_{\mathcal{E}/\mathcal{I}} = \sqrt{\tau K_{\mathcal{E}} J^2(1 + \gamma g^2)\nu + \sigma_{\text{ext}}^2}$ following from Eq. (3.7). Due to homogeneity the spike trains of the excitatory and inhibitory neurons can be summarized into populations with activity $\boldsymbol{s} = \begin{pmatrix} s_{\mathcal{E}}(t), & s_{\mathcal{I}}(t) \end{pmatrix}^T$, where $s_\alpha(t) = \frac{1}{N_\alpha} \sum_{i \in \alpha} s_i(t),\ \alpha \in \{\mathcal{E}, \mathcal{I}\}$. A sketch of the network architecture is shown in Fig. 5.1A.

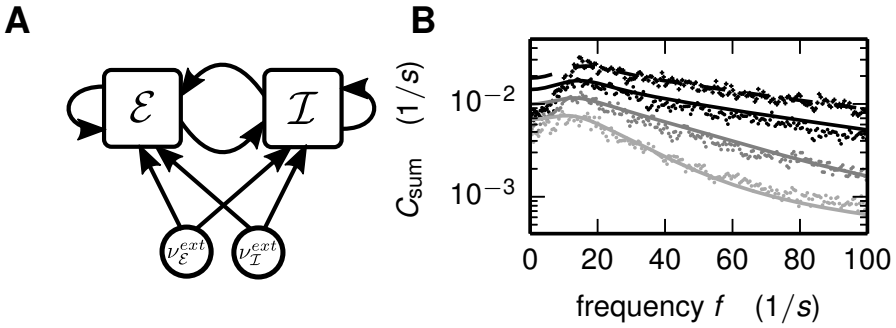



**Figure 5.1: Effect of synaptic filtering on power spectra of the population activity in a balanced random network model. A** Illustration of network architecture. **B** Theoretical prediction Eq. (5.4) (curves) vs. simulations (symbols) for white noise (black dashes, crosses) and different values of $\tau_{\text{s}} \in [0.5, 2, 4]$ ms (from black to gray, dots). The network operates at the set point $\mu = 16.42$ mV, $\sigma = 4$ mV. Neuron parameters are identical to the ones used in Fig. 2.3, except a reset of $V_{\text{R}} = 0$. Other network parameters: $N = 10,000,\ \gamma = 0.25,\ g = 5,\ J = 0.05$ mV, $d = 1.5$ ms. Power spectra obtained from a simulation of $T = 100\,s$ and smoothed by moving average (frame size 1 Hz).

The autospectra and cross-spectra $\boldsymbol{C}(\omega)$ of the population activity can be determined by mapping the system to a noisy linear rate model (Grytskyy et al., 2013)

$$r_\alpha(t) = [h_{\alpha\beta} * (r_\beta + \sqrt{\nu_\beta}\xi)(\circ - d)](t), \tag{5.1}$$

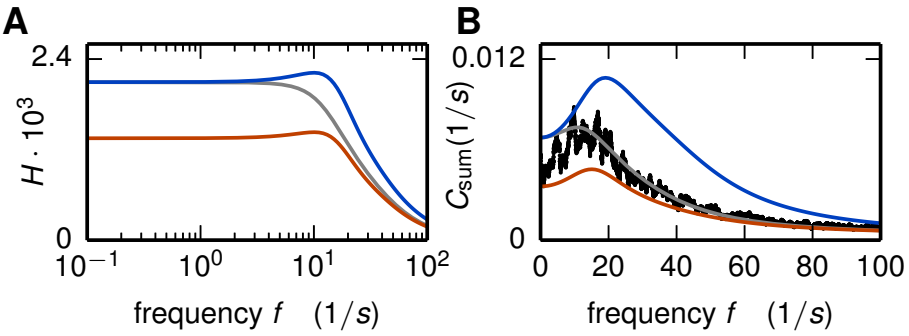



**Figure 5.2: Comparison of colored-noise and white-noise transfer function.** Parameters as in Fig. 5.1 with $\tau_s = 4\,\mathrm{ms}$. **A** Weighted transfer function Eq. (5.3) evaluated with colored-noise $\tilde{n}$ (gray), white noise $n$ (red) and white noise $n$ with corrected DC-limit, i.e., $\frac{n(\omega)\tilde{n}(0)}{n(0)}$ (blue) **B** Corresponding theoretical prediction Eq. (5.4) for power spectra. Gray curve is identical to the gray curve in Fig. 5.1B.

where we use $f(\circ - d)$ to denote the function $f$ shifted right by the synaptic delay $d$. The noise term $\xi$ originates from an approximation of the irregular spiking activity of the neurons by a modulated Poisson processes with rate $r(t)$, where the realization of the spikes amounts to a source of noise added to the output of the neurons.

The solution for $\boldsymbol{C}(\omega)$ is obtained in Fourier domain (Grytskyy et al., 2013)

$$\boldsymbol{C}(\omega) = \boldsymbol{P}(\omega)\,\boldsymbol{D}\,\boldsymbol{P}^T(-\omega) \tag{5.2}$$

with the diagonal matrix $\boldsymbol{D}_{\alpha\beta} = \delta_{\alpha\beta}\,\nu_\alpha\,N_\alpha^{-1}$ and the propagator matrix

$$P_{\alpha\beta}(\omega) = e^{-i\omega d} K_{\alpha\beta} H_{\alpha\beta}(\omega).$$

Here $H_{\alpha\beta}$ is the weighted transfer function

$$H_{\alpha\beta}(\omega) = \tau\nu_\alpha J_{\alpha\beta}\tilde{n}_G(\omega)\,\frac{1}{1+i\omega\tau_s} + \tau\nu_\alpha J_{\alpha\beta}^2\tilde{n}_H(\omega), \tag{5.3}$$

with $\nu_\alpha$ given by Eq. (2.44) and identifying $G = \sqrt{2}/\sigma$, $H = 1/\sigma^2$ in $\tilde{n}_G$ (Eq. (2.49)) and the corresponding expression for $\tilde{n}_H$. The latter takes into account the contribution to the colored-noise transfer function originating from a modulation of the variance. Its form is again obtained from a shift of the boundaries in the white-noise solution $n_H$ (Eq. (2.48)), since the general method of reduction (Sec. 2.2) holds true also for a time-dependent variance $\sigma^2(t) = \sigma^2 + H\sigma^2 e^{i\omega t}$ as shown in A.4.

From Eq. (5.2) we obtain the autospectra of the network activity $s_{\mathrm{sum}} = \frac{1}{N}\sum_{i\in\mathcal{E},\mathcal{I}} s_i$ by a weighted average over the matrix entries

$$C_{\mathrm{sum}} = \frac{1}{N}(N_\mathcal{E}^2 C_{\mathcal{E}\mathcal{E}} + N_\mathcal{E} N_\mathcal{I} C_{\mathcal{E}\mathcal{I}} + N_\mathcal{I} N_\mathcal{E} C_{\mathcal{I}\mathcal{E}} + N_\mathcal{I}^2 C_{\mathcal{I}\mathcal{I}}). \tag{5.4}$$

The analytical prediction Eq. (5.4) is in excellent agreement with the spectra obtained in direct simulations of the network (Fig. 5.1). The shape of the power spectra is well captured for different values of $\tau_s$ and the theory predicts the suppression of fluctuations at frequencies in the low (30−60Hz) and the high gamma range (60−200Hz) caused by synaptic filtering. These frequency bands are related to task dependent activity studied in animals (Ray et al., 2008) and humans (Ball et al., 2008). Although the synaptic time constants are small compared to the membrane time constant ($k = \sqrt{\tau_s/\tau_m} \leq 0.44$), the influence on the network dynamics is strong. We here present for the first time an analytical argument explaining the effect.

To further strengthens the importance of our results we also show the spectra obtained with the already known white-noise transfer function (Fig. 5.2). The latter is not able to accurately predict the spectra, e.g. the height of the peak is strongly underestimated. Even if we correct the DC-limit to the colored-noise case, the prediction of the spectra is far off. This considerations show that it is important to correctly capture the effect of colored noise on the DC-limit as well as on the frequency dependence by using our newly derived expression for the colored-noise transfer function.

The observed deviations at low frequencies are expected: The theory in (Grytskyy et al., 2013) assumes the autocorrelation of single neurons to be $\delta$-shaped due to the Poisson assumption on the spike-train statistics. This is an approximation likely to be violated, since after reset the membrane potential needs time to recharge, resulting in a dip of the autocorrelation around zero corresponding to a reduction of power at low frequencies. However, the spectrum of the population activity is dominated by cross-correlations resulting in a good overall agreement between theory and simulation. Dummer et al. (2014) show that the influence of the cross-correlations on the single neuron's auto-correlation is negligible and developed an iterative procedure which numerically solves for the self-consistent solution of the auto-correlation. One could incorporate this approach into the theory presented here, replacing the simplified assumption of a $\delta$-shaped auto-correlation with its self-consistent solution.

## 5.2 Application to complex spiking networks

In the last section we have illustrated how the power spectra of the network activity are accurately predicted within the linear-response framework by Eq. (5.2). As the latter expression is analytical, this could be used to almost instantaneously scan the behavior of the network for different parameters without performing time-consuming simulation of the spiking network.

Moreover, there is another advantage of having an analytical description at hand, which especially comes into play when considering the dynamics of complex network: Since the analytic expression relates the network connectivity $\boldsymbol{K}$ to the dynamical measure $\boldsymbol{C}$ it allows the identifications of mechanisms by which the connectivity shapes the dynamics. For example the cortical microcircuit model (Potjans & Diesmann, 2014) shows a strong peak in the power spectrum, which can be seen as a collective oscillation of the neurons. However, the simulation alone does not reveal its anatomical origin. Bos et al. (2016) develop a sensitivity measure

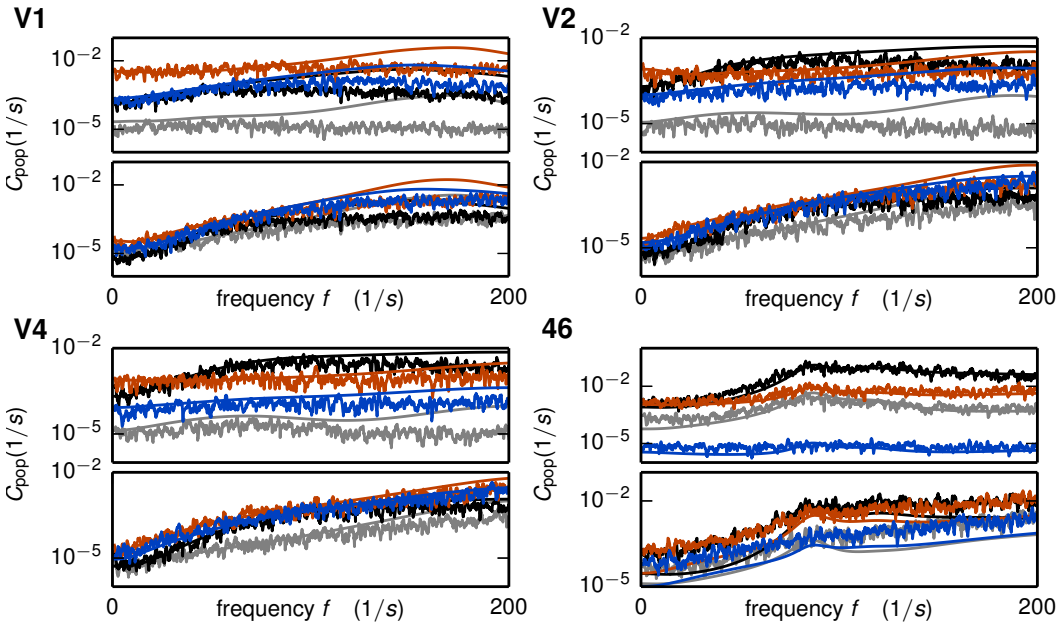



**Figure 5.3: Population power spectra in the multi-area model.** Upper left: Power spectra (diagonal elements of **C** Eq. (5.2), solid lines) in area V1 for excitatory populations (upper panel) and inhibitory populations (lower panel). Layers $2/3, 4, 5, 6$ encoded in colors grey, black, red, blue. Power spectra obtained from a simulation of $T = 10\,s$ and smoothed by moving average (frame size 1 Hz) shown as dots. Other panels for areas V2, V4 and 46, respectively. Network parameters specified in Tab. 4.1.

based on the analytical framework described in Sec. 5.1 including the novel expression for the transfer function (Chap. 2). The measure shows that the peak is generated by the subcircuit of Layer 2/3 and Layer 4 and thus establishes a direct link between structure and dynamics.

It is believed that collective oscillations play a major role in the communication between areas. The so called communication through coherence hypothesis (Fries, 2005) states that oscillations control neuronal excitability and thereby open and close time windows for communication. In recent studies it was further found that areas, in particular focusing on areas V1,V2 and V4, communicate on different frequency bands depending on the hierarchical relation between the involved areas (Bastos et al., 2015b; van Kerkoerle et al., 2014; Bastos et al., 2015) (see also Chap. 10). As the multi-area network (Sec. 4.3) models cortical areas along the hierarchy it serves as an ideal test environment for understanding the mechanism leading to these experimental observations. To this end it is very important to understand the emergence of oscillations in the network of areas.

Fig. 5.3 shows the power spectra in the population activity Eq. (5.2) for the multi-area model in the relevant areas. The analytic prediction is in excellent agreement with the spectra obtained with the full spiking simulation for frequencies below 100 Hz. We also include area 46 to show that the prediction also works for the areas higher in the hierarchy. Minor deviations occur for low frequencies in area 46 for the inhibitory populations in terms of an underestimation

of the spectra, which was observed and explained already in Sec. 5.1. In general, we expect that the analytic prediction breaks down for larger frequencies, because our expression for the transfer function is not valid in this frequency regime (Sec. 2.3.3). Indeed, we observe that the analytical prediction for the power spectrum for frequencies larger than 100 Hz occasionally overestimates the power especially in the excitatory populations in V1 and V2. Furthermore, one obvious observation from Fig. 5.3 is that the spectra in V1,V2 and V4 do not feature prominent peaks. A further study of the spectra and their origin is out of scope of the current thesis and left for future work. We here form the basis allowing for these important investigations, which is further discussed in Chap. 10.

# Part III

# Dynamic mean-field theory for stochastic rate dynamics

# Chapter 6

# Breakdown of mean-field theory for spiking dynamics

In Chap. 3 and Chap. 4 we have seen how the stationary state in spiking networks can be accurately captured by mean-field theory. In particular, we developed a mean-field based method which allowed us to find a realistic level of activity in the multi-area-model (Fig. 4.9). The state is realistic on the level of firing rates. To further improve the agreement between model activity and fMRI measurements the inter-area coupling weights need to be strengthened (Schmidt et al. (2016)). However, in the regime with increased coupling the mean-field predictions for the stationary-state firing rates break down (Fig. 6.1B); the rates in the spiking simulation are higher compared to the mean-field prediction. A closer examination of the spiking behavior reveals that populations that show large deviations between the simulated and the theoretically predicted rate, exhibit bursty activity: neurons emit several spikes within a short time interval followed by a long time interval of silence (horizontal stripes in Fig. 6.1A, upper panel). In the following we will investigate the underlying reason for the deviations between theory and simulation.

Let us for a moment recapitulate the underlying assumptions: In order to arrive at a self-consistent description we assumed Poisson statistics for the spike-times and neglected cross-correlations between neurons. Furthermore we assumed the synaptic weights $J$ to be small justifying the diffusion approximation. In addition, we assume stationary activity as we consider the stationary solution of the Fokker Planck equation.

To investigate the breakdown we turn back to the two-dimensional (number of populations) random balanced network model with homogeneous drive (Sec. 3.2.2). Fig. 6.2 shows the mean-field prediction for the stationary rate as a function of the coupling weight $J$. For small values of $J$ the prediction is accurate corresponding to the regimes we have considered in the previous chapters. The corresponding raster plot (Fig. 6.2A, upper panel) indicates that the neurons emit spikes at a relatively constant rate. For increased weights the mean-field solution strongly underestimates the time-averaged rates, which was found already by Tetzlaff et al. (2012) and by Ostojic (2014). As in the case of the multi-area model, the spike emission of the neurons is bursty in this regime (Fig. 6.2B, lower panel).

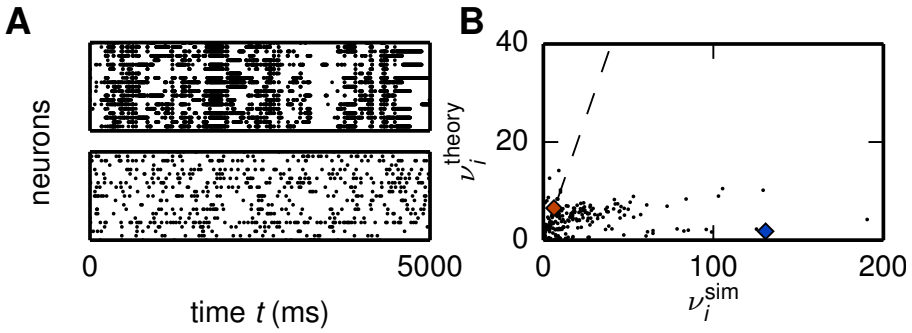



**Figure 6.1: Breakdown of mean-field theory in the multi-area model with increased inter-area coupling.** **A** Upper panel: Raster plot for 20 neurons in area $7a$ population $5E$. Lower panel. Raster plot for 20 neurons in area $AITv$ population $5I$. **B** Scatter plot of analytical prediction of firing rates $\nu_i$ versus rates $\nu_i^{\text{sim}}$ obtained from a spiking simulation of the multi-area network model. Populations corresponding to upper/lower panel in A marked in blue/red. Network parameters as in Fig. 4.9, except with increased cortico-cortical coupling weights $w_{cc} \to 1.9 \cdot w_{cc}$. Data for spiking simulation obtained from Maximilian Schmidt, personal communication.

In other words, in the strongly coupled regime the rates of the individual neurons show strong temporal fluctuations, which can not be captured by the stationary solution of the Fokker-Planck equation. Furthermore, Wieland et al. (2015) showed that the fluctuations come along with a Fano factor an order of magnitude larger than for the assumed Poisson process. Third, as the weights increase the diffusion approximation becomes worse.

In principle, the emerging fluctuations can only be captured by a dynamical mean-field theory (DMFT). The latter, a well known concept in physics (e.g., Sénéchal et al., 2006, Part III), has been brought to neuroscience by the seminal work by Sompolinsky et al. (1988) showing that a network of randomly coupled rate neurons display a transition from a fixed point to chaotic fluctuations at a critical coupling strength, which will be explained in detail in Chap. 7. Ostojic (2014) claims that the observed fluctuations in the network of spiking neurons originate from an analog transition to chaotic dynamics. However, it is controversially discussed, whether the instability of deterministic rate dynamics explains a transition to strongly fluctuating firing rates in networks of spiking neurons (Ostojic, 2014; Engelken et al., 2015; Ostojic, 2015). In a recent study Mastroguiseppe & Ostojic (2016) investigate rate networks using DMFT and show how the emerging fluctuations lead to an increase in the averaged activity, where the increase is due to an interplay between the fluctuations and and asymmetry in the gain function prohibiting negative values for the rates. However, they find that these results do not make a quantitative prediction for the increase in the rates found in spiking networks.

For the analysis of oscillations and correlations a mapping between a network of spiking model neurons to a network of noisy linear rate models is used (Chap. 5, Eq. (5.1)). In this mapping the irregular spiking activity of the neurons is approximated by Poisson processes, where the realization of the spikes amounts to the source of noise, also termed spiking noise (Kadmon &

Sompolinsky, 2015b). This constitutes a striking difference to the deterministic rate networks considered in Sompolinsky et al. (1988) and in Ostojic (2014), where any noise term is absent. In the next chapter we will therefore investigate the effect of noise on the transition to chaos in neural networks.

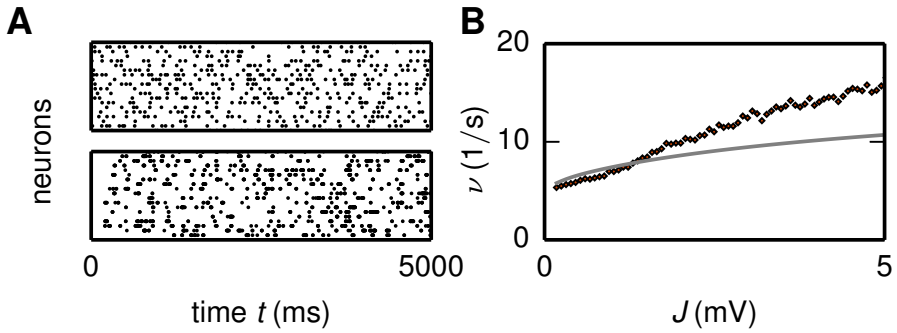



**Figure 6.2: Breakdown of mean-field theory.** **A** Upper panel: Raster plot for 20 excitatory neurons in random balanced network with coupling weight $J = 0.1756$ mV. Lower panel: Raster plot for $J = 5.23$ mV. **B** Analytic prediction (solid curves, obtained from Eq. (3.7)) vs. population- and time-averaged rates obtained from a spiking simulation (diamonds). Network and simulation parameters as in Fig. 3.2 with $g = -6$ $\nu_{ext} = 8$ Hz and adapted weights.

# Chapter 7

# Noise dynamically suppresses chaos in random neural networks

---

The following chapter is partly based on the following publications:

[1] Goedeke, S., Schuecker, J., & Helias, M. (2016a). Noise dynamically suppresses chaos in neural networks. arXiv preprint arXiv:1603.01880v2.
[2] Schuecker, J., Goedeke, S., Dahmen, D., & Helias, M. (2016). Functional methods for disordered neural networks. arXiv. 1605.06758 [cond-mat.dis-nn].

**Author contributions:**

[1] Sven Goedeke and the author performed the work concertedly with equal shares on all parts of the study. Moritz Helias supervised the work. All Authors contributed to the writing of the manuscript.

[2] The following chapter is partly based on Section II of this publication. Under the supervision of M Helias, S Goedeke, D Dahmen and the author jointly worked on sections II A,B of the publication. Under the supervision of M Helias, S Goedeke and the author jointly worked on sections II C,D,E,F of the publication. All authors contributed to the writing of the manuscript.

## 7.1 Introduction

Networks of randomly coupled neurons with time-continuous and deterministic dynamics display a transition from a fixed point to chaotic fluctuations at a critical coupling strength (Sompolinsky et al., 1988). This transition is illustrated in Fig. 7.1A. The transition is well understood by a dynamical mean-field theory, originally developed for spin glasses (Sompolinsky & Zippelius, 1981, 1982). The onset of chaos is equivalent to the emergence of a decaying autocorrelation function, whose decay time diverges at the transition. This equivalence has been used in several subsequent studies (Rajan et al., 2010; Aljadeff et al., 2015; Harish & Hansel, 2015). Furthermore a tight relationship to random matrix theory exists: the transition happens precisely when the fixed point becomes linearly unstable, which identifies the spectral radius of the random connectivity matrix (Sommers et al., 1988; Rajan & Abbott, 2006) as the parameter controlling the transition.

In the presence of noise, a decaying autocorrelation function does not indicate chaos, since the noisy fluctuations cause perpetual decorrelation also for stable dynamics (Fig. 7.1b). Therefore, the transition to chaos, if existent at all, must be qualitatively different from the noiseless case. Recently, Kadmon & Sompolinsky (2015b) showed that already weak noise smooths the transition to chaos in networks of stochastically spiking neurons. However, the mechanism by which noise of arbitrary strength affects the maximum Lyapunov exponent and the location of the transition is not understood.

In random neural networks with simpler discrete-time dynamics chaos is suppressed by noise (Molgedey et al., 1992). However, the neuron dynamics does not possess temporal memory and the corresponding mean-field theory shows that the chaotic fluctuations are temporally uncorrelated even without noise. Therefore, these results can not be transferred to continuous-time dynamics with memory, where the temporal correlations play an important role.

## 7.2 Dynamic Mean-field equation

We study the continuous-time dynamics of a random network of $N$ neurons, whose states $x_i(t) \in \mathbb{R}$ evolve according to the system of stochastic differential equations

$$\frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} J_{ij}\phi(x_j) + \xi_i(t)\,, \tag{7.1}$$

$i = 1, \ldots, N$. The $J_{ij}$ are independent and identically Gaussian distributed random coupling weights with zero mean and variance $g^2/N$, where the intensive gain parameter $g$ controls the recurrent coupling strength or, equivalently, the weight heterogeneity of the network. We further exclude self-coupling, setting $J_{ii} = 0$. The $\xi_i(t)$ are independent Gaussian white noise processes with autocorrelation function $\langle \xi_i(t)\xi_i(s)\rangle = 2\sigma^2\delta(t-s)$. We choose the sigmoidal transfer function $\phi(x) = \tanh(x)$, so that without noise, for $\sigma = 0$, the model agrees with the

one studied in (Sompolinsky et al., 1988).

The dynamical system (7.1) contains two sources of randomness: the quenched disorder due to the random coupling weights and temporally fluctuating noise. A particular realization of the random couplings $J_{ij}$ defines a fixed network configuration and its dynamical properties usually vary between different realizations. For large network size $N$, however, certain quantities are self-averaging, meaning that their values for a typical realization can be obtained by an average over network configurations (Fischer & Hertz, 1991). An important example is the population-averaged autocorrelation function.

We here derive a dynamical mean-field theory that describes the statistical properties of the system under the joint distribution of disorder, noise and possibly random initial conditions in the limit of large network size $N \to \infty$. The theory can be derived via a heuristic "local chaos" assumption (Amari, 1972) or using a generating functional formulation (Sompolinsky & Zippelius, 1982; Crisanti & Sompolinsky, 1987). We here follow the latter approach, because it casts the problem into the established language of statistical field theory for which a wealth of approximation techniques are available (Zinn-Justin, 1996). A mathematically rigorous proof uses large deviation techniques (Cabana & Touboul, 2013). The general idea is that for large network size $N$ the local recurrent input $\sum_{j=1}^{N} J_{ij}\phi(x_j)$ in Eq. (7.1) approaches a Gaussian process with self-consistently determined statistics.

We formulate the problem Eq. (7.1) in terms of a generating functional $Z$ from which we can derive all moments of the activity. Using the Martin-Siggia-Rose-De Dominicis-Janssen path integral formalism (Martin et al., 1973; De Dominicis & Peliti, 1978; Altland & B., 2010) we obtain

$$Z[\mathbf{l}](\mathbf{J}) = \int \mathcal{D}\mathbf{x} \int \mathcal{D}\tilde{\mathbf{x}} \exp\left(S_0[\mathbf{x},\tilde{\mathbf{x}}] - \tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{J}\phi\left(\mathbf{x}\right) + \mathbf{l}^{\mathrm{T}}\mathbf{x}\right) \tag{7.2}$$

$$\text{with } S_0[\mathbf{x},\tilde{\mathbf{x}}] = \tilde{\mathbf{x}}^{\mathrm{T}}\left(\partial_t + 1\right)\mathbf{x} + \frac{D}{2}\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}, \tag{7.3}$$

where $\mathbf{x}^T\mathbf{y} = \sum_i \int x_i(t)y_i(t)\,dt$ denotes the scalar product in time and in space and $\tilde{\mathbf{x}}$ represents a so called response field. The measures are defined as $\int \mathcal{D}\mathbf{x} = \lim_{M\to\infty} \Pi_{l=1}^{N}\Pi_{k=1}^{M} \int_{-\infty}^{\infty} dx_k^l$ and $\int \mathcal{D}\tilde{\mathbf{x}} = \lim_{M\to\infty} \Pi_{k=1}^{N}\Pi_{l=0}^{M-1} \int_{-i\infty}^{i\infty} \frac{d\tilde{x}_k^l}{2\pi i}$ with the subscript $k$ denoting the $k$-th unit and the superscript $l$ denoting the $l$-th time slice. The action $S_0$ is defined to contain all single unit properties, therefore excluding the coupling term $-\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{J}\phi\left(\mathbf{x}\right)$, which is written explicitly.

Motivated by the self-averaging property, we average over the quenched disorder and perform a saddle-point approximation (Sec. B.1). The resulting functional factorizes into $N$ terms, reducing the network to $N$ non-interacting units on a background of an independent Gaussian noise with common self-consistently determined statistics. At this level of approximation, the problem is hence equivalent to a single unit system. The effective equation of motion corresponding to this system reads

$$\frac{dx}{dt} \quad = \quad -x + \eta(t) + \xi(t). \tag{7.4}$$

Here, $\xi(t)$ is a Gaussian white noise process as in Eq. (7.1), independent of $\eta(t)$. Because the random couplings $J_{ij}$ have zero mean, the Gaussian process $\eta(t)$ is centered, $\langle\eta(t)\rangle = 0$, and thus fully specified by its autocorrelation function

$$\begin{aligned}\langle\eta(t)\eta(s)\rangle &= g^2\langle\phi(x(t))\phi(x(s))\rangle \\ &=: g^2 C_{\phi(x)\phi(x)}(t,s),\end{aligned} \tag{7.5}$$

where we defined the average autocorrelation function $C_{\phi(x)\phi(x)}(t,s)$ of the non-linearly transformed activity of the units (Eq. (B.9)).

## 7.3 Effective motion of particle in a potential

Our goal is to determine the mean-field autocorrelation function $\langle x(t)x(s)\rangle$, which, by self-averaging, also describes the population-averaged autocorrelation function. Assuming that $x(t)$ is a stationary process, $c(\tau) = \langle x(t+\tau)x(t)\rangle$ obeys the differential equation (Sec. B.2)

$$\ddot{c} = \frac{d^2c}{d\tau^2} = c - g^2 f_\phi(c, c_0) - 2\sigma^2\delta(\tau) \tag{7.6}$$

with $c_0 = c(0)$. The Dirac-$\delta$ inhomogeneity originates from the white-noise autocorrelation function and is absent in Sompolinsky et al. (1988). The same inhomogeneity arises from Poisson spiking noise with $2\sigma^2 = g^2 r$ (Kadmon & Sompolinsky, 2015b), where $r$ is the population-averaged firing rate. In Eq. (7.6) we write $f_\phi(c(\tau), c_0) = C_{\phi(x)\phi(x)}(t+\tau, t)$, introducing the notation

$$f_u(c(\tau), c_0) = \iint u\left(\sqrt{c_0 - \tfrac{c(\tau)^2}{c_0}}z_1 + \tfrac{c(\tau)}{\sqrt{c_0}}z_2\right) u\left(\sqrt{c_0}z_2\right)\, Dz_1 Dz_2 \tag{7.7}$$

for an arbitrary function $u(x)$ and the Gaussian integration measure $Dz = \exp(-z^2/2)/\sqrt{2\pi}\, dz$. This representation holds since $x(t)$ is itself a Gaussian process. Note that Eq. (7.7) reduces to a one-dimensional integral for $f_u(c_0, c_0) = \langle u(x)^2\rangle$ and $f_u(0, c_0) = \langle u(x)\rangle^2$, where $x$ has zero mean and variance $c_0$.

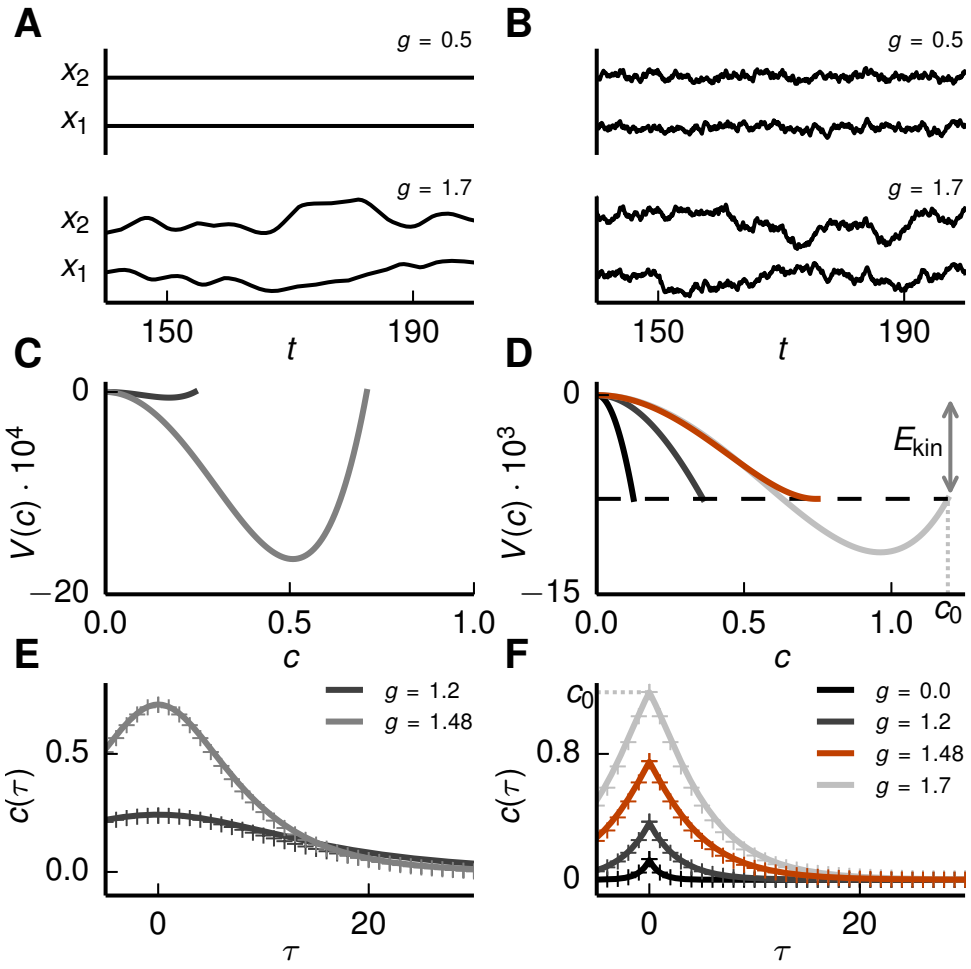



**Figure 7.1: Solution of the mean-field theory.** Noiseless case $\sigma = 0$ (left column) and noisy case $\sigma = \sqrt{0.125}$ (right column). **Upper row**: Simulated trajectories of two example neurons for sub-critical $g = 0.5$ (upper part of vertical axis) and super-critical $g = 1.7$ (lower part of vertical axis). **Middle row**: Classical potential Eq. (7.9) with self-consistent variance $c_0$ following from energy conservation Eq. (7.10) for different coupling strength $g$ (corresponding legends in lower row); in the noisy case the critical value $g_c = 1.48$ from Eq. (7.19) is shown in red. Dashed line represents initial kinetic energy $E_{\text{kin}} = \sigma^4/2$. **Lower row**: Corresponding self-consistent autocorrelation function (solid line) compared to simulations (crosses). The variance $c_0$ is given by the largest value of $c$ at which the potential (middle row) is defined, indicated for $g = 1.7$ with gray dotted lines in D and F. Network size in simulations is $N = 10000$.

We formulate Eq. (7.6) as the one-dimensional motion of a classical particle in a potential:

$$\ddot{c} = -V'(c) - 2\sigma^2\delta(\tau)\,, \tag{7.8}$$

where we define

$$V(c) = V(c;c_0) = -\frac{1}{2}c^2 + g^2 f_\Phi(c,c_0) - g^2 f_\Phi(0,c_0)\,, \tag{7.9}$$

with $\Phi(x) = \int_0^x \phi(y)\,dy$ and $\partial/\partial c\, f_\Phi(c,c_0) = f_\phi(c,c_0)$ following from Price's theorem (Papoulis, 1991). The autocorrelation $c(\tau)$ here plays the role of the position of the particle and the time lag $\tau$ the role of time. The potential Eq. (7.9) depends on the initial value $c_0$, which has to be determined self-consistently. We obtain $c_0$ from classical energy conservation $\dot{c}^2/2 + V(c) =$ constant. Considering $\tau \geq 0$ and the symmetry of $c(\tau)$, the noise term in Eq. (7.8) amounts to an initial velocity $\dot{c}(0+) = -\sigma^2$ and thus to the kinetic energy $\dot{c}^2(0+)/2 = \sigma^4/2$. Since $|c(\tau)| \leq c_0$, the solution $c(\tau)$ and its first derivative must approach zero as $\tau \to \infty$. Thus we obtain the self-consistency condition for $c_0$ as

$$\frac{1}{2}\sigma^4 + V(c_0;c_0) = V(0;c_0) = 0. \tag{7.10}$$

For the noiseless case, Fig. 7.1C,E shows the resulting potential and the corresponding self-consistent autocorrelation function $c(\tau)$ in the chaotic regime. Approaching the transition from above, $g \to g_c = 1$, the amplitude $c_0$ vanishes and the decay time of $c(\tau)$ diverges (Sompolinsky et al., 1988). This picture breaks down in the noisy case (Fig. 7.1D,F), where $c_0$ is always nonzero and $c(\tau)$ decays and has a kink at zero. The mean-field prediction is in excellent agreement with the population-averaged autocorrelation function obtained from numerical simulations showing that the self-averaging property is fulfilled. In the following we derive a condition for the transition from stable to chaotic dynamics in the presence of noise.

## 7.4 Effect of noise on the transition to chaos

The maximum Lyapunov exponent quantifies how sensitive the dynamics depends on the initial conditions (Eckmann & Ruelle, 1985). It measures the asymptotic growth rate of infinitesimal perturbations. For stochastic dynamics the stability of the solution for a fixed realization of the noise is also characterized by the maximum Lyapunov exponent [1]: If it is negative, trajectories with different initial conditions converge to the same time-dependent solution; the dynamics is stable. If it is positive, the distance between two initially arbitrary close trajectories grows exponentially in time; the dynamics exhibits sensitive dependence on initial conditions and is hence chaotic.

We derive the maximum Lyapunov exponent by using dynamical mean-field theory. To this end, we consider two copies of the network distinguished by superscripts $\alpha \in \{1, 2\}$. These copies, or replicas, have identical coupling matrix $\mathbf{J}$ and, for $\sigma > 0$, are subject to the same realization of the noise $\xi_i(t)$. The maximum Lyapunov exponent can be defined as the asymptotic growth rate of the Euclidean distance between trajectories of the two copies

$$\lambda_{\max} = \lim_{t\to\infty} \lim_{||\mathbf{x}^1(0)-\mathbf{x}^2(0)||\to 0} \frac{1}{2t} \ln\left(\frac{||\mathbf{x}^1(t) - \mathbf{x}^2(t)||^2}{||\mathbf{x}^1(0) - \mathbf{x}^2(0)||^2}\right) .$$

We now follow an idea by Derrida & Pomeau (1986) and exploit the self-averaging property of population-averaged correlation functions, i.e, $\frac{1}{N}\sum_{i=1}^{N} x_i^\alpha(t)x_i^\beta(s) \approx c^{\alpha\beta}(t, s)$, where $c^{\alpha\beta}(t, s)$ denote the correlation functions averaged over the realization of the couplings. We express the mean squared Euclidean distance as

$$\frac{1}{N}\sum_{i=1}^{N}\left(x_i^1(t) - x_i^2(t)\right)^2 \approx c^{11}(t, t) + c^{22}(t, t) - 2c^{12}(t, t) \equiv d(t) , \tag{7.11}$$

where we defined the mean-field squared distance $d(t)$. Thus the asymptotic growth rate of $d(t)$ provides us with a mean-field description of the maximum Lyapunov exponent. To obtain the growth rate we first define

$$d(t, s) = c^{11}(t, s) + c^{22}(t, s) - c^{12}(t, s) - c^{21}(t, s) \tag{7.12}$$

with the obvious property $d(t) = d(t, t)$. We then consider the temporal evolution of $d(t, s)$ for infinitesimally perturbed initial conditions $||\mathbf{x}^1(0) - \mathbf{x}^2(0)|| = \epsilon$. To this end it is again convenient to use a generating functional that captures the joint statistics of the two systems and in addition allows averaging over the quenched disorder (see also Zinn-Justin, 1996, Appendix 23, last remark). The generating functional describing the two copies is defined

[1] The theory of random dynamical systems makes this more precise; a brief overview is given in (Lajoie et al., 2013).

analog to the single system Eq. (7.3) as

$$Z[\{\mathbf{l}^\alpha\}_{\alpha\in\{1,2\}}](\mathbf{J}) = \Pi_{\alpha=1}^{2}\Big\{\int \mathcal{D}\mathbf{x}^\alpha \int \mathcal{D}\tilde{\mathbf{x}}^\alpha \, \exp\Big(S_0[\mathbf{x}^\alpha,\tilde{\mathbf{x}}^\alpha] - \tilde{\mathbf{x}}^{\alpha\mathrm{T}}\mathbf{J}\phi\left(\mathbf{x}^\alpha\right) + \mathbf{l}^{\alpha\mathrm{T}}\mathbf{x}^\alpha\Big)\Big\}$$
$$\times \exp\left(2\sigma^2\tilde{\mathbf{x}}^{1\mathrm{T}}\tilde{\mathbf{x}}^2\right). \tag{7.13}$$

The factor in the last line results from the common white noise in the two copies and effectively couples the two systems. We also note that the coupling matrix $\mathbf{J}$ is the same in both copies.

Averaging Eq. (7.13) over the quenched disorder and performing the saddle-point approximation we obtain a pair of effective dynamical equations (Sec. B.3),

$$(\partial_t + 1)\, x^\alpha(t) = \xi(t) + \eta^\alpha(t)\,, \quad \alpha \in \{1,2\}\,, \tag{7.14}$$

together with a set of self-consistency equations for the statistics of the noises $\eta^\alpha$

$$\langle \eta^\alpha(s)\, \eta^\beta(t)\rangle = g^2\, \langle \phi(x^\alpha(s))\phi(x^\beta(t))\rangle. \tag{7.15}$$

There are two terms which effectively couple the two copies. First the noise $\xi$ represents common temporal fluctuations injected into both systems. Second the effective noises $\eta^\alpha$ and $\eta^\beta$ are correlated between replicas Eq. (7.15), arising from the two systems having the same coupling $\mathbf{J}$ in each realization. The origin of the latter coupling is hence of static nature.

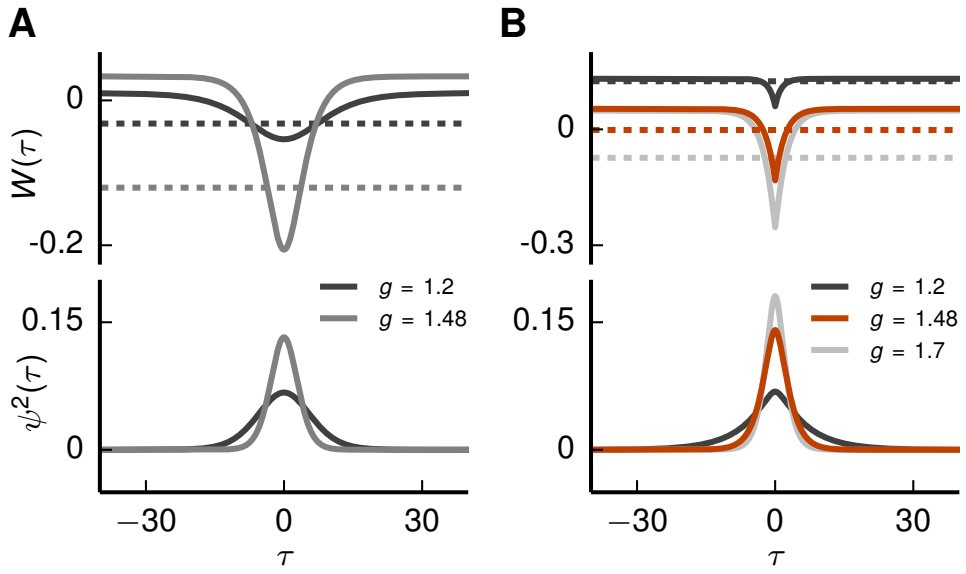



**Figure 7.2: Ground state of Schrödinger equation.** Upper part of vertical axis: Quantum potential $W$ (solid curve) and ground state energy $E_0$ (dashed line) for noiseless case **A** and noisy case **B** for $\sigma = \sqrt{0.125}$. Lower part of vertical axis: Corresponding squared ground state wave function. Parameters as in Fig. 7.1 (noisy case for $g = 0$ left out).

The distance Eq. (7.12) between the two copies is given by the auto-correlations of the single systems and the cross-correlations between them. We consider the case where both copies are prepared with identical initial conditions and thus are fully synchronized; the cross-correlation $c^{12}$ equals the auto-correlations $c^{11}$, $c^{22}$. The latter are identical to the single-system autocorrelation function $c$ as the the marginal statistics of each subsystem cannot be affected by the mere presence of the respective other system. Moreover we consider the stationary state, as the Lyapunov exponent is given by an asymptotic limit. An increase of the distance $d(t)$, by Eq. (7.12), amounts to the departure of $c^{12}$ from its initial value $c$. To determine the growth rate in the limit of small distances $d(0) \propto \epsilon$ between the two copies we therefore expand the cross-correlation around its stationary solution $c^{12}(t,s) = c(t-s) + \epsilon\, k^{(1)}(t,s)\,,\ \epsilon \ll 1$, which leads to an equation of motion for the first order deflection (Sec. B.3.1)

$$(\partial_t + 1)\,(\partial_s + 1)\,k^{(1)}(t,s) = g^2 f_{\phi'}(c(t-s), c_0)\,k^{(1)}(t,s) \tag{7.16}$$

with $d(t) = -\epsilon\, k^{(1)}(t,t)$.

A separation ansatz in the coordinates $\tau = t - s$ and $T = t + s$ then yields an eigenvalue problem in the form of a time-independent Schrödinger equation (Sompolinsky et al., 1988; Kadmon & Sompolinsky, 2015b) (Sec. B.3.2)

$$\left[-\partial_\tau^2 + W(\tau)\right]\psi(\tau) = E\,\psi(\tau), \tag{7.17}$$

where now $\tau$ plays the role of a spatial coordinate. Here, the quantum potential $W(\tau) = -V''(c(\tau)) = 1 - g^2 f_{\phi'}(c(\tau), c_0)$ is given by the negative second derivative of the classical potential $V(c)$ evaluated along the self-consistent autocorrelation function $c(\tau)$. The ground state energy $E_0$ of Eq. (7.17) determines the asymptotic growth rate of $k^{(1)}(t,t)$ as $t \to \infty$ and, hence, the maximum Lyapunov exponent via $\lambda_{\max} = -1 + \sqrt{1 - E_0}$ (see Eq. (B.24)). Therefore, the dynamics is predicted to become chaotic if $E_0 < 0$. This is shown together with the solution for the ground state energy and wave function in Fig. 7.2. The latter are obtained as solutions of a finite difference discretization of Eq. (7.17).

In the noiseless case, a decaying autocorrelation function corresponds to a positive maximum Lyapunov exponent (Sompolinsky et al., 1988). This follows from the observation that for $g > 1$ the derivative of the self-consistent autocorrelation function $\dot{c}(\tau)$ solves the Schrödinger equation with $E = 0$. But as $\dot{c}(\tau)$ is an eigenfunction with a single node, there must also exists a ground state without a node that has energy $E_0 < 0$. Hence, the dynamics is chaotic and $\lambda_{\max}$ crosses zero at $g = 1$ (Fig. 7.3A).

In the presence of noise, the maximum Lyapunov exponent becomes positive at a critical coupling strength $g_c > 1$; depending on the noise intensity the transition is shifted to larger values (Fig. 7.3A). The mean-field prediction $\lambda_{\max} = -1 + \sqrt{1 - E_0}$ shows excellent agreement with the maximum Lyapunov exponent obtained in simulations using a standard algorithm (Eckmann & Ruelle, 1985). Since the ground state energy $E_0$ must be larger than the minimum

$W(0) = 1 - g^2\langle[\phi'(x)]^2\rangle$ of the quantum potential, an upper bound for $\lambda_{\max}$ is provided by $-1 + g\sqrt{\langle[\phi'(x)]^2\rangle}$ leading to a necessary condition for chaotic dynamics. However, close to the transition $\lambda_{\max}$ is clearly smaller than the upper bound, which is a good approximation only for small $g$ (Fig. 7.3A, inset): the actual transition occurs at substantially larger coupling strengths. In contrast, for memoryless discrete-time dynamics the necessary condition found here is also sufficient for the transition to chaos (Molgedey et al., 1992, eq. 13).

The local linear stability of the dynamical system Eq. (7.1) is analyzed via the variational equation

$$\frac{d}{dt}y_i(t) = -y_i(t) + \sum_{j=1}^{N} J_{ij}\phi'(x_j(t))\, y_j(t)\,, \tag{7.18}$$

$i = 1, \ldots, N$, describing the temporal evolution of an infinitesimal deviation $y_i(t)$ about a reference trajectory $x_i(t)$. Interestingly, $\rho = g\sqrt{\langle[\phi'(x)]^2\rangle}$ is also the radius of the disk formed by the eigenvalues of the Jacobian matrix in the variational equation Eq. (7.18) estimated by random matrix theory (Sommers et al., 1988; Rajan & Abbott, 2006). Therefore, the dynamics is expected to become locally unstable if this radius exceeds unity, as shown in the inset in Fig. 7.3B displaying $\rho$ and the eigenvalues at an arbitrary point in time. But even for the case with $\rho > 1$ the system is not chaotic. Hence, contrary to the noiseless case (Sommers et al., 1988; Sompolinsky et al., 1988), the transition to chaos is not predicted by random matrix theory.

To derive an exact condition for the transition we determine a ground state with vanishing energy $E_0 = 0$. As in the noiseless case, $\dot{c}(\tau)$ solves Eq. (7.17) for $E = 0$, except at $\tau = 0$ where it exhibits a jump caused by the noise. However, due to linearity $|\dot{c}(\tau)|$ would be a continuous and symmetric solution with zero nodes. Therefore, if its derivative is continuous as well, requiring $\ddot{c}(0+) = 0$, it constitutes the searched for ground state. This is in contrast to the noiseless case, where $\dot{c}(\tau)$ corresponds to the first excited state. Consequently, with Eq. (7.6) we find the condition for the transition

$$g_c^2\, f_\phi(c_0, c_0) - c_0 = 0\,, \tag{7.19}$$

in which $c_0$ is determined by the self-consistency condition Eq. (7.10) resulting in the transition curve $(g_c, \sigma_c)$ in parameter space (Fig. 7.3B). This reveals the relationship between the onset of chaos, the statistics of the random coupling matrix, and the noise intensity.

At the transition the classical self-consistent potential $V(c; c_0)$ has a horizontal tangent at $c_0$, while in the chaotic regime a minimum emerges (Fig. 7.1D). This implies that the curvature $\ddot{c}(0+)$ of the autocorrelation function at zero changes sign from positive to negative (Fig. 7.1F). Furthermore, according to Eq. (7.19) the system becomes chaotic precisely when the variance $c_0 = \langle x^2\rangle$ of a typical neuron equals the variance $g^2\langle\phi^2(x)\rangle$ of its recurrent input. Close to the transition a standard perturbative approach shows that $\lambda_{\max}$ is proportional to $g^2\langle\phi^2(x)\rangle - c_0$, indicating a self-stabilizing effect: since both terms grow with $g$, the growth of their difference

is attenuated, explaining why $\lambda_{\max}(g)$ bends down as the transition is approached (Fig. 7.3A).

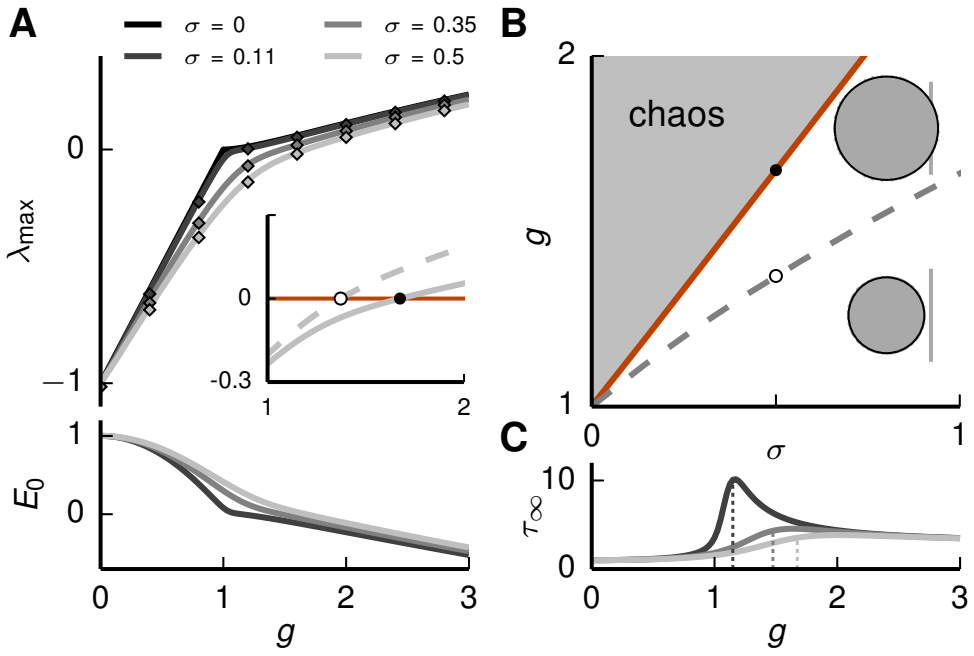



**Figure 7.3: Transition to chaos.** **A** Upper part of vertical axis: Maximum Lyapunov exponent $\lambda_{\max}$ as a function of the coupling strength $g$ for different noise levels. Mean-field prediction (solid curve) and simulation (diamonds). Comparison to the upper bound $-1+g\sqrt{\langle[\phi'(x)]^2\rangle}$ (dashed) for $\sigma = 0.5$ in inset. Zero crossings marked with dots. Lower part of vertical axis: Ground state energy $E_0$ as function of $g$. **B** Phase diagram with transition curve (solid red curve) obtained from (7.19) and necessary condition $g^2\langle\phi'^2(x)\rangle > 1$ (gray dashed curve). Dots correspond to zero crossings in inset in (a). Disk of eigenvalues of the Jacobian matrix in Eq. (7.18) for $\sigma = 0.8$ and $g = 1.25$ (lower) and $g = 2.0$ (upper) centered at $-1$ in the complex plane (gray). Radius $\rho = g\sqrt{\langle[\phi'(x)]^2\rangle}$ from random matrix theory (black). Vertical line at zero. **C** Asymptotic decay time $\tau_\infty$ of autocorrelation function. Vertical dashed lines mark the transition to chaos. Color code as in (a). Network size of simulations $N = 5000$.

The condition Eq. (7.19) predicts the transition at significantly larger coupling strengths compared to the necessary condition $g\sqrt{\langle[\phi'(x)]^2\rangle} > 1$ (Fig. 7.3B), which is explained as follows. For continuous-time dynamics the effect of noise is twofold: First, because $\phi'(x)$ is maximal at the origin, it reduces the averaged squared slope in $g^2\langle[\phi'(x)]^2\rangle$, thereby stabilizing the dynamics. This is an essentially static effect as it can be fully attributed to the increase of the instantaneous variance $c_0$ by the noise, which could similarly result from static heterogeneous inputs. Second, noise sharpens the autocorrelation function (Fig. 7.1E,F) and hence the quantum potential (Fig. 7.2). This shifts the ground-state energy to larger values, further decreasing the maximum Lyapunov exponent. Because this effect depends on the temporal correlations, noise suppresses chaos by a dynamic mechanism yielding stable dynamics even in the presence of local linear instability.

To understand this dynamic mechanism we return to the variational equation Eq. (7.18): its fundamental solution can be regarded as a product of short-time propagator matrices, where each factor has the same stability properties with unstable directions given by the local Jacobian matrix at the respective time. Even though the fraction of eigenvalues with positive real part stays approximately constant, the corresponding unstable directions vary in time. The sharpening of the autocorrelation function suggests that noise causes a faster variation such that perturbations asymptotically decay.

In low-dimensional systems the suppression of chaos by external fluctuations is understood: Noise forces the system to visit regions of the phase space with locally contracting dynamics more frequently (Zhou & Kurths, 2002) so that contraction dominates expansion, in total yielding stable asymptotic behavior. This mechanism is similar to the static stabilization effect described above, where noise causes the system to sample regions of the phase space with smaller eigenvalues of the Jacobian. The self-averaging high-dimensional system, however, has a constant spectral radius over time and hence the dynamics is either locally contracting or locally expanding for all times. While the previous effects are explained by local stability, the dynamic suppression of chaos found here is a genuinely time-dependent mechanism, explained by the time evolution of the Jacobian.

Finally, we consider the effect of noise on the asymptotic decay time $\tau_\infty = 1/\sqrt{1 - g^2\langle\phi'(x)\rangle^2}$ of the autocorrelation function (Fig. 7.3C). For weak noise, the decay time peaks at the transition, reflecting the diverging time scale in the noiseless case. For larger noise intensities, the peak is strongly reduced and the maximum decay time is attained above the transition.

# Part IV

# Simulation of rate-based neuron models

# Chapter 8

# Integration of rate-models in a spiking neural network simulator

---

The following chapter is based on the following publication:

Hahne, J., Dahmen, D., Schuecker, J., Frommer, A., Bolten, M., Helias, M., & Dies- mann, M. (2016). Integration of continuous-time dynamics in a spiking neural network simulator. arXiv. 1610.09990 [q–bio.NC].

**Author contributions:**

Under the supervision of M Helias and M Diesmann, the author jointly worked with J Hahne and D Dahmen on all parts of the above publication. Together with D Dahmen, the author focused on the neuroscientific applications and the implementation of neuron models. J Hahne focused on the NEST infrastructure, numerical schemes and performance benchmarks. All authors contributed to the writing of the manuscript.

In the mean-field theories considered in the previous chapters various rate-based models have occurred, namely a deterministic nonlinear rate model (Chap. 3, Chap. 4), a noisy linear rate model (Chap. 5) and a noisy non-linear rate model (Chap. 6). The cascade of simplifications from the original spiking network to the rate-based model, involves a combination of approximations which are routinely benchmarked in comparative simulations of the two models. In this chapter we develop a unified code base that features both models, which simplifies these validations rendering duplication of code obsolete.

## 8.1 Introduction

Simulation of rate-based models goes back to the works by Grossberg (1973), McClelland & Rumelhart (1981), Feldman & Ballard (1982), and the PDP group (Rumelhart et al., 1986). Various specialized tools have developed since then (O'Reilly, 2014), such as *PDP++* (McClelland & Rumelhart, 1989; O'Reilly et al., 2000), the *Neural Simulation Language* (Weitzenfeld et al., 2002), *emergent* (O'Reilly et al., 2012), the *MIIND* simulator (De Kamps et al., 2008), the simulation platform *DANA* (Rougier & Fix, 2012), *TheVirtualBrain* (Sanz Leon et al., 2013), *Topographica* (Bednar, 2009) and the *Neural Field Simulator* (Nichols & Hutt, 2015). Similarly, efficient simulators for spiking neural networks have evolved with different foci ranging from detailed neuron morphology (NEURON: Carnevale & Hines 2006, GENESIS: Bower & Beeman 2007) to an abstraction of neurons to a single point in space (NEST: Bos et al. 2015, BRIAN: Goodman & Brette 2013, see Brette et al. 2007 for a review). Such open-source software supports maintainability, reproducibility, and exchangeability of models and code, as well as community driven development. However, these tools are restricted to either rate-based or spike-based models only.

Besides the mentioned cross-check of mean-field predictions, there is a more general motivation to develop a unified framework: Bottom-up (typically spike-based models) and top-down (typically rate-based) strategies are still mostly disjoint and a major challenge in neuroscience is to form a bridge between the spike- and rate-based models (Abbott et al., 2016), and, more generally, between the fields of computational neuroscience and cognitive science. From a practical point of view, a common simulation framework would allow the exchange and the combination of concepts and code between the two descriptions and trigger interaction between the corresponding communities. This is in particular important since rate-based and spike-based models can describe the same spatial scales. For example the spiking multi-area model (Sec. 4.3) describes the same macroscopic system as rate-based descriptions (Chaudhuri et al., 2015).

Rate neurons typically represent populations of spiking neurons. Thus, a hybrid model, employing both types of neuron models in a multi-scale modeling approach, would contain a relatively large number of spiking neurons compared to the number of rate units. Furthermore a downscaling of a spiking network cannot be performed without changing the dynamics (van Albada et al., 2015) and thus it is crucial that a common simulation framework is able

to handle real-sized spiking networks. In addition, the employed mean-field theories exploit the large number of neurons in biological networks. In fact, they are strictly valid only in the thermodynamic limit $N \to \infty$ (Helias et al., 2014). Therefore, in the above mentioned validation studies (Chap. 3, Chap. 4), the spiking networks are typically large. Thus, a common simulation framework should be optimized for spiking neurons rather than rate-based models.

Current spiking network simulators solve the neuronal dynamics in a distributed and parallel manner. They exploit the point-event like nature of the spike interaction between neurons, for example in event-based simulation schemes. The latter, however, cannot be employed in the context of rate-based models which require continuous interactions between units. Spiking point-neuron models furthermore interact in a delayed fashion. The delays mimic the synaptic transmission and the propagation times along axons and dendrites. For the duration of the minimal delay $d_{\min}$ in a network, the dynamics of all neurons is decoupled. Hence, during $d_{\min}$, the neurons can be updated independently without requiring information from other neurons. Distributed processes therefore need to communicate spikes only after this period (Morrison et al., 2005). Due to considerable latencies associated with each communication, this scheme significantly improves performance and scalability of current simulators. In contrast, rate based-models (see Bressloff, 2012, and references therein) consider instantaneous interactions between neurons. A priori, this requires communication of continuous state variables between neurons at each time step.

The present study provides the concepts and a reference implementation for the embedding of continuous-time dynamics in a spiking network simulator. In order to exploit existing functionality we choose as a platform the open source simulation code NEST (Gewaltig & Diesmann, 2007; Bos et al., 2015) which is scalable software that can be used on machines ranging from laptops to supercomputers. The software is utilized by a considerable user community and equipped with a Python interface, support for the construction of complex networks, and mechanisms to shield the neuroscientist from the difficulties of handling a model description, potentially including stochastic components, in a distributed setting (Morrison et al., 2005; Plesser et al., 2015). Within this framework we introduce an iterative numerical solution scheme that reduces communication between compute nodes. The scheme builds on the waveform-relaxation technique (Lelarasmee, 1982) already employed for gap-junction interactions (Hahne et al., 2015).

This chapter begins with a brief review of numerical solution schemes for ordinary and stochastic (delay) differential equations in Sec. 8.2 and their application to neural networks in Sec. 8.2.2. Subsequently, we develop the concepts for embedding rate-based network models into a simulation code for spiking networks, adapt the waveform-relaxation scheme, and detail an extendable implementation framework for neuron models in terms of `C++` templates (Sec. 8.2.3). In Sec. 8.3, different numerical schemes are evaluated as well as the scalability of our reference implementation. We illustrate the applicability of the framework to a broad class of network models on the examples of a linear network model (Grytskyy et al., 2013), a nonlinear network model (Sompolinsky et al., 1988; Goedeke et al., 2016), a neural field

model (Roxin et al., 2005), and a mean-field description (Wong & Wang, 2006) of the stationary activity in a model of the cortical microcircuit (Potjans & Diesmann, 2014; Schuecker et al., 2016a). Straight-forward generalizations are briefly mentioned at the end of the Results section. The technology described in the present article will be made available with one of the next major releases of the simulation software NEST as open source. The conceptual and algorithmic work is a module in our long-term collaborative project to provide the technology for neural systems simulations (Gewaltig & Diesmann, 2007).

## 8.2 Methods

Rate-based single neuron and population models are described in terms of differential equations that often include delays and stochastic elements. Before we turn to the implementation of such models in computer code (Sec. 8.2.3) we review how such systems are mathematically solved and in particular how the stochastic elements are commonly interpreted with the aim to avoid an ad hoc design. A stochastic differential equation (SDE) is defined by the corresponding stochastic integral equation. Let $W(t)$ denote a Wiener process, also called Standard Brownian motion. For the initial condition $X(t_0) = X_0$ an Itô-SDE in its most general form satisfies

$$X(t) = X_0 + \int_{t_0}^{t} a(s, X(s))\, ds + \int_{t_0}^{t} b(s, X(s))\, dW(s)\,, \tag{8.1}$$

where the second integral is an Itô integral

$$\int_{t_0}^{t} Y(s)\, dW(s) := \lim_{n\to\infty} \sum_{i=1}^{n} Y_{i-1} \cdot (W_i - W_{i-1})$$

with $Y_i = Y(t_0 + i \cdot \frac{t-t_0}{n})$ and $W_i = W(t_0 + i \cdot \frac{t-t_0}{n})$. Alternatively, the second integral can be chosen as a Stratonovich integral, indicated by the symbol $\circ$,

$$\int_{t_0}^{t} Y(s) \circ dW(s) := \lim_{n\to\infty} \sum_{i=1}^{n} \frac{Y_{i-1} + Y_i}{2} (W_i - W_{i-1})$$

which approximates $Y(s)$ with the mid-point rule. In this case, the corresponding SDE is called a Stratonovich-SDE. We refer to Kloeden & Platen (1992) and Gardiner (2004) for a derivation and a deeper discussion on the differences between the two types of stochastic integrals. In the case of additive noise ($b(t, X(t)) = b(t)$) the Itô and Stratonovich integrals coincide. If furthermore the noise is constant ($b(t, X(t)) = \sigma = \text{const.}$) the integrals can be solved analytically

$$\int_{t_0}^{t} \sigma\, dW(s) = \int_{t_0}^{t} \sigma \circ dW(s) = \lim_{n\to\infty} \sigma \cdot \sum_{i=1}^{n} (W_i - W_{i-1}) = \sigma \cdot (W(t) - W(t_0))$$

with $W(t) - W(t_0) \sim \mathcal{N}(0, t - t_0)$. In the following, we focus on Itô-SDEs only.

The differential notation corresponding to Eq. (8.1) reads

$$dX(t) = a(t, X(t))\, dt + b(t, X(t))\, dW(t) \tag{8.2}$$

and denotes an informal way of expressing the integral equation. Another widely used differential notation, called the Langevin form of the SDE, is mostly employed in physics. It reads

$$\frac{dX(t)}{dt} = a(t, X(t)) + b(t, X(t))\, \xi(t)\,, \tag{8.3}$$

where $\xi(t)$ is a Gaussian white noise with $\langle \xi(t) \rangle = 0$ and $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$. Using the Fokker-Planck equation one obtains

$$\int_0^t \xi(t')dt' = W(t)\,,$$

which is a paradox, as one can also show that $W(t)$ is not differentiable (Gardiner, 2004, Chapter 4). Mathematically speaking this means that Eq. (8.3) is not strictly well-defined. The corresponding stochastic integral equation

$$X(t) = X_0 + \int_{t_0}^t a(s, X(s))\, ds + \int_{t_0}^t b(s, X(s))\, \xi(s)\, ds\,,$$

however, can be interpreted consistently with Eq. (8.1) as $dW(t) \equiv \xi(t)dt$.

### 8.2.1 Approximate numerical solution of SDEs

Similar to ordinary differential equations most stochastic differential equations cannot be solved analytically. Neuroscience therefore relies on approximate numerical schemes to obtain the solution of a given SDE. This section presents some basic numerical methods. Let $\Delta t$ denote the fixed step size, $t_k = t_0 + k\Delta t$ the grid points of the discretization for $k = 0, \ldots, n$, and $X_k$ the approximation for $X(t_k)$ obtained by the numerical method, at which $X_0$ is the given initial value. We consider systems of $N$ stochastic differential equations

$$dX(t) = a(t, X(t))\, dt + b(t, X(t))\, dW(t) \tag{8.4}$$

with initial condition $X(t_0) = X_0$. Here, $X(t) = (X^1(t), \ldots, X^N(t))$ and $W(t) = (W^1(t), \ldots, W^N(t))$ denote $N$-dimensional vectors and $a : \mathbb{R}^N \to \mathbb{R}^N$ and $b : \mathbb{R}^N \to \mathbb{R}^N$ are $N$-dimensional functions. $W(t)$ is an $N$-dimensional Wiener process, i.e., the components $W^i(t)$ are independent and identically distributed.

**Euler-Maruyama** The Euler-Maruyama method is a generalization of the forward Euler method for ordinary differential equations (ODE). Accordingly, it approximates the integrands

in Eq. (8.1) with their left-sided values. The update formula reads

$$X_{k+1} = X_k + a(t_k, X_k) \cdot \Delta t + b(t_k, X_k) \cdot \Delta W_k \tag{8.5}$$

with $\Delta W_k = W(t_{k+1}) - W(t_k) \sim \mathcal{N}(0, \Delta t)$ for $k = 0, \ldots, n-1$.

**Semi-implicit Euler** The (semi-)implicit Euler method is a generalization of the backwards Euler method for ODEs. The update formula reads

$$X_{k+1} = X_k + a(t_{k+1}, X_{k+1}) \cdot \Delta t + b(t_k, X_k) \cdot \Delta W_k. \tag{8.6}$$

The resulting scheme requires the solution of a system of nonlinear algebraic equations. Standard techniques for the solution of the system are Newton iteration and fixed-point iteration (Kelley, 1995). The method is sometimes called semi-implicit, because the function $b$ is still evaluated at $(t_k, X_k)$ instead of $(t_{k+1}, X_{k+1})$. However, a fully implicit Euler scheme for SDEs is not practicable (see Kloeden & Platen (1992), Chapter 9.8) and thus the term implicit Euler usually refers to the semi-implicit method.

**Exponential Euler** The exponential Euler method relies on the assumption that $a(t, X(t))$ consists of a linear part and a nonlinear remainder, i.e.,

$$a(t, X(t)) = A \cdot X(t) + f(t, X(t))$$

with $A \in \mathbb{R}^{N \times N}$. The idea is to solve the linear part exactly and to approximate the integral of the nonlinear remainder and the Itô integral with an Euler-like approach. Variation of constants for Eq. (8.4) yields

$$X(t) = e^{A(t-t_0)} X_0 + \int_{t_0}^{t} e^{A(t-s)} f(s, X(s))\, ds + \int_{t_0}^{t} e^{A(t-s)} b(s, X(s))\, dW(s)\,.$$

There are several versions of stochastic exponential Euler methods that differ in the approximation of the integral. Unfortunately a standardized nomenclature to distinguish the methods is so far missing. The simplest approach, sometimes named stochastic Lawson-Euler scheme (e.g. in Komori & Burrage, 2014), approximates the integrands with their left-sided values:

$$X_{k+1} = e^{A\Delta t} X_k + e^{A\Delta t} f(t_k, X_k) \cdot \Delta t + e^{A\Delta t} b(t_k, X_k) \cdot \Delta W_k\,.$$

More advanced schemes approximate the nonlinear part by keeping $f(s, X(s))$ constant for $[t_0, t)$ and solving the remaining integral analytically

$$\int_{t_0}^{t} e^{A(t-s)} f(s, X(s))\, ds \approx \int_{t_0}^{t} e^{A(t-s)} f(t_0, X(t_0))\, ds = A^{-1}(e^{A(t-t_0)} - I) \cdot f(t_0, X(t_0))\,.$$

Here $I$ denotes the $N \times N$ identity matrix. The same technique can be used for the Itô integral

$$\int_{t_0}^{t} e^{A(t-s)} b(s, X(s))\, dW(s) \approx \int_{t_0}^{t} e^{A(t-s)} b(t_0, X(t_0))\, dW(s)\,. \tag{8.7}$$

For a single SDE, Shoji (2011) proposed a method where the remaining integral $\int_{t_0}^{t} e^{a(t-s)}$ $dW(s)$ with $a \in \mathbb{R}$ is approximated by $\int_{t_0}^{t} \alpha\, dW(s)$, such that $\alpha \in \mathbb{R}$ is chosen to minimize the mean-square error. This results in a similar approximation as for the nonlinear part. Komori & Burrage (2014) adapted this approach for systems of SDEs. The scheme reads

$$X_{k+1} = e^{A\Delta t} X_k + A^{-1}(e^{A\Delta t} - I) \cdot f(t_k, X(t_k)) + \frac{1}{\Delta t} \cdot A^{-1}(e^{A\Delta t} - I) \cdot b(t_k, X_k) \cdot \Delta W_k\,.$$

Alternatively, calculating the variance of $X(t)$ within the approximation Eq. (8.7), amounts to (Adamu, 2011)

$$\mathrm{Var}\left(X(t)\right) = b(t_0, X(t_0))^2 \cdot \mathrm{Var}\left(\int_{t_0}^{t} e^{A(t-s)}\, dW(s)\right) = b(t_0, X(t_0))^2 \cdot A^{-1}\left(\frac{e^{2A(t-t_0)} - I}{2}\right).$$

The corresponding scheme reads

$$X_{k+1} = e^{A\Delta t} X_k + A^{-1}(e^{A\Delta t} - I) \cdot f(t_k, X(t_k)) + \sqrt{A^{-1}\left(\frac{e^{2A\Delta t} - I}{2}\right)} \cdot b(t_k, X_k) \cdot \eta_k \tag{8.8}$$

with $\eta_k \sim \mathcal{N}(0, 1)$ and yields the exact solution of the system if $a(t, X(t)) = A \cdot X(t)$ and $b(t, X(t)) = \text{const.}$, since $X(t)$ has Gaussian statistics in this case (Risken, 1996). Therefore in the following we exclusively employ Eq. (8.8) and just refer to it as the stochastic exponential Euler scheme. For more detailed reviews on the different stochastic exponential Euler methods we refer to Adamu (2011) and Komori & Burrage (2014).

### 8.2.2 Network of rate models

We now consider networks of $N$ rate-based model neurons where each neuron receives recurrent input from the network. The system fulfills the Itô-SDEs for $i = 1, \ldots, N$

$$\tau^i dX^i(t) = \left[-X^i(t) + \mu^i + \phi\left(\sum_{j=1}^{N} w^{ij} \psi\left(X^j(t - d^{ij})\right)\right)\right] dt + \sqrt{\tau^i}\sigma^i\, dW^i(t) \tag{8.9}$$

with possibly nonlinear input-functions $\phi(x)$ and $\psi(x)$, connection weights $w^{ij}$, mean input $\mu^i$, and optional delays $d^{ij} \geq 0$. The corresponding Fokker-Planck equation shows that the parameter $\sigma^i \geq 0$ controls the variance of $X^i(t)$ and the time constant $\tau^i > 0$ its temporal evolution. For readability, from here on we omit neuron indices for $\sigma, \tau, \mu$, and $d$. The considered class of rate models only contains additive noise. Therefore, as noted above, the system Eq. (8.9) can be written as Stratonovich-SDEs without the need for change in the

employed numerical methods. For an illustrative purpose we explicitly state the different explicit solution schemes for the network dynamics Eq. (8.9) with $d = 0$. The Euler-Maruyama update step reads

$$X^i_{k+1} = X^i_k + \left[ -X^i_k + \mu + \phi\left( \sum_{j=1}^{N} w^{ij} \psi\left(X^j_k\right) \right) \right] \frac{1}{\tau}\Delta t + \frac{1}{\sqrt{\tau}} \sigma \Delta W^i_k \,. \tag{8.10}$$

For nonlinear $\phi(x)$ or $\psi(x)$ the exponential Euler update step is

$$X^i_{k+1} = e^{-\Delta t/\tau} X^i_k + \left(1 - e^{-\Delta t/\tau}\right) \left[ \mu + \phi\left( \sum_{j=1}^{N} w^{ij} \psi\left(X^j_k\right) \right) \right] + \sqrt{\frac{1}{2}(1 - e^{-2\Delta t/\tau})}\, \sigma \eta^i_k \tag{8.11}$$

with $\eta^i_k \sim \mathcal{N}(0,1)$. As $A = -I$ is a diagonal matrix, the exponential Euler scheme does not rely on a matrix exponential, but decomposes into $N$ equations with scalar exponential functions. Note that with a linear choice, $\phi(x) = \psi(x) = x$, the system of SDEs can be written in matrix notation

$$\tau dX(t) = \left[A \cdot X(t) + \mu\right] dt + \sqrt{\tau}\sigma dW(t) \qquad i = 1, \ldots, N \tag{8.12}$$

with $A = -I + W$ and $W = (w^{ij})_{N \times N}$. Here the stochastic exponential Euler scheme Eq. (8.8) yields the exact solution of the system.

The numerical schemes presented in Sec. 8.2.1 are developed for SDEs ($d = 0$), but can analogously be used for stochastic delay differential equations (SDDEs) ($d > 0$), if the delay $d$ is a multiple of the step size $\Delta t$. For the calculation of the approximation $X^i_{k+1}$ in time step $k + 1$ the recurrent input is then evaluated from $X^j_{k - \frac{d}{\Delta t}}$, i.e. from $\frac{d}{\Delta t}$ steps earlier.

### 8.2.3 Implementation in spiking network simulation code

This section describes the embedding of rate-based models (Sec. 8.2.2) in a simulation code for spiking neuronal networks. Examples of how to create, connect and record activity from rate models in our reference implementation are illustrated in the appendix (Sec. C.2).

The software architecture for rate models is based on existing concepts: Morrison et al. (2005) describe distributed buffers for the storage of delayed interactions and the technique to consistently generate random numbers in a distributed setting, and Hahne et al. (2015) introduce so called `SecondaryEvent`s, that allow the communication of any kind of data between pairs of neurons. These components are designed to be compatible with the parallel and distributed operation of a simulation kernel for spiking neuronal networks, ensuring an efficient use of clusters and supercomputers (Helias et al., 2012). This allows researchers to easily scale up network sizes to more realistic number of neurons. The highly parallelizable structure of modern simulation codes for spiking neuronal networks, however, also poses restrictions on the utilizable numerical methods.

#### 8.2.3.1 Restrictions

Parallelization for spiking neuronal networks is achieved by distributing neurons over compute nodes. Since the dynamics of spiking neurons (in the absence of gap junctions) is decoupled for the duration of the minimal synaptic delay $d_{\min}$ of the connections in the network, the states of the neurons can be propagated independently for this time interval. Thus it is sufficient to specify solvers on the single-neuron level. The spike times, i.e. the mediators of interaction between neurons, are then communicated in steps of $d_{\min}$. In contrast, in the case of interactions via gap junctions (Hahne et al., 2015) or rates, the single-neuron dynamics depends on continuous state variables, membrane potential or rate, of other neurons. These continuous variables need to be communicated and the mechanism for this is the `SecondaryEvent` introduced in the gap-junction framework by Hahne et al. (2015).

Furthermore, the global connectivity of the network is unknown to the single neuron. The neuron object sends and receives events handled by the network manager on the compute node harboring the neuron. However, the network manager only knows the incoming connections of the neurons on the compute node.

This structure makes it impossible to employ the implicit Euler scheme Eq. (8.6) with Newton iteration, which would require the simultaneous solution of a system of nonlinear algebraic equations with information distributed over all compute nodes. It is however possible to use implicit schemes with fixed-point iteration. To this end, the corresponding scheme needs to be formulated as a fixed-point iteration on the single-neuron level and the updated influences of other neurons have to be communicated in every iteration. The gap junction framework by Hahne et al. (2015) already specifies an iterative method to advance the state of the network by one time step with global accuracy control. Therefore, we investigate in Sec. 8.3.1 if for rate-based network models the payoff of an implicit scheme is large enough to justify the additional effort of an iterative solver.

The restricted knowledge of connectivity also limits the usage of the exponential Euler method. In the case of a linear rate model, we are unable to add the influence from all other rate neurons to the matrix $A$ in Eq. (8.12), because most of these connections are unknown at the single-neuron level. Therefore, we use the exponential Euler method with $A = -I$ resulting in the update formula Eq. (8.11). This also has the benefit of avoiding the need to numerically evaluate a general matrix exponential as $A$ is a diagonal matrix (see Sec. 8.2.2 for details).

#### 8.2.3.2 Implementation

This section describes the additional data structure required for the implementation of rate-based models. As a result of the previous section and our later shown analysis of the numerical schemes (see particular Sec. 8.3.1) we restrict the discussion to the exponential Euler method where we assume $A = -I$ and identify $\Delta t = h$ with $h$ denoting the global computation step size (Morrison et al., 2005). We have to distinguish the cases of connections with delay ($d > 0$) and connections without delay ($d = 0$). The former case is similar to spiking interaction:

assuming a connection from neuron $i$ to neuron $j$, the rate of neuron $i$ needs to be available at neuron $j$ after $\frac{d}{h}$ additional time steps. This can be ensured if the delay of the connection is considered in the calculation of the minimal delay $d_{\min}$ that determines the communication interval. After communication the rate values are stored in a ring buffer of neuron $j$ until they are due (Morrison & Diesmann, 2008). In the case of an instantaneous connection, the rate of neuron $i$ at time $t_0$ needs to be known at time $t_0$ at the process which updates neuron $j$ from $t_0$ to $t_0 + h$. Therefore, communication in every step is required for instantaneous rate connections, i.e. setting $d_{\min} = h$.

Due to the conceptual differences between instantaneous and delayed interactions (for the conceptual difference in the case of spiking interaction see Morrison & Diesmann, 2008) we define two different connection types and associated events: The connection type for connections with delay is called `delay_rate_connection` and is associated with the new `SecondaryEvent` type `DelayRateNeuronEvent`. Connections without delay are implemented with the connection type `rate_connection` with the corresponding secondary event `RateNeuronEvent`.

The template class `rate_neuron_ipn` provides a base implementation for rate models of category Eq. (8.9). In this implementation neurons can handle both `DelayRateNeuronEvent` and `RateNeuronEvent` allowing for simultaneous use of instantaneous and delayed connections. To represent the nonlinearites $\phi(x)$ and $\psi(x)$ the class contains an object `gain_` the type of which is determined by the template parameter `TGainfunction`. The ending `ipn` indicates input noise, as the noise directly enters the r.h.s. of (8.9). A constant boolean class member `linear_summation_` of `rate_neuron_ipn` determines if the nonlinearity expressed by the `operator()` of the object `gain_` should be interpreted as $\phi(x)$ (true) or $\psi(x)$ (false). The respective other function is assumed to be the identity function and the default setting for `linear_summation_` is true. While to our knowledge this implementation covers the majority of neuron models, the evaluation of the boolean parameter `linear_summation_` in every update step of each neuron could be improved in terms of efficiency if the type of nonlinearity would be decided upon at compile time. In the present architecture this would, however, result in twice as many template instances for a given set of gain functions. With the future capabilities of code generation (Plotnikov et al., 2016) in mind it might be beneficial to elevate the constant boolean member object to a constant template parameter to allow compilers efficient preprocessing and at the same time profit from the code reliability achievable by modern C++ syntax. The present base implementation reduces the effort of creating a specific rate model of category Eq. (8.9) to the specification of an instance of the template class `TGainfunction`. Afterwards the actual neuron model can be defined with a simple typedef like e.g.

```
typedef rate_neuron_ipn< nest::gainfunction_lin_rate > lin_rate_ipn;
```

Tab. 8.1 gives an overview of template-derived rate models of the NEST reference implementation. These models serve as a reference for the implementation of customized neuron models. Activity of rate neurons can be recorded using the `multimeter` and the recordable `rate`.

**Table 8.1: Template-derived rate-based neuron models.** Gain functions of the rate-based neuron models available in the NEST reference implementation. The name of a particular neuron model is formed by `<gain model>_ipn`.

| gain model | $\phi(x)$ or $\psi(x)$ |
|---|---|
| `lin_rate` | $x$ |
| `tanh_rate` | $\tanh(g \cdot x)$ with $g \in \mathbb{R}$ |
| `thresholdlin_rate` | $g \cdot (x - \theta) \cdot H(x - \theta)$ with $g \in \mathbb{R}$ |

In addition to these template-derived models of category Eq. (8.9) the reference implementation also contains a neuron model called `siegert_neuron`. This model, described by Eq. (3.6) is used for mean-field analysis of complex networks (Chap. 3, Chap. 4) and constitutes a special case with respect to the recurrent input from the network. Firstly, it requires a numerically stable implementation of the Siegert formula (see Sec. C.1). Secondly, Eq. (3.4) and Eq. (3.7) demonstrate that for this model the input rates are weighted by two factors. Thus for connections between neurons of this model two different weights need to be specified and the neuron model must be able to handle this anomaly. Therefore the `siegert_neuron` is not derived from our base class `rate_neuron_ipn`, but constitutes an independent class. It comes with the connection type `diffusion_connection` that provides the weight parameters. Sec. C.2 motivates the parameter names and shows the usage of the model in the reference implementation.

#### 8.2.3.3 Waveform-relaxation techniques

The instantaneous connections between rate-based models requires communication in every time step, which impairs the performance and scalability. On supercomputers communication is particularly expensive, because it is associated with a considerable latency. Therefore, we also study an alternative iterative approach based on waveform-relaxation techniques that allows us to use communication on a coarser time grid. As outlined above, in a simulator for spiking neuronal networks the communication intervals are defined by the minimal delay $d_{\min}$ in the network. For simulations with instantaneous connections only, we attempt to reduce the communication load by setting the minimal delay to an arbitrary user specified value given by the parameter `wfr_comm_interval` (see Tab. 8.2). In case of additional delayed connections, the actual communication interval for waveform relaxation then follows as $\min(d_{\min}, \texttt{wfr_comm_interval})$. For details on waveform-relaxation methods and their application in the neuroscience context we refer to Hahne et al. (2015). Originally these methods were developed (Lelarasmee, 1982) and investigated (see e.g. Miekkala & Nevanlinna, 1987) for ODEs. More recently waveform-relaxation methods have also been analyzed for SDEs (Schurz & Schneider, 2005) and successfully applied to large systems of SDEs (Fan, 2013).

Fig. 8.1B illustrates the concept of the iterative approach in contrast to the standard procedure in panel A. The iterative approach requires the repeated solution of all time steps in the

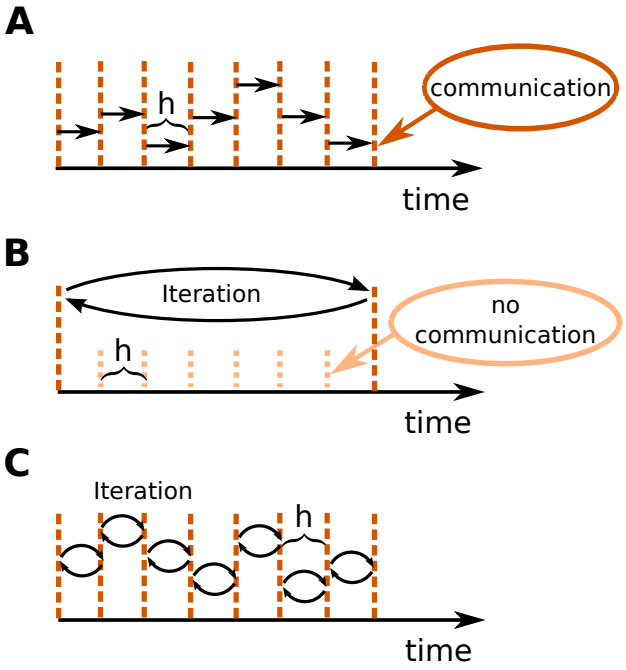



**Figure 8.1: Different communication strategies for distributed simulations.** Distance between neighboring dotted orange lines indicates computation time step of size $h$. Distance between neighboring dashed red lines symbolize one communication interval where rates (and other events like spike events) are communicated at the end of the interval. **A** Straight-forward solution for rate-based models: rates are communicated in every time step. **B** Iterative approach using waveform relaxation: rates are communicated only after $\frac{d_{\min}}{h}$ steps and the entire interval is solved repeatedly. **C** Iterative approach with communication in every step.

communication interval and converges to the solution obtained with the standard approach (Fig. 8.1A). The iteration terminates when a user chosen convergence tolerance `wfr_tol` (see Tab. 8.2) is met. If the method needs less than $d_{min}/h$ iterations, the approach reduces the overall number of communications required to obtain the solution. In conclusion, the avoidance of communication in every step comes for the price of additional computational load.

The coupling of neurons via gap junctions is instantaneous and continuous in time and thus constitutes a very similar problem to the rate dynamics. In order to combine gap junctions with spiking dynamics Hahne et al. (2015) already devised an iterative technique. The dynamics of a neuron model supporting gap junctions is solved with an adaptive step-size ODE-solver, routinely carrying out several steps of the employed numerical method within one global computation time step $h$. The communication of a cubic interpolation of the membrane potential provides the solver with additional information, resulting in a more accurate solution than the one obtained from the standard approach. For rate-based models this approach is however impossible. The combination of an iterative method with an adaptive step-size solver is not applicable to SDEs, where the noise in each time step constitutes a random number.

However, an iterative approach with fixed step size $\Delta t = h$ is applicable, as long as we ensure that the random noise applied to the neurons remains the same in every iteration. In Sec. 8.3.2 we investigate the performance of the iterative (Fig. 8.1B) and the standard approach (Fig. 8.1A) with a focus on large network simulations on supercomputers. In our reference implementation waveform relaxation can be enabled or disabled by a parameter `use_wfr`. Note that in the traditional communication scheme for spiking neuronal networks (Morrison et al., 2005) the first communication occurs earliest at the end of the first update step. Therefore, in the absence of waveform relaxation, the initial input to neurons from the network is omitted.

Fig. 8.1C shows an alternative iterative approach also feasible within our framework. While this scheme is not needed for the exponential Euler method investigated in the present work, it can be employed to perform a fixed-point iteration in order to obtain the solution of the implicit Euler method. However, Sec. 8.3.1 demonstrates that this is not an efficient option for the integration of rate-based models in distributed simulations.

Tab. 8.2 summarizes the parameters of our reference implementation of the waveform-relaxation technique. A subset (`wfr_interpolation_order`, `wfr_max_iterations`, `wfr_tol`) was previously introduced by Hahne et al. (2015), but we rename them here to arrive at more descriptive names. The remaining parameters (`use_wfr`, `wfr_comm_interval`) result from the generalization to rate-based models.

**Table 8.2: Parameters of the waveform relaxation algorithm.** The different parameters of the waveform relaxation algorithm together with their C++ data-type, default value, and a brief description.

| parameter name | type | default | description |
|---|---|---|---|
| `use_wfr` | `bool` | `true` | Boolean parameter to enable (`true`) or disable (`false`) the use of the waveform relaxation technique. If disabled and any rate-based neurons (or neurons supporting gap junctions) are present, communication in every step is automatically activated ($d_{\min} = h$). |
| `wfr_comm_interval` | `double` | 1.0 ms | Instantaneous rate connections (and gap junctions) contribute to the calculation of the minimal network delay with $\min(d_{\min}, \texttt{wfr_comm_interval})$. This way the length of the iteration interval of the waveform relaxation can be regulated. |
| `wfr_tol` | `double` | $10^{-4}$ | Convergence criterion for waveform relaxation. The iteration is stopped if the rates of all neurons change less than `wfr_tol` from one iteration to the next. |
| `wfr_max_iterations` | `int` | 15 | Maximum number of iterations performed in one application of the waveform relaxation. If the maximum number of iterations has been carried out without reaching the accuracy goal the algorithm advances system time and the reference implementation issues a warning. Additional speed-up in the simulation of rate-based neurons can only be achieved by `wfr_max_iterations` $< d_{min}/h$. |
| `wfr_interpolation_order` | `int` | 3 | This parameter is exclusively used for gap junctions (see Hahne et al., 2015, Sec. 2.1.2) and has no influence on the simulation of rate-based models. |

## 8.3 Results

In the following, we assess the stability of the different numerical solution schemes and benchmark their performance on large-scale machines. Furthermore, we illustrate the application of the simulation framework to different models relevant in the neuroscientific literature.

### 8.3.1 Stability and accuracy of integration methods

To investigate the accuracy and stability of the different numerical methods (see Sec. 8.2.1) we consider an exactly solvable network of linear rate neurons with $\mu = 0$ (see also Sec. 8.2.2)

$$\tau dX(t) = A \cdot X(t)\, dt + \sqrt{\tau}\sigma\, dW(t)\,. \tag{8.13}$$

The exact solution of the system of SDEs coincides with the exponential Euler scheme and involves a matrix exponential and a matrix square root (Eq. (8.8)). This exact solution cannot be obtained with a distributed representation of $A$ as it is typically employed in the distributed simulation scheme of a spiking network simulation code (see Sec. 8.2.3.1). However, using the methods for numerical matrix computations in MATLAB or Python (both provide an implementation of the same state-of-the-art algorithms, see Al-Mohy & Higham, 2009; Deadman et al., 2012), we obtain an approximation to the exact solution, close to the general limits of floating point numerics, and use this as a reference to compute the root mean square error of the different approximative methods. In order to employ the root mean square error in the context of stochastic differential equations we compute the reference solution for every tested step size and use the same random numbers for both the reference solution and the approximative schemes. Furthermore, we consider analytical stability criteria for some of the employed methods.

In the following we assume that $A$ is diagonalizable, i.e. $A = T^{-1}DT$ with $T = (t^{ij})_{N\times N} \in \mathbb{C}^{N\times N}$ and $D = \mathrm{diag}(\lambda_1, \ldots, \lambda_N)$, and transform the system of SDEs with $Z(t) = T\,X(t)$. It follows

$$\tau dZ(t) \quad = \quad D \cdot Z(t)\, dt + \sqrt{\tau}\sigma T\, dW(t)$$

and $Z(t_0) = TX_0$. The transformed system consists of $N$ equations of the form

$$\tau dZ^i(t) = \lambda_i \cdot Z^i(t)\, dt + \sum_{j=1}^{N} \sqrt{\tau}\sigma t^{ij}\, dW^j(t) \qquad i = 1, \ldots, n \tag{8.14}$$

that depend on the eigenvalues of $A$ and are independent of each other except for the contribution of the Wiener processes $W^j(t)$ . For eigenvalues $\lambda_i \in \mathbb{C}$ with negative real part $\mathrm{Re}(\lambda_i) < 0$, the solution of the $i$-th transformed equation satisfies

$$|Z^i(t) - \tilde{Z}^i(t)| = e^{\lambda_i (t-t_0)/\tau} |Z_0^i - \tilde{Z}_0^i| < |Z_0^i - \tilde{Z}_0^i|$$

for two different initial values $Z_0^i$ and $\tilde{Z}_0^i$. It is a desirable stability criterion that a numerical method applied to Eq. (8.13) conserves this property. This is closely related to the concept of A-stability for SDEs (see Kloeden & Platen (1992), Chapter 9.8) and A- respectively B-stability for ODEs (Hairer & Wanner, 1991). A straight-forward calculation shows that the implicit Euler method and the exponential Euler scheme retain this condition regardless of the step size $\Delta t$ and that the Euler-Maruyama method retains the condition if $|1 + \lambda_i \cdot \Delta t/\tau| < 1$ holds. For $\lambda_i \in \mathbb{R}$ we obtain the step size restriction $\Delta t < \frac{2\tau}{|\lambda_i|}$ and for complex eigenvalues the condition is conserved if $\lambda_i \cdot \Delta t/\tau$ is located inside a unit circle centered at $-1$ in the complex plane.

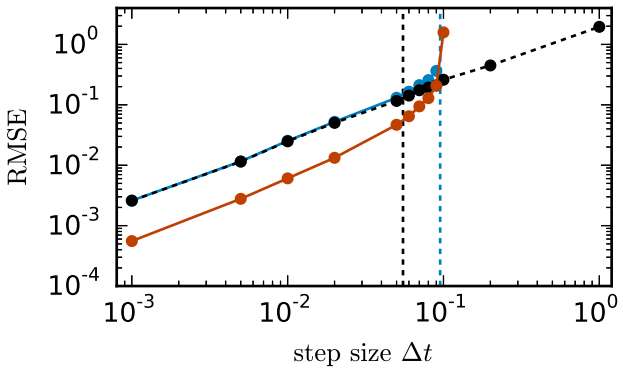



**Figure 8.2: Comparison of numerical methods for an all-to-all connected inhibitory network.** RMSE $= \sqrt{\frac{1}{N(t_n - t_0)} \sum_{i=1}^{N} \sum_{j=1}^{n} (X_j^i - \hat{X}_j^i)^2}$ of the solution $X$ obtained by the approximate solvers (black dashed curve: implicit Euler method solved with Newton iteration, blue curve: Euler-Maruyama method, red curve: exponential Euler method) with respect to the exact solution $\hat{X}$ as a function of step size in double logarithmic representation. The black vertical line marks the largest step size for which the implicit Euler method solved with fixed-point iteration converges against the one obtained with Newton iteration. The blue vertical line corresponds to the analytical stability restriction $\Delta t \leq \frac{2}{21}$ of the Euler-Maruyama method. Network parameters: size $N = 400$, all-to-all connectivity with $w^{ij} = \frac{-1}{\sqrt{N}}$, $\mu = 0$, $\sigma = 10$ and $\tau = 1\,\text{ms}$.

We investigate the accuracy of the numerical methods for an all-to-all connected network with inhibitory connections of weight $w^{ij} = \frac{-1}{\sqrt{N}}$. In this example $A$ takes the form $A = -I + \frac{-1}{\sqrt{N}} \cdot \mathbb{1}$, with $\mathbb{1}$ denoting a $N \times N$ all-ones matrix. From its eigenvalues $\lambda_1 = -1 - \sqrt{N}$ and $\lambda_2 = \ldots = \lambda_N = -1$ it follows that the Euler-Maruyama scheme satisfies the stability criterion for $\Delta t \leq \frac{2\tau}{\sqrt{N}+1}$. Fig. 8.2 shows the root mean square error of the different numerical schemes for an example network. With decreasing step size all investigated methods converge towards the exact solution with convergence order 1, which is consistent with the established theory for SDEs with additive noise (Kloeden & Platen, 1992). The Euler-Maruyama scheme only works within the calculated stability region. The exponential Euler with $A = -I$ and $f(X) = \frac{-1}{\sqrt{N}} \cdot \mathbb{1} \cdot X$, in the following called scalar exponential Euler, shows a similar stability, as the whole

input from the network is approximated with an (explicit) Euler-like approach. Within the stability region of Euler-Maruyama, however, the scalar exponential Euler yields more accurate results than the two other methods. The implicit Euler scheme solved with Newton iteration shows no stability issues, but it is not applicable in the distributed simulation framework for spiking neuronal networks (see Sec. 8.2.3.1). For completeness, we also test the implicit Euler scheme with a parallelizable Jacobi fixed-point iteration. The convergence properties of the fixed-point iteration demand the scheme to be contractive (see e.g. Kelley, 1995, Sec. 4.2). Therefore, in our test case the step size is restricted to roughly $\Delta t < 0.05\,\text{ms}$ and accordingly to a region where the scalar exponential Euler yields better results. In addition, an iterative scheme in each single time step is expected to be more time consuming than using the scalar exponential Euler scheme. Based on these results we employ the scalar exponential Euler to solve rate-based neuron dynamics Eq. (8.9), as it is the most accurate, stable and efficient scheme compatible with the constraints of the distributed simulation scheme for spiking neural networks. Nevertheless, the inevitable restrictions on the step size $\Delta t$ need to be taken into account in order to obtain an accurate solution. An appropriate step size can be estimated with the analytical stability criterion of the Euler-Maruyama method. For an all-to-all connected inhibitory network the restriction $\Delta t \leq \frac{2\tau}{\sqrt{N}+1}$ shows that, with increasing network size $N$ or decreasing time constant $\tau$ the step size $\Delta t$ needs to be reduced.

The fully connected network constitutes the worst case test for the class of rate-based models Eq. (8.9), as the absolute value of the negative eigenvalue quickly increases with the number of neurons $N$. A network which does not suffer from this problem is a perfectly balanced network of excitatory and inhibitory neurons (Rajan & Abbott, 2006). For these models the conditions on the step size $\Delta t$ of the Euler-Maruyama method are less restrictive and the same is expected for the scalar exponential Euler method. As an example we employ a sparse balanced excitatory-inhibitory-network. In a scaling of the connection weights as $\frac{1}{\sqrt{N}}$, the spectral radius of $A$ and therefore the subsequent stability analysis is independent of $N$. Fig. 8.3B demonstrates that for this test case the Euler-Maruyama method is stable for $\Delta t < 1.2\tau$. Given a commonly used simulation step size of $h = \Delta t = 0.1$, networks of this kind can be safely simulated if the time constant $\tau$ fulfills $\tau \geq 0.085$.

Random networks with incomplete balance exhibit both types of stability issues discussed above. In this case the matrix $A$ contains an eigenvalue $\lambda_1 = -1 - \rho\sqrt{N}$ which scales with the network size, however, with a proportionality constant $|\rho| < 1$ which is reduced compared to the fully connected inhibitory network and determined by the sparseness and the partial balance. Nevertheless, the network size needs to be taken into account for the choice of the step size. The latter also needs to ensure that the cloud of eigenvalues determined by the randomness in the connectivity meets the stability criterion.

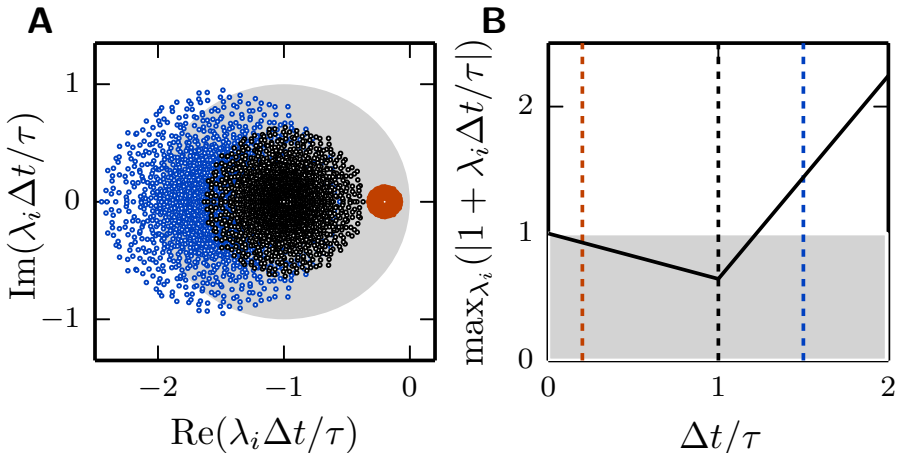



**Figure 8.3: Stability analysis for a balanced excitatory-inhibitory network.** The network contains in total $N$ neurons where the number of excitatory neurons is four times larger then the number of inhibitory neurons. Each neuron receives input from a fixed number of $0.8 \cdot p \cdot N$ excitatory and $0.2 \cdot p \cdot N$ inhibitory randomly chosen source neurons with connection probability $p$ and connection weights $\frac{1}{\sqrt{N}}$ and $\frac{-4}{\sqrt{N}}$, respectively. **A** Black circles show the eigenvalues $\lambda_i$ of the matrix $A$ for a network of $N = 2000$ neurons. Blue and red circles show the rescaled eigenvalues $\lambda_i \cdot \Delta t/\tau$ for $\Delta t/\tau = 1.5$ and $\Delta t/\tau = 0.2$. The filled gray circle indicates the region where the rescaled eigenvalues $\lambda_i \cdot \Delta t/\tau$ meet the stability criterion $|1 + \lambda_i \cdot \Delta t/\tau| < 1$ of the Euler-Maruyama method. **B** The curve shows the maximum of $|1 + \lambda_i \cdot \Delta t/\tau|$ over all eigenvalues $\lambda_i$ dependent on $\Delta t/\tau$. The gray area again indicates the region where the stability criterion of the Euler Maruyama method is met. Colored vertical lines correspond to the rescaled eigenvalues displayed in panel A.

### 8.3.2 Performance of iteration schemes

This section investigates the performance of the rate model implementation. We are interested in i) the scalability of the rate model framework and ii) the comparison between the straight-forward implementation with communication in every computation time step and the iterative approach using waveform relaxation (see Sec. 8.2.3.3 for details). We perform the simulations on the JUQUEEN BlueGene/Q supercomputer (Jülich Supercomputing Centre, 2015) at the Jülich Research Centre in Germany. It comprises 28,672 compute nodes, each with a 16-core IBM PowerPC A2 processor running at 1.6 GHz. For our benchmarks we use 8 OpenMP threads per JUQUEEN compute node and denote by VP $= 8 \cdot$ #nodes the total number of virtual processes employed.

As a test case we employ the excitatory-inhibitory-network of linear rate neurons ($\phi(x) = \psi(x) = x$) introduced in Sec. 8.3.1, but with a fixed number of inputs (2000) independent of the number of neurons to allow for an unbiased weak scaling.

A weak scaling (Fig. 8.4A) shows that the scalability of the straight-forward computation is impaired by the massive amount of communication. While for perfect scaling the simulation time should be constant over the number of virtual processes, the actual simulation time is increased by $15 - 25\%$ when the number of virtual processes is doubled for VP $< 256$

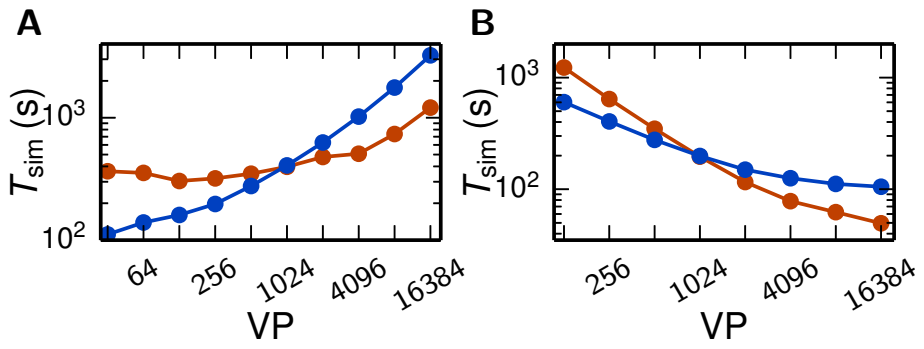



**Figure 8.4: Scaling behavior of an excitatory-inhibitory network.** Simulation time with waveform relaxation (red curves, `wfr_comm_interval`: 1.0 ms, `wfr_tol`: $10^{-4}$) and without waveform relaxation (blue curves) as a function of the number of virtual processes in double logarithmic representation. The simulations span $T = 100$ ms of biological time at a computation step size of $h = 0.1$ ms. Network as described in Fig. 8.3 but with a fixed number of 2000 inputs per neuron. Other parameters: $\mu = 0$, $\sigma = 1$ and $\tau = 10$ ms **A** Weak scaling with 100 neurons per virtual process VP. **B** Strong scaling with a total number of $N = 51,200$ neurons.

and even up to 83% from 8,192 to 16,384 virtual processes. For the iterative method, the scaling behavior is close to constant up to 1,024 virtual processes. When more processes are employed, the simulation time is increasing. However, the iterative method shows a better scaling behavior as the increase is weaker compared to the straight-forward computation due to the lower total number of communication steps. Due to the higher computational load of the iterative method (see Sec. 8.2.3.3) the simulation time is larger compared to the straight forward approach for a small number of VP, where communication is not that crucial. For VP $\geq$ 1024, the iterative approach is superior with a speed up factor close to three for 16,384 virtual processes (1209 s vs. 3231 s).

The strong scaling scenario with a fixed total number of $N = 51,200$ neurons in Fig. 8.4B constitutes a similar result. The iterative approach is beneficial for more than 1,024 virtual processes and the scaling behavior of the iterative method outperforms that of the straight-forward computation. Starting at 4,096 virtual processes the savings in computation time decrease, which is explained by the very low workload of each single compute node. Again, for a smaller number of virtual processes the amount of additional computations is too high to outperform the straight-forward computation.

Despite the overall good scaling behavior, the performance in terms of absolute compute time is inferior to a simulator specifically designed for rate-based models alone (not shown). In the latter case it increases performance to collect the states of all neurons in one vector. If further the connectivity is available in form of a matrix and the delays are zero or homogeneous, the network can be efficiently updated with a single matrix-vector multiplication. Thus the increased functionality and flexibility of having rate- and spiking model neurons unified in one simulator comes for the price of a loss of performance for the rate-based models. However,

as noted in the introduction, the number of neurons in rate-based network models is usually small and therefore performance is not as critical as for spiking network models.

### 8.3.3 Applications

First, we discuss a balanced random network of linear rate units, then include nonlinear neuron dynamics in a random network, and spatially structured connectivity in a functional neural-field model. In each case, simulation results are compared to analytical predictions. Furthermore, we simulate a mean-field model of a spiking model of a cortical microcircuit and discuss possible generalizations.

#### 8.3.3.1 Linear model

In the asynchronous irregular regime which resembles cortical activity, the dominant contribution to correlations in networks of nonlinear units is given by effective interactions between linear response modes (Grytskyy et al., 2013; Trousdale et al., 2012; Pernice et al., 2011; Dahmen et al., 2016) (see also Chap. 5). Networks of such noisy linear rate models have been investigated to explain features such as oscillations (Bos et al., 2016) or the smallness of average correlations (Tetzlaff et al., 2012; Helias et al., 2013). We here consider a prototypical network model of excitatory and inhibitory neurons following the linear dynamics given by Eq. (8.9) with $\phi(x) = \psi(x) = x$, $\mu = 0$, and noise amplitude $\sigma$,

$$\tau dX^i(t) = \left( -X^i + \sum_{j=1}^{N} w^{ij} X^j(t) \right) dt + \sqrt{\tau}\sigma dW^i(t)\,. \tag{8.15}$$

Due to the linearity of the model, the cross-covariance between neurons $i$ and $j$ can be calculated analytically and is given by (Ginzburg & Sompolinsky, 1994; Risken, 1996; Gardiner, 2004; Dahmen et al., 2016)

$$c(t) = \sum_{i,j} \frac{v^{i\mathrm{T}} \sigma^2 v^j}{\lambda_i + \lambda_j} u^i u^{j\mathrm{T}} \left( \theta(\Delta) \frac{1}{\tau} e^{-\lambda_i \frac{t}{\tau}} + \theta(-\Delta) \frac{1}{\tau} e^{\lambda_j \frac{t}{\tau}} \right), \tag{8.16}$$

where $\theta$ denotes the Heaviside function. The $\lambda_i$ indicate the eigenvalues of the matrix $1 - W$ corresponding to the $i$-th left and right eigenvectors $v^i$ and $u^i$ respectively. Non-zero delays yield more complex analytical expressions for cross-correlations. In the population-averaged case, theoretical predictions are still analytically tractable (eq. 18 in Grytskyy et al., 2013). Fig. 8.5 shows the cross-covariance functions for pairs of instantaneously coupled neurons in a large network, as well as population-averaged covariance functions in a network of excitatory and inhibitory neurons with delayed interactions. In both cases, simulations are in good agreement with the theoretical predictions.

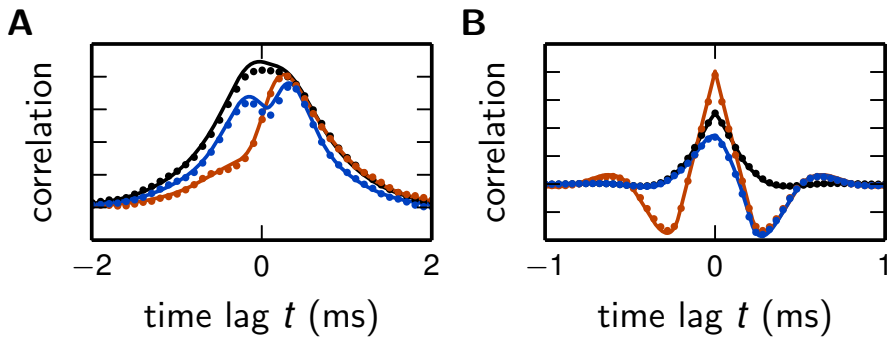



**Figure 8.5: Linear rate model of a random excitatory-inhibitory network. A** Cross-correlation functions of two pairs of excitatory neurons (black, red) and an excitatory-inhibitory neuron pair (blue) in a network without delay. The variability across correlation functions arises from heterogeneity in network connections (difference between black and red curves) and from different combinations of cell types (e.g. difference between black and blue curves). **B** Population-averaged autocorrelation function for excitatory (black) and inhibitory neurons (red), and cross-correlation function between excitatory and inhibitory neurons (blue) in a network with delay $d = 2\,\text{ms}$. Symbols denote simulation results, curves show theoretical predictions. Parameters: $N_E = 80$ excitatory and $N_I = 20$ inhibitory neurons, random connections with fixed out-degree, connection probability $p = 0.1$, excitatory weight $w_E = 1/\sqrt{N_E + N_I}$, inhibitory weight $w_I = -6\,w_E$, $\tau = 1\,\text{ms}$, $\mu = 0$, $\sigma = 1$.

#### 8.3.3.2 Nonlinear model

So far we considered a network with linear couplings between the units. Qualitatively new features appear in the presence of nonlinearities. One of the most prominent examples is the emergence of chaotic dynamics (Sompolinsky et al., 1988) in a network of non-linearly coupled rate units. The original model is deterministic and we have extended the model to stochastic dynamics in Chap. 7. While the results in Chap. 7 were obtained with FORTRAN we here use the new implementation of rate-based models in the unifying simulation framework developed in this chapter. The model definition Eq. (7.1) in the notation of this chapter follows from Eq. (8.9) with $\mu = 0$, $\phi(x) = \text{x}$, $\psi(x) = \tanh(x)$, i.e.

$$\tau dX^i(t) = \left( -X^i(t) + \sum_{j=1}^{N} w^{ij} \tanh\left(X^j(t)\right) \right) dt + \sqrt{\tau}\sigma\, dW^i(t)\,, \tag{8.17}$$

where $w^{ij} \approx \mathcal{N}(0, g^2/N)$ are Gaussian random couplings. We have seen in Chap. 6 that in the thermodynamic limit $N \to \infty$, the population averaged autocorrelation function $c(t)$ can be determined within dynamic mean-field theory. Comparing $c(t)$ obtained by simulation of a network Eq. (8.17) with the analytical result Eq. (7.8) and Eq. (7.10) demonstrates excellent agreement (Fig. 8.6).

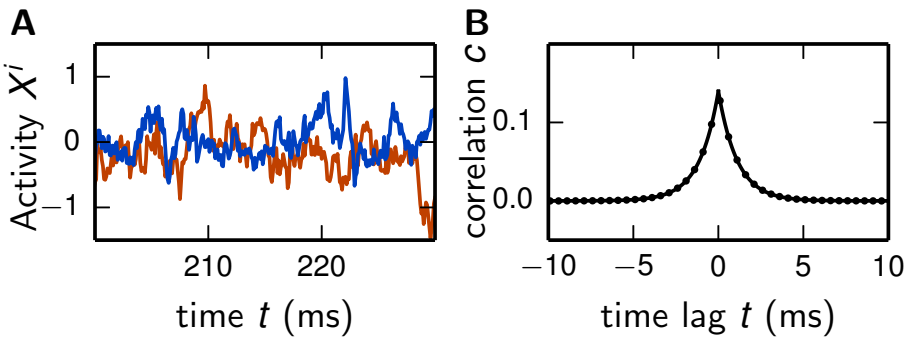



**Figure 8.6: Nonlinear network model.** Simulation of the network specified by Eq. (8.17) with $N = 1000$ neurons. **A** Noisy example trajectories of two neurons. **B** Autocorrelation function obtained by simulation averaged over all neurons (dots) and theory (solid curve). Other parameters: $\tau = 1\,\text{ms}$, $\sigma = 0.5$, $g = 0.5$. Step size as in Fig. 8.4.

#### 8.3.3.3 Functional model

Complex dynamics not only arises from nonlinear single-neuron dynamics, but also from structured network connectivity (Yger et al., 2011). One important nonrandom feature of brain connectivity is the spatial organization of connections (Malach et al., 1993; Voges et al., 2010). In spatially structured networks, delays play an essential role in shaping the collective dynamics (Roxin et al., 2005; Voges & Perrinet, 2012). Patterns of activity in such networks are routinely investigated using neural-field models. In contrast to the models discussed above, field models require a discretization of space for numerical simulation. Such discretization can be done in the real space, leading effectively to a network of neurons at discrete positions in space, or alternatively, for particular symmetries in the couplings, in k-space (Roxin et al., 2006). Here, we follow the more general approach of discretization in real space.

A prototypical model of a spatial network is given by (Roxin et al., 2005), where the authors consider the neural-field model

$$\tau dX(\varphi,t) = \left(-X(\varphi,t) + \phi\left[I_{\text{ext}} + \int_{-\pi}^{\pi} d\varphi'\, w(|\varphi-\varphi'|)X(\varphi',t-d)\right]\right)dt \tag{8.18}$$

with delayed (delay $d$) interactions, constant input $I_{\text{ext}}$, threshold-linear activation function $\phi = x \cdot \theta(x)$ and periodic Mexican-hat shaped connectivity

$$w(|\varphi-\varphi'|) = w_0 + w_1\cos(\varphi-\varphi'). \tag{8.19}$$

The spatial variable $\varphi$ can also be interpreted as the preferred orientation of a set of neurons, thus rendering Eq. (8.18) a model in feature space (Hansel & Sompolinsky, 1998). Discretizing space into $N$ segments yields the following set of coupled ODEs

$$\tau dX^i = \left(-X^i + \phi\left[I_{\text{ext}} + \sum_{j=1}^{N} w^{ij}X^j(t-d)\right]\right)dt \tag{8.20}$$

with connectivity $w^{ij} = \frac{2\pi}{N} w(|\varphi^i - \varphi^j|)$, $\varphi_i = -\pi + \frac{2\pi}{N} \cdot i$ for $i \in [1, N]$ and discretization factor $\frac{2\pi}{N}$ that scales the space constants $w_0$ and $w_1$ with the neuron density. The spatial connectivity together with a delay in the interaction introduce various spatial activity patterns depending on the shape of the Mexican-hat connectivity.

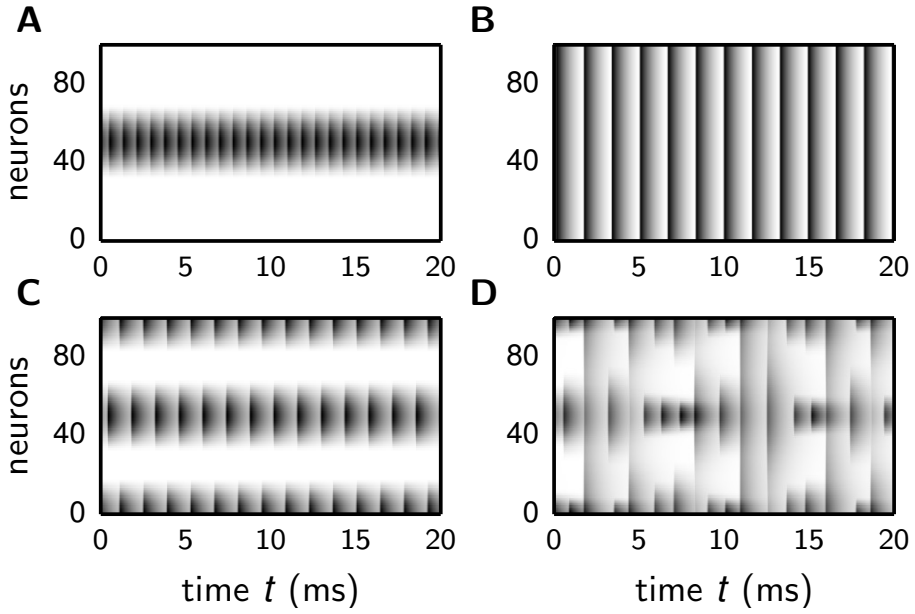



**Figure 8.7: Spatial patterns in functional neural-field model.** Vertical axis shows neuron indices organized according to ascending angle $\varphi \in [-\pi, \pi)$. Neuronal activity $X_i(t) = X(\varphi_i, t)$ encoded by gray scale with white denoting no activity. Initial transients not shown. Patterns reproduce the analytically derived phase diagram in the original study by Roxin et al. (2005). Parameters: $N = 100$, $d = 0.1\,\text{ms}$, $\tau = 1\,\text{ms}$, $I_{\text{ext}} = 1$, $w_0 = -80$, $w_1 = 15$ **A**, $w_1 = 5$ **B**, $w_1 = -46$ **C**, $w_1 = -86$ **D**. Initial condition: $X_i(0) = X(\varphi_i, 0) = \pi^2 - \varphi_i^2$.

To illustrate applicability of the simulation framework to neural-field models, we reproduce various patterns (Fig. 8.7) observed by Roxin et al. (2005). Although the discrete and continuous networks strictly coincide only in the thermodynamic limit $N \to \infty$, numerically obtained patterns shown in Fig. 8.7 well agree with the analytically derived phase diagram of the continuous model (Roxin et al., 2005) already for network sizes of only $N = 100$ neurons.

#### 8.3.3.4 Mean-field analysis of complex networks

In Chap. 3 we have shown how the stationary state of a spiking network can be analyzed by solving the dynamics of a network of deterministic rate neurons (Eq. (3.6) together with Eq. (3.5) or Eq. (3.7)). The ability to represent spiking as well as rate dynamics by the same simulation framework presented here, allows a straight-forward analysis of the spiking network by replacing the spiking neuron populations by single rate-based model neurons. An example

has already been shown in Sec. 3.3.

#### 8.3.3.5 Further neuron models

**Nonlinear neuron dynamics** A characteristic feature of rate neurons considered so far is the leaky neuron dynamics, i.e the linear term $-X^i(t)$ in Eq. (8.9). However the presented framework can be extended to nonlinear neuron dynamics as used for example by Stern et al. (2014). In a more general form Eq. (8.9) reads

$$\tau dX^i(t) = \left[ a(X^i(t)) + \phi\left( \sum_{j=1}^{N} w^{ij} \psi\left( X^j(t-d) \right) \right) \right] dt + \sqrt{\tau}\sigma dW^i(t) \tag{8.21}$$

where $a$ characterizes the intrinsic neuron dynamics. If $a$ does not contain a linear part, the Euler-Maruyama scheme can be used for the update, i.e,

$$X^i_{k+1} = X^i_k + \left[ a(X^i_k) + \mu + \phi\left( \sum_{j=1}^{N} w^{ij} \psi\left( X^j_{k-\frac{d}{\Delta t}} \right) \right) \right] \frac{1}{\tau}\Delta t + \frac{1}{\sqrt{\tau}}\sigma \Delta W^i_k. \tag{8.22}$$

If $a$ also contains a linear part, so that $a(X^i) = -X^i + f(X^i)$, one can use an exponential Euler update approximating the non-linear part as constant during the update. This leads to

$$\begin{aligned} X^i_{k+1} = e^{-\Delta t/\tau} X^i_k + \left(1 - e^{-\Delta t/\tau}\right) \left[ f(X^i_k) + \phi\left( \sum_{j=1}^{N} w^{ij} \psi\left( X^j_{k-\frac{d}{\Delta t}} \right) \right) \right] \\ + \sqrt{\frac{1}{2}(1 - e^{-2\Delta t/\tau})}\, \sigma \eta^i_k, \end{aligned} \tag{8.23}$$

with $\eta^i_k \sim \mathcal{N}(0,1)$.

**Multiplicative coupling** Another possible extension is a multiplicative coupling between units as for example employed in Gancarz & Grossberg (1998) or the original works of Wilson & Cowan (1972, 1973). In the most general form, this amounts to

$$\tau dX^i(t) = \left[ -X^i(t) + H(X^i) \cdot \phi\left( \sum_{j=1}^{N} w^{ij} \psi\left( X^j(t-d) \right) \right) \right] dt + \sqrt{\tau}\sigma dW^i(t)\,, \tag{8.24}$$

which, again assuming the coupling term to be constant during the update, can be solved using the exponential Euler update

$$\begin{aligned} X^i_{k+1} = e^{-\Delta t/\tau} X^i_k + \left(1 - e^{-\Delta t/\tau}\right) \left[ H(X^i_k) \cdot \phi\left( \sum_{j=1}^{N} w^{ij} \psi\left( X^j_{k-\frac{d}{\Delta t}} \right) \right) \right] \\ + \sqrt{\frac{1}{2}(1 - e^{-2\Delta t/\tau})}\, \sigma \eta^i_k \end{aligned} \tag{8.25}$$

with $\eta_k^i \sim \mathcal{N}(0,1)$.

**Multiplicative noise** So far, we have considered rate models subject to additive noise corresponding to $b(t, x(t)) = b(t)$ in Eq. (8.4). The linear rate model considered in Sec. 8.3.3.1 describes the dynamics around a stationary state and due to the stationary baseline, the noise amplitude is constant. However, one might relax the stationarity assumption which would render the noise amplitude proportional to the time dependent rate, i.e. a multiplicative noise amplitude. The presented framework covers the latter since the exponential Euler update is also valid for multiplicative noise (Eq. (8.8)).

**Output noise** Grytskyy et al. (2013) show that there is a mapping between a network of leaky integrate-and-fire models and a network of linear rate models with so-called output noise (see Chap. 5). Here the noise is added to the output rate of the afferent neurons

$$\tau \frac{dX^i(t)}{dt} = -X^i(t) + \mu + \phi\left(\sum_{j=1}^{N} w^{ij}\psi\left(X^j(t-d) + \sqrt{\tau}\sigma\xi^j(t)\right)\right) \qquad i = 1, \ldots, N \tag{8.26}$$

and we cannot write the system as a SDE of type Eq. (8.2), as the nonlinearities $\phi(x)$ and $\psi(x)$ are also applied to the white noise $\xi^j$. In addition to the implementation `rate_neuron_ipn` for the rate-based models Eq. (8.9) discussed in the present work, our reference implementation also contains a base implementation `rate_neuron_opn` for models with output noise. For these models, the stochastic exponential Euler method can not be employed. Instead the solver assumes the noise $\xi^j$ to be constant over the update interval which leads to the update formula

$$X_{k+1}^i = e^{-\Delta t/\tau} X_k^i + \left(1 - e^{-\Delta t/\tau}\right)\left[\mu + \phi\left(\sum_{j=1}^{N} w^{ij}\psi\left(X_k^j + \sqrt{\frac{\tau}{\Delta t}}\sigma\eta_k^j\right)\right)\right]. \tag{8.27}$$

The term $X_k^j + \sqrt{\frac{\tau}{\Delta t}}\sigma\eta_k^j$ with $\eta_k^j \sim \mathcal{N}(0,1)$ is calculated beforehand in the sending neuron $j$, which results in the same amount of communicated data as in the case of models with input noise.

## 8.4 Interim discussion

In the following we discuss specific issues, regarding this section. A general summary and a broader discussion of the results follow in Chap. 9.

The current reference implementation uses an exponential Euler scheme (Adamu, 2011; Komori & Burrage, 2014) with a diagonal matrix $A$ (scalar exponential Euler): The additive noise as well as the leaky dynamics of single neurons are exactly integrated while the network input to the neurons is approximated as piecewise constant. The analysis in Sec. 8.3.1 demonstrates

that the scalar exponential Euler is the most accurate and stable standard-method for SDEs that is applicable to a distributed spiking simulator. In particular the distributed design renders implicit methods less feasible, as the convergence of the involved fixed-point iteration requires small time-steps in a region where the exponential Euler is already stable and more accurate. As the computation step size needs to be compared against the time constant $\tau$, stable solutions for small values $\tau \ll 1$ may require to decrease the step size below a default value.

The reference implementation provides an optional iterative method, the waveform relaxation (Lelarasmee, 1982), for networks with instantaneous rate connections. This method improves scalability by reducing communication at the cost of additional computations. As a consequence, the optimal method (standard vs. iterative) depends on the numbers of compute nodes and virtual processes. In our test case the use of the waveform-relaxation technique is beneficial for 1024 or more virtual processes. It is therefore recommended to employ the iterative scheme for large-scale simulations on supercomputers, but to disable it for smaller rate-model simulations on local workstations or laptops. This can easily be expressed by the parameter `use_wfr` (see Sec. 8.2.3.3 for details) of the algorithm. In general, the scalability for simulations of rate models is worse than for spiking network simulations (Kunkel et al., 2014) and comparable to simulations with gap junctions (Hahne et al., 2015). This is expected since for rate neurons as well as for gap junctions a large amount of data needs to be communicated compared to a spiking simulation. Future work should assess whether this bottleneck can be overcome by a further optimized communication scheme.

While our reference implementation uses the simulation software NEST as a platform, the employed algorithms can be ported to other parallel spiking network simulators. Furthermore, the implementation of the example neuron models as C++ templates allows customization to arbitrary gain functions. Researchers can create additional models, without in-depth knowledge of simulator specific data structures or numerical methods. In addition, the infrastructure is sufficiently general to allow for extensions to other categories of neuron models as shown explicitly for nonlinear neuron dynamics, multiplicative coupling, and other types of noise. This design enables the usage of the framework for a large body of rate-based neuronal network models. Furthermore, the generality of the model equations supports applications beyond neuronal networks, such as in computational gliascience (Amiri et al., 2012) or artificial intelligence (Haykin, 2009).

Some design decisions for the reference implementation come with an up- and a downside and may at the present state of knowledge and experience constitute judgment calls: The choice to determine the type of nonlinearity of the recurrent network input with a boolean parameter is based on the assumption that this implementation covers the majority of neuron models used in neuroscience today. The solution has an advantage in maintainability as it results in half as many template instances for a given set of gain functions than the alternative solution discussed above. It also avoids the introduction of potentially confusing names of neuron models encoding the nature of the nonlinearity. On the downside, models that do actually

employ both nonlinearities at once cannot be expressed. Furthermore, a decision that can already be made at the time when the model instance is created, is delayed to the simulation phase. Moreover, the choice to create a separate connection type for mean-field models of the `siegert` type is led by the ambition to avoid memory overhead. This comes at the price that neurons of this type cannot be connected to instances of rate models using the generic rate connection. Adapter elements like the `parrot_neuron` (see Kunkel et al., 2011, for a recent application) are one way to overcome this problem. Only the experience of researchers with the present implementation will inform us on whether characteristics and user interface serve the purpose of the community or if particular decisions need revision.

# Part V

# Discussion

# Chapter 9

# Summary and conclusions

In this thesis we contribute to the theoretical understanding of complex spiking neural network models, a model class studied in computational neuroscience. We extend the current scientific knowledge with respect to three main problems hitherto hindering the application of state-of-the-art mean-field theory to these network models. In particular we advance mean-field theory by studying

- the impact of colored noise on dynamic properties of single neurons
- the impact of complex network structure on the global stability of the models.

In addition, we develop dynamical mean-field theory further by considering

- the impact of spiking noise on network-generated fluctuations.

Our main results are summarized in the following.

In Chap. 2, we present a general method to reduce a colored-noise to a white-noise system by capturing the color of the noise by effective and time-dependent boundary conditions. We apply this formalism to the LIF neuron model, revealing a novel analytical expression for its transfer function. This result allows the assessment of small fluctuations around the stationary state in a network of neurons as illustrated in Chap. 5.

In Chap. 4 we devise a mean-field method allowing us to shape the phase space of a multi-scale network model of the vision-related areas of macaque cortex by systematically refining its connectivity. Fundamental constraints on the activity, i.e., prohibiting quiescence and requiring global stability, prove sufficient to obtain realistic layer- and area-specific activity.

In Chap. 6 we develop the dynamical mean-field theory for a stochastic system, where the noise mimics the spiking behavior of neurons. We derive the self-consistent statistics of the activity and the maximum Lyapunov exponent. Our results show that noise suppresses chaos by a dynamic mechanism, shifting the transition to significantly larger coupling strengths than predicted by local stability analysis.

The quantitative validation of the various mean-field approaches requires simulations of rate-based as well as spike-based neural networks. In Chap. 8 we extend a spike-based network simulator and thus develop a unified simulation framework supporting spiking and rate-based

models which facilities this cross check and increases reliability by using of the same simulation code and the same network model specifications for both model classes.

In the following we discuss the individual results in detail and in the next chapter we provide an overall outlook.

**Impact of colored noise** In Chap. 2 we investigate the impact of realistic colored noise on the escape of a Brownian particle from a meta-stable state. Independent of the physical system, specified by the potential, the leading order correction to the escape time solely comes about by the effect of the colored noise in conjunction with the absorbing boundary; there is no correction at the leading order to the Fokker-Planck equation describing the bulk of the density away from the boundary. This is reflected in the main result: A time-dependent system driven by colored-noise can, to first order, be equivalently described by a white-noise system with shifted boundary locations. The modification of the boundary condition can be understood on physical grounds. An absorbing boundary in a white-noise driven system results in a vanishing density at the boundary, because the infinitely fast noise causes any state sufficiently close to the boundary to immediately cross this threshold. The situation is different if the noise has a band-limited spectrum. The diffusive motion induced by the noise is slower, reducing the rate of escape for states close to the boundary. Since the density follows a continuity equation (it obeys conservation of probability), its stationary magnitude results from an equilibrium between inflow and outflow at any instant. A reduced outflow over the threshold is therefore accompanied by elevated density at threshold, which is exactly what is formally described by displacing the perfectly absorbing white-noise boundary beyond the point of the physical threshold.

In Sec. 2.3 we apply the central result to the LIF neuron model and obtain an approximation for its transfer function in the presence of colored noise. We show that for biologically relevant parameters the theory delivers a viable approximation. In deriving the underlying white-noise transfer function we exploit the analogy between diffusion processes and quantum mechanics by transforming the Fokker-Planck equation of the LIF model to the Hamiltonian of the quantum harmonic oscillator. The approximation for colored noise is obtained by simply shifting the boundaries in the white-noise solution. This simple application shows the strength of the new theory as it does not require lengthy calculations to obtain approximation for the mathematically difficult problem of colored noise.

With respect to the overall goal of this thesis, the transfer function constitutes the basic module for a theoretical understanding of fluctuations and collective oscillations in the activity of neural network models composed of LIF neuron models. On the one hand this is exemplified in Chap. 5, where we show that the analytic prediction for the power spectra of the population activity, which relies on our results for the transfer function, accurately describes the fluctuations in a simple network as well as in the complex multi-area model. On the other hand our expression for the transfer function has been successfully used in the work by Bos et al. (2016) to determine the anatomical origin of oscillations in the cortical microcircuit model.

**Stationary activity in complex spiking networks** Chap. 3 describes the framework to predict the time-averaged activity in spiking networks using standard mean-field theory. Due to a possibly large number of neurons or populations, the fixed-point solution of the stationary activity can generally not be determined analytically. We show how it can be found by numerically evolving a pseudo-time dynamics.

This forms the basis for Chap. 4 where we devise a mean-field based method which reveals and controls the global stability properties of complex spiking networks. With this method we address the following concrete problem: Initially, the multi-area model exhibits an unrealistic activity state with partly vanishing firing rates. An increase in the external drive to the model leads to a transition to a high activity state with rates orders of magnitude higher than what is actually observed in cortex. Thus the bistability prevents a realistic set point of activity.

We use mean-field theory to first understand the origin of this behavior. An interplay between the inherent neuron's non-linearity and excitatory connections leads to a bistable phase space with a low- and a high-activity attractor. The separatrix divides their basins of attraction and thus determines the global stability of the individual fixed points. An increase in the drive shifts the separatrix which in turn reduces the global stability of the low-activity fixed point. The reduced stability then leads to the transition to the state with unrealistically high activity. The basic idea of the method is to control the global stability of the low activity fixed point by controlling the location of the unstable fixed points on the separatrix.

The complexity, i.e. the large number of populations of the analyzed model, comes along with two main difficulties with respect to the control of the unstable fixed points. First, in this high-dimensional system it is difficult to find the unstable fixed points. To solve this problem, we follow trajectories along the separatrix and identify the point at which the trajectories are repelled from the separatrix as the unstable fixed point. Second, it is not a priori clear in which direction the separatrix has to be shifted to ensure stable dynamics. We therefore choose a pragmatic approach and shift the separatrix back to its initial location, inverting the shift caused by the increased drive which has reduced the global stability of the low-activity fixed point. The preserved stability allows us to increase the drive to the model without causing the transition to the high activity-state, which finally leads to a realistic set point of activity.

We control the locations of the unstable fixed points by expressing their dependencies on the anatomical connectivity of the model. To this end we perform a linearization around the unstable fixed points. This identifies a set of connections to be adjusted in order to ensure the desired stability properties.

The change of the structural connectivity of the model is justified due to the large uncertainties in the experimental data underlying the models structural connectivity. Thus, the method contributes to the data integration process by constraining the experimentally obtained connectivity map to a realization that is compatible with physiological experiments. To this end we use only weak constraints, demanding the activity to neither vanish nor be pathologically high. The resulting heterogeneity of the firing rates across areas is thus not imposed by the

method, but rather arises from the connectivity that remains strongly informed by anatomical data.
This interplay between model construction and theory serves as a prime example showing the importance of theoretically understanding the activity of spiking network models. Without the analytic description it would have been extremely difficult to find the necessary modifications due to the complexity of the system.

Connections are modified both inside and between cortical areas, on average well within the known uncertainties of the underlying data. The necessary changes can be classified into three main findings discussed in the following.

The model areas are based on the early sensory cortex model presented in Potjans & Diesmann (2014). This circuit is adapted to individual areas by taking into account neuronal densities and laminar thicknesses. The model definition renders areas with equal architectural type similar in their internal connectivity, a drastic but inevitable simplification due to the lack of more detailed experimental data. The proposed method softens this assumption: small adaptations of the internal connectivity distinguish the architectural types into further subtypes. These modifications are significant for the global stability of the network. Thus, our approach enables purely anatomy-based area categorizations to be refined with dynamical information.

Second, connections between areas are changed in terms of total strength and laminar patterns. Overall, the changes are small, but significant for specific connections. The loop formed by areas 46 and FEF is critical to the global stability of the network. Both areas have been investigated in Markov et al. (2014a), albeit in a different parcellation scheme than the scheme used here (Felleman & Van Essen, 1991). Our method suggests a weaker coupling of these two areas than found in the anatomical data set. Uncertainties, partly due to the mapping between parcellations, leave room for this interpretation. Areas 46 and FEF belong to prefrontal cortex and are multimodal, indicating that the influence of other parts of cortex could stabilize them, a mechanism outside the scope of the present model of vision-related areas. Both explanations can be tested either in an experimental study or with an extended model.

Third, our analysis reveals that layer 5 excitatory cells play a critical role in the model's dynamics, in line with the observed ongoing activity in mammalian neocortex (Sanchez-Vives & McCormick, 2000; Beltramo et al., 2013). This critical role is often attributed to single-neuron properties, with a subset of layer 5 neurons displaying pacemaker activity (Le Bon-Jego & Yuste, 2007; Lőrincz et al., 2015; Neske et al., 2015). Since these single-neuron properties are neglected in the model, we here find that in addtion the network architecture itself already explains the strong impact of layer 5 on the phase space of the network, suggesting that single-neuron properties and network structure jointly enable layer 5 to exert its dominant influence.

In conclusion, the method establishes a path from experimentally observed activity to specific hypotheses about the anatomy. In the following we discuss possible generalizations of the procedure. The method can be applied to networks with more complex sets of attractors

compared to the bistable case considered here. Though in high-dimensional systems such as the multi-area model, a larger number of attractors would make it more challenging to find all relevant unstable fixed points, the underlying idea of preserving the location of a separatrix is general. In contrast to the model considered here, transitions between fixed points can have a functional meaning in certain neuronal networks with multiple attractors. The specific location of the separatrix is then functionally relevant. Our method exposes the sensitivity of the location of the separatrix to certain model parameters and allows controlling its location in a specific manner.

**Dynamical mean-field theory for stochastic rate dynamics** In Chap. 6 we illustrate the break-down of mean-field predictions for increased coupling strength, a phenomenon already shown by Tetzlaff et al. (2012); Ostojic (2014). In particular, the stationary firing rates in the multi-area model can not be analytically predicted due to large fluctuations in the activity, which cannot be captured by the static mean-field theory.

Consequently, we turn the focus of Chap. 7 to dynamic mean-field theory (DMFT), which provides an effective description of fluctuating activity. To date, a dynamic mean-field formulation for spiking neural networks does not exist. One of the steps towards applying dynamic mean-field theory to spiking networks is the inclusion of noise representing the spiking activity. Our dynamic mean-field theory for stochastic dynamics allows us to investigate the generic effect of noise. In particular, we study the effect of noise on the transition to chaos, which is believed to be related to the emergence of fluctuations in the activity of spiking networks (Ostojic, 2014; Mastroguiseppe & Ostojic, 2016; Harish & Hansel, 2015).

Adding white noise to the seminal model by Sompolinsky et al. (1988) yields a stochastic continuous-time dynamical system. By developing the DMFT for this system we reduce the interacting system to the dynamics of a single unit. The self-consistent solution of the effective equation yields a standard physics problem: the population-averaged autocorrelation function is given by the motion of a classical particle in a potential where the noise amounts to the initial kinetic energy.

We then determine the maximum Lyapunov exponent which follows from the ground-state of a quantum mechanical problem. This leads to an exact condition for the transition from stable to chaotic dynamics. Our main results is that noise suppresses chaos by a dynamic effect: by sharpening the autocorrelation function, noise decreases the maximum Lyapunov exponent. Qualitatively this can be understood by considering that noise strengthens the temporal variability of unstable directions and thus hinders a coherent growth of a perturbation. This dynamic supression of chaos leads to the emergence of a novel dynamical regime, which combines local instability with long term stability.

The seminal works (Sompolinsky et al., 1988; Sommers et al., 1988) established a tight relationship between the fields of random matrix theory and neural networks with random topology: Deterministic chaos emerges if the spectral radius of the coupling matrix exceeds unity. In

contrast, we find that in the situation of stochastic dynamics the spectral radius only yields a necessary condition. Thus our main result, the relation between temporal noise, the statistics of the random matrix, and the onset of chaos (Eq. (7.19)) generalizes this well known link to stochastic dynamics.

Furthermore, we show that the decay time of the autocorrelation function does not diverge at the transition. The peak of the decay time is strongly reduced by noise and occurs slightly above the critical coupling strength. Thus, noise weakens the critical slowing down at the transition. The absence of a diverging time scale was also found by a scaling analysis in a parallel and independent study (Kadmon & Sompolinsky, 2015b). However this work is valid only for weak noise.

In networks of spiking neurons the correlation time does not peak at the point where the transition is predicted by the instability of the corresponding deterministic rate dynamics (Engelken et al., 2015; Ostojic, 2015; Mastroguiseppe & Ostojic, 2016). This could be attributed to the effect of spiking noise: even for the simple rate dynamics considered in Chap. 6, noise leads to the absence of a diverging time scale and shifts the transition to larger coupling strengths than those predicted by local linear stability analysis.

The theoretical analysis of stochastic and hence non-autonomous dynamics lays the foundation for a better understanding of the computational properties of random neural networks: the maximum Lyapunov exponent for nonautonomous dynamics indicates the echo state property, i.e., the reliability of the network (Wainrib & Galtier, 2016). Since noise shapes the temporal correlations, we further expect that it significantly affects information processing capabilities such as short-term memory (Ganguli et al., 2008; Toyoizumi & Abbott, 2011). Thus, our results can facilitate the design and control of functional networks. More generally, the coexistence of local instability and hence expansive short-term dynamics with stable long-term behavior could serve as a computationally powerful substrate.

In order to show the generic effect of spiking noise we extended the seminal model by Sompolinsky et al. (1988) to stochastic dynamics. However, the connectivity in the original model is an all-to-all coupled Gaussian network, which does not correspond to the sparse coupling employed in the random balanced network of spiking neurons. Furthermore, the individual neurons have positive and negative outgoing synapses which violates Dayles' law and is in contrast to the $\mathcal{E} - \mathcal{I}$ connectivity pattern. In addition, the sigmoidal gain function is not realistic, since it is symmetric and enables negative firing rates, while the actual input-output relationship of the LIF neuron model is strictly positive with a threshold-like behavior (Chap. 4). Recently, extensions of the dynamic mean-field theory to more realistic networks structures and gain functions have been developed (Mastroguiseppe & Ostojic, 2016; Kadmon & Sompolinsky, 2015b; Harish & Hansel, 2015), which show that the transition to chaos is also present in these settings. However, the effect of noise on the critical coupling strength had not been considered and thus our results are complementary to these works.

We have motivated the study of dynamic mean-field theory by the observed breakdown of standard mean-field theory for spiking neural networks for large coupling strengths. However,

the results found in Chap. 7 only constitute one first step in order to apply DMFT to spiking networks. Considering the other works on DMFT as well (Mastroguiseppe & Ostojic, 2016; Kadmon & Sompolinsky, 2015b; Harish & Hansel, 2015), a theory which quantitatively explains the increase in the mean firing rates and the emerging fluctuations in strongly coupled spiking networks is still lacking. The underlying reason is that the starting point of DMFT are rate dynamics. The latter can be formulated as a field theory in which the DMFT is developed. However, a general formulation of spiking dynamics as a field theory is missing, hindering a quantitative agreement between DMFT results and spiking network dynamics.

**Simulation of rate models** Chap. 8 presents an efficient way to integrate rate-based models in a neuronal network simulator that is originally designed for models with delayed spike-based interactions. The advantage of the latter is a decoupling of neuron dynamics between spike events. This is used by current parallel simulators for large-scale networks of spiking neurons to reduce communication between simulation processes and significantly increase performance and scaling capabilities up to supercomputers (Morrison et al., 2005). In contrast, rate-based models interact continuously. For delayed interactions, neuronal dynamics are still decoupled during the minimal delay of the network such that information can be exchanged on a coarse time-grid. For instantaneous coupling, communication in every time step is required. This is feasible for small networks that can be simulated on small machines and thus require only a small amount of communication. For improved efficiency of simulations of large networks on supercomputers, we implement an iterative numerical solution scheme (Lelarasmee, 1982). Furthermore, we investigate several standard methods for the solution of rate model equations and demonstrate that the scalar exponential Euler method is the best choice in the context of a neuronal network simulator that is originally designed for models with delayed spike-based interactions. Afterwards, we show the applicability of the numerical implementation to a variety of well-known and widely-used neuron models and illustrate possible generalizations to other categories of rate-based neuron models.

The mean-field predictions for the time-averaged activity (Chap. 3, Chap. 4) in spiking neural networks are easily obtained with the presented unifying framework. Only minor adaptations to a script defining a spiking network simulation are needed to yield the corresponding mean-field solution: The neuron model and the connection type have to be replaced by the `siegert_neuron` and the `diffusion_connection` (with effective coupling parameters), respectively. Without the unifying framework one would have to rewrite the mean-field solution from scratch in a different simulator. Thus, the unified framework strongly reduces the risk of introducing errors while rewriting or duplicating simulation code.

# Chapter 10

# Outlook

The presented analytical method (Chap. 4) combining anatomy and activity data into a consistent model is not restricted to stationary firing rates. In future studies, also higher-order statistical measures of activity can be used as constraints. Resting-state fMRI, for example, provides information on the functional connectivity and thus the coherence between areas as a second-order measure. When combined with analytical predictions of functional connectivity, the method may shed light on the anatomical connection patterns underlying inter-area communication.

Further possible extensions rely on the fact that the method applies to any model parameter. It would be possible to modify, for example, synaptic weights. Since experimental data on synaptic weights are sparse, this is another natural choice. Moreover, such an analysis may provide hints about suitable synaptic plasticity rules that dynamically stabilize the model.

Another direct application of the analytic description of the fixed points of the multi-area model is to study the behavior of the model with respect to constant changes in the model's external input similar to the study by Cain et al. (2016) on the microcircuit model. This could represent an external stimuli or changing input from other parts of the brain.

The results obtained in Chap. 5 combined with the methods developed in Bos et al. (2016) constitute the prequisite for understanding and controlling the emergence of oscillations in the multi-area model. In future this approach can be used to study the cross-spectra between different areas in the model in light of the communication through coherence hypothesis (Fries, 2005; Bastos et al., 2015; van Kerkoerle et al., 2014). This hypothesis states that areas communicate with each other by synchronizing their activity on different frequency bands depending on the hierarchical relationship between the areas: While high frequencies are involved in the feed-forward direction (e.g. $V1 \rightarrow V4$), low frequencies dominate feed back communication (e.g. $V4 \rightarrow V1$). Bastos et al. (2015b) further hypothesize that the distinct frequency bands originate from direction-specific laminar connection pattern. As the model features both hierarchically ordered areas and laminar-specific connectivity between the areas combined with simplified neuron dynamics, one can investigate in how far the structural connectivity could explain the experimental observations. The analytical description of the oscillations in the model, developed in this thesis, could support these investigations by revealing direct links

between structure and oscillations (Bos et al., 2016).

The presented joint platform for spike-based and rate-based models can trigger new research questions by facilitating collaboration and translation of ideas between scientists working in the two fields. The unifying framework allows researchers to easily switch between rate-based and spiking neurons in a particular network model requiring only minimal changes to the simulation script. This facilitates an evaluation of the different model types against each other. In particular, it is instructive to study how certain measures, e.g. firing rates, change with the neuron model (Brette, 2015). Functional networks being able to perform a given task are typically designed with rate-based neurons. With our framework, their validity can now be more easily evaluated by going from a more abstract rate-based model to a biologically more realistic spiking neuron model.

The unification of spike and rate-based approaches is a promising route for future research also from a theoretical point of view. A formulation of spiking dynamics as a field theory could finally unify the two modeling approaches, which are so far connected only via ad-hoc mean-field theories. This could reveal how to model interactions between spiking and rate-based units, which can then be implemented in our unifying simulation framework in a straight forward manner. A possible application is a multi-scale modeling approach combining large-scale spiking networks with functionally inspired rate-based elements. We believe that such a combination of the two fields, both from a theoretical point of view or with multi-scale models, is an important step towards deciphering the working principles of the brain.

# Part VI

# Appendices

# Appendix A

# Supplementary calculations for Fokker-Planck theory

## A.1 N-dimensional Fokker-Planck equation

We now set out to derive the Fokker-Planck equation describing the time-evolution of the density of states $P$, loosely following Risken (1996) and (Helias et al., 2015). We start from the Chapman Kolmogorov equation Risken (1996) for a general $N_x$-dimensional state vector $\boldsymbol{x}$, which reads

$$P(\boldsymbol{x},t) \quad = \quad \int P(\boldsymbol{x},t|\boldsymbol{x}',s)\,P(\boldsymbol{x}',s)\,d^{N_x}\boldsymbol{x}'\,, \tag{A.1}$$

where $P$ denotes the probability density and the state variables $\boldsymbol{x}$ is assigned to time point $t$ and $\boldsymbol{x'}$ to time point $s$ earlier in time. We now aim for an approximation of the integrand. We consider a small time step $\delta t$. Taking the limit $\delta t \to 0$, we note that $P(\boldsymbol{x},t+\delta t|\boldsymbol{x}',t)$ approaches a $N_x-$dimensional Dirac $\delta$-distribution $\prod_i \delta(\boldsymbol{x}_i - \boldsymbol{x}_i')$. We further assume that the transition probability $P(\boldsymbol{x},t+\delta t|\boldsymbol{x}',t)$ only couples nearby states, namely those where $\boldsymbol{\Delta} \equiv \boldsymbol{x}-\boldsymbol{x'}$ is small. Therefore, the joint probability $P(\boldsymbol{x},t+\delta t,\boldsymbol{x}',t) = P(\boldsymbol{x},t+\delta t|\boldsymbol{x}',t)P(\boldsymbol{x}',t)$ must have a maximum near $\boldsymbol{x} \simeq \boldsymbol{x'}$ for small $\delta t$. Thus, we approximate the dependence of the joint probability density on $\boldsymbol{\Delta}$ as a parabola, i.e we perform a Taylor series up to second order in $\boldsymbol{\Delta}$.

First, however, we eliminate one of the states from the equation. Eliminating $\boldsymbol{x}$ would result in the backward equation Risken (1996), since the final equation would depend on the less advanced point in time. Instead, we here eliminate $\boldsymbol{x'}$, which in the end will result in the so called forward equation. We therefore replace $\boldsymbol{x}' = \boldsymbol{x} - \boldsymbol{\Delta}$ and obtain

$$P(\boldsymbol{x},t+\delta t,\boldsymbol{x}-\boldsymbol{\Delta},t) \quad = \quad P(\boldsymbol{x},t+\delta t|\,\boldsymbol{x}-\boldsymbol{\Delta},t)\,P(\boldsymbol{x}-\boldsymbol{\Delta},t).$$

As we will see later, it is beneficial to introduce a $\boldsymbol{\Delta}$-dependence in the first argument of $P$. We do this by defining a function $f(\boldsymbol{\Delta},\boldsymbol{\Delta}')$ introducing $\boldsymbol{\Delta}'$ which we first treat as independent of $\boldsymbol{\Delta}$ as

$$f(\boldsymbol{\Delta}, \boldsymbol{\Delta}') \quad = P(\boldsymbol{x} + \boldsymbol{\Delta}' - \boldsymbol{\Delta}, t + \delta t | \boldsymbol{x} - \boldsymbol{\Delta}, t)\, P(\boldsymbol{x} - \boldsymbol{\Delta}, t) \quad , \tag{A.2}$$

and set $\boldsymbol{\Delta}' = \boldsymbol{\Delta}$ in the end. The multi-dimensional Taylor expansion of Eq. (A.2) to second order in $\boldsymbol{\Delta}$ yields

$$\begin{aligned} f(\boldsymbol{\Delta}, \boldsymbol{\Delta}') &= \left(1 + \sum_i \Delta_i \partial_{\Delta_i} + \frac{1}{2} \sum_{i,j} \Delta_i \Delta_j \partial_{\Delta_i} \partial_{\Delta_j} + O(\Delta^3)\right) f(\boldsymbol{\Delta}, \boldsymbol{\Delta}')|_{\boldsymbol{\Delta}=0} \\ &= \left(1 - \sum_i \Delta_i \partial_i + \frac{1}{2} \sum_{i,j} \Delta_i \Delta_j \partial_i \partial_j\right) P(\boldsymbol{x} + \boldsymbol{\Delta}, t + \delta t | \boldsymbol{x}, t)\, P(\boldsymbol{x}, t) + O(\boldsymbol{\Delta}^3) \end{aligned} \tag{A.3}$$

where we replace the derivative $\partial_{\boldsymbol{\Delta}} \to -\partial_{\boldsymbol{x}}$. This replacement is possible due to the introduction of the $\boldsymbol{\Delta}$-dependence in the first argument of $P$ (Eq. (A.2)). In the last line we further set $\boldsymbol{\Delta}' = \boldsymbol{\Delta}$. Inserting Eq. (A.3) into Eq. (A.1) yields

$$P(\boldsymbol{x}, t + \delta t) \quad = \int \left(1 + - \sum_i \Delta_i \partial_i + \frac{1}{2} \sum_{i,j} \Delta_i \Delta_j \partial_i \partial_j\right) P(\boldsymbol{x} + \boldsymbol{\Delta}, t + \delta t | \boldsymbol{x}, t)\, P(\boldsymbol{x}, t)\, d^{N_x}\boldsymbol{\Delta}, \tag{A.4}$$

where we used $d^{N_x}\boldsymbol{\Delta} = -d^{N_x}\boldsymbol{x'}$. The Taylor expansion of the integrand is also called Kramers-Moyal expansion and in principle also higher order terms could have been taken into account. Instead, the truncating at the second order yields the Fokker-Planck equation. We arrive at the latter by defining the first and second infinitesimal moments as

$$\begin{aligned} a_i^1(\boldsymbol{x}, t + \delta t, t) &= \int_{-\infty}^{\infty} \Delta_i P(\boldsymbol{x} + \boldsymbol{\Delta}, t + \delta t | \boldsymbol{x}, t)\; d^N\boldsymbol{\Delta} \\ a_{ij}^2(\boldsymbol{x}, t + \delta t, t) &= \int_{-\infty}^{\infty} \Delta_i \Delta_j P(\boldsymbol{x} + \boldsymbol{\Delta}, t + \delta t | \boldsymbol{x}, t)\; d^N\boldsymbol{\Delta}\,. \end{aligned} \tag{A.5}$$

We insert these definitions into Eq. (A.4), rearrange the term with the 1 to the LHS and use that $\int P(\boldsymbol{x} + \boldsymbol{\Delta}, t + \delta t | \boldsymbol{x}, t)\, d^{N_x}\boldsymbol{\Delta} = 1$ which follows from conversation of probability. Finally, dividing by $\delta t$ and taking the limit $\delta t \to 0$ yields the Fokker-Planck equation

$$\frac{\partial P(\boldsymbol{x}, t)}{\partial t} \quad = \left(-\sum_i^{N_x} \partial_i A_i^1(\boldsymbol{x}, t) + \frac{1}{2} \sum_{i,j}^{N_x} \partial_i \partial_j A_{ij}^2(\boldsymbol{x}, t)\right) P(\boldsymbol{x}, t),$$

where we defined $A^n(\boldsymbol{x}, t) = \lim_{\delta t \to 0} \frac{1}{\delta t}\, a^n(\boldsymbol{x}, t + \delta t, t)$ for $n \in 1, 2$.

### A.1.1 LIF neuron model driven by Gaussian noise

We consider Eq. (2.33), however, with the mean $\mu$ shifted to the equation for the current $I$ so that $\mu$ represents synaptic input, i.e,

$$\begin{aligned} \tau_{\mathrm{m}}\dot{V} &= -V + I \\ \tau_{\mathrm{s}}\dot{I} &= -I + \mu + \sigma\sqrt{\tau_{\mathrm{m}}}\,\xi\,. \end{aligned}$$

In order to derive the Fokker-Planck for a specific system one needs to calculate the moments of the time evolution of the system for infinitesimal $\delta t$ (A.5). The state vector for the LIF neuron model is $\boldsymbol{x} = (V, I)^T$ and the time evolution for infinitesimal $\delta t$ amount to

$$\begin{aligned} \delta V &= (-V + I)\frac{\delta t}{\tau_{\mathrm{m}}} \\ \delta I &= (-I + \mu + \sigma\sqrt{\tau_{\mathrm{m}}}\,\xi_{t/\delta t})\frac{\delta t}{\tau_{\mathrm{s}}}\,. \end{aligned} \tag{A.6}$$

$\xi_{k/\delta t}$ denotes a realization of the white noise in Eq. (2.33) as a limit of as a piece-wise constant process, where time is discretized in bins of length $\delta t$ and at each time point $k \in \mathbb{Z}$ the value of $\xi_k := \xi(\delta t\, k)$ is drawn independently from a Gaussian distribution $\mathcal{N}(0, \frac{1}{\delta t})$. In the limit $\delta t \to 0$ this results in the white noise process $\xi$ (Helias et al., 2015). The infinitesimal moments follow as

$$\begin{aligned} A_1^1 &= \lim_{\delta t\to 0}\frac{1}{\delta t}\langle(-V+I)\frac{\delta t}{\tau_{\mathrm{m}}}\rangle = (-V+I)\frac{1}{\tau_{\mathrm{m}}} \\ A_2^1 &= \lim_{\delta t\to 0}\frac{1}{\delta t}\langle(-I+\mu+\sigma\sqrt{\tau_{\mathrm{m}}}\,\xi_{t/\delta t})\frac{\delta t}{\tau_{\mathrm{s}}}\rangle = -\frac{I+\mu}{\tau_{\mathrm{s}}} \\ A_{11}^2 = A_{12}^2 = A_{21}^2 &= \lim_{\delta t\to 0}\frac{1}{\delta t}O(\delta t^2) = 0 \\ A_{22}^2 &= \lim_{\delta t\to 0}\frac{1}{\delta t}\langle(\sigma\sqrt{\tau_{\mathrm{m}}}\,\xi_{t/\delta t}\frac{\delta t}{\tau_{\mathrm{s}}})^2\rangle = \sigma^2\frac{\tau_{\mathrm{m}}}{\tau_{\mathrm{s}}^2}, \end{aligned}$$

which yields the Fokker Planck equation

$$\frac{\partial P(V,I,t)}{\partial t} = \left(-\frac{1}{\tau_{\mathrm{m}}}\partial_V(-V+I) + \frac{1}{\tau_{\mathrm{s}}}\partial_I(I+\mu) + \frac{\sigma^2\tau_{\mathrm{m}}}{2\tau_{\mathrm{s}}^2}\partial_I^2\right) P(V,I,t)\,. \tag{A.7}$$

Note, that higher orders in the Kramers-Moyal expansion vanish: Since we take $\lim_{\delta t\to 0}$ in the end, only a term $\sim \xi_{t/\delta t}$ can contribute to the higher moments. However, one has $\langle(\delta t\xi_{t/\delta t})^n\rangle = \begin{cases} 0 & \text{for } n \text{ odd} \\ \propto \delta t^{n/2} & \text{for } n \text{ even} \end{cases}$ so that all moment higher than order two vanish for $\lim_{\delta t\to 0}$ .

### A.1.2 LIF neuron driven by Poissonian spike trains

We consider the situation of one a LIF model neuron receiving a Poissonian spike-train with rate $\nu_{\text{ext}}$ and synaptic efficacy $J$. The time evolution for infinitesimal $\delta t$ is

$$\begin{aligned} \delta V &= (-V+I)\frac{\delta t}{\tau_{\text{m}}} \\ \delta I &= \begin{cases} (-I\frac{\delta t}{\tau_{\text{s}}} + \frac{\tau_{\text{m}}}{\tau_{\text{s}}}J) & \text{for a spike in the } \delta t\text{-interval} \\ (-I\frac{\delta t}{\tau_{\text{s}}}) & \text{for no spike in the } \delta t\text{-interval} \end{cases} \end{aligned} \tag{A.8}$$

where a spike occurs with probability $\nu_{\text{ext}}\delta t$ and no spike with probability $1-\nu_{\text{ext}}\delta t$. The time evolution for the membrane potential is exactly the same as before (Eq. (A.6)) and thus the moment $A_1^1$ is identical. To calculate the moments for the current $I$ we need to sum over the two possibilities in Eq. (A.8) which yields

$$\begin{aligned} A_2^1(I,V,t+\delta t,t) &= \lim_{\delta t\to 0}\frac{1}{\delta t}\left((J\frac{\tau_{\text{m}}}{\tau_{\text{s}}} - \frac{\delta t}{\tau_{\text{s}}}I)\nu_{\text{ext}}\delta t - \frac{\delta t}{\tau_{\text{s}}}I(1-\nu_{\text{ext}}\delta t)\right) \\ &= J\frac{\tau_{\text{m}}}{\tau_{\text{s}}}\nu_{\text{ext}} - \frac{I}{\tau_{\text{s}}} \\ &\equiv -(I+\mu)\frac{1}{\tau_{\text{s}}}, \end{aligned}$$

where we defined $\mu = \tau_{\text{m}} J \nu_{\text{ext}}$. The second moment is

$$\begin{aligned} A_{22}^2(I,V,t+\delta t,t) &= \lim_{\delta t\to 0}\frac{1}{\delta t}\left[\left(J\frac{\tau_{\text{m}}}{\tau_{\text{s}}} - \frac{\delta t}{\tau_{\text{m}}}I\right)^2\nu_{\text{ext}}\delta t + \left(-\frac{\delta t}{\tau_{\text{m}}}I\right)^2(1-\nu_{\text{ext}}\delta t))\right]. \\ &= (J\frac{\tau_{\text{m}}}{\tau_{\text{s}}})^2\nu_{\text{ext}} \\ &\equiv \frac{\tau_{\text{m}}}{\tau_{\text{s}}^2}\sigma^2, \end{aligned}$$

where we defined $\sigma^2 = \tau_{\text{m}} J^2 \nu_{\text{ext}}$. As in the Gaussian noise case all other second moments $A_{11}^2, A_{12}^2, A_{21}^2$ vanish in the $\lim_{\delta t\to 0}$ as they are of $O(\delta t)$, which leaves us with the Fokker-Planck equation identical to the Gaussian case Eq. (2.35).

## A.2 White-noise stationary rate

We here derive the solution to Eq. (2.41) leading to the stationary firing rate Eq. (2.42). The homogeneous solution $a\,q_h = (\partial_x + \frac{1}{2}x)\,q_h(x) = 0$ is $q_h(x) = u(x)$. Hence, the full solution satisfying the white-noise boundary condition $q_0(x_\theta) = 0$ is

$$\begin{aligned} q_0(x) &= \tau_{\text{m}}\nu_0\,u(x)\int_{\max(x,x_R)}^{x_\theta} u^{-1}(x')u^{-1}(x')\,dx' \\ &= \tau_{\text{m}}\nu_0\,e^{-\frac{x^2}{4}}\int_{\max(x,x_R)}^{x_\theta} e^{\frac{x'^2}{2}}\,dx'. \end{aligned} \tag{A.9}$$

Consequently, the solution in terms of the density $\rho$ is

$$\rho_0(x) = u(x)\, q_0(x) = \tau_{\mathrm{m}}\nu_0\, e^{-\frac{x^2}{2}} \int_{\max(x,x_R)}^{x_\theta} e^{\frac{x'^2}{2}}\, dx'$$

in agreement with Brunel (2000, eq. 19). We determine the (as yet arbitrary) constant $\nu_0$ from the normalization condition $1 = \int \rho(x)\, dx$ as

$$\begin{aligned}(\tau_{\mathrm{m}}\nu_0)^{-1} &= \int_{-\infty}^{x_\theta} \rho(x)\, dx \\ &= \int_{-\infty}^{x_\theta} e^{-\frac{x}{2}^2} \int_{\max(x,x_R)}^{x_\theta} e^{\frac{x'^2}{2}}\, dx'\, dx \\ &= 2\int_{-\infty}^{y_\theta} \underbrace{e^{-y^2}}_{f'} \underbrace{\int_{\max(y,y_R)}^{y_\theta} e^{u^2}\, du}_{g}\, dy,\end{aligned}$$

where in the last step we substituted $y = x/\sqrt{2}$ and $u = x'/\sqrt{2}$. Using integration by parts with $f(y) = \int_{-\infty}^{y} e^{-x^2} dx = \frac{\sqrt{\pi}}{2}(1 + \mathrm{erf}(y))$ and $g'(y) = -e^{y^2} H(y - y_R)$ and noting that the boundary term vanishes, because $g(y_\theta) = 0$ and $f(-\infty) = 0$ we obtain the final result (Siegert, 1951; Ricciardi, 1977) for the firing rate

$$(\tau_{\mathrm{m}}\nu_0)^{-1} = \int_{x_R}^{x_\theta} u^{-2}(x) F(x)\, dx,$$

with $F(x) = \sqrt{\frac{\pi}{2}}(1 + \mathrm{erf}(\frac{x}{\sqrt{2}}))$.

## A.3 White-noise transfer function

The membrane potential $V$ of the LIF neuron model without synaptic filtering (white noise) evolves according to the differential equation

$$\tau_{\mathrm{m}}\dot{V} \;=\; -V + \mu + \sigma\sqrt{\tau_{\mathrm{m}}}\xi(t). \tag{A.10}$$

The corresponding Fokker-Planck equation is

$$\begin{aligned}\partial_t P(V,t) &= -\partial_V \varphi(V,t) \\ \varphi(V,t) &\equiv \left(-\frac{1}{\tau_{\mathrm{m}}}(V-\mu) - \frac{1}{2}\frac{\sigma^2}{\tau_{\mathrm{m}}}\partial_V\right) P(V,t).\end{aligned} \tag{A.11}$$

We consider a periodic modulation of the mean input in Eq. (A.10)

$$\begin{aligned}\mu(t) &= \mu + \delta\mu(t) \\ \delta\mu(t) &= \epsilon\mu e^{i\omega t}\end{aligned} \tag{A.12}$$

and of the variance

$$\begin{aligned}\sigma^2(t) &= \sigma^2 + \delta\sigma(t)^2 \\ \delta\sigma(t)^2 &= H\sigma^2 e^{i\omega t}.\end{aligned}$$

To linear order this will result in a modulation of the firing rate $\nu_0(t) = \nu_0(1 + n(\omega)e^{i\omega t})$, where $n(\omega)$ is the transfer function. Note that both modulations $\delta\mu$ and $\delta\sigma$ have their own contribution to $n(\omega)$ and in principle could be treated separately since we only determine the linear response here. For brevity we consider them simultaneously here. The time-dependent Fokker-Planck equation takes the form

$$\begin{aligned}\partial_t P(V,t) &= -\partial_V(\varphi(V,t) + \delta\varphi(V,t)) \\ \delta\varphi(V,t) &= \left(\frac{\delta\mu(t)}{\tau_\mathrm{m}} - \frac{\delta\sigma^2(t)}{2\tau_\mathrm{m}}\partial_V\right) P(V,t)\end{aligned}$$

or, in the natural coordinates

$$\begin{aligned}\partial_s\rho(x,s) &= \mathcal{L}_0(x)\rho(x,s) \\ &+ e^{i\omega\tau_\mathrm{m}s}\underbrace{(-G\,\partial_x + H\partial_x^2)}_{\equiv\mathcal{L}_1(x)}\rho(x,s).\end{aligned} \tag{A.13}$$

Here, $G = \sqrt{2}\epsilon\mu/\sigma$ and we defined the perturbation operator $\mathcal{L}_1(x)$.

### A.3.1 Perturbative treatment of the time-dependent Fokker-Planck equation

For small amplitudes $n(\omega) \ll 1$, so weak modulations of the rate compared to the stationary baseline rate, we employ the ansatz of a perturbation series, namely that the time-dependent solution of Eq. (A.13) is in the vicinity of the stationary solution, $\rho(x,s) = \rho_0(x) + \rho_1(x,s)$, with the correction $\rho_1$ of linear order in the perturbing quantities $\delta\mu(t)$ and $\delta\sigma(t)$. Inserting this ansatz into Eq. (A.13) and using the property of the stationary solution $\mathcal{L}_0\rho_0 = 0$ we are left with an inhomogeneous partial differential equation for the unknown function $\rho_1$

$$\begin{aligned}\partial_s\rho_1(x,s) &= \mathcal{L}_0(x)\rho_1(x,s) + e^{i\omega\tau_\mathrm{m}s}\,\mathcal{L}_1(x)\rho_0(x) \\ &+ e^{i\omega\tau_\mathrm{m}s}\mathcal{L}_1(x)\rho_1(x,s).\end{aligned} \tag{A.14}$$

Neglecting the third term that is of second order in the perturbed quantities, the separation ansatz $\rho_1(x,s) = \rho_1(x)\,e^{i\omega\tau_\mathrm{m}s}$ (for brevity we drop the $\omega$-dependence of $\rho_1(x)$) then leads to the linear ordinary inhomogeneous differential equation of second order

$$i\omega\tau_\mathrm{m}\,\rho_1 = \mathcal{L}_0\rho_1 + \mathcal{L}_1\rho_0. \tag{A.15}$$

From here the operator representation introduced in Sec. 2.3.1 guides us to the solution. Writing $\rho_1(x) = u(x)\, q_1(x)$ and using the commutation relation Eq. (2.37) the transformed inhomogeneity takes the form

$$\begin{aligned} \mathcal{L}_1 u\, q_0 &= -\partial_x \underbrace{(G - H\,\partial_x)}_{\equiv S_1}\, u\, q_0 \\ &= u\, a^\dagger (G + H\, a^\dagger)\, q_0, \end{aligned} \tag{A.16}$$

where we defined the contribution of the perturbation to the flux operator as $S_1$. With $\mathcal{L}_0 u q_1 = -u a^\dagger a q_1$ we need to solve the equation

$$(i\omega\tau_\mathrm{m} + a^\dagger a)\, q_1 = (G\, a^\dagger + H\, (a^\dagger)^2)\, q_0. \tag{A.17}$$

Since the equation is linear in $q_1$, its solution is a superposition of a particular solution and a homogeneous solution. The latter needs to be chosen such that the full solution complies with the boundary conditions but we first need to find the particular solution. To this end we will use the property Eq. (2.40). For $n = 1$ and $n = 2$ we have

$$\begin{aligned} a^\dagger a\, (a^\dagger q_0) &= a^\dagger q_0 \\ \text{and} & \\ a^\dagger a\, ((a^\dagger)^2 q_0) &= 2(a^\dagger)^2 q_0. \end{aligned}$$

Hence a term proportional to $a^\dagger q_0$ reproduces the first part of the inhomogeneity in Eq. (A.17) and a term proportional to $(a^\dagger)^2 q_0$ generates the second term. We therefore use $q_p = (\gamma a^\dagger + \beta (a^\dagger)^2)\, q_0$ as the ansatz for the particular solution and determine the coefficients $\gamma$ and $\beta$ by inserting into Eq. (A.17), which yields

$$i\omega\tau_\mathrm{m}\, (\gamma a^\dagger + \beta (a^\dagger)^2)\, q_0 + (\gamma a^\dagger + 2\beta (a^\dagger)^2)\, q_0 = (G\, a^\dagger + H\, (a^\dagger)^2)\, q_0.$$

Sorting by terms according to powers of $a^\dagger$, we obtain two equations determining $\gamma, \beta$

$$\begin{aligned} (i\omega\tau_\mathrm{m}\gamma + \gamma - G)\, a^\dagger q_0 &= 0 \\ (i\omega\tau_\mathrm{m}\beta + 2\beta - H)\, (a^\dagger)^2 q_0 &= 0, \end{aligned}$$

where the factor in parenthesis must be nought, because neither $a^\dagger q_0(x)$ nor $(a^\dagger)^2 q_0(x)$ vanish for all $x$. This leaves us with the particular solution

$$q_p = \left( \frac{G}{1 + i\omega\tau_\mathrm{m}}\, a^\dagger q_0 + \frac{H}{2 + i\omega\tau_\mathrm{m}}\, (a^\dagger)^2 q_0 \right). \tag{A.18}$$

This equation together with Eq. (2.40) shows that the perturbed solution consists of the first and the second excited state above the ground state, because the two terms are proportional to $a^\dagger q_0$ and $\left(a^\dagger\right)^2 q_0$. Thus the modulation of the input to the neuron is equivalent to exciting

the harmonic oscillator to higher energy states. Since only the ground state is a stationary solution, it is intuitively clear that the response of the neuron relaxes back after some time, in analogy to the return from the exited states.

### A.3.2 Homogeneous solution

The homogeneous equation follows from Eq. (A.17)

$$(i\omega\tau_\mathrm{m} + a^\dagger a)\, q_h \;=\; 0. \tag{A.19}$$

Evaluating $a^\dagger a$ yields

$$(-\partial_x^2 + \frac{1}{4}x^2 + i\omega\tau_\mathrm{m} - \frac{1}{2})\, q_h \;=\; 0,$$

which can be rearranged to the form

$$\begin{aligned} \partial_x^2 q_h - (\frac{1}{4}x^2 + m)\, q_h &= 0 \\ \text{with} \quad m &= i\omega\tau_\mathrm{m} - \frac{1}{2}, \end{aligned} \tag{A.20}$$

the solution of which can be written as a linear combination of two parabolic cylinder functions (Olver et al., 2010). The function $U(m,x) = D_{-m-\frac{1}{2}}(x)$ of Whittaker (Abramowitz & Stegun, 1974, 19.3.1/2) has the asymptotic behavior $U(m,x) \propto e^{-\frac{1}{4}x^2}|x|^{-m-\frac{1}{2}}$ for $x \to -\infty$ (Abramowitz & Stegun, 1974, 19.8.1). The other independent solution $V(m,x) \propto e^{\frac{1}{4}x^2}|x|^{m-\frac{1}{2}}$ is divergent for $|x| \to \infty$, so that $u(x)\,V(m,x) \propto |x|^{-1+i\omega\tau_\mathrm{m}}$ is due to the logarithmic divergence not integrable on $(-\infty, 0)$. The contribution of $V(m,x)$ therefore needs to vanish in order to arrive at a normalizable density. Due to the boundary conditions, we distinguish two different domains

$$q_1(x) \;=\; q_p(x) + \begin{cases} c_{1-}U(x) & \text{for } x < x_R \\ c_{1+}U(x) + c_{2+}V(x) & \text{for } x_R \le x < x_\theta. \end{cases} \tag{A.21}$$

In the following we skip the dependence of the parabolic cylinder function on $m$ for brevity of the notation. The homogeneous solution (i.e. the coefficients $c_{1,2\pm}$) adjusts the complete solution $q_1 = q_p + q_h$ to the boundary conditions dictated by the physics of the problem.

### A.3.3 Boundary condition for the modulated density

The complete solution $q_1 = q_h + q_p$ of Eq. (A.17) must fulfill the white noise boundary conditions

$$\begin{aligned} q_1(x_\theta) &= 0 \quad \text{at threshold} \\ q_1(x_{R+}) - q_1(x_{R-}) &= 0 \quad \text{at reset}\ , \end{aligned} \tag{A.22}$$

whereby it must vanish at threshold to ensure a finite probability flux and be continuous at reset for the same reason. Introducing the short hand

$$f(x)|_{x'} = \begin{cases} f(x') & \text{for } x' = x_\theta \\ f(x'+) - f(x'-) & \text{for } x' = x_R \end{cases} \tag{A.23}$$

we state these two conditions compactly as

$$q_1(x)|_{\{x_R,x_\theta\}} = 0.$$

To determine the boundary values of the homogeneous solution we need the boundary values of the particular solution first. The latter follow with the stationary flux Eq. (2.41)

$$a^\dagger q_0\Big|_{\{x_R,x_\theta\}} = (x-a)\, q_0|_{\{x_R,x_\theta\}} = u^{-1}(\{x_R,x_\theta\})\,\tau_\mathrm{m}\nu_0, \tag{A.24}$$

where the term $xq_0$ vanishes because of the continuity of $q_0$. Along the same lines follows the term proportional to $(a^\dagger)^2 q_0$

$$\begin{aligned} (a^\dagger)^2 q_0 &= (x-a)a^\dagger q_0 \\ &= (xa^\dagger - [a,a^\dagger] - a^\dagger a)q_0 \\ &= (xa^\dagger - 1)q_0. \end{aligned}$$

With Eq. (A.24) and the continuity of $q_0$ we therefore have

$$(a^\dagger)^2 q_0\Big|_{\{x_R,x_\theta\}} = \{x_R,x_\theta\}u^{-1}(\{x_R,x_\theta\})\,\tau_\mathrm{m}\nu_0. \tag{A.25}$$

From the explicit expression of the particular solution Eq. (A.18) with the term proportional to $a^\dagger q_0$ specified by Eq. (A.24), the term proportional to $(a^\dagger)^2 q_0$ by Eq. (A.25) and the continuity of the complete solution Eq. (A.22) then follows the initial value for the homogeneous solution as

$$\begin{aligned} -q_h|_{\{x_R,x_\theta\}} &= q_p|_{\{x_R,x_\theta\}} \\ &= \left(\frac{G}{1+i\omega\tau_\mathrm{m}} + \frac{H}{2+i\omega\tau_\mathrm{m}}\{x_R,x_\theta\}\right)\tau_\mathrm{m}\nu_0\, u^{-1}(\{x_R,x_\theta\}). \end{aligned} \tag{A.26}$$

### A.3.4 Boundary condition for the derivative of the density

The boundary condition for the first derivative of $q_1$ follows considering the probability flux: The flux at threshold must be equal to the flux re-inserted at reset. Given the firing rate follows the periodic modulation $\nu_0(t) = \nu_0(1+n(\omega)e^{i\omega t})$, we can express the flux $\tau_\mathrm{m}\nu_0\, n(\omega)$ due to the perturbation (the stationary solution fulfills $S_0 u q_0|_{\{x_R,x_\theta\}} = -ua\, q_0|_{\{x_R,x_\theta\}} = \tau_\mathrm{m}\nu_0$)

as a sum of two contributions, corresponding to the first two terms in Eq. (A.14)

$$\begin{aligned} \tau_\mathrm{m}\nu_0\, n(\omega) &= S_0 u q_1 + S_1 u q_0|_{\{x_R,x_\theta\}} \\ &= -ua\, q_1 + u(G + Ha^\dagger)q_0\Big|_{\{x_R,x_\theta\}}. \end{aligned} \tag{A.27}$$

Again we first evaluate the contribution of the particular solution Eq. (A.18) considering

$$\begin{aligned} a\, a^\dagger q_0 &= (1 + a^\dagger a)q_0 \\ &= q_0 + \underbrace{a^\dagger a q_0}_{=0}. \end{aligned}$$

Analogously follows

$$a\,(a^\dagger)^2 q_0 = 2a^\dagger q_0,$$

so that the flux due to the particular solution Eq. (A.18) can be written as

$$-ua\, q_p = -u\left(\frac{G}{1 + i\omega\tau_\mathrm{m}} + \frac{2H}{2 + i\omega\tau_\mathrm{m}}\, a^\dagger\right) q_0. \tag{A.28}$$

As the stationary solution vanishes at threshold $q_0(x_\theta) = 0$ and is continuous at reset, the first term vanishes when inserted into Eq. (A.27). Hence with Eq. (A.24) the contribution to the flux Eq. (A.27) yields

$$-ua\, q_p|_{\{x_R,x_\theta\}} = -\frac{2H\tau_\mathrm{m}\nu_0}{2 + i\omega\tau_\mathrm{m}}.$$

With Eq. (A.24) the term due to the perturbed flux operator $S_1$ in Eq. (A.27) is

$$u\,(G + Ha^\dagger)q_0\Big|_{\{x_R,x_\theta\}} = H\,\tau_\mathrm{m}\nu_0.$$

Inserting the previous two expressions into Eq. (A.27) we obtain

$$\begin{aligned} \tau_\mathrm{m}\nu_0\, n(\omega) &= \tau_\mathrm{m}\nu_0 \frac{i\omega\tau_\mathrm{m}H}{2 + i\omega\tau_\mathrm{m}} - uaq_h|_{\{x_R,x_\theta\}} \\ \Leftrightarrow\ u(\frac{1}{2}x + \partial_x)q_h\Big|_{\{x_R,x_\theta\}} &= \tau_\mathrm{m}\nu_0 \left(\frac{i\omega\tau_\mathrm{m}H}{2 + i\omega\tau_\mathrm{m}} - n(\omega)\right), \end{aligned} \tag{A.29}$$

where we used the explicit form of $a = \frac{1}{2}x + \partial_x$. The derivative then follows as

$$\partial_x q_h|_{\{x_R,x_\theta\}} = \tau_\mathrm{m}\nu_0 \left(\frac{i\omega\tau_\mathrm{m}H}{2 + i\omega\tau_\mathrm{m}} - n(\omega)\right) u^{-1}(\{x_R, x_\theta\}) - \frac{1}{2}xq_h(x)\Big|_{\{x_R,x_\theta\}} \tag{A.30}$$

and with Eq. (A.26) we obtain

$$\begin{aligned}\partial_x q_h|_{\{x_R,x_\theta\}} &= \tau_\mathrm{m}\nu_0\left(\frac{i\omega\tau_\mathrm{m}H}{2+i\omega\tau_\mathrm{m}} - n(\omega) + \frac{1}{2}\{x_R,x_\theta\}\left(\frac{G}{1+i\omega\tau_\mathrm{m}} + \frac{H}{2+i\omega\tau_\mathrm{m}}\{x_R,x_\theta\}\right)\right)\\ &\times\ u^{-1}(\{x_R,x_\theta\}) \qquad (A.31)\\ &= \tau_\mathrm{m}\nu_0\left(-n(\omega) + \frac{1}{2}\{x_R,x_\theta\}\frac{G}{1+i\omega\tau_\mathrm{m}} + \left(\frac{1}{2}\{x_R,x_\theta\}^2 + i\omega\tau_\mathrm{m}\right)\frac{H}{2+i\omega\tau_\mathrm{m}}\right)\\ &\times\ u^{-1}(\{x_R,x_\theta\}).\end{aligned}$$

### A.3.5 Transfer function

Having found the function value and the derivative at threshold, the homogeneous solution (of the second order differential equation) is uniquely determined on $x_R < x < x_\theta$. Writing the solution on this interval as

$$\begin{aligned}U(x_\theta)\quad c_{1+} \quad+\quad V(x_\theta)\quad c_{2+} \quad=\quad q_1^h(x_\theta)\\ U'(x_\theta)\quad c_{1+} \quad+\quad V'(x_\theta)\quad c_{2+} \quad=\quad \partial_x q_1^h(x_\theta),\end{aligned}$$

the coefficients follow as the solution of this linear system of equations, which is in matrix form

$$\begin{pmatrix} U(x_\theta) & V(x_\theta)\\ U'(x_\theta) & V'(x_\theta)\end{pmatrix}\begin{pmatrix} c_{1+}\\ c_{2+}\end{pmatrix} = \begin{pmatrix} q_h(x_\theta)\\ \partial_x q_h(x_\theta)\end{pmatrix}.$$

The solution is

$$\begin{pmatrix} c_{1+}\\ c_{2+}\end{pmatrix} = \frac{1}{W(x_\theta)}\begin{pmatrix} V' & -V\\ -U' & U\end{pmatrix}\begin{pmatrix} q_1^h(x_\theta)\\ \partial_x q_1^h(x_\theta)\end{pmatrix} \qquad (A.32)$$

with

$$W = \det\begin{pmatrix} U & V\\ U' & V'\end{pmatrix},$$

where the function $W(x)$ is the Wronskian and for the given functions $U, V$ is a constant $W = \sqrt{\frac{2}{\pi}}$ (Abramowitz & Stegun, 1974, 19.4.1). The coefficients follow from the previous expression using Eq. (A.26) and Eq. (A.31) and $c_{2+}$ is hence

$$\begin{aligned}c_{2+} &= \sqrt{\frac{\pi}{2}}u^{-1}(x_\theta)\tau_\mathrm{m}\nu_0\left(U'(x_\theta)\left(\frac{G}{1+i\omega\tau_\mathrm{m}} + x_\theta\frac{H}{2+i\omega\tau_\mathrm{m}}\right)\right. \qquad (A.33)\\ &\left. +U(x_\theta)\left(-n(\omega) + \left(\frac{1}{2}x_\theta\frac{G}{1+i\omega\tau_\mathrm{m}} + (\frac{1}{2}x_\theta^2 + i\omega\tau_\mathrm{m})\frac{H}{2+i\omega\tau_\mathrm{m}}\right)\right)\right).\end{aligned}$$

An analog expression holds for $c_{1+}$, which is, however, not needed in the following, because we just need a condition for the solvability. As the function $V$ is absent in the lower interval $x < x_R$, the boundary condition at reset also determines $c_{2+}$, as seen in the following. Expressing the solution in terms of $U$ and $V$ and subtracting the solutions above and below $x_R$, leads to

the linear system of equations

$$\begin{array}{rcrcrcl} U(x_R) & (c_{1+}-c_{1-}) & + & V(x_R) & c_{2+} & = & q_1^h\Big|_{x_R} \\ U'(x_R) & (c_{1+}-c_{1-}) & + & V'(x_R) & c_{2+} & = & \partial_x q_1^h\Big|_{x_R}. \end{array}$$

The coefficients $c_{1+}-c_{1-}$ and $c_{2+}$ are determined as above as the solution of this system of linear equations

$$\begin{pmatrix} c_{1+}-c_{1-} \\ c_{2+} \end{pmatrix} = \frac{1}{W(x_R)} \begin{pmatrix} V' & -V \\ -U' & U \end{pmatrix} \begin{pmatrix} q_1^h\Big|_{x_R} \\ \partial_x\, q_1^h\Big|_{x_R} \end{pmatrix}. \tag{A.34}$$

Using the Wronskian $W(x_R) = \sqrt{\frac{2}{\pi}}$ and the expressions Eq. (A.26) and Eq. (A.31) for the boundary values we obtain

$$\begin{aligned} c_{2+} &= \sqrt{\frac{\pi}{2}} u^{-1}(x_R)\tau_{\mathrm{m}}\nu_0 \left( U'(x_R) \left( \frac{G}{1+i\omega\tau_{\mathrm{m}}} + x_R \frac{H}{2+i\omega\tau_{\mathrm{m}}} \right) \right. \\ &\quad \left. + U(x_R) \left( -n(\omega) + \left( \frac{1}{2} x_R \frac{G}{1+i\omega\tau_{\mathrm{m}}} + (\frac{1}{2}x_R + i\omega\tau_{\mathrm{m}}) \frac{H}{2+i\omega\tau_{\mathrm{m}}} \right) \right) \right). \end{aligned} \tag{A.35}$$

Equating Eq. (A.33) and Eq. (A.35) determines the transfer function

$$\begin{aligned} n(\omega) &= \frac{G}{1+i\omega\tau_{\mathrm{m}}} \frac{u^{-1}\ (U' + \frac{x}{2}U)\big|_{x_\theta}^{x_R}}{u^{-1}U\big|_{x_\theta}^{x_R}} \\ &\quad + \frac{H}{2+i\omega\tau_{\mathrm{m}}} \frac{u^{-1}\ \left( x\,U' + (\frac{x^2}{2} + i\omega\tau_{\mathrm{m}})\,U \right)\Big|_{x_\theta}^{x_R}}{u^{-1}U\big|_{x_\theta}^{x_R}}. \end{aligned} \tag{A.36}$$

With the definition (using $m = i\omega\tau_{\mathrm{m}} - \frac{1}{2}$ as defined in Eq. (A.20))

$$\Phi_\omega(x) \equiv \Phi(m,x) = u^{-1}(x)\,U(m,x) \tag{A.37}$$

follows

$$\Phi'_\omega(x) = u^{-1}(x)(U'(m,x) + \frac{x}{2}U(m,x)) \tag{A.38}$$

and

$$\begin{aligned} \Phi''_\omega(x) &= \frac{x}{2}u^{-1}(U' + \frac{x}{2}U) + u^{-1}(U'' + \frac{1}{2}U + \frac{x}{2}U') \\ &= \frac{x}{2}u^{-1}(U' + \frac{x}{2}U) + u^{-1}((\frac{1}{4}x^2 + i\omega\tau_{\mathrm{m}} - \frac{1}{2})U + \frac{1}{2}U + \frac{x}{2}U') \\ &= u^{-1}\,(xU' + (\frac{x^2}{2} + i\omega\tau_{\mathrm{m}})U). \end{aligned}$$

Inserting into Eq. (A.36) we obtain the final result

$$\begin{aligned} n(\omega) &= \frac{G}{1+i\omega\tau_{\rm m}} \frac{\Phi'_\omega(x)|_{x_\theta}^{x_R}}{\Phi_\omega(x)|_{x_\theta}^{x_R}} \\ &+ H\left(\frac{1}{2+i\omega\tau_{\rm m}} \frac{\Phi''_\omega(x)|_{x_\theta}^{x_R}}{\Phi_\omega(x)|_{x_\theta}^{x_R}}\right). \end{aligned}$$

This is the white-noise transfer function of the LIF model neuron (Brunel & Hakim (1999); Lindner & Schimansky-Geier (2001); Brunel et al. (2001)).

## A.4 Modulation of noise amplitude

We here show that the reduction from the colored-noise system to the white-noise system presented in Sec. 2.2 can be performed analogously in the case of a temporally modulated noise. To this end we start with the system of equations

$$\begin{aligned} \frac{dy}{ds} &= f(y,s) + \frac{z}{k} \\ k\frac{dz}{ds} &= -\frac{z}{k} + \eta(s)\xi, \end{aligned}$$

where $\eta(s)$ is a time-dependent positive ($\eta(s) > 0 \quad \forall s$) prefactor modulating the amplitude of the noise. Rescaling both coordinates with the amplitude of the noise $\tilde{y}(s) = y(s)/\eta(s)$ and $\tilde{z}(s) = z(s)/\eta(s)$, the temporal derivatives transform to $\frac{dy}{ds} = \frac{d}{ds}(\tilde{y}\eta) = \eta\frac{d\tilde{y}}{ds} + \tilde{y}\frac{d\eta}{ds}$ and $\frac{dz}{ds} = \frac{d}{ds}(\tilde{z}\eta) = \eta\frac{d\tilde{z}}{ds} + \tilde{z}\frac{d\eta}{ds}$. After division by $\eta(s)$ and using $\frac{1}{\eta}\frac{d\eta}{ds} = \frac{d\ln\eta}{ds}$, we get

$$\begin{aligned} \frac{d\tilde{y}}{ds} &= \underbrace{f(\eta\tilde{y},s) - \frac{d\ln\eta}{ds}\tilde{y}}_{\equiv\tilde{f}(\tilde{y},s)} + \frac{\tilde{z}}{k} \\ k\frac{d\tilde{z}}{ds} &= -\frac{\tilde{z}}{k} - k\frac{d\ln\eta}{ds}\,\tilde{z} + \xi, \end{aligned}$$

where we defined the function $\tilde{f}$ appearing in the modified drift term for $\tilde{y}$. The corresponding Fokker-Planck equation (cf. Eq. (2.2) for case with noise of constant amplitude) is

$$\begin{aligned} k^2\partial_s P = \partial_{\tilde{z}}\left(\frac{1}{2}\partial_{\tilde{z}} + \tilde{z} + k^2\,\tilde{z}\,\frac{d\ln\eta}{ds}\right) P - k^2\partial_{\tilde{y}}\tilde{S}_{\tilde{y}}\,P \\ \tilde{S}_{\tilde{y}} = \tilde{f}(\tilde{y},s) + \frac{\tilde{z}}{k}. \end{aligned} \tag{A.39}$$

By comparing to Eq. (2.2) we observe that the only formal difference is the additional term $k^2\,\tilde{z}\,\frac{d\ln\eta}{ds}$. Factoring-off the stationary solution of the fast part $P = Q\,\frac{e^{-\tilde{z}^2}}{\sqrt{\pi}}$ with

$$\begin{aligned} & \partial_{\tilde{z}}\left(\tilde{z}\frac{d\ln\eta}{ds}\frac{e^{-\tilde{z}^2}}{\sqrt{\pi}}Q\right) \\ = \quad & \frac{e^{-\tilde{z}^2}}{\sqrt{\pi}}\,\frac{d\ln\eta}{ds}\left(1-2\tilde{z}^2+\tilde{z}\partial_{\tilde{z}}\right)Q \end{aligned}$$

we obtain (cf. Eq. (2.5))

$$\begin{aligned} k^2\partial_s Q \quad = \quad & LQ - k\tilde{z}\partial_{\tilde{y}}Q \\ & -k^2\left(\partial_{\tilde{y}}\,\tilde{f}(\tilde{y},s) - \frac{d\ln\eta}{ds}\left(1-2\tilde{z}^2+\tilde{z}\partial_{\tilde{z}}\right)\right)Q. \end{aligned} \tag{A.40}$$

The additional term affects the perturbation expansion Eq. (2.7) at second order. Since $Q^{(0)}$ is independent of $\tilde{z}$ we have $\partial_{\tilde{z}}Q^{(0)} = 0$ which leads to

$$\begin{aligned} LQ^{(2)} \quad = \quad & \partial_s Q^{(0)} + \tilde{z}\partial_y Q^{(1)} \\ & + \left(\partial_{\tilde{y}}\,\tilde{f}(\tilde{y},s) - \frac{d\ln\eta}{ds}\left(1-2\tilde{z}^2\right)\right)Q^{(0)}. \end{aligned}$$

Due to $L\tilde{z}^2 = 1 - 2\tilde{z}^2$, the additional term in the particular solution of the second order is $-\frac{d\ln\eta}{ds}\tilde{z}^2$. This term does not contribute to the first order correction of the marginalized probability flux in $\tilde{y}$-direction $\nu_{\tilde{y}}(\tilde{y},s) \equiv \int d\tilde{z}\,\frac{e^{-\tilde{z}^2}}{\sqrt{\pi}}S_{\tilde{y}}Q(\tilde{y},\tilde{z},s) = \sum_{n=0}^{1} k^n\,\nu_{\tilde{y}}^{(n)}(\tilde{y},s) + O(k^2)$, because the factor $\frac{\tilde{z}}{k}$ in $\tilde{S}_{\tilde{y}}$ causes a point-symmetric function in $\tilde{z}$ that vanishes after integration over $\tilde{z}$ while the factor $\tilde{f}(\tilde{y})$ yields a contribution that is second order in $k$. Therefore the effective flux has the same form as before Eq. (2.10) meaning that the one-dimensional Fokker-Planck equation Eq. (2.11) with $f$ replaced with $\tilde{f}(\tilde{y},s)$

$$\partial_s\tilde{P} = \partial_{\tilde{y}}\left(-\tilde{f}(\tilde{y},s) + \frac{1}{2}\partial_{\tilde{y}}\right)\,\tilde{P}$$

holds for the case of time-modulated noise. The corresponding SDE is

$$\frac{d\tilde{y}}{ds} = \tilde{f}(\tilde{y},s) + \xi.$$

Transforming back to the original coordinate $y = \eta\tilde{y}$ we get

$$\frac{dy}{ds} = f(y,s) + \eta(s)\xi,$$

so that we obtain a SDE with time modulated white noise $\eta(s)\xi$.

To obtain the boundary condition, we follow the same calculation as in Sec. 2.2 and transform the FPE Eq. (A.40) to the shifted and scaled coordinates $\tilde{r} = \frac{\tilde{y}-\theta}{k}$, and $\tilde{z} + kf(\theta,s) \to \tilde{z}$ which

yields

$$\begin{aligned} k^2\partial_s Q^B = {} & LQ^B - \tilde{z}\partial_{\tilde{r}} Q^B \\ & + G(\theta,\tilde{r},s,\tilde{z})\, k\, Q^B \\ & + k^2 \frac{d\ln\eta}{ds}\left(1 - 2\tilde{z}^2 + \tilde{z}\partial_{\tilde{z}}\right) Q^B \\ & + O(k^3), \end{aligned} \tag{A.41}$$

with $Q^B(\tilde{r},\tilde{z},s) \equiv Q(\tilde{y}(\tilde{r}),\tilde{z},s)$ and the operator $G(\theta,\tilde{r},s,\tilde{z}) = f(\theta,s)\partial_{\tilde{z}} - \partial_{\tilde{r}}\left(f(k\tilde{r}+\theta,s) - f(\theta,s)\right)$. Since the additional term $k^2\frac{d\ln\eta}{ds}\left(1-2\tilde{z}^2+\tilde{z}\partial_{\tilde{z}}\right)$ is of second order in $k$ it only affects the perturbative solution of the second order in Eq. (2.19), so the effective boundary conditions Eq. (2.28) and Eq. (2.29) also hold in the case of time-dependent noise. The reduction of the colored-noise to the corresponding white-noise problem can therefore be performed in the same way as for constant noise.

# Appendix B

# Supplementary calculations for dynamical mean-field theory

## B.1 Derivation of mean-field equation

The generating functional $Z[\mathbf{l}](\mathbf{J})$ is properly normalized independent of the realization of $\mathbf{J}$. This property allows us to follow De Dominicis & Peliti (1978) and to introduce the disorder-averaged generating functional

$$\begin{aligned} \bar{Z}[\mathbf{l}] &:= \langle Z[\mathbf{l}](\mathbf{J})\rangle_{\mathbf{J}} \\ &= \int \Pi_{ij} dJ_{ij}\, \mathcal{N}(0, \frac{g^2}{N}, J_{ij})\, Z[\mathbf{l}](\mathbf{J}). \end{aligned} \tag{B.1}$$

The coupling term $\exp\left(-\sum_{i\neq j} J_{ij}\tilde{x}_i^{\mathbf{T}}\phi(x_j)\right)$ in Eq. (7.2) factorizes and the random weights $J_{ij}$ appear linear in the exponent. Thus we can separately integrate over the independently distributed $J_{ij}$, $i \neq j$, by completing the square and obtain

$$\begin{aligned} &\int dJ_{ij}\mathcal{N}(0, \frac{g^2}{N}, J_{ij})\, \exp\left(-J_{ij}\tilde{x}_i^{\mathbf{T}}\phi(x_j)\right) \\ = \;& \exp\left(\frac{g^2}{2N}\left(\tilde{x}_i^{\mathbf{T}}\phi(x_j)\right)^2\right). \end{aligned} \tag{B.2}$$

We reorganize the resulting sum in the exponent of the coupling term as

$$\begin{aligned}
&\frac{g^2}{2N}\sum_{i\neq j}\left(\int \tilde{x}_i(t)\phi(x_j(t))\,dt\right)^2 \\
=\;& \frac{g^2}{2N}\sum_{i\neq j}\int\int \tilde{x}_i(t)\phi(x_j(t))\,\tilde{x}_i(t')\phi(x_j(t'))\,dt\,dt' \\
=\;& \frac{1}{2}\sum_i \int\int \tilde{x}_i(t)\tilde{x}_i(t')\,\left(\frac{g^2}{N}\sum_j \phi(x_j(t))\phi(x_j(t'))\right)\,dt\,dt' \\
&-\frac{1}{2}\sum_i\int\int \tilde{x}_i(t)\tilde{x}_i(t')\,\frac{g^2}{N}\phi(x_i(t))\phi(x_i(t'))\,dt\,dt',
\end{aligned} \tag{B.3}$$

where we used $\left(\int f(t)dt\right)^2 = \int\int f(t)f(t')\,dt\,dt'$ in the first step and $\sum_{ij} x_i y_j = \sum_i x_i \sum_j y_j$ in the second. The second term is the diagonal to be skipped in the double sum. It is a correction of order $N^{-1}$ and will be neglected in the following. The disorder-averaged generating functional Eq. (B.1) therefore takes the form

$$\begin{aligned}
\bar{Z}[\mathbf{l}] \;=\; & \int \mathbf{D}\mathbf{x}\int \mathbf{D}\tilde{\mathbf{x}}\,\exp\left(S_0[\mathbf{x},\tilde{\mathbf{x}}]+\mathbf{l}^{\mathbf{T}}\mathbf{x}\right) \\
& \times\exp\left(\frac{1}{2}\tilde{\mathbf{x}}^{\mathbf{T}}\,Q_1\tilde{\mathbf{x}}\right),
\end{aligned} \tag{B.4}$$

where we extended our notation with $x^{\mathbf{T}}Ay := \iint x(t)\,A(t,t')\,y(t')\,dt\,dt'$ and defined

$$Q_1(t,t') := \frac{g^2}{N}\sum_j \phi(x_j(t))\phi(x_j(t'))\,. \tag{B.5}$$

The field $Q_1$ is an empirical average over $N$ contributions, which in the case of weak correlations will converge to its expectation value for large $N$ by the law of large numbers. This heuristic argument is shown in the following in a formal setting: A saddle-point approximation leads to the replacement of $Q_1$ by its expectation value. We first decouple the interaction term by inserting the Fourier representation of the Dirac-$\delta$ functional:

$$\begin{aligned}
&\delta[-\frac{N}{g^2}Q_1(s,t)+\phi(\mathbf{x}(s))^{\mathbf{T}}\,\phi(\mathbf{x}(t))] \\
=&\int \mathbf{D}Q_2\,\exp\left(-\frac{N}{g^2}Q_1^{\mathbf{T}}Q_2+\phi(\mathbf{x})^{\mathbf{T}}Q_2\phi(\mathbf{x})\right),
\end{aligned} \tag{B.6}$$

where we further extended our notation with $A^{\mathbf{T}}B := \iint\, A(s,t)\,B(s,t)\,ds\,dt$ and $\phi(\mathbf{x}(s))^{\mathbf{T}}\phi(\mathbf{x}(t)) = \sum_{i=1}^N \phi(x_i(s))\phi(x_i(t))$. We note that the conjugate field $Q_2 \in i\mathbb{R}$ is purely imaginary. We hence

rewrite Eq. (B.4) as

$$\begin{aligned}\bar{Z}[j,\tilde{j}] &= \int \mathbf{D}Q_1 \int \mathbf{D}Q_2 \\ &\quad \exp\left(-\frac{N}{g^2}Q_1^{\mathbf{T}}Q_2 + N\,\ln\,Z[Q_1,Q_2] + j^{\mathbf{T}}Q_1 + \tilde{j}^{\mathbf{T}}Q_2\right) \\ Z[Q_1,Q_2] &= \int \mathbf{D}x \int \mathbf{D}\tilde{x}\,\exp\left(S_0[x,\tilde{x}] + \frac{1}{2}\tilde{x}^{\mathbf{T}}Q_1\tilde{x} + \phi(x)^{\mathbf{T}}Q_2\phi(x)\right),\end{aligned} \tag{B.7}$$

where we introduced source terms $j, \tilde{j}$ to derive self-consistency equations for the auxiliary fields. The integral measures $\mathbf{D}Q_{1,2}$ must be defined suitably. In writing $N\,\ln\,Z[Q_1,Q_2]$ we have used that the auxiliary fields couple only to sums of fields $\sum_i \phi^2(x_i)$ and $\sum_i \tilde{x}_i^2$, so that the generating functional for the fields $\mathbf{x}$ and $\tilde{\mathbf{x}}$ factorizes into a product of $N$ factors $Z[Q_1,Q_2]$.

The remaining problem can be considered a field theory for the auxiliary fields $Q_1$ and $Q_2$. The form Eq. (B.7) clearly exposes the $N$ dependence of the action for these latter fields in Eq. (B.7): It is of the form $\int dQ \exp(Nf(Q))$, which, for large $N$, suggests a saddle point approximation, which neglectings fluctuations in the auxiliary fields and hence sets them equal to their expectation value; this point is the dominant contribution to the probability mass. To obtain the saddle point equations we consider the Legendre-Fenchel transform of $\ln\,\bar{Z}$ as

$$\Gamma(q_1,q_2) := \sup_{j,\tilde{j}}\; j^{\mathbf{T}}q_1 + \tilde{j}^{\mathbf{T}}q_2 - \ln\bar{Z}[j,\tilde{j}],$$

called the vertex generating functional or effective action (Zinn-Justin, 1996; Negele & Orland, 1998). It holds that $\frac{\delta\Gamma}{\delta q_1} = j$ and $\frac{\delta\Gamma}{\delta q_2} = \tilde{j}$, called equations of state. The leading order mean-field approximation amounts to the approximation $\Gamma[q_1,q_2] \simeq -S[q_1,q_2]$, where $S[Q_1,Q_2] = -\frac{N}{g^2}Q_1^{\mathbf{T}}Q_2 + N\,\ln\,Z[Q_1,Q_2]$ is the action for the auxiliary fields $Q_1$ and $Q_2$. We insert this tree-level approximation into the equations of state and further set $j = \tilde{j} = 0$ since the source fields have no physical meaning and thus must vanish. We get the saddle point equations

$$\begin{aligned}0 &= \frac{\delta S[Q_1,Q_2]}{\delta Q_{\{1,2\}}} \\ &= \frac{\delta}{\delta Q_{\{1,2\}}}\left(-\frac{N}{g^2}Q_1^{\mathbf{T}}Q_2 + N\,\ln Z[Q_1,Q_2]\right)\end{aligned} \tag{B.8}$$

from which we obtain a pair of equations

$$\begin{aligned}0 &= -\frac{N}{g^2}\,Q_1^*(s,t) + \frac{N}{Z}\,\frac{\delta Z[Q_1,Q_2]}{\delta Q_2(s,t)}\bigg|_{Q^*} \\ \leftrightarrow Q_1^*(s,t) &= g^2\,\langle\phi(x(s))\phi(x(t))\rangle_{Q^*} =: g^2 C_{\phi(x)\phi(x)}(s,t) \\ 0 &= -\frac{N}{g^2}\,Q_2^*(s,t) + \frac{N}{Z}\,\frac{\delta Z[Q_1,Q_2]}{\delta Q_1(s,t)}\bigg|_{Q^*} \\ \leftrightarrow Q_2^*(s,t) &= \frac{g^2}{2}\langle\tilde{x}(s)\tilde{x}(t)\rangle_{Q^*} = 0,\end{aligned} \tag{B.9}$$

where we defined the average autocorrelation function $C_{\phi(x)\phi(x)}(s,t)$ of the non-linearly transformed activity of the units. The second saddle point $Q_2^* = 0$ vanishes, because it was introduced to represent a Dirac $\delta$ constraint in Fourier domain. One can show that consequently $\int \mathbf{D}Q \, \exp(S[Q_1, Q_2])Q_2 = 0$, which is the true mean value $Q_2^* = \langle Q_2 \rangle = 0$.

Here, $\langle\rangle_{Q^*}$ denotes the expectation value with respect to $x$ evaluated at the saddle point $Q^*$. The expectation value must be computed self-consistently, since the values of the saddle points, by Eq. (B.7), influence the statistics of the field $\mathbf{x}$, which in turn determines the function $Q_1^*$ by Eq. (B.9).

Inserting the saddle point solution into the generating functional Eq. (B.7) we get

$$\bar{Z}^* \propto \int \mathbf{D}x \int \mathbf{D}\tilde{x} \, \exp\left(S_0[x,\tilde{x}] + \frac{g^2}{2}\tilde{x}^{\mathbf{T}} C_{\phi(x)\phi(x)}\tilde{x}\right). \tag{B.10}$$

The action has the important property that it decomposes into a sum of actions for individual, non-interacting units that feel a common field with self-consistently determined statistics, characterized by its second cumulant $C_{\phi(x)\phi(x)}$. Prior to the saddle point approximation (Eq. (B.7)) the fluctuations in the field $Q_1$ are common to all the single units, which effectively couples them. The saddle-point approximation replaces the fluctuating field $Q_1$ by its mean (Eq. (B.9)) which reduces the network to $N$ non-interacting units, or, equivalently, a single unit system. The second term in Eq. (B.10) is a Gaussian noise with a two point correlation function $C_{\phi(x)\phi(x)}(s,t)$. The physical interpretation is the noisy signal each unit receives due to the input from the other $N$ units. Its autocorrelation function is given by the summed autocorrelation functions of the output activities $\phi(x_i(t))$ weighted by $g^2 N^{-1}$, which incorporates the Gaussian statistics of the couplings.

The interpretation of the noise can be appreciated by explicitly considering the moment generating functional of a Gaussian noise with a given autocorrelation function $C(s,t)$, which leads to the cumulant generating functional $\ln Z_\eta[\tilde{x}]$ that appears in the exponent of Eq. (B.10) and has the form

$$\begin{aligned} \ln Z_\eta[-\tilde{x}] &= \ln\langle \exp\left(-\tilde{x}^T \eta\right)\rangle \\ &= \frac{1}{2}\tilde{x}^{\mathbf{T}} C \, \tilde{x}. \end{aligned}$$

Note that the effective noise term only has a non-vanishing second cumulant as the cumulant generating function is quadratic. This means the effective noise is Gaussian and only couples pairs of time points in proportion to the correlation function. Finally, we can read off the mean-field equation Eq. (7.4) from Eq. (B.10).

## B.2 Stationary process

We rewrite equation Eq. (7.4) as

$$(\partial_t + 1)\, x(t) \quad = \quad \tilde{\eta}(t), \tag{B.11}$$

where we combined the two independent Gaussian processes $\eta$ and $\xi$ appearing in Eq. (7.4) into $\tilde{\eta}(t)$. We then multiply Eq. (B.11) for time points $t$ and $s$ and take the expectation value with respect to the noise $\tilde{\eta}$ on both sides, which leads to

$$(\partial_t + 1)\,(\partial_s + 1)\, C_{xx}(t,s) = g^2\, C_{\phi(x)\phi(x)}(t,s) + D\delta(t-s), \tag{B.12}$$

where we defined the covariance function of the activities $C_{xx}(t,s) := \langle x(t)x(s)\rangle$. We are now interested in the stationary statistics of the system, i.e. $C_{xx}(t,s) =: c(t-s)$. The inhomogeneity in Eq. (B.12) is then also time-translation invariant; $C_{\phi(x)\phi(x)}(t+\tau,t)$ is only a function of $\tau$. Therefore the differential operator $(\partial_t + 1)\,(\partial_s + 1)\, c(t-s)$, with $\tau = t-s$, simplifies to $(-\partial_\tau^2 + 1)\, c(\tau)$ so we get

$$(-\partial_\tau^2 + 1)\, c(\tau) \quad = \quad g^2\, C_{\phi(x)\phi(x)}(t+\tau,t) + 2\sigma^2\, \delta(\tau), \tag{B.13}$$

given as Eq. (7.6) in the main text.

## B.3 Assessing chaos by a pair of identical systems

We start from the generating functional for the pair of systems Eq. (7.13) and perform the average over realizations of the connectivity $\mathbf{J}$, as in Eq. (B.2). We therefore need to evaluate the Gaussian integral

$$\begin{aligned}
&\int dJ_{ij}\mathcal{N}(0,\frac{g^2}{N},J_{ij})\, \exp\left(-J_{ij}\sum_{\alpha=1}^{2}\tilde{x}_i^{\alpha\mathbf{T}}\phi(x_j^\alpha)\right)\\
&= \exp\left(\frac{g^2}{2N}\sum_{\alpha=1}^{2}\left(\tilde{x}_i^{\alpha\mathbf{T}}\phi(x_j^\alpha)\right)^2\right)\\
&\times \exp\left(\frac{g^2}{N}\,\tilde{x}_i^{1\mathbf{T}}\phi(x_j^1)\,\tilde{x}_i^{2\mathbf{T}}\phi(x_j^2)\right).
\end{aligned} \tag{B.14}$$

The first exponential factor only includes variables of a single subsystem and is identical to the term appearing in Eq. (B.4). The second exponential factor is a coupling term between the two systems arising from the identical coupling $\mathbf{J}$ in each realization entering the expectation value. We treat the former terms as before and here concentrate on the mixed coupling term.

Analogous to Eq. (B.3), the exponent of the mixed coupling term can be rewritten as

$$\begin{aligned}&\frac{g^2}{N}\sum_{i\neq j}\tilde{x}_i^{1\mathbf{T}}\phi(x_j^1)\,\tilde{x}_i^{2\mathbf{T}}\phi(x_j^2) \\ &=\iint\sum_i \tilde{x}_i^1(s)\tilde{x}_i^2(t)\frac{g^2}{N}\sum_j \phi(x_j^1(s))\,\phi(x_j^2(t))\,ds\,dt+O(N^{-1}),\end{aligned} \tag{B.15}$$

where we included the self coupling term $i=j$, which is again a subleading correction of order $N^{-1}$.

We now follow the steps in Sec. B.1 and introduce three pairs of auxiliary variables. The pairs $Q_1^\alpha, Q_2^\alpha$ are defined as before in Eq. (B.5) and Eq. (B.6), but for each subsystem, while the pair $T_1, T_2$ decouples the mixed term Eq. (B.15) by defining

$$T_1(s,t) := \frac{g^2}{N}\sum_j \phi(x_j^1(s))\,\phi(x_j^2(t)).$$

Taken together, we can therefore rewrite the generating functional Eq. (7.13) averaged over the couplings as

$$\begin{aligned}\bar{Z}[\{l^\alpha\}_{\alpha\in\{1,2\}}] &:= \langle Z[\{l^\alpha\}_{\alpha\in\{1,2\}}](\mathbf{J})\rangle_{\mathbf{J}} \\ &= \Pi_{\alpha=1}^2\left\{\int \mathbf{D}Q_1^\alpha\int\mathbf{D}Q_2^\alpha\right\}\int\mathbf{D}T_1\int\mathbf{D}T_2\,\exp\left(\Omega[\{Q_1^\alpha,Q_2^\alpha\}_{\alpha\in\{1,2\}},T_1,T_2]\right)\end{aligned} \tag{B.16}$$

$$\Omega[\{Q_1^\alpha,Q_2^\alpha\}_{\alpha\in\{1,2\}},T_1,T_2] := -\sum_{\alpha=1}^2\frac{N}{g^2}Q_1^{\alpha\mathbf{T}}Q_2^\alpha-\frac{N}{g^2}T_1^{\mathbf{T}}T_2+N\,\ln\,Z^{12}[\{Q_1^\alpha,Q_2^\alpha\}_{\alpha\in\{1,2\}},T_1,T_2]$$

$$\begin{aligned}Z^{12}[\{Q_1^\alpha,Q_2^\alpha\}_{\alpha\in\{1,2\}},T_1,T_2] &= \Pi_{\alpha=1}^2\Big\{\int\mathbf{D}x^\alpha\int\mathbf{D}\tilde{x}^\alpha \\ &\quad\times\exp\left(S_0[x^\alpha,\tilde{x}^\alpha]+l^{\alpha\mathbf{T}}x^\alpha+\frac{1}{2}\tilde{x}^{\alpha\mathbf{T}}Q_1^\alpha\tilde{x}^\alpha+\phi(x^\alpha)^{\mathbf{T}}Q_2^\alpha\phi(x^\alpha)\right)\Big\} \\ &\quad\times\exp\left(\tilde{x}^{1\mathbf{T}}\left(T_1+2\sigma^2\right)\tilde{x}^2+\phi(x^1)^{\mathbf{T}}T_2\phi(x^2)\right),\end{aligned}$$

where we used that the generating functional factorizes into a product of $2N$ factors.

Analogously to Sec. B.1 we could introduce sources for the auxiliary fields $Q_1^\alpha$, $Q_2^\alpha$, $T_1$, $T_2$. Then the equations of state are obtained from the vertex-generating functional $\Gamma$ as before, which, in the tree-level approximation $\Gamma=-S$ and for vanishing sources leads to the saddle-point equations $\frac{\delta\Omega}{\delta Q_{1,2}^\alpha}=\frac{\delta\Omega}{\delta T_{1,2}}\stackrel{!}{=}0$. From the latter we obtain the set of equations

$$
\begin{aligned}
Q_1^{\alpha *}(s,t) &= g^2 \, \frac{1}{Z^{12}} \, \frac{\delta Z^{12}}{\delta Q_2^{\alpha}(s,t)} = g^2 \, \langle \phi(x^{\alpha}(s)) \phi(x^{\alpha}(t)) \rangle_{Q^*,T^*} \\
Q_2^{\alpha *}(s,t) &= 0 \\
T_1^{*}(s,t) &= g^2 \, \frac{1}{Z^{12}} \, \frac{\delta Z^{12}}{\delta T_2(s,t)} = g^2 \, \langle \phi(x^{1}(s)) \phi(x^{2}(t)) \rangle_{Q^*,T^*} \\
T_2^{*}(s,t) &= 0.
\end{aligned}
\tag{B.17}
$$

The generating functional at the saddle point is therefore

$$
\begin{aligned}
\bar{Z}^*[\{l^{\alpha}\}_{\alpha \in \{1,2\}}] = \iint \Pi_{\alpha=1}^{2} \mathbf{D}x^{\alpha} \mathbf{D}\tilde{x}^{\alpha} \, \exp\Big( \sum_{\alpha=1}^{2} S_0[x^{\alpha}, \tilde{x}^{\alpha}] + l^{\alpha \mathbf{T}} x^{\alpha} + \frac{1}{2} \tilde{x}^{\alpha \mathbf{T}} Q_1^{\alpha *} \tilde{x}^{\alpha} \Big) \\
\times \exp\left( \tilde{x}^{1\mathbf{T}} \left( T_1^* + 2\sigma^2 \right) \tilde{x}^2 \right).
\end{aligned}
\tag{B.18}
$$

We make the following observations:

1. The two subsystems $\alpha = 1, 2$ in the first line of Eq. (B.18) have the same form as in Eq. (B.10). This has been expected, because the absence of any physical coupling between the two systems implies that the marginal statistics of the activity in one system cannot be affected by the mere presence of the second, hence also their saddle points $Q_{1,2}^{\alpha}$ must be the same as in Eq. (B.10).
2. If the term in the second line of Eq. (B.18) was absent, the statistics in the two systems would be independent. Two sources, however, contribute to the correlations between the systems: The common Gaussian white noise that gives rise to the term $\propto 2\sigma^2$ and the non-white Gaussian noise due to a non-zero value of the auxiliary field $T_1^*(s,t)$.
3. Only products of pairs of fields appear in Eq. (B.18), so that the statistics of the $x^{\alpha}$ is Gaussian.

From Eq. (B.18) and Eq. (B.17) we can read off the pair of effective dynamical equation Eq. (7.14) with self-consistent statistics Eq. (7.15).

### B.3.1 Derivation of the variational equation

We multiply the equation Eq. (7.14) for $\alpha = 1$ and $\alpha = 2$ and take the expectation value on both sides, so we get for $\alpha, \beta \in \{1, 2\}$

$$
(\partial_t + 1)(\partial_s + 1) \, c^{\alpha\beta}(t,s) = D\delta(t-s) + g^2 F_{\phi}\left( c^{\alpha\beta}(t,s), c^{\alpha\alpha}(t,t), c^{\beta\beta}(s,s) \right), \tag{B.19}
$$

where the function $F_{\phi}$ is defined as the Gaussian expectation value

$$
F_{\phi}(c^{12}, c^1, c^2) := \mathrm{E}\phi(x^1)\phi(x^2)
$$

for the bi-variate Gaussian

$$\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \sim \mathcal{N}_2 \left( 0, \begin{pmatrix} c^1 & c^{12} \\ c^{12} & c^2 \end{pmatrix} \right).$$

First, we observe that the equations for the autocorrelation functions $c^{\alpha\alpha}(t,s)$ decouple and can each be solved separately, leading to the same equation Eq. (7.8) as before. This formal result could have been anticipated, because the marginal statistics of each subsystem cannot be affected by the mere presence of the respective other system. Their solutions

$$c^{11}(s,t) = c^{22}(s,t) = c(t-s)$$

then provide the "background", i.e., the second and third argument of the function $F_\phi$ on the right-hand side, for the equation for the crosscorrelation function between the two copies. Hence it remains to determine the equation of motion for $c^{12}(t,s)$.

We first determine the stationary solution $c^{12}(t,s) = k(t-s)$. We see immediately from Eq. (B.19) that $k(\tau)$ obeys the same equation of motion as $c(\tau)$, so $k(\tau) = c(\tau)$. The distance Eq. (7.12) therefore vanishes. Let us now study the stability of this solution. We hence need to expand $c^{12}$ around the stationary solution

$$c^{12}(t,s) = c(t-s) + \epsilon\, k^{(1)}(t,s)\,,\ \epsilon \ll 1\,.$$

We develop the right hand side of Eq. (B.19) into a Taylor series using Prices theorem and Eq. (7.7)

$$\begin{aligned} F_\phi\left(c^{12}(t,s), c_0, c_0\right) &= f_\phi\left(c^{12}(t,s), c_0\right) \\ &= f_\phi\left(c(t-s), c_0\right) + \epsilon\, f_{\phi'}\left(c(t-s), c_0\right)\, k^{(1)}(t,s) + O(\epsilon^2). \end{aligned}$$

Inserted into Eq. (B.19) and using that $c$ solves the lowest order equation, we get the linear equation of motion for the first order deflection Eq. (7.16). By Eq. (7.12) the first order deflection $k^{(1)}(t,s)$ is linked to the distance between the two subsystems as

$$\begin{aligned} d(t) &= \underbrace{c^{11}(t,t)}_{c_0} + \underbrace{c^{22}(s,s)}_{c_0} \underbrace{-c^{12}(t,t) - c^{21}(t,t)}_{-2c_0 - \epsilon\, k^{(1)}(t,t)} \\ &= -\epsilon\, k^{(1)}(t,t). \end{aligned} \tag{B.20}$$

The negative sign makes sense, since we expect in the chaotic state that $c^{12}(t,s) \overset{t,s\to\infty}{=} 0$, so $k^{(1)}$ must be of opposite sign than $c > 0$.

### B.3.2 Schrödinger equation for the maximum Lyapunov exponent

We here want to reformulate the equation for the variation of the cross-system correlation Eq. (7.16) into a Schrödinger equation, as in the original work (Sompolinsky et al., 1988, eq. 10).

First, noting that $C_{\phi'\phi'}(t,s) = f_{\phi'}\left(c(t-s), c_0\right)$ is time translation invariant, it is advantageous to introduce the coordinates $T = t + s$ and $\tau = t - s$ and write the covariance $k^{(1)}(t,s)$ as $k(T,\tau)$ with $k^{(1)}(t,s) = k(t+s, t-s)$. The differential operator $(\partial_t + 1)(\partial_s + 1)$ with the chain rule $\partial_t \to \partial_T + \partial_\tau$ and $\partial_s \to \partial_T - \partial_\tau$ in the new coordinates is $(\partial_T + 1)^2 - \partial_\tau^2$. A separation ansatz $k(T,\tau) = e^{\frac{1}{2}\kappa T}\,\psi(\tau)$ then yields the eigenvalue equation

$$(\frac{\kappa}{2} + 1)^2 \psi(\tau) - \partial_\tau^2 \psi(\tau) = g^2 f_{\phi'}\left(c(\tau), c_0\right) \psi(\tau)$$

for the growth rates $\kappa$ of $d(t) = -k^{(1)}(t,t) = -k(2t,0)$. We can express the right hand side by the second derivative of the potential Eq. (7.9) so that with

$$V''(c(\tau); c_0) \;=\; -1 + g^2 f_{\phi'}\left(c(\tau), c_0\right) \tag{B.21}$$

we get the time-independent Schrödinger equation

$$\left(-\partial_\tau^2 - V''(c(\tau); c_0)\right) \psi(\tau) \;=\; \underbrace{\left(1 - \left(\frac{\kappa}{2} + 1\right)^2\right)}_{=:E} \psi(\tau), \tag{B.22}$$

where the time lag $\tau$ plays the role of a spatial coordinate for the Schrödinger equation. The eigenvalues ("energies") $E_n$ determine the exponential growth rates $\kappa_n$ of the solutions $k(2t,0) = e^{\kappa_n t}\,\psi_n(0)$ at $\tau = 0$ with

$$\kappa_n^{\pm} \;=\; 2\left(-1 \pm \sqrt{1 - E_n}\right). \tag{B.23}$$

We can therefore determine the growth rate of the mean-square distance of the two subsystems by Eq. (B.20). The fastest growing mode of the distance is hence given by the ground state energy $E_0$ and the plus in Eq. (B.23). The deflection between the two subsystems therefore growth with the rate

$$\begin{aligned} \lambda_{\max} &= \frac{1}{2}\kappa_0^{+} \\ &= -1 + \sqrt{1 - E_0}, \end{aligned} \tag{B.24}$$

where the factor $1/2$ in the first line is due to $d$ being the squared distance, hence the length $\sqrt{d}$ growth with half the exponent as $d$.

# Appendix C

# Simulation of rate-based neuron models

## C.1 Numerical evaluation of the Siegert formula

We here describe how to numerically calculate Eq. (3.5), frequently called Siegert formula (for $\tau_\mathrm{s} = 0$) in the literature (for a recent textbook see Gerstner et al., 2014), which is not straight forward due to numerical instabilities in the integral. We use the abbreviations $y_\theta = (\theta - \mu)/\sigma + \gamma\sqrt{\tau_\mathrm{s}/\tau_\mathrm{m}}$ and $y_r = (V_r - \mu)/\sigma + \gamma\sqrt{\tau_\mathrm{s}/\tau_\mathrm{m}}$ and rewrite the integral as

$$\begin{aligned}
&\sqrt{\pi}\int_{y_r}^{y_\theta} e^{u^2}(1+\mathsf{erf}(u))du \\
=&2\int_{y_r}^{y_\theta} e^{u^2}\int_{-\infty}^{u} e^{-v^2}\,dv\,du \\
=&2\int_{y_r}^{y_\theta}\int_{-\infty}^{u} e^{(u+v)(u-v)}\,dv\,du\,.
\end{aligned}$$

Here, the numerical difficulty arises due to a multiplication of a divergent ($e^{u^2}$) and a convergent term ($1+\mathrm{erf}(u)$) in the integrand. We therefore use the variable transform $w = v - u$ and obtain

$$\begin{aligned}
=&2\int_{y_r}^{y_\theta}\int_{-\infty}^{u} e^{(u+v)(u-v)}\,dv\,du \\
=&2\int_{y_r}^{y_\theta}\int_{-\infty}^{0} e^{(2u+w)(-w)}\,dw\,du \\
=&\int_0^\infty e^{-w^2}\frac{e^{2y_\theta w} - e^{2y_r w}}{w}\,dw\,,
\end{aligned}$$

where we performed the integral over $u$ in the last line.

For $y_r, y_\Theta < 0$ the integrand can be integrated straightforwardly as

$$\int_0^\infty \frac{e^{2y_\Theta w - w^2} - e^{2y_r w - w^2}}{w}\,dw\,, \tag{C.1}$$

where the two terms in the integrand converge separately and where, in approximation, the upper integration bound is chosen, such that the integrand has dropped to a sufficiently small

value (here chosen to be $10^{-12}$). Here, for $w = 0$, the integrand has to be replaced by $\lim_{w\to 0} \frac{e^{2y_\Theta w} - e^{2y_r w}}{w} = 2\,(y_\Theta - y_r)$.

For $y_\theta > 0$ and $y_r < 0$ only the combination of the two terms in Eq. (C.1) converges. So we rewrite

$$\begin{aligned}
&\int_0^\infty e^{-w^2} \frac{e^{2y_\Theta w} - e^{2y_r w}}{u}\, dw \\
=&\int_0^\infty e^{2y_\theta w - w^2} \frac{1 - e^{2(y_r - y_\Theta)w}}{w}\, dw \\
=&e^{y_\Theta^2} \int_0^\infty e^{-(w-y_\Theta)^2} \frac{1 - e^{2(y_r - y_\Theta)w}}{w}\, dw\,. \qquad \text{(C.2)}
\end{aligned}$$

The integrand has a peak near $y_\Theta$. Therefore, in approximation, the lower and the upper boundary can be chosen to be left and right of $y_\Theta$, respectively, such that the integrand has fallen to a sufficiently low value (here chosen to be $10^{-12}$). For $w = 0$ we replace the integrand by its limit, which is $\lim_{w\to 0} e^{-(w-y_\Theta)^2} \frac{1-e^{2(y_r-y_\Theta)w}}{w} = e^{-y_\Theta^2} 2\,(y_\Theta - y_r)$.

We actually switch from Eq. (C.1) to Eq. (C.2) when $y_\theta > 0.05|\tilde{V}_\theta|/\sigma_\alpha$ with $\tilde{V}_\theta = V_\theta + \gamma\sqrt{\tau_\mathrm{s}/\tau_\mathrm{m}}$. This provides a numerically stable solution in terms of a continuous transition between the two expressions. Our reference implementation numerically evaluates the integrals using the adaptive GSL implementation `gsl_integration_qags` (Galassi et al., 2006) of the Gauss-Kronrod 21-point integration rule.

## C.2 Usage of the NEST reference implementation

We here give a brief description of how to use our reference implementation of the continuous-time dynamics in the simulation code NEST with the syntax of the PyNEST interface (Eppler et al., 2009). Complete simulation scripts will be made available with one of the next major releases of NEST. The description focuses on rate-neuron specific aspects. A general introduction to PyNEST can be found on the simulator website (http://www.nest-simulator.org).

C.1 shows the creation of an excitatory-inhibitory network of linear rate-model neurons as used in our example in Sec. 8.3.3.1. Researchers already familiar with the PyNEST interface notice that there is no fundamental difference to scripts for the simulation of spiking neural networks. Line 4 illustrates how to disable the usage of the waveform-relaxation method. This is advisable for simulations on local workstations or clusters, where the waveform-relaxation method does typically not improve performance (see Sec. 8.3.2). The iterative method is enabled by default, as it is also employed for simulations with gap junctions, where it improves performance and even accuracy of the simulation, regardless of network size and parallelization (Hahne et al., 2015). Instances of the linear rate-based neuron models are created by calling `nest.Create` in the usual way with model type `lin_rate_ipn`. The parameter `linear_summation` characterizes the type of nonlinearity ($\phi$ or $\psi$, see Sec. 8.2.3.2) of the neuron model. In this particular example the explicit specification of the parameter is added for

**Algorithm C.1 Simulation of an excitatory-inhibitory network of linear rate neurons.** Here and in the following scripts we use the syntax of the PyNEST interface (Eppler et al., 2009) of the NEST simulation software as of version 2.10.0 (Bos et al., 2015). The script excludes the definitions of the parameters (`h,NE,NI,mu,sigma,tau,T_start,w,d,g,KE,KI,T`).

```
1  import nest
2
3  # Disable usage of waveform relaxation method
4  nest.SetKernelStatus({'resolution': h, 'use_wfr': False})
5
6  # Create neurons and recording device
7  n_e = nest.Create('lin_rate_ipn', NE,
8                    params = {'linear_summation': True,
9                    'mean': mu, 'std': sigma, 'tau': tau})
10 n_i = nest.Create('lin_rate_ipn', NI,
11                   params = {'linear_summation': True,
12                   'mean': mu, 'std': sigma, 'tau': tau})
13 mm = nest.Create('multimeter', params = {'record_from': ['rate'
      ],
14                  'interval': h, 'start': T_start})
15
16 # Specify synapse and connection dictionaries
17 syn_e = {'weight': w, 'delay': d,
18          'model': 'delay_rate_connection'}
19 syn_i = {'weight': -g * w, 'delay': d,
20          'model': 'delay_rate_connection'}
21 conn_e = {'rule': 'fixed_outdegree', 'outdegree': KE}
22 conn_i = {'rule': 'fixed_outdegree', 'outdegree': KI}
23
24 # Connect neurons
25 nest.Connect(n_e, n_e, conn_e, syn_e)
26 nest.Connect(n_i, n_i, conn_i, syn_i)
27 nest.Connect(n_e, n_i, conn_i, syn_e)
28 nest.Connect(n_i, n_e, conn_e, syn_i)
29
30 # Connect recording device to neurons
31 nest.Connect(mm, n_e + n_i)
32
33 # Start simulation
34 nest.Simulate(T)
```

illustrative purposes only as i) the employed model is a linear rate model, where regardless of the choice of the parameter $\phi(x) = \psi(x) = x$ holds and ii) the default value is `True` anyway. Lines 13-14 and 31 demonstrate how to record the rate activity with the `multimeter`. The `record_from` parameter needs to be set to `rate` to pick up the corresponding state variable. As this particular network model includes delayed rate connections the synapse model

`delay_rate_connection` is chosen (lines 17-20). In order to create instantaneous rate connections instead one changes the synapse model to `rate_connection` and removes the parameter `delay` from the synapse dictionary. For the simultaneous use of delayed and instantaneous connections one duplicates lines 17-28 and adapts the synapse and connection dictionaries of the copy according to the needs of the additional instantaneous connections.

**Algorithm C.2 Connecting neurons of type `siegert_neuron`.** The script shows how connections between neurons of type Eq. (4.1) are created in PyNEST. Again the code snippet does not contain the definitions of the parameters (`tau_m`,`K`,`w`) for brevity.

```
1  import nest
2
3  # Create one siegert_neuron each representing
4  # ... the excitatory population
5  s_ex = nest.Create("siegert_neuron", 1)
6  # ... the inhibitory population
7  s_in = nest.Create("siegert_neuron", 1)
8
9  [...]
10
11 # Create connections originating from the excitatory neuron
12 syn_e = {'drift_factor': tau_m * K * w,
13           'diffusion_factor': tau_m * K * w * w,
14           'model': 'diffusion_connection'}
15 nest.Connect(s_ex, s_ex + s_in, 'all_to_all', syn_e)
16
17 [...]
```

C.2 shows a code snippet from a simulation script of a rate-network Eq. (3.6) with Eq. (3.7) used for mean-field analysis of complex networks in Sec. 3.1. Here single rate-model neurons of type `siegert_neuron` represent an entire population of spiking neurons (lines 3-7). The neurons are coupled by connections of type `diffusion_connection`. This connection type is identical to type `rate_connection` for instantaneous rate connections except for the two parameters `drift_factor` and `diffusion_factor` substituting the parameter `weight` (lines 12-14). These two parameters reflect the prefactors in front of the rate variable in Eq. (3.7). In general the prefactors differ from these well known forms; for example in case of distributed connection weights (see Helias et al., 2008, their eq. 33) or in case of the mean-field solution on the single neuron level (Eq. (3.4), indegree factor $K$ absent). Therefore we prefer a generic parameterization over more specific alternatives like the pair `weight` and `convergence`.

# Bibliography


Abbott, L., DePasquale, B., & Memmesheimer, R.-M. (2016). Building functional networks of spiking model neurons. *Nat. Neurosci. 19*(3), 350–355.

Abeles, M. (1991a). *Corticonics: Neural Circuits of the Cerebral Cortex* (1st ed.). Cambridge: Cambridge University Press.

Abeles, M. (1991b). *Corticonics: Neural Circuits of the Cerebral Cortex*. Cambridge: Cambridge University Press.

Abramowitz, M., & Stegun, I. A. (1974). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications.

Adamu, I. A. (2011). *Numerical approximation of SDEs and stochastic Swift-Hohenberg equation*. Ph. D. thesis, Heriot-Watt University.

Al-Mohy, A. H., & Higham, N. J. (2009). A New Scaling and Squaring Algorithm for the Matrix Exponential. *SIAM J. Matrix Analysis Applications 31*(3), 970–989.

Alijani, A. K., & Richardson, M. J. E. (2011). Rate response of neurons subject to fast or frozen noise: from stochastic and homogeneous to deterministic and heterogeneous populations. *Physical Review E 84*(1), 011919.

Aljadeff, J., Stern, M., & Sharpee, T. (2015). Transition to chaos in random networks with cell-type-specific connectivity. *Phys. Rev. Lett. 114*, 088101.

Altland, A., & B., S. (2010). *Concepts of Theoretical Solid State Physics*. Cambridge university press.

Amari, S.-I. (1972). Characteristics of random nets of analog neuron-like elements. *Systems, Man and Cybernetics, IEEE Transactions on* (5), 643–657.

Amari, S.-I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern. 27*, 77–87.

Amiri, M., Bahrami, F., & Janahmadi, M. (2012). Functional contributions of astrocytes in synchronization of a neuronal network model. *Journal of Theoretical Biology 292*, 60 – 70.

Amit, D. J., & Brunel, N. (1997a). Dynamics of a recurrent network of spiking neurons before and following learning. *Network: Comput. Neural Systems 8*, 373–404.

Amit, D. J., & Brunel, N. (1997b). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex 7*, 237–252.

Axer, M., Grässel, D., Kleiner, M., Dammers, J., Dickscheid, T., Reckfort, J., Hütz, T., Eiben, B., Pietrzyk, U., Zilles, K., et al. (2011). High-resolution fiber tract reconstruction in the human brain by means of three-dimensional polarized light imaging. *Front. Neuroinformatics 5*(34).

Bakker, R., Thomas, W., & Diesmann, M. (2012). CoCoMac 2.0 and the future of tract-tracing databases. *Front. Neuroinformatics 6*(30).

Ball, T., Demandt, E., Mutschler, I., Neitzel, E., Mehring, C., Vogt, K., Aertsen, A., & Schulze-Bonhage, A. (2008). Movement related activity in the high gamma range of the human eeg. *NeuroImage 41*(2), 302–310.

Barbas, H., & Rempel-Clower, N. (1997). Cortical structure predicts the pattern of corticocortical connections. *Cereb. Cortex 7*(7), 635–646.

Barnes, C. L., & Pandya, D. N. (1992). Efferent cortical connections of multimodal cortex of the superior temporal sulcus in the rhesus monkey. *J. Comp. Neurol. 318*(2), 222–244.

Bastos, A., Litvak, V., Moran, R., Bosman, C., Fries, P., & Friston, K. (2015a). A dcm study of spectral asymmetries in feedforward and feedback connections between visual areas v1 and v4 in the monkey. *Neuroimage 108*, 460–475.

Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J. R., De Weerd, P., Kennedy, H., & Fries, P. (2015b). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron 85*(2), 390–401.

Bastos, A. M., Vezoli, J., & Fries, P. (2015). Communication through coherence with inter-areal delays. *Current opinion in neurobiology 31*, 173–180.

Bednar, J. A. (2009). Topographica: Building and analyzing map-level simulations from python, c/c++, matlab, nest, or neuron components. *Front Neuroinform. 24*, 8.

Beltramo, R., D'Urso, G., Maschio, M. D., Farisello, P., Bovetti, S., Clovis, Y., Lassi, G., Tucci, V., Tonelli, D. D. P., & Fellin, T. (2013). Layer-specific excitatory circuits differentially control recurrent network dynamics in the neocortex. *Nat. Neurosci. 16*(2), 227–234.

Benda, J., & Herz, V. (2003). A universal model for spike-frequency adaptation. *Neural Comput. 15*, 2523–2546.

Berger, D., Borgelt, C., Louis, S., Morrison, A., & Grün, S. (2010). Efficient identification of assembly neurons within massively parallel spike trains. *2010*, Article ID 439648.

Beul, S. F., Barbas, H., & Hilgetag, C. C. (2015). A predictive structural model of the primate connectome. *arXiv preprint arXiv:1511.07222*.

Binzegger, T., Douglas, R. J., & Martin, K. A. C. (2004). A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci. 39*(24), 8441–8453.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review 113*(4), 700.

Bogacz, R., Usher, M., Zhang, J., & McClelland, J. L. (2007). Extending a biologically inspired model of choice: multi-alternatives, nonlinearity and value-based multidimensional choice. *Phil. Trans. R. Soc. B 362*, 1655–1670.

Bos, H., Diesmann, M., & Helias, M. (2016). Identifying anatomical origins of coexisting oscillations in the cortical microcircuit. *PLoS Comput. Biol. 12*(10), 1–34.

Bos, H., Morrison, A., Peyser, A., Hahne, J., Helias, M., Kunkel, S., Ippen, T., Eppler, J. M., Schmidt, M., Seeholzer, A., Djurfeldt, M., Diaz, S., Morén, J., Deepu, R., Stocco, T., Deger, M., Michler, F., & Plesser, H. E. (2015). Nest 2.10.0.

Bos, H., Schuecker, J., & Helias, M. (2016). How network structure shapes responses to oscillatory stimuli. *in preparation*.

Bower, J. M., & Beeman, D. (2007). GENESIS (simulation environment). *Scholarpedia 2*(3), 1383.

Braitenberg, V., & Schüz, A. (1991). *Anatomy of the Cortex: Statistics and Geometry*. Berlin, Heidelberg, New York: Springer-Verlag.

Bressloff, P. C. (2012). Spatiotemporal dynamics of continuum neural fields. *Journal of Physics A: Mathematical and Theoretical 45*(3), 033001.

Brette, R. (2015). Philosophy of the spike: Rate-based vs. spike-based theories of the brain. *Frontiers in Systems Neuroscience 9*, 151.

Brette, R., & Destexhe, A. (2012). *Handbook of neural activity measurement*. Cambridge University Press.

Brette, R., Rudolph, M., Carnevale, T., Hines, M., Beeman, D., Bower, J. M., Diesmann, M., Morrison, A., Goodman, P. H., Harris Jr., F. C., Zirpe, M., Natschläger, T., Pecevski, D., Ermentrout, B., Djurfeldt, M., Lansner, A., Rochel, O., Vieville, T., Muller, E., Davison, A. P., El Boustani, S., & Destexhe, A. (2007). Simulation of networks of spiking neurons: A review of tools and strategies. *J. Comput. Neurosci. 23*(3), 349–398.

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Leipzig: Johann Ambrosius Barth.

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci. 8*(3), 183–208.

Brunel, N., Chance, F. S., Fourcaud, N., & Abbott, L. F. (2001). Effects of synaptic noise and filtering on the frequency response of spiking neurons. *Phys. Rev. Lett. 86*(10), 2186–2189.

Brunel, N., & Hakim, V. (1999). Fast global oscillations in networks of integrate-and-fire neurons with low firing rates. *Neural Comput. 11*(7), 1621–1671.

Buzsáki, G., & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science 304*, 1926–1929.

Cabana, T., & Touboul, J. (2013). Large deviations, dynamics and phase transitions in large stochastic and disordered neural networks. *J. statist. Phys. 153*(2), 211–269.

Cabral, J., Kringelbach, M. L., & Deco, G. (2014). Exploring the network dynamics underlying brain activity during rest. *Prog. Neurobiol. 114*, 102–131.

Cain, N., Iyer, R., Koch, C., & Mihalas, S. (2016). The computational properties of a simplified cortical column model. *PLoS Comput. Biol. 12*(9), e1005045.

Camperi, M., & Wang, X.-J. (1998). A model of visuospatial working memory in prefrontal cortex: recurrent network and cellular bistability. *J. Comput. Neurosci. 5*(4), 383–405.

Carnevale, T., & Hines, M. (2006). *The NEURON Book*. Cambridge: Cambridge University Press.

Chaudhuri, R., Knoblauch, K., Gariel, M.-A., Kennedy, H., & Wang, X.-J. (2015). A large-scale circuit mechanism for hierarchical dynamical processing in the primate cortex. *Neuron 88*(2), 419–431.

Cichocki, A., Zdunek, R., Phan, A. H., & Amari, S.-i. (2009). *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons.

Cohen, M. R., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nat. Rev. Neurosci. 14*(7), 811–819. doi:10.1038/nn.2842.

Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., & Petersen, S. E. (2014). Intrinsic and task-evoked network architectures of the human brain. *Neuron 83*(1), 238–251.

Connors, B. W., & Long, M. A. (2004). Electrical synapses in the mammalian brain. *Annu. Rev. Neurosci. 27*(1), 393–418.

Coombes, S. (2005). Waves, bumps, and patterns in neural field theories. *Biol. Cybern. 93*, 91–108.

Crisanti, A., & Sompolinsky, H. (1987). Dynamics of spin systems with randomly asymmetric bonds: Langevin dynamics and a spherical model. *Phys. Rev. A 36*(10), 4922–4939.

Dahmen, D., Bos, H., & Helias, M. (2016). Correlated fluctuations in strongly coupled binary networks beyond equilibrium. *Phys. Rev. X 6*, 031024.

Dayan, P., & Abbott, L. F. (2001). *Theoretical Neuroscience*. Cambridge: MIT Press.

De Dominicis, C., & Peliti, L. (1978). Field-theory renormalixation and critical dynamics above $t_c$: Helium, antiferromagnets, and liquid-gas systems. *Phys. Rev. B 18*(1), 353–376.

De Kamps, M., Baier, V., Drever, J., Dietz, M., Mösenlechner, L., & van der Felde, F. (2008). The state of miind. *Neural Networks 21*(8), 1164–1181.

de Kock, C. P. J., & Sakmann, B. (2009). Spiking in primary somatosensory cortex during natural whisking in awake head-restrained rats is cell-type specific. *Proc. Natl. Acad. Sci. USA 106*(38), 16446–16450.

De la Rocha, J., Doiron, B., Shea-Brown, E., Kresimir, J., & Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature 448*(16), 802–807.

Deadman, E., Higham, N. J., & Ralha, R. (2012). Blocked Schur Algorithms for Computing

the Matrix Square Root. In P. Manninen & P. Öster (Eds.), *PARA*, Volume 7782 of *Lecture Notes in Computer Science*, pp. 171–182. Springer.

Deco, G., & Jirsa, V. K. (2012). Ongoing cortical activity at rest: Criticality, multistability, and ghost attractors. *J. Neurosci. 32*(10), 3366–3375.

Deco, G., Jirsa, V. K., & McIntosh, A. R. (2011). Emerging concepts for the dynamical organization of resting-state activity in the brain. *natrevnsci 12*, 43–56.

Dehghani, N., Peyrache, A., Telenczuk, B., Le Van Quyen, M., Halgren, E., Cash, S. S., Hatsopoulos, N. G., & Destexhe, A. (2016). Dynamic balance of excitation and inhibition in human and monkey neocortex. *Scientific reports 6*.

Derrida, B., & Pomeau, Y. (1986). Random networks of automata: a simple annealed approximation. *EPL (Europhysics Letters) 1*(2), 45.

Destexhe, A., & Paré, D. (1999). Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. *J. Neurophysiol. 81*(4), 1531–1547.

Doering, C. R., Bagley, R. J., Hagan, P. S., & Levermore, C. D. (1988). Doering et al. reply. *Phys. Rev. Lett. 60*, 2805.

Doering, C. R., Hagan, P. S., & Levermore, C. D. (1987). Bistability driven by weakly colored gaussian noise: The fokker-planck boundary layer and mean first-passage times. *Phys. Rev. Lett. 59*(19), 2129–2132.

Dummer, B., Wieland, S., & Lindner, B. (2014). Self-consistent determination of the spike-train power spectrum in a neural network with sparse connectivity. *Front. Comput. Neurosci. 8*, 104.

Ecker, A. S., Berens, P., Keliris, G. A., Bethge, M., & Logothetis, N. K. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science 327*(5965), 584–587.

Eckmann, J.-P., & Ruelle, D. (1985). Ergodic theory of chaos and strange attractors. *Reviews of modern physics 57*(3), 617.

Engelken, R., Farkhooi, F., Hansel, D., van Vreeswijk, C., & Wolf, F. (2015). Comment on "two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons". *bioRxiv*, 017798.

Eppler, J. M., Helias, M., Muller, E., Diesmann, M., & Gewaltig, M. (2009). PyNEST: a convenient interface to the NEST simulator. *Front. Neuroinformatics 2*, 12.

Eppler, J. M., Pauli, R., Peyser, A., Ippen, T., Morrison, A., Senk, J., Schenck, W., Bos, H., Helias, M., Schmidt, M., Kunkel, S., Jordan, J., Gewaltig, M.-O., Bachmann, C., Schuecker, J., van Albada, S. J., Zito, T., Deger, M., Michler, F., Hagen, E., Setareh, H., Riquelme, L., Shirvani, A., Duarte, R., Deepu, R., & Plesser, H. E. (2015). NEST 2.8.0.

Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological review 109*(3), 545.

Fan, Z. (2013). Sor waveform relaxation methods for stochastic differential equations. *Appl. Math. Comput. 219*(10), 4992–5003.

Farkas, L. (1927). Keimbildungsgeschwindigkeit in übersättigten dämpfen. *Z. phys. Chem 125*, 236–242.

Farkhooi, F., & van Vreeswijk, C. (2015). Renewal approach to the analysis of the asynchronous state for coupled noisy oscillators. *Phys. Rev. Lett. 115*, 038103.

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive science 6*(3), 205–254.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex 1*, 1–47.

Fischer, K., & Hertz, J. (1991). *Spin glasses*. Cambridge University Press.

Fourcaud, N., & Brunel, N. (2002). Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural Comput. 14*, 2057–2110.

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

Fox, R. F. (1986). Functional-calculus approach to stochastic differential equations. *Phys. Rev. A 33*, 467–476.

Fries, P. (2005). A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci. 9*(10), 474–480.

Frisch, U., & Morf, R. (1981). Intermittency in nonlinear dynamics and singularities at complex times. *Phys. Rev. A 23*, 2673–2705.

Galassi, M., Davies, J., Theiler, J., Gough, B., Jungman, G., Booth, M., & Rossi, F. (2006). *GNU Scientific Library Reference Manual (2nd Ed.)*. Network Theory Limited.

Gancarz, G., & Grossberg, S. (1998). A neural model of the saccade generator in the reticular formation. *IEEE Trans. Neural Netw. 11*(7), 1159–1174.

Ganguli, S., Huh, D., & Sompolinsky, H. (2008). Memory traces in dynamical systems. *Proc. Natl. Acad. Sci. USA 105*(48), 18970–18975.

Gardiner, C. W. (1983). *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Number 13 in Springer Series in Synergetics. Berlin: Springer-Verlag.

Gardiner, C. W. (2004). *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (3rd ed.). Springer Series in Synergetics. Springer.

Gerstner, W., & Kistler, W. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press.

Gerstner, W., Kistler, W. M., Naud, R., & Paninski, L. (2014). *Neuronal Dynamics. From single Neurons to Networks and Models of Cognition*. Cambridge University Press.

Gewaltig, M.-O., & Diesmann, M. (2007). NEST (NEural Simulation Tool). *Scholarpedia 2*(4), 1430.

Ginzburg, I., & Sompolinsky, H. (1994). Theory of correlations in stochastic neural networks. *Phys. Rev. E 50*(4), 3171–3191.

Goedeke, S., Schuecker, J., & Helias, M. (2016). Noise dynamically suppresses chaos in neural networks. *arXiv preprint arXiv:1603.01880v2*.

Goodman, D., & Brette, R. (2013). Brian simulator. *Scholarpedia 8*(1), 10883.

Goulas, A., Bastiani, M., Bezgin, G., Uylings, H. B., Roebroeck, A., & Stiers, P. (2014). Comparative analysis of the macroscale structural connectivity in the macaque and human brain. *PLoS Comput. Biol. 10*(3), e1003529.

Goychuk, I., & Hänggi, P. (2002). Ion channel gating: A first-passage time analysis of the kramers type. *Proc. Natl. Acad. Sci. USA 99*(6), 3552–3556.

Griffith, J. S., & Horn, G. (1966). An analysis of spontaneous impulse activity of units in the striate cortex of unrestrained cats. *J. Physiol. (Lond.) 186*(3), 516–534.

Grigolini, P. (1986). The projection approach to the problem of colored noise. *Phys. Lett. 119*(4), 157 – 162.

Grossberg, S. (1973). Contour enhancement, short term memory, and constancies in reverberating neural networks. *Stud. Appl. Math. 52*(3), 213–257.

Grytskyy, D., Tetzlaff, T., Diesmann, M., & Helias, M. (2013). A unified view on weakly correlated recurrent networks. *Front. Comput. Neurosci. 7*, 131.

Hahne, J., Dahmen, D., Schuecker, J., Frommer, A., Bolten, M., Helias, M., & Diesmann, M. (2016). Integration of continuous-time dynamics in a spiking neural network simulator. *arXiv*. 1610.09990 [q–bio.NC].

Hahne, J., Helias, M., Kunkel, S., Igarashi, J., Bolten, M., Frommer, A., & Diesmann, M. (2015). A unified framework for spiking and gap-junction interactions in distributed neuronal network simulations. *Front. Neuroinform. 9*(22).

Hairer, E., & Wanner, G. (1991). *Solving Ordinary Differential Equations II*. Berlin: Springer.

Hales, J., Zhukov, A., Roy, R., & Dykman, M. (2000). Dynamics of activated escape and its observation in a semiconductor laser. *Phys. Rev. Lett. 85*(1), 78.

Hänggi, P., & Jung, P. (1995). Colored noise in dynamical systems. *Advances in chemical physics 89*, 239–326.

Hänggi, P., Jung, P., & Talkner, P. (1988). Comment on "bistability driven by weakly colored gaussian noise: The fokker-planck boundary layer and mean first-passage times". *Phys. Rev. Lett. 60*, 2804–2804.

Hanggi, P., Mroczkowski, T. J., Moss, F., & McClintock, P. V. E. (1985). Bistability driven by colored noise: Theory and experiment. *Phys. Rev. A 32*, 695–698.

Hänggi, P., Talkner, P., & Borkovec, M. (1990). Reaction-rate theory: fifty years after kramers. *Rev. Mod. Phys. 62*, 251–341.

Hansel, D., & Sompolinsky, H. (1998). Modeling feature selectivity in local cortical circuits. In C. Koch & I. Segev (Eds.), *Methods in Neuronal Modeling*. Cambridge, Massachusetts: MIT Press.

Harish, O., & Hansel, D. (2015). Asynchronous rate chaos in spiking neuronal circuits. *PLoS Comput Biol 11*(7), e1004266.

Haykin, S. S. (2009). *Neural Networks and Learning Machines* (3rd ed.). Prentice Hall.

Helias, M., Deger, M., Rotter, S., & Diesmann, M. (2010). Instantaneous non-linear processing by pulse-coupled threshold units. *PLoS Comput. Biol. 6*(9), e1000929.

Helias, M., Kunkel, S., Masumoto, G., Igarashi, J., Eppler, J. M., Ishii, S., Fukai, T., Morrison, A., & Diesmann, M. (2012). Supercomputers ready for use as discovery machines for neuroscience. *Front. Neuroinform. 6*, 26.

Helias, M., Rotter, S., Gewaltig, M., & Diesmann, M. (2008). Structural plasticity controlled by calcium based correlation detection. *Front. Comput. Neurosci. 2*(7), doi:10.3389/neuro.10.007.2008.

Helias, M., Tetzlaff, T., & Diesmann, M. (2013). Echoes in correlated neural systems. *New J. Phys. 15*, 023002.

Helias, M., Tetzlaff, T., & Diesmann, M. (2014). The correlation structure of local cortical networks intrinsically results from recurrent dynamics. *PLoS Comput. Biol. 10*(1), e1003428.

Helias, M., Tetzlaff, T., & M, D. (2015). Lecture notes on "theoretical neuroscience: Correlation structure of neuronal networks", held at university of aachen.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation*. Perseus Books.

Hilgetag, C. C., Medalla, M., Beul, S., & Barbas, H. (2016). The primate connectome in context: principles of connections of the cortical visual system. *NeuroImage*, (in press).

Hines, M., Eichner, H., & Schürmann, F. (2008). Neuron splitting in compute-bound parallel network simulations enables runtime scaling with twice as many processors. *J Comput Neurosci. 25*(1), 203–210.

Hines, M., Kumar, S., & Schürmann, F. (2011). Comparison of neuronal spike exchange methods on a Blue Gene/P supercomputer. *Front. Comput. Neurosci. 5*(49). 10.3389/fncom.2011.00049.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol. (Lond) 117*, 500–544.

Horsthemke, W. (1984). *Noise Induced Transitions*, Volume 27 of *Springer Series in Synergetics*. Springer Berlin Heidelberg.

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *J Physiol 148*, 574–591.

Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging*, Volume 1. Sinauer Associates Sunderland.

Johnson, J. B. (1928). Thermal agitation of electricity in conductors. *Phys. Rev. 32*, 97–109.

Jolivet, R., Rauch, A., Lüscher, H.-R., & Gerstner, W. (2006). Predicting spike timing of

neocortical pyramidal neurons by simple threshold models. *J. Comput. Neurosci. 21*(1), 35–49.

Jones, E., Coulter, J., & Hendry, S. (1978). Intracortical connectivity of architectonic fields in the somatic sensory, motor and parietal cortex of monkeys. *J. Comp. Neurol. 181*(2), 291–347.

Jülich Supercomputing Centre (2015). JUQUEEN: IBM Blue Gene/Q® supercomputer system at the Jülich Supercomputing Centre. *Journal of large-scale research facilities 1*.

Kadmon, J., & Sompolinsky, H. (2015a). Transition to chaos in random neuronal networks. *arXiv*, 1508.06486.

Kadmon, J., & Sompolinsky, H. (2015b). Transition to chaos in random neuronal networks. *Phys. Rev. X 5*, 041030.

Kandel, E. R., Schwartz, J. H., & Jessel, T. M. (2000). *Principles of Neural Science* (4 ed.). New York: McGraw-Hill. ISBN 978-0838577011.

Kelley, C. T. (1995). *Iterative Methods for Linear and Nonlinear Equations*. Number 16 in Frontiers in Applied Mathematics. SIAM.

Kilavik, B. E., Roux, S., Ponce-Alvarez, A., Confais, J., Gruen, S., & Riehle, A. (2009). Long-term modifications in motor cortical dynamics induced by intensive practice. *J. Neurosci. 29*, 12653–12663.

Kilpatrick, Z. P. (2014). Wilson-Cowan model. In *Encyclopedia of Computational Neuroscience*. Springer.

Kirkpatrick, S., & Sherrington, D. (1978). Infinite-ranged models of spin-glasses. *Phys. Rev. B 17*(11), 4384.

Kloeden, P. E., & Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Berlin: Springer.

Kłosek, M. M., & Hagan, P. S. (1998). Colored noise and a characteristic level crossing problem. *J. Math. Phys. 39*, 931–953.

Koch, K. W., & Fuster, J. M. (1989). Unit activity in monkey parietal cortex related to haptic perception and temporary memory. *Exp. Brain Res. 76*(2), 292–306.

Komori, Y., & Burrage, K. (2014). A stochastic exponential Euler scheme for simulation of stiff biochemical reaction systems. *BIT Numerical Mathematics 54*(4), 1067–1085.

Kramers, H. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica 7*(4), 284–304.

Kriener, B., Enger, H., Tetzlaff, T., Plesser, H. E., Gewaltig, M.-O., & Einevoll, G. T. (2014a). Dynamics of self-sustained asynchronous-irregular activity in random networks of spiking neurons with strong synapses. *Front Comput Neurosci 8*, 136.

Kriener, B., Helias, M., Rotter, S., Diesmann, M., & Einevoll, G. (2014b). How pattern formation in ring networks of excitatory and inhibitory spiking neurons depends on the input current regime. *Front. Comput. Neurosci. 7*(187), 1–21.

Kuhn, A., Aertsen, A., & Rotter, S. (2004). Neuronal integration of synaptic input in the fluctuation-driven regime. *J. Neurosci. 24*(10), 2345–2356.

Kumar, S., Heidelberger, P., Chen, D., & Hines, M. (2010). Optimization of applications with non-blocking neighborhood collectives via multisends on the blue gene/p supercomputer. *IPDPD*, 1–11.

Kunkel, S., Diesmann, M., & Morrison, A. (2011). Limits to the development of feed-forward structures in large recurrent neuronal networks. *Front. Comput. Neurosci. 4*.

Kunkel, S., Schmidt, M., Eppler, J. M., Masumoto, G., Igarashi, J., Ishii, S., Fukai, T., Morrison, A., Diesmann, M., & Helias, M. (2014). Spiking network simulation code for petascale computers. *Front. Neuroinformatics 8*, 78.

Laing, C. R., Troy, W. C., Gutkin, B., & Ermentrout, B. G. (2002). Multiple bumps in a neuronal model of working memory. *SIAM J. Appl. Math. 63*, 62–97.

Lajoie, G., Lin, K. K., & Shea-Brown, E. (2013). Chaos and reliability in balanced spiking networks with temporal drive. *Phys. Rev. E 87*(5), 052901.

Lapicque, L. (1907). Recherches quantitatives sur l'excitation electrique des nerfs traitee comme une polarization. *J. Physiol. Pathol. Gen. 9*, 620–635.

Le Bon-Jego, M., & Yuste, R. (2007). Persistently active, pacemaker-like neurons in neocortex. *Front. Neurosci. 1*(1), 123–129.

Lelarasmee, E. (1982). The waveform relaxation method for time domain analysis of large scale integrated circuits: theory and applications. *Memorandum*, No. UCB/ERL M82/40.

Lőrincz, M. L., Gunner, D., Bao, Y., Connelly, W. M., Isaac, J. T., Hughes, S. W., & Crunelli, V. (2015). A distinct class of slow ( 0.2–2 Hz) intrinsically bursting layer 5 pyramidal neurons determines UP/DOWN state dynamics in the neocortex. *J. Neurosci. 35*(14), 5442–5458.

Lindén, H., Tetzlaff, T., Potjans, T. C., Pettersen, K. H., Grün, S., Diesmann, M., & Einevoll, G. T. (2011). Modeling the spatial reach of the LFP. *Neuron 72*(5), 859–872.

Lindenberg, K., & West, B. J. (1983). Finite correlation time effects in nonequilibrium phase transitions: I. dynamic equation and steady state properties. *Physica A 119*(3), 485 – 503.

Lindner, B. (2004). Interspike interval statistics of neurons driven by colored noise. *Phys. Rev. E 69*, 0229011–0229014.

Lindner, B., Doiron, B., & Longtin, A. (2005). Theory of oscillatory firing induced by spatially correlated noise and delayed inhibitory feedback. *Phys. Rev. E 72*, 061919.

Lindner, B., & Schimansky-Geier, L. (2001). Transmission of noise coded versus additive signals through a neuronal ensemble. *Phys. Rev. Lett. 86*, 2934–2937.

Magnus, J. R., & Neudecker, H. (1995). *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons.

Malach, R., Amir, Y., Harel, M., & Grinvald, A. (1993). Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. *Proc. Natl. Acad. Sci. USA 90*, 10469–10473.

Malakhov, A., & Pankratov, A. (1996). Influence of thermal fluctuations on time characteristics of a single josephson element with high damping exact solution. *Physica C: Superconductivity 269*(1), 46–54.

Mantegna, R. N., & Spagnolo, B. (1996). Noise enhanced stability in an unstable system. *Phys. Rev. Lett. 76*, 563–566.

Mantegna, R. N., & Spagnolo, B. (2000). Experimental investigation of resonant activation. *Phys. Rev. Lett. 84*(14), 3025.

Markov, N. T., Ercsey-Ravasz, M., Van Essen, D. C., Knoblauch, K., Toroczkai, Z., & Kennedy, H. (2013). Cortical high-density counterstream architectures. *Science 342*(6158).

Markov, N. T., Ercsey-Ravasz, M. M., Ribeiro Gomes, A. R., Lamy, C., Magrou, L., Vezoli, J., Misery, P., Falchier, A., Quilodran, R., Gariel, M. A., Sallet, J., Gamanut, R., Huissoud, C., Clavagnier, S., Giroud, P., Sappey-Marinier, D., Barone, P., Dehay, C., Toroczkai, Z., Knoblauch, K., Van Essen, D. C., & Kennedy, H. (2014a). A weighted and directed interareal connectivity matrix for macaque cerebral cortex. *Cereb. Cortex 24*(1), 17–36.

Markov, N. T., Misery, P., Falchier, A., Lamy, C., Vezoli, J., Quilodran, R., Gariel, M. A., Giroud, P., Ercsey-Ravasz, M., Pilaz, L. J., Huissoud, C., Barone, P., Dehay, C., Toroczkai, Z., Van Essen, D. C., Kennedy, H., & Knoblauch, K. (2011). Weight consistency specifies regularities of macaque cortical networks. *Cereb. Cortex 21*(6), 1254–1272.

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., Lamy, C., Misery, P., Giroud, P., Ullman, S., Barone, P., Dehay, C., Knoblauch, K., & Kennedy, H. (2014b). Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. *J. Comp. Neurol. 522*(1), 225–259.

Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., et al. (2015). Reconstruction and simulation of neocortical microcircuitry. *Cell 163*(2), 456–492.

Martin, P., Siggia, E., & Rose, H. (1973). Statistical dynamics of classical systems. *Phys. Rev. A 8*(1), 423–437.

Mastroguiseppe, F., & Ostojic, S. (2016). Intrinsically-generated fluctuating activity in excitatory-inhibitory networks. *arXiv*, 1605.04221.

Mattia, M., & Del Guidice, P. (2002). Population dynamics of interacting spiking neurons. *Phys. Rev. E 66*, 051917.

Mattia, M., & Del Guidice, P. (2004). Finite-size dynamics of inhibitory and excitatory interacting spiking neurons. *Phys. Rev. E 70*, 052903.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. an account of basic findings. *Psychological review 88*(5), 375.

McClelland, J. L., & Rumelhart, D. E. (1989). *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*. MIT press.

Miekkala, U., & Nevanlinna, O. (1987). Convergence of dynamic iteration methods for initial value problems. *SIAM J. Sci. and Stat. Comput. 8*(4), 459–482.

Molgedey, L., Schuchhardt, J., & Schuster, H. (1992). Suppressing chaos in neural networks by noise. *Phys. Rev. Lett. 69*(26), 3717.

Morel, A., & Bullier, J. (1990). Anatomical segregation of two cortical visual pathways in the macaque monkey. *Visual neuroscience 4*(06), 555–578.

Moreno-Bote, R., & Parga, N. (2006). Auto- and crosscorrelograms for the spike response of leaky integrate-and-fire neurons with slow synapses. *Phys. Rev. Lett. 96*, 028101.

Morrison, A., & Diesmann, M. (2008). Maintaining causality in discrete time neuronal network simulations. In P. beim Graben, C. Zhou, M. Thiel, & J. Kurths (Eds.), *Lectures in Supercomputational Neuroscience: Dynamics in Complex Brain Networks*, Understanding Complex Systems, pp. 267–278. Springer.

Morrison, A., Diesmann, M., & Gerstner, W. (2008). Phenomenological models of synaptic plasticity based on spike-timing. *Biol. Cybern. 98*, 459–478.

Morrison, A., Mehring, C., Geisel, T., Aertsen, A., & Diesmann, M. (2005). Advancing the boundaries of high connectivity network simulation with distributed computing. *Neural Comput. 17*(8), 1776–1801.

Moss, F., & McClintock, P. V. E. (Eds.) (1989a). *Noise in Nonlinear Dynamical Systems*, Volume 2. Cambridge University Press.

Moss, F., & McClintock, P. V. E. (Eds.) (1989b). *Noise in Nonlinear Dynamical Systems*, Volume 3. Cambridge University Press.

Naundorf, B., Geisel, T., & Wolf, F. (2003). The intrinsic time scale of transient neuronal responses. *arXiv*, arXiv:physics/0307135 [physics.bio–ph].

Negele, J. W., & Orland, H. (1998). *Quantum Many-Particle Systems*. Perseus Books.

Neske, G. T., Patrick, S. L., & Connors, B. W. (2015). Contributions of diverse excitatory and inhibitory neurons to recurrent network activity in cerebral cortex. *J. Neurosci. 35*(3), 1089–1105.

Nichols, E. J., & Hutt, A. (2015). Neural field simulator: two-dimensional spatio-temporal dynamics involving finite transmission speed. *Front. Neuroinform. 9*, 25.

Nyquist, H. (1928). Thermal agitation of electric charge in conductors. *Phys. Rev. 32*, 110–113.

Ohbayashi, M., Ohki, K., & Miyashita, Y. (2003). Conversion of working memory to motor sequence in the monkey premotor cortex. *Science 301*(5630), 233–236.

Okun, M., & Lampl, I. (2008). Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nat. Neurosci. 11*(5), 535–537.

Olver, F. W. J., Lozier, D. W., Boisvert, R. F., & Clark, C. W. (Eds.) (2010). *NIST Handbook of Mathematical Functions*. New York, NY: Cambridge University Press.

O'Reilly, R. C. (2014). Comparison of neural network simulators. `https://grey.`

colorado.edu/emergent/index.php/Comparison_of_Neural_Network_Simulators. Accessed: 2016-10-14.

O'Reilly, R. C., Munakata, Y., Frank, M. J., Hazy, T. E., & Contributors (2012). *Computational Cognitive Neuroscience*. Wiki Book, 1st Edition, URL: http://ccnbook.colorado.edu.

O'Reilly, R. C., Munakata, Y., & McClelland, J. L. (2000). *Computational Explorations in Cognitive Neuroscience: Understanding the Mind by Simulating the Brain* (1 ed.). MIT Press.

Ostojic, S. (2014). Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nat. Neurosci. 17*, 594–600.

Ostojic, S. (2015). Response to comment on "two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons". *bioRxiv*, 020354.

Ostojic, S., & Brunel, N. (2011). From spiking neuron models to linear-nonlinear models. *PLoS Comput. Biol. 7*(1), e1001056.

Ostojic, S., Brunel, N., & Hakim, V. (2009). How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. *J. Neurosci. 29*(33), 10234–10253.

Pankratov, A. L., & Spagnolo, B. (2004). Suppression of timing errors in short overdamped josephson junctions. *Phys. Rev. Lett. 93*(17), 177001.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes* (3 ed.). Boston, Massachusetts: McGraw-Hill.

Perkel, D. J., Bullier, J., & Kennedy, H. (1986). Topography of the afferent connectivity of area 17 in the macaque monkey: A double-labelling study. *J. Comp. Neurol. 253*(3), 374–402.

Pernice, V., Staude, B., Cardanobile, S., & Rotter, S. (2011). How structure determines correlations in neuronal networks. *PLoS Comput. Biol. 7*(5), e1002059.

Plesser, H., Diesmann, M., Gewaltig, M.-O., & Morrison, A. (2015). Nest: the neural simulation tool. In D. Jaeger & R. Jung (Eds.), *Encyclopedia of Computational Neuroscience*, pp. 1849–1852. Springer New York.

Plesser, H. E., & Diesmann, M. (2009). Simplicity and efficiency of integrate-and-fire neuron models. *Neural Comput. 21*, 353–359.

Plotnikov, D., Blundell, I., Ippen, T., Eppler, J. M., Morrison, A., & Rumpe, B. (2016). NESTML: a modeling language for spiking neurons. In *Modellierung 2016 Conference*, Volume 254 of *LNI*, pp. 93–108. Bonner Köllen Verlag.

Potjans, T. C., & Diesmann, M. (2014). The cell-type specific cortical microcircuit: Relating structure and activity in a full-scale spiking network model. *Cereb. Cortex 24*(3), 785–806.

Rajan, K., Abbott, L., & Sompolinsky, H. (2010). Stimulus-dependent suppression of chaos in recurrent neural networks. *Phys. Rev. E 82*(1), 011903.

Rajan, K., & Abbott, L. F. (2006). Eigenvalue spectra of random matrices for neural networks. *Phys. Rev. Lett. 97*, 188104.

Rauch, A., La Camera, G., Lüscher, H., Senn, W., & Fusi, S. (2003). Neocortical pyramidal cells respond as integrate-and-fire neurons to in vivo like input currents. *J. Neurophysiol. 90*, 1598–1612.

Ray, S., Crone, N. E., Niebur, E., Franaszczuk, P. J., & Hsiao, S. S. (2008). Neural correlates of high-gamma oscillations (60-200 hz) in macaque local field potentials and their potential implications in electrocorticography. *J. Neurosci. 28*(45), 11526–11536.

Renart, A., De La Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., & Harris, K. D. (2010). The asynchronous state in cortical circuits. *Science 327*, 587–590.

Ricciardi, L. M. (1977). *Diffusion Processes and Related Topics on Biology*. Berlin: Springer-Verlag.

Ricciardi, L. M., Di Crescenzo, A., Giorno, V., & Nobile, A. G. (1999). An outline of theoretical and algorithmic approaches to first passage time problems with applications to biological modeling. *MathJaponica 50*(2), 247–322.

Risken, H. (1996). *The Fokker-Planck Equation*. Springer Verlag Berlin Heidelberg.

Rockland, K. S., & Pandya, D. N. (1979). Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Res. 179*, 3–20.

Rougier, N. P., & Fix, J. (2012). Dana: Distributed numerical and adaptive modelling framework. *Network: Computation in Neural Systems 23*(4), 237–253.

Roxin, A., Brunel, N., & Hansel, D. (2005). The role of delays in shaping spatio-temporal dynamics of neuronal activity in large networks. *Phys. Rev. Lett. 94*(23), 238103.

Roxin, A., Brunel, N., & Hansel, D. (2006). Rate Models with Delays and the Dynamics of Large Networks of Spiking Neurons. *Progress of Theoretical Physics Supplement 161*, 68–85.

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

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (1986). *Parallel Distributed Processing, Explorations in the Microstructure of Cognition: Foundations*, Volume 1. Cambridge, Massachusetts: MIT Press.

Sadeh, S., & Rotter, S. (2015). Orientation selectivity in inhibition-dominated networks of spiking neurons: effect of single neuron properties and network dynamics. *PLoS Comput Biol 11*(1), e1004045.

Salin, P.-A., & Bullier, J. (1995). Corticocortical connections in the visual system: structure and function. *Physiol. Rev. 75*(1), 107–154.

Samsonovich, A., & McNaughton, B. L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci. 17*(15), 5900–5920.

Sanchez-Vives, M. V., & McCormick, D. (2000). Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nat. Neurosci. 3*, 1027–1034.

Sancho, J. M., SanMiguel, M., Katz, S. L., & Gunton, J. D. (1982). Analytical and numerical studies of multiplicative noise. *Phys. Rev. A 26*, 1589–1609.

Sanz Leon, P., Knock, S., Woodman, M., Domide, L., Mersmann, J., McIntosh, A., & Jirsa, V. (2013). The virtual brain: a simulator of primate brain network dynamics. *Front. Neuroinform. 7*, 10.

Scannell, J. W., Grant, S., Payne, B. R., & Baddeley, R. (2000). On variability in the density of corticocortical and thalamocortical connections. *Phil. Trans. R. Soc. B 355*(1393), 21–35.

Schmidt, M., Bakker, R., Diesmann, M., & van Albada, S. (2016). Full-density multi-scale account of structure and dynamics of macaque visual cortex. *arXiv preprint arXiv:1511.09364v4*.

Schmitt, O., Eipert, P., Kettlitz, R., Leßmann, F., & Wree, A. (2014). The connectome of the basal ganglia. *Brain Structure and Function*, 1–62.

Schöner, G., Spencer, J., & Group, D. (2015). *Dynamic Thinking: A Primer on Dynamic Field Theory*. Oxford Series in Developmental Cognitive Neuroscience. Oxford University Press.

Schuecker, J., Diesmann, M., & Helias, M. (2014). Reduction of colored noise in excitable systems to white noise and dynamic boundary conditions. *arXiv preprint arXiv:1410.8799*.

Schuecker, J., Diesmann, M., & Helias, M. (2015). Modulated escape from a metastable state driven by colored noise. *Phys. Rev. E 92*, 052119.

Schuecker, J., Goedeke, S., Dahmen, D., & Helias, M. (2016). Functional methods for disordered neural networks. *arXiv*. 1605.06758 [cond-mat.dis-nn].

Schuecker, J., Schmidt, M., van Albada, S., Diesmann, M., & Helias, M. (2016a). Fundamental activity constraints lead to specific interpretations of the connectome. *arXiv preprint arXiv:1509.03162v3*.

Schuecker, J., Schmidt, M., van Albada, S., Diesmann, M., & Helias, M. (2016b). Fundamental activity constraints lead to specific interpretations of the connectome. *PLoS Comput. Biol.*. "in press, preprint available at http://arxiv.org/abs/1509.03162v3".

Schultze-Kraft, M., Diesmann, M., Gruen, S., & Helias, M. (2013). Noise suppression and surplus synchrony by coincidence detection. *PLoS Comput. Biol. 9*(4), e1002904.

Schurz, H., & Schneider, K. R. (2005). Waveform relaxation methods for stochastic differential equations. *International Journal of Numerical Analysis and Modeling 3*(2), 232–254.

Seltzer, B., & Pandya, D. N. (1991). Post-rolandic cortical projections of the superior temporal sulcus in the rhesus monkey. *J. Comp. Neurol. 312*(4), 625–640.

Seltzer, B., & Pandya, D. N. (1994). Parietal, temporal, and occipita projections to cortex of the superior temporal sulcus in the rhesus monkey: A retrograde tracer study. *J. Comp. Neurol. 343*(3), 445–463.

Sénéchal, D., Tremblay, A.-M., & Bourbonnais, C. (2006). *Theoretical methods for strongly correlated electrons*. Springer Science & Business Media.

Senk, J., Hagen, E., van Albada, S., & Diesmann, M. (2015). From randomly connected to spatially organized multi-layered cortical network models. In *11th Goettingen Meeting of the German Neuroscience Society*.

Shadlen, M. N., & Newsome, W. T. (1998). The variable discharge of cortical neurons: Implications for connectivity, computation, and information coding. *J. Neurosci. 18*(10), 3870–3896.

Shea-Brown, E., Josic, K., de la Rocha, J., & Doiron, B. (2008). Correlation and synchrony transfer in integrate-and-fire neurons: basic properties and consequences for coding. *Phys. Rev. Lett. 100*, 108102.

Shen, K., Hutchison, R. M., Bezgin, G., Everling, S., & McIntosh, A. R. (2015). Network structure shapes spontaneous functional connectivity dynamics. *J. Neurosci. 35*(14), 5579–5588.

Shoji, I. (2011). A note on convergence rate of a linearization method for the discretization of stochastic differential equations. *Communications in Nonlinear Science and Numerical Simulation 16*(7), 2667–2671.

Shriki, O., Hansel, D., & Sompolinsky, H. (2003). Rate models for conductance-based cortical neuronal networks. *Neural Comput. 15*, 1809–1841.

Siegert, A. J. (1951). On the first passage time probability problem. *Phys. Rev. 81*(4), 617–623.

Silberberg, G., Bethge, M., Markram, H., Pawelzik, K., & Tsodyks, M. (2004). Dynamics of population codes in ensembles of neocortical neurons. *J. Neurophysiol. 91*, 704–709.

Softky, W. R., & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci. 13*(1), 334–350.

Sommers, H., Crisanti, A., Sompolinsky, H., & Stein, Y. (1988). Spectrum of large random asymmetric matrices. *Phys. Rev. Lett. 60*(19), 1895–1898.

Sompolinsky, H., Crisanti, A., & Sommers, H. J. (1988). Chaos in random neural networks. *Phys. Rev. Lett. 61*, 259–262.

Sompolinsky, H., & Zippelius, A. (1981). Dynamic theory of the spin-glass phase. *Phys. Rev. Lett. 47*, 359–362.

Sompolinsky, H., & Zippelius, A. (1982). Relaxational dynamics of the edwards-anderson model and the mean-field theory of spin-glasses. *Phys. Rev. B 25*(11), 6860–6875.

Spagnolo, B., Dubkov, A., Pankratov, A., Pankratova, E., Fiasconaro, A., & Ochab-Marcinek, A. (2007). Lifetime of metastable states and suppression of noise in interdisciplinary physical models. *Acta Phys. Pol. B 38*, 1925.

Stein, R. B. (1967). Some models of neuronal variability. *Biophys. J. 7*, 37–68.

Stephan, K., Kamper, L., Bozkurt, A., Burns, G., Young, M., & Kötter, R. (2001). Advanced database methodology for the collation of connectivity data on the macaque brain (CoCoMac). *Phil. Trans. R. Soc. B 356*, 1159–1186.

Steriade, M., Nuñez, A., & Amzica, F. (1993). A novel slow (< 1 hz) oscillation of neocortical neurons in vivo. *J. Neurosci. 13*, 3252–3265.

Stern, M., Sompolinsky, H., & Abbott, L. F. (2014). Dynamics of random neural networks with bistable units. *Phys. Rev. E 90*, 062710.

Strogatz, S. H. (1994). *Nonlinear Dynamics and Chaos: with Applications to Physics, Biology, Chemistry, and Engineering.* Reading, Massachusetts: Perseus Books.

Suzuki, W. L., & Amaral, D. G. (1994). Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *J. Comp. Neurol. 350*(4), 497–533.

Swadlow, H. A. (1988). Efferent neurons and suspected interneurons in binocular visual cortex of the awake rabbit: Receptive fields and binocular properties. *J. Neurophysiol. 59*(4), 1162–1187.

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

Tomioka, R., & Rockland, K. S. (2007). Long-distance corticocortical GABAergic neurons in the adult monkey white and gray matter. *J. Comp. Neurol. 505*(5), 526–538.

Torre, E., Quaglio, P., Denker, M., Brochier, T., Riehle, A., & Grün, S. (2016). Synchronous spike patterns in macaque motor cortex during an instructed-delay reach-to-grasp task. *J. Neurosci. 36*(32), 8329–8340.

Toyoizumi, T., & Abbott, L. F. (2011). Beyond the edge of chaos: Amplification and temporal integration by recurrent networks in the chaotic regime. *Phys. Rev. E 84*(5), 051908.

Tripp, B., & Eliasmith, C. (2007). Neural populations can induce reliable postsynaptic currents without observable spike rate changes or precise spike timing. *Cereb. Cortex 17*(8), 1830–1840.

Trousdale, J., Hu, Y., Shea-Brown, E., & Josic, K. (2012). Impact of network structure and cellular response on spike time correlations. *PLoS Comput. Biol. 8*(3), e1002408.

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

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review 108*(3), 550.

van Albada, S. J., Helias, M., & Diesmann, M. (2015). Scalability of asynchronous networks is limited by one-to-one mapping between effective connectivity and correlations. *PLoS Comput. Biol. 11*(9), e1004490.

Van Essen, D. C. (2002). Windows on the brain: the emerging role of atlases and databases in neuroscience. *Curr. Opin. Neurobiol. 12*(5), 574–579.

Van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D., & Anderson, C. H. (2001). An integrated software suite for surface-based analyses of cerebral cortex. *Journal of the American Medical Informatics Association 8*(5), 443–459.

van Kerkoerle, T., Self, M. W., Dagnino, B., Gariel-Mathis, M.-A., Poort, J., van der Togt, C., & Roelfsema, P. R. (2014). Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex. *Proc. Natl. Acad. Sci. USA 111*(40), 14332–14341.

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

Vogel, K., Leiber, T., Risken, H., Hänggi, P., & Schleich, W. (1987). Locking equation with colored noise: Continued fraction solution versus decoupling theory. *Phys. Rev. A 35*, 4882–4885.

Voges, N., & Perrinet, L. U. (2012). Complex dynamics in recurrent cortical networks based on spatially realistic connectivities. *Front. Comput. Neurosci. 6*, 41.

Voges, N., Schüz, A., Aertsen, A., & Rotter, S. (2010). A modeler's view on the spatial structure of intrinsic horizontal connectivity in the neocortex. *Prog. Neurobiol. 92*(3), 277–292.

Voorhees, E. M. (1986). Implementing agglomerative hierarchic clustering algorithms for use in document retrieval. *Information Processing & Management 22*(6), 465–476.

Wainrib, G., & Galtier, M. N. (2016). A local echo state property through the largest lyapunov exponent. *Neural Networks 76*, 39 – 45.

Wedeen, V. J., Wang, R., Schmahmann, J. D., Benner, T., Tseng, W., Dai, G., Pandya, D., Hagmann, P., D'Arceuil, H., & de Crespigny, A. J. (2008). Diffusion spectrum magnetic resonance imaging (dsi) tractography of crossing fibers. *NeuroImage 41*(4), 1267–1277.

Weitzenfeld, A., Arbib, M. A., & Alexander, A. (2002). *The neural simulation language: A system for brain modeling.* MIT Press.

Wieland, S., Bernardi, D., Schwalger, T., & Lindner, B. (2015). Slow fluctuations in recurrent networks of spiking neurons. *Phys. Rev. E 92*(4), 040901.

Wilson, H. R., & Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J. 12*(1), 1–24.

Wilson, H. R., & Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik 13*, 55–80.

Wong, K.-F., & Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *J. Neurosci. 26*(4), 1314–1328.

Yger, P., El Boustani, S., Destexhe, A., & Frégnac, Y. (2011). Topologically invariant macroscopic statistics in balanced networks of conductance-based integrate-and-fire neurons. *J. Comput. Neurosci. 31*, 229–245.

Zhou, C., & Kurths, J. (2002). Noise-induced phase synchronization and synchronization transitions in chaotic oscillators. *Phys. Rev. Lett. 88*(23), 230602.

Zinn-Justin, J. (1996). *Quantum field theory and critical phenomena.* Clarendon Press, Oxford.

Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature 370*, 140–143.

# Author's list of publications

The work presented in this thesis is in parts based on the following publications:

### Modulated escape from a metastable state driven by colored noise

***by Jannis Schuecker, Markus Diesmann and Moritz Helias***
published in Physical Review E (Schuecker et al., 2015).

Parts of this publication enter Chapter 2 and 5 and the Discussion of the thesis.

### Reduction of colored noise in excitable systems to white noise and dynamic boundary conditions

***by Jannis Schuecker, Markus Diesmann and Moritz Helias***
published on arXiv (Schuecker et al., 2014).

Parts of this publication enter Chapter 2.

### Fundamental activity constraints lead to specific interpretations of the connectome

***by Jannis Schuecker*, Maximilian Schmidt*, Sacha J. van Albada, Markus Diesmann and Moritz Helias***
published in PloS Comput Biol (in press) (Schuecker et al., 2016b).

* These authors contributed equally.

Parts of this publication enter Chapter 4 and the Discussion of the thesis.

### Functional methods for disordered neural networks

***by Jannis Schuecker, Sven Goedeke, David Dahmen, and Moritz Helias***
published on arXiv (Schuecker et al., 2016).

Parts of this publication enter Chapter 7.

### Noise dynamically suppresses chaos in neural networks

***by Sven Goedeke*, Jannis Schuecker*, and Moritz Helias***

published on arXiv (Goedeke et al., 2016).

* These authors contributed equally.

Parts of this publication enter Chapter 7.

### Integration of continuous-time dynamics in a spiking neural network simulator

***by Jan Hahne*, David Dahmen*, Jannis Schuecker*, Andreas Frommer, Matthias Bolten, Moritz Helias, and Markus Diesmann***

published on arXiv (Hahne et al., 2016).

* These authors contributed equally.

Parts of this publication enter Chapter 8 and the Discussion of the thesis.

# Acknowledgment

First of all I would like to thank Moritz Helias, who guided me through the recent years with all his knowledge and experience, while leaving me enough freedom to explore my own research ideas. In particular, I enjoyed all the interactive meetings and fruitful discussions.
Furthermore, I would like to thank Markus Diesmann for sharing his endless experience in the science business with all its hidden mechanisms and unpredictabilities. Moreover, I thank him and Sonja Gruen for giving me the opportunity to work in such an interactive research group as the INM-6.
A special thanks goes to Maximilian Schmidt, Sven Goedeke, David Dahmen, Jan Hahne, and Hannah Bos with whom I worked together on various collaborative projects in a very creative, complementary, effective and exciting manner.
Moreover, I have benefited from endless discussions with Jakob Jordan on the technical but also on the philosophical aspects of neuroscience, which helped a lot for developing a deeper understanding and scientific intuition. I would like to thank Pietro Qualio, Vahid Rostami and Phillip Weidel for discussions on the train, especially about the philosophy of the spike and coding in the brain.
Many parts of this thesis rely on simulations and therefore I would like to thank Bernd Wiebelt, Tiziano Zito and Michael Bontenackels for all their help with hardware and software problems as well as taking good care of our computer clusters. Furhermore, I thank Verena, Janine, Steffi and Petra for helping me to deal with all the bureaucratic hurdles.
A Special thanks also goes to Lena for all her patience and empathy especially in the final writing period of this thesis.

# Funding

This work was supported by the Helmholtz young investigator's group VH-NG-1028, the Helmholtz Portfolio Supercomputing and Modeling for the Human Brain (SMHB), the Jülich Aachen Research Alliance (JARA) and the Next-Generation Supercomputer Project of MEXT. The work further received funding from the European Union under grant agreement No 269912 (BrainScaleS), the Seventh Framework Programme (FP7/2007– 2013) under grant agreement No 604102 (HBP RUP), and the Horizon 2020 research and innovation programme under grant agreement No 720270 (HBP SGA1).

Computing time on JUQUEEN was granted by the JARA-HPC Vergabegremium and provided on the JARA-HPC Partition part of the supercomputer JUQUEEN (Jülich Super- computing Centre, 2015) at Forschungszentrum Jülich (VSR computation time grant JINB33).