Correlated fluctuations in neural networks on the microscopic and mesoscopic scale

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
David Dahmen
aus Geilenkirchen

Berichter: Univ.-Prof. Dr. rer. nat. Markus Diesmann
Univ.-Prof. Dr. rer. nat. Andreas Offenhäusser

Tag der mündlichen Prüfung: 07.07.2017

Zusammenfassung

Die Untersuchung der Multiskalen-Dynamik kortikaler Netzwerke ist ein wesentlicher Schritt zum Verständnis der Funktionsweise des Gehirns. Moderne Multielektrodentechnologie ermöglicht die Ableitung von massiv-parallelen Hirnsignalen auf der Ebene von Aktionspotentialen (Spikes) und lokalen Feldpotentialen (LFPs). Beide Signale sind verwandt und zeigen die Hirndynamik auf verschiedenen räumlichen Skalen. Während Aktionspotentiale die Aktivität einzelner Nervenzellen darstellen, ist das LFP ein großflächiges Summensignal mit Beiträgen von Tausenden oder gar Millionen von Neuronen. Sowohl die Spikes einzelner Zellen als auch das LFP resultieren aus der koordinierten Aktivität vieler Neuronen in komplexen Netzwerken. Diese Aktivität ist sowohl durch die Verbindungsstruktur als auch durch die intrinsische Dynamik der einzelnen Zellen bestimmt.

Die vorliegende Arbeit beruht auf einem konnektionistischen Ansatz für die mathematische Modellierung von Spike-Aktivität und LFPs. Eine formale Reduktion von Nervenzellen auf einfache Berechnungseinheiten erlaubt die Fragestellung, welche Aspekte von experimentell beobachtbarer Aktivität sich aus deren Zusammenschaltung ergeben. Im Fokus steht die Untersuchung von korrelierten Fluktuationen in Netzwerken, die grundlegende Eigenschaften von Verbindungsstrukturen des Gehirns aufweisen. Diese Netzwerke bestehen aus exzitatorischen und inhibitorischen Neuronen, welche jeweils die Aktivität von Hunderten bis Tausenden von anderen Neuronen des Netzwerks als Eingangssignal empfangen und verarbeiten. Ein Gleichgewicht von Exzitation und Inhibition in solchen Netzwerken resultiert in einer Aktivität der Neurone, die durch Fluktuationen im Eingangssignal bestimmt ist. Molekularfeldtheorien nutzen dies und nähern letzteres als ein effektives Feld mit Gaußscher Statistik um die Aktivität und Korrelationen auf der Ebene von Populationen von Neuronen zu beschreiben. Diese Arbeit geht über die Populationsbeschreibung und die Approximation des Gaußschen Eingangssignals hinaus und erweitert die angewandte Molekularfeldtheorie um Korrekturen, die aus der endlichen Netzwerkgröße resultieren. Diese Korrekturen erklären die experimentell beobachtete Statistik der Korrelationen, insbesondere deren Mittelwert und Varianz über Neuronenpaare. Als Ausgangspunkt dafür dient ein zufällig verbundenes Netzwerk aus Ising Spins, welches ein klassisches Modell für kollektive Phänomene in ungeordneten Systemen darstellt. Es folgt die Herleitung von äquivalenten ratenbasierten Netzwerkmodellen, welche mit Funktional-Methoden aus der statistischen Physik und Feldtheorie analysiert werden. Zur Validierung der theoretischen Reduktion Spike-basierter auf ratenbasierte Neuronenbeschreibungen wird eine Methode

hergeleitet, welche die effiziente Integration von ratenbasierter Dynamik in einen Spike-basierten Netzwerk Simulator erlaubt. Diese schafft die Grundvoraussetzung für die Entwicklung von Multiskalen-Modellen von Hirnaktivität, die beide Beschreibungsebenen beinhalten. Den Abschluss der Arbeit bildet ein Modell eines bestimmten balancierten Netzwerks, des kortikalen Mikroschaltkreises, welches hier der Fragestellung dient, wie sich mikroskopische Dynamik für spontane und stimulierte Aktivität im LFP widerspiegelt. Das Modell dient desweiteren zur Illustration eines Hybrid-Schemas zur Modellierung von LFPs aus Netzwerken von Spike-basierten Punkt-Neuronen und dessen Implementation im frei verfügbaren Software-Paket `hybridLFPy`. Zusammenfassend verbindet diese Arbeit molekularfeldtheoretische Methoden mit Vorwärtsmodellen um eine Brücke zwischen mikroskopischer und mesoskopischer Hirnaktivität zu schlagen und Modellvorhersagen zu machen für in Experimenten zunehmend messbare Aspekte kortikaler Dynamik auf verschiedenen Skalen.

Summary

There is broad consent that understanding the brain's function relies on the investigation of the multi-scale dynamics of cortical networks. Rapidly advancing multi-electrode technology for recordings of extracellular brain activity allows monitoring of massively parallel spiking activity and local field potentials (LFPs). Both signals are related, in fact, they show brain dynamics on different spatial scales. While the spiking activity represents the output signal of individual neurons, the LFP is a spatially coarse-grained population measure with contributions from thousands up to millions of cells. The spike output as well as the LFP largely result from coordinated activity between neurons in complex networks. This activity is determined by the structure in network connections as well as intrinsic single-neuron dynamics.

Proper understanding of the spiking activity and the LFP in recurrent networks requires mathematical modeling. In this thesis, we employ a connectionist modeling approach by regarding neurons as simplified computational units, and investigate which features of experimentally observed activity arise from the connectivity between such units. In particular, we concentrate on correlated fluctuations as arising in networks which fulfill the basic features of brain connectivity, that is, we consider networks of excitatory and inhibitory neurons, each receiving inputs from hundreds to thousands of other neurons in the network. In this setting, a balance between excitation and inhibition causes activity which is driven by fluctuations in the total input. Mean-field theories have employed this fact and approximated the input by effective Gaussians to explain activity and correlations on the level of neuronal populations. Here, we go beyond the population level and the approximation of Gaussian input by extending the applied mean-field theory to finite-size corrections that capture the experimentally observed first- and second-order statistics of correlations across neurons. Starting from randomly coupled Ising spins which constitute the classical model of collective phenomena in disordered systems, we derive equivalent networks of rate-based models, which can be analyzed using functional methods from statistical physics and field theory. In order to validate mean-field theoretical mappings between spiking and rate-based descriptions of neurons, and to foster the development of multi-scale models of brain activity combining both, we provide a simulation framework to efficiently integrate rate-based neuron dynamics in a spiking network simulator. Finally, we discuss a particular type of balanced network, the cortical microcircuit, and show how the microscopic dynamics for spontaneous and evoked activity are reflected in the LFP. Thereby we illustrate a generally applicable hybrid scheme for modeling local field potentials from spiking

point-neuron networks which we derive and implement in the open-source software package `hybridLFPy`. In conclusion, this thesis combines mean-field theoretical and forward-modeling methods to build a bridge between microscopic and mesoscopic brain activity and derive model predictions for cortical dynamics that become increasingly accessible in experiments.

Author's list of publications

The work presented in this thesis is in parts based on the following publications:

Correlated Fluctuations in Strongly Coupled Binary Networks Beyond Equilibrium

by David Dahmen, Hannah Bos, and Moritz Helias
published in Physical Review X (2016a).

Preliminary results have been published in Dahmen et al. (2016b).

Parts of this publication enter Chapter 1 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 (2016).

Parts of this publication enter Chapter 2.

Distributions of covariances as a window into the operational regime of neuronal networks

by David Dahmen, Markus Diesmann, and Moritz Helias
published on arXiv (2016b).

Preliminary results have been published in Dahmen et al. (2015, 2016a).

Parts of this publication enter Chapters 2 and 3.

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 in Frontiers of Neuroinformatics (2017) and on arXiv (2016).

* These authors contributed equally.

Parts of this publication enter Chapter 4 and the Discussion of the thesis.

Hybrid Scheme for Modeling Local Field Potentials from Point-Neuron Networks

by Espen Hagen[*]*, David Dahmen*[*]*, Maria L. Stavrinou, Henrik Lindén, Tom Tetzlaff, Sacha J. van Albada, Sonja Grün, Markus Diesmann, and Gaute T. Einevoll*
published in Cerebral Cortex (2016).

[*] Both authors contributed equally.

Preliminary results have been published in Hagen et al. (2013); Dahmen et al. (2013); Stavrinou et al. (2014); Dahmen et al. (2014); Hagen et al. (2015a,b).

Parts of this publication enter Chapters 5 and 6, and the Discussion of the thesis.

Acknowledgements

I would like to thank those who contributed to this thesis, in particular Markus Diesmann and Sonja Grün for giving me the opportunity to enter computational neuroscience and work at the INM-6. Their vision, support, and trust in me always encouraged me throughout this work. I am very grateful to Markus for his great commitment in the doctoral supervision and thank Prof. Andreas Offenhäusser for being the second assessor of this dissertation. I would like to express my special gratitude to Tom Tetzlaff and Moritz Helias who reliably guided me during the recent years. Without their immense knowledge, constant support and their contagious passion for science, this dissertation would have never been possible. I thank them for entrusting me with important and interesting projects, but also for giving me the freedom to explore the field, develop my own ideas and find my personal research profile. I am very grateful for sharing their intuition and enthusiasm, and for countless insightful discussions. I owe a lot to my colleagues Espen Hagen, Hannah Bos and Jannis Schücker, whose expertise, persistence, thoroughness, and helpful open manner was essential in promoting our collaborative research projects. I further thank Jan Hahne for introducing me to the insides of NEST and the numerics of stochastic differential equations. I have benefited greatly from countless discussions with Jakob Jordan, Maximilian Schmidt, Sven Goedeke and Johanna Senk, and a considerable amount of credit also goes to Sacha van Albada, Henrik Lindén and Maria Stavrinou whose work was an essential help in shaping my ideas. I am grateful in particular to Gaute Einevoll whose humorous way and view on neuroscientific problems from a physicist's perspective inspired me a lot. Many parts of this thesis are based on numerical computations which would not have been possible without the efforts of the NEST initiative, the NEURON and LFPy simulation software, and the assistance of Jochen Eppler, Bernd Wiebelt, Tiziano Zito, Michael Bontenackels and Alper Yegenoglu. Furthermore, I would like to thank Verena, Janine, Steffi, Petra and Martina for helping me to deal with all sorts of bureaucratic matters that arise in a German research center which is involved in EU funded large-scale initiatives. I thank good old friends and all colleagues at INM-6 who became new friends. I enjoyed the nice and lively working atmosphere a lot, especially the Kicker matches after lunch. I am particularly thankful to my family, who always gave me confidence, courage and freedom, and who supported me in every possible way. Special thanks go to my beloved wife Alexandra and my children for their endurance and mental support, especially in the final phase of this dissertation.

Funding

This work was supported by the Initiative and Networking Fund of the Helmholtz Association, 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), the Next-Generation Supercomputer Project of MEXT, the German Research Foundation (DFG; grant DI 1721/3-1 [KFO219-TP9]), the Danish Council for Independent Research and FP7 Marie Curie Actions CO-FUND (grant id: DFF-1330–00226), and the Research Council of Norway (NFR, through ISP, NOTUR -NN4661 K). 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 the supercomputer JUQUEEN at Forschungszentrum Jülich was granted by firstly the JARA-HPC Vergabegremium (provided on the JARA-HPC partition, jinb33) and secondly by the Gauss Centre for Supercomputing (GCS) (provided by the John von Neumann Institute for Computing (NIC) on the GCS share, hwu12).

Contents

Part I

Introduction

The cerebral cortex is an interacting many-body system with a complex structural architecture and strong diversity in the dynamics of its constituents (Abeles, 1991; Connors & Gutnick, 1990; Braitenberg & Schüz, 1998; Gibson et al., 1999). The interplay between structure and dynamics gives rise to collective behavior on various spatio-temporal scales which underlies many functions of the brain such as perception, learning, language, and motor initiation (Arbib et al., 1998; Bressler & Kelso, 2001; Jirsa, 2004; Gerstner et al., 2014; Wang & Kennedy, 2016). Understanding coordinated brain dynamics in the healthy brain proves useful for technological applications such as brain-machine interfaces (Nicolas-Alonso & Gomez-Gil, 2012) and artificial intelligence (Haykin, 2009), and is a prerequisite to comprehend brain diseases (Holt & Netoff, 2013).

Anatomical and electrophysiological studies reveal a hierarchical organization of cortex (Abeles, 1991), from subcellular structures, single cells, small circuits up to whole cortical areas (Fig. 1). Each scale relates to neuroscience data, from subcellular patch-clamp recordings, single-unit and multi-unit spiking activity (EAP), voltage-sensitive dye (VSD) and calcium imaging, local field potentials (LFP) to functional magnetic resonance imaging (fMRI), the electroencephalogram (EEG), and the magnetoencephalogram (MEG) (Brette & Destexhe, 2012). These signals show complex dynamics, such as synchronizations (Torre et al., 2016), oscillations (Cabral et al., 2014) and waves (Rubino et al., 2006). Their characterization typically rests on the quantification of correlations which aims at an interpretation of interactions of neural systems on the respective scale.

It is widely believed that the primary information exchange in cortex is given by discrete events between nerve cells (*neurons*) via anatomical connections (*synapses*). These events are electric pulses called action potentials (*spikes*). The cell membrane of neurons acts as an electrical insulator which creates a difference in electrical potential between the intracellular and extracellular space, the *membrane potential*. Spikes are generated as large excursions of the membrane potential at the soma or axon hillock, propagated along the axon and translated into chemical signals at synapses via release of neurotransmitters. At the postsynaptic terminal of the chemical synapses, the neurotransmitters cause a deflection of the membrane potential which is propagated along the dendrites to the soma where the cell integrates signals from many of these events to finally fire an action potential if the membrane gets sufficiently depolarized (Fig. 2). Sequences of spike events of a single neuron are referred to as *spike trains,* the frequency of spike events as the *firing rate*.

In addition to this interaction mediated by chemical synapses, there is evidence for direct electrical synapses (*gap junctions*) between neurons in cortex (Fukuda, 2007). While gap junctions are the main communication means between a second type of brain cells (*glial cells*) that are involved in plasticity and homeostatic processes (Amiri et al., 2012; De Pittà et al., 2016), their role for cortical neurons is less clear. Gap junctions between dendrites are mostly attributed a modulatory effect in synchronizing the neuronal dynamics in pathological cortical network states (Jefferys, 1995; Dere &

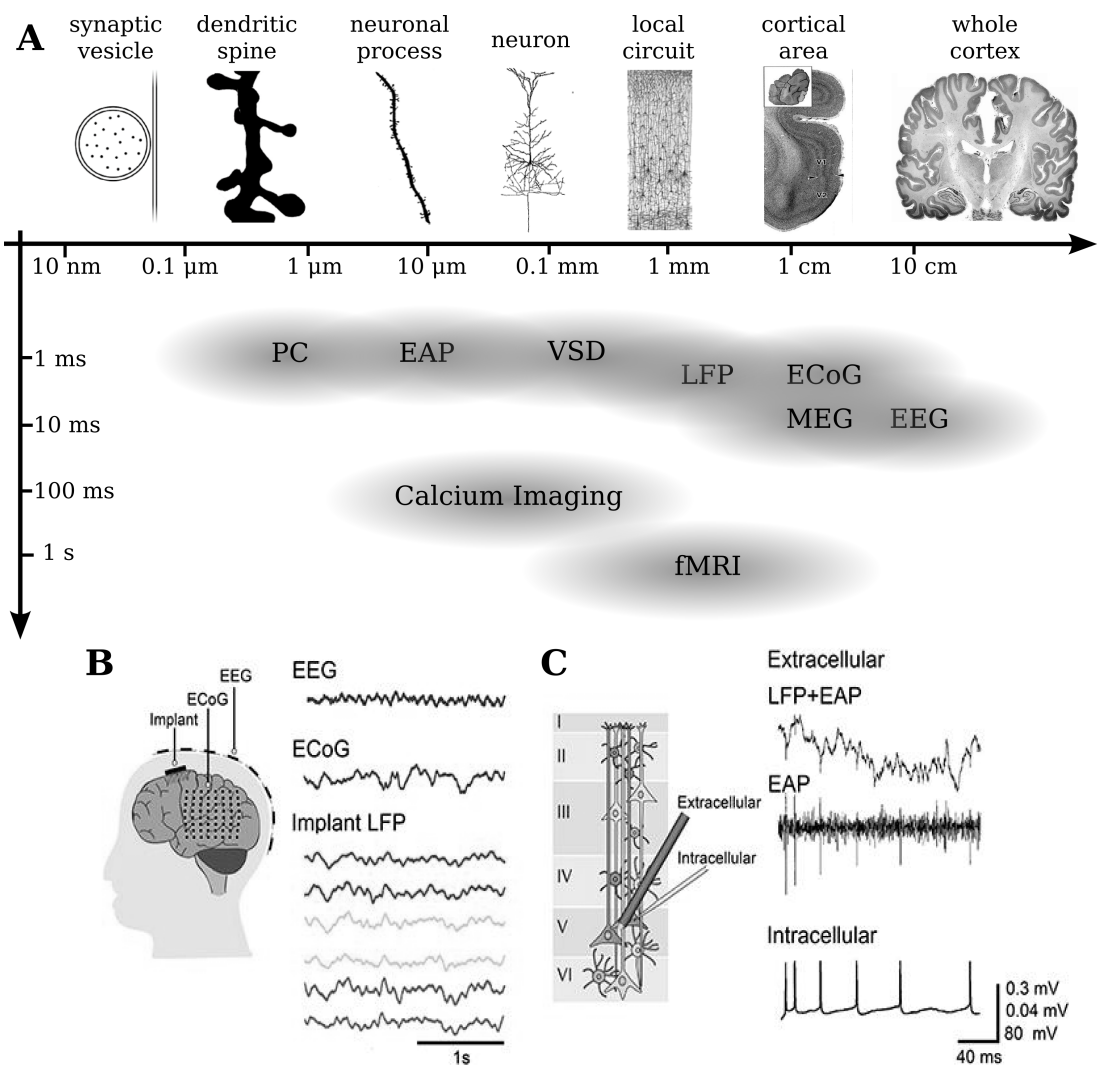


Figure 1: **Multi-scale structure of cortex and measurement modalities. A** Top: biological structures at different spatial scales (horizontal axis). Bottom: spatial (horizontal axis) and temporal resolution (vertical axis) of different measurement modalities of brain activity. Intracellular electric potentials (postsynaptic potentials, synaptic currents, action potentials) are measured with patch-clamp (PC) or sharp electrode recordings. Extracellular electric potentials are measured inside the cortical tissue (extracellular action potentials: EAP, local field potential: LFP), on the brain surface (electrocorticography: ECoG) or at the scalp (electroencephalography: EEG). Non-electric signals include the calcium concentration (Calcium Imaging), the luminance of voltage-sensitive dyes (VSD), magnetic fields (Magnetoencephalography: MEG) or the blood-oxygen-level-dependent (BOLD) signal in functional magnetic-resonance imaging (fMRI). **B** Sketch of electrode arrays in the cortical tissue (Implant LFP), at the brain surface (ECoG) or the scalp (EEG) with time traces of the respective signals (for details on measurement locations see Obien et al. (2015)). **C** Sketch of intracellular recording using sharp electrode and extracellular recording with time traces of intracellular potential (Intracellular), raw extracellular potential (LFP+EAP) and high-frequency component of extracelluar potential (EAP) (for details on measurement locations see Obien et al. (2015)). Figure panel **A** adapted with permission from Mascaro et al. (2015); Preuss & Coleman (2002); DeFelipe (2011), **B** and **C** adapted with permission from Obien et al. (2015).

Zlomuzica, 2011). Axo-axonal gap junctions, that are electrical synapses between the axons of different neurons, could potentially cause fast and strong coupling independent of somatic integration of inputs (Schmitz et al., 2001). The existence and extent of such coupling in cortex, however, remains unclear (Feldmeyer & Lübke, 2010).

An additional coupling between neurons arises from the signal propagation along axons, dendrites and synapses which involves ionic and capacitive currents across the membrane. These currents generate electric and magnetic fields that influence the dynamics of other cells on short time scales (Anastassiou et al., 2011). The extent of such *ephaptic couplings* between neurons under physiological activity is still debated (Jefferys, 1995; Anastassiou et al., 2010), and mostly assumed to be a weak global perturbation in comparison to direct communications via spikes (Fröhlich & McCormick, 2010). Furthermore, at longer time scales, diffusion of ions and activity-associated chemical cotransmitters (*neuromodulation*) in the extracellular medium lead to non-local feedback between larger groups of neurons (Halnes et al., 2013; Katz & Frost, 1996).

In this thesis, we only consider physiological neuronal activity in cortex at short time scales and therefore focus on spike interactions between neurons. The single-cell level can then be regarded as the lowest level where information is exchanged. All other measures of cortical activity follow as epiphenomena that are determined by the spiking activity of the network. Full knowledge of the spikes of all neurons would therefore allow to determine any other brain signals using forward models (Einevoll et al., 2013).

Cortical networks can be represented as directed graphs of dynamical nodes and weighted edges representing neurons and anatomical connections, respectively. The nodes are embedded in physical space and interactions are delayed due to signal transmission times (Thomson & Bannister, 2003). In general, connections are plastic with changes related to learning (Konorski, 1948; Hebb, 1949). In this thesis, we, however, restrict ourselves to study static networks. The multi-scale organization of cortex is reflected by clusters in the graph on various levels which result from the dynamical properties of nodes, their spatial arrangement and the patterns of connections. While there is large heterogeneity in the dynamics of nodes (Connors & Gutnick, 1990), there is a striking classification of neurons based on their outgoing connections given by Dale's law which states that neurons either excite or inhibit all of their targets (Eccles et al., 1954). While the majority of neurons (roughly 80%) in cortex is excitatory (Braitenberg & Schüz, 1991), a balance between excitation and inhibition in cortical networks has been observed experimentally (Haider et al., 2006; Dehghani et al., 2016). Cortical connectomics further reveal a preference of connections between cells of similar spatial locations. This leads to a clustering of the network graph which reflects the laminar, lateral and areal organization of cortex (Potjans & Diesmann, 2014; Senk et al., 2015; Schmidt et al., 2016a; Wang & Kennedy, 2016).

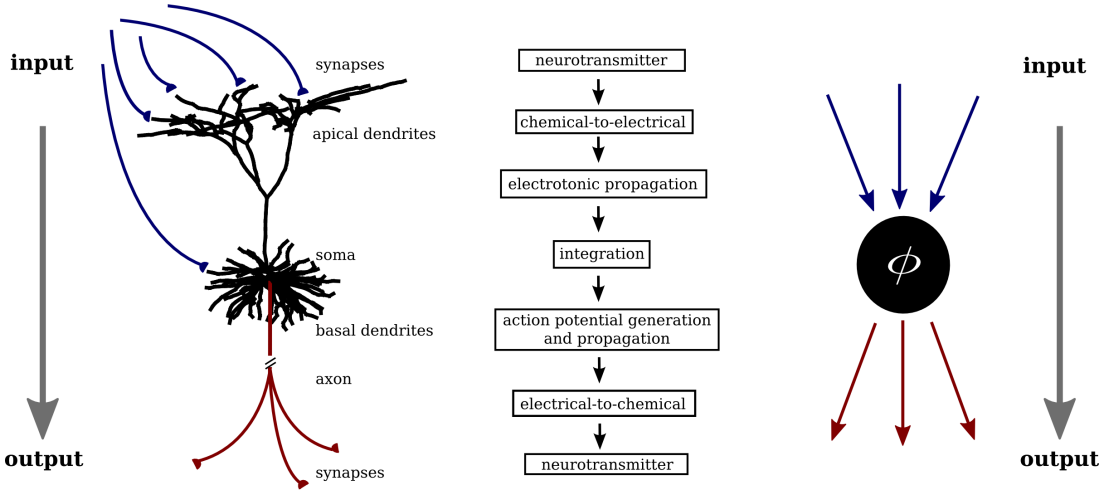


Figure 2: **Information transfer within and between neurons in spike communication.** Left: illustration of neuron (black morphology from Mainen & Sejnowski 1996) receiving synaptic input which gets propagated to the soma and transformed into output spikes that propagate along axons to synapses of postsynaptic targets. Middle: cascade of input-to-output conversion of single neurons. Right: simplification of neuron as connectionist information processing unit with predefined input-to-output conversion ϕ. Figure idea based on Marino (1994) and McLeod et al. (1998).

Computational models at different scales aim at understanding the fundamental mechanisms of neural processes and their relation to neuroscience data. While microscopic models consider the dynamics of individual neurons embedded in complex networks, mesoscopic and macroscopic models focus on the emergent coarse-grained dynamics as resulting from effective interactions between clusters of neurons in small circuits, cortical columns or even whole cortical regions (Deco et al., 2008). Top-down modeling approaches thereby impose a desired function on a neural system and infer how the system needs to be organized such that this function can be implemented with biologically-grounded components (Griffiths et al., 2010). In contrast, bottom-up approaches define the dynamics of the neural constituents and let collective phenomena underlying functions emerge from their complex interplay (McClelland et al., 2010). While both modeling approaches are critical in integrating theory and experiments, this thesis focuses on the bottom-up approach (Rumelhart et al., 1986). Considering individual neurons with biologically-grounded, but simplified dynamics as the basic units, we assess which features in experimental data on the microscopic and mesoscopic scale can be explained as arising from the network structure.

Neuron models

In a connectionist modeling approach (Hebb, 1949; Rumelhart et al., 1986; McLeod et al., 1998), neurons are considered to be simple computational units which receive inputs, integrate information and produce output (Fig. 2). Since spikes are brief events with rather fixed waveforms (Fee et al., 1996), neurons essentially have two different states in terms of their output, the active state when a spike is elicited and the inactive state when no spike is generated. The biophysics of neurons which determines the dynamics that lead to transitions between these two states is approximated differently in a variety of neuron models (Gerstner & Kistler, 2002).

A simple neuron model is the stochastic binary neuron (Hinton & Sejnowski, 1983; van Vreeswijk & Sompolinsky, 1996). The binary states of neurons are updated asynchronously with transition probabilities determined by the activity of presynaptic neurons at the time of the random updates. The resulting Markovian network dynamics, called Glauber dynamics (Glauber, 1963), has originally been proposed for the Ising model of spins, and later successfully employed in neuroscience (Ginzburg & Sompolinsky, 1994; van Vreeswijk & Sompolinsky, 1996; Renart et al., 2010). A detailed description of the model is given in Chap. 1.

More biologically realistic neuron models derive spike times from the dynamics of the somatic membrane potential, which, in turn, is determined by synaptic inputs and the intrinsic cell dynamics as derived from the biophysics of the synapse and the membrane (Hille, 1992; Koch, 1999). Biologically detailed models, such as the *Hodgkin-Huxley neuron model* (Hodgkin & Huxley, 1952), consider the dynamics of specific ion channels through gating variables which determine conductances in analogous electrical circuits. The resulting membrane dynamics, which follows from the law of current conservation in these circuits containing capacitances, leak resistances, as well as sodium (Na^+), potassium (K^+), chloride (Cl^-), and calcium (Ca^+) conductances, generates action potentials of realistic waveform.

An intermediate level of complexity is obtained by linearization of the channel dynamics around the equilibrium potential, or Nernst potential, which is obtained in the absence of synaptic input (Abbott & Kepler, 1990). The resulting *leaky integrate-and-fire neuron model* (LIF, Lapicque (1907); Stein (1967)) has linear subthreshold dynamics and relies on an artificial threshold at which spikes are elicited and the membrane potential is reset to a lower value. Spike after-depolarization is often modeled as a refractory period in which the membrane potential is clamped to a fixed value below threshold. Details of this model are given in Tab. 5.1. A summary of other neuron models can be found, for example, in Gerstner & Kistler (2002), Izhikevich (2004) or Gerstner & Naud (2009). Synaptic inputs are generally modeled as changes in currents or conductances (see e.g. Burkitt, 2006). While current-based synapse models have a fixed temporal activation pattern following input spikes, the shape of the current response in conductance-based synapse models depends on the membrane potential. This dependence causes nonlinear interactions between inputs, even for models with linear membrane dynamics.

The above mentioned neuron models assume an isopotential membrane and thereby neglect the spatial aspects of neuronal morphologies, effectively describing neurons as points in space (*point-neuron models*). Despite their simplicity, point-neuron-network models explain a variety of salient features of neural activity observed *in vivo,* such as spike-train irregularity (Softky & Koch, 1993; van Vreeswijk & Sompolinsky, 1996; Amit & Brunel, 1997; Shadlen & Newsome, 1998), membrane-potential fluctuations (Destexhe & Paré, 1999), asynchronous firing (Ecker et al., 2010; Renart et al., 2010; Ostojic, 2014; Brunel, 2000), correlations in neural activity (Gentet et al., 2010; Okun & Lampl, 2008; Helias et al., 2013), self-sustained activity (Ohbayashi et al., 2003; Kriener et al., 2014), rate distributions across neurons (Griffith & Horn, 1966; Koch & Fuster, 1989; Roxin et al., 2011) and across laminar populations (Potjans & Diesmann, 2014), as well as resting state activity (Deco & Jirsa, 2012). Point-neuron networks are amenable to mathematical analysis (see, e.g., Brunel (2000); Deco et al. (2008); Tetzlaff et al. (2012); Helias et al. (2013); de Kamps (2013); Bos et al. (2015); Schuecker et al. (2015)) and can be efficiently evaluated numerically (Plesser et al., 2007; Brette et al., 2007; Helias et al., 2012; Kunkel et al., 2014). Modeling large-scale neural-network dynamics with individual spiking neurons is challenging due to the memory required to represent the large number of synapses. With current technology and using the largest supercomputers available today, simulations of neural networks comprising up to 10^9 neurons and 10^{13} synapses (roughly corresponding to the size of a cat brain) are feasible for point-neuron models (Diesmann, 2013; Kunkel et al., 2014).

An alternative modeling approach focuses more on the level of individual biophysically detailed cells (Segev et al., 1989). The dynamics within specific subcellular domains can be integrated using neuronal cable theory for *multicompartment-neuron models* (Rall, 1964), where the morphology is divided into small compartments, each modeled as an electrical circuit with nearest-neighbor resistive coupling (Fig. 3). In the continuous limit of infinitesimal compartment size, the dynamics is governed by the cable equation, which is a one dimensional diffusion equation (see e.g. De Schutter & Van Geit, 2009). The discretized models allow the incorporation of specific ion channel concentrations across the morphology and take into account the filtering properties of dendrites (Lindén et al., 2010). Similarly to the linearization of conductances from the Hodgkin-Huxley model to the leaky integrate-and-fire model, the dynamics of each compartment can be approximated. Passive dendrites thereby refer to fully linear dynamics. Active dendrites consider nonlinear dynamics arising from variable ion channel conductances. Multicompartment-neuron models have been used, for example, to study dendritic computation (London & Häusser, 2005; Hay et al., 2011) and information processing in single cells (Poirazi & Mel, 2001; Poirazi, 2009), as well as in network studies (Reimann et al., 2013; Migliore et al., 2014; Markram et al., 2015) and studies of extracellular potentials (see below). The increase in model complexity, however, comes with certain disadvantages. The mechanisms governing networks of biophysically detailed multicompartment model neurons are less accessible to theoretical analysis and these models are more prone to overfitting.

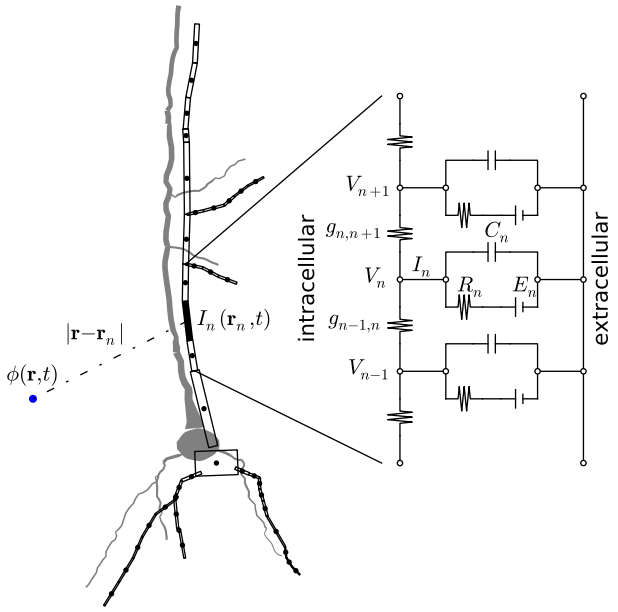


Figure 3: **Illustration of the cable and forward-modeling formalism.** In multicompartment-neuron modeling, the neuronal morphology (gray) is divided into small compartments (black rectangles) which are represented by electrical circuits that are connected in parallel. For passive dendrites, these circuits are composed of capacitances generating capacitive transmembrane currents, as well as conductances and voltage sources connected in series, which mimic leak currents across the membrane and reversal potentials, respectively. In the forward-modeling scheme of extracellular potentials, the sum of all transmembrane currents in each compartment n is weighted with the inverse of the distance $|\mathbf{r} - \mathbf{r}_n|$ to an extracellular electrode at position $\mathbf{r}$ (blue dot) to predict the LFP ϕ. Figure adapted with permission from Lindén et al. (2014).

Due to the increase in dynamical variables per neuron, existing multicompartment-neuron-network models accounting for realistic cell morphologies are restricted to sizes of $\sim 10^4$–10^5 neurons (Hines et al., 2008; Markram et al., 2015).

Forward modeling of meso-scale observables

Recent advances in electrode technology in combination with spike-sorting algorithms allow the inference of single-unit spiking activity from extracellular recordings (Buzsaki, 2004). The latter can be performed with multi-electrodes that measure extracellular potentials at different spatial locations. Prominent examples are linear multi-electrodes inserted perpendicular to the brain surface to record activity from all layers of cortex, or two-dimensional arrays that record signals from multiple locations in the lateral direction (Fig. 4). Implantations of these multi-electrodes are more

stable than intracellular recordings, which record the membrane potential from individual neurons. Modern "Utah" electrode arrays (Blackrock Microsystems, Salt Lake City, UT, USA) allow detection of spike events in the high-frequency component ($\gtrsim 500$ Hz) of the extracellular potential and wave-form based sorting into spike trains of up to hundreds of neurons (Rossant et al., 2016). However, the result is still a massive undersampling of the neural activity of the hundred thousands up to millions of neurons that form the network in the tissue surrounding the electrodes.

The local field potential (LFP), the low-frequency component ($\lesssim 500$ Hz) of the extracellular potential recorded in the brain, has also been commonly used as a measure of neuronal activity (Buzsáki et al., 2012; Einevoll et al., 2013). The LFP mainly originates from transmembrane currents across dendrites and somas (Nicholson & Freeman, 1975), and at the single-cell level the biophysical origin of such extracellular potentials is well understood (see, e.g., Buzsáki et al. (2012); Einevoll et al. (2013) and Chap. 5). Biophysically plausible LFP models (Einevoll et al., 2007; Pettersen et al., 2008; Lindén et al., 2010) employ a forward-modeling scheme based on volume conduction theory (Rall & Shepherd, 1968; Holt & Koch, 1999) that allows to predict LFPs from transmembrane currents of multicompartment-neuron models in response to spike input (see e.g. Lindén et al., 2014). This formalism has also been used for predictions of extracellular signatures of spikes (Holt & Koch, 1999; Gold et al., 2006, 2007; Schomburg et al., 2012; Reimann et al., 2013). In particular, the spatial reach of the high-frequency components of the extracellular potential was shown to be a few tens of micrometers (Pettersen & Einevoll, 2008) explaining the large subsampling of spike data in multi-electrode recordings. In contrast and in line with experimental findings (Kajikawa & Schroeder, 2011), the LFP, the low-frequency components, was shown to extend to several millimeters, depending on the amount of correlations in the spiking activity (Lindén et al., 2011; Łęski et al., 2013). This renders the LFP a mesoscopic signal. Similar forward models have been developed for other mesoscopic measurements, such as the ECoG and EEG (Nunez & Ramesh, 2006), the MEG (Hämäläinen et al., 1993), and the VSD (Chemla & Chavane, 2010).

Despite the forward-modeling scheme, the interpretation of the LFP remains difficult due to the large number of neurons contributing to the recorded signal. In cortex, the thousands or even millions of neurons near the recording electrode form microcircuits, that are multi-layered networks of excitatory and inhibitory neuron populations, which are considered to be the basic computational circuits in the cortex (Douglas et al., 1989; Douglas & Martin, 2004). Moreover, the LFP reflects synaptic input also generated by remote populations, e.g., inputs from other cortical or subcortical areas in addition to local network interactions (Herreras et al., 2015). The LFP is therefore a meso-scale measure of the dynamics of microcircuits and their directed interaction with other cortical regions.

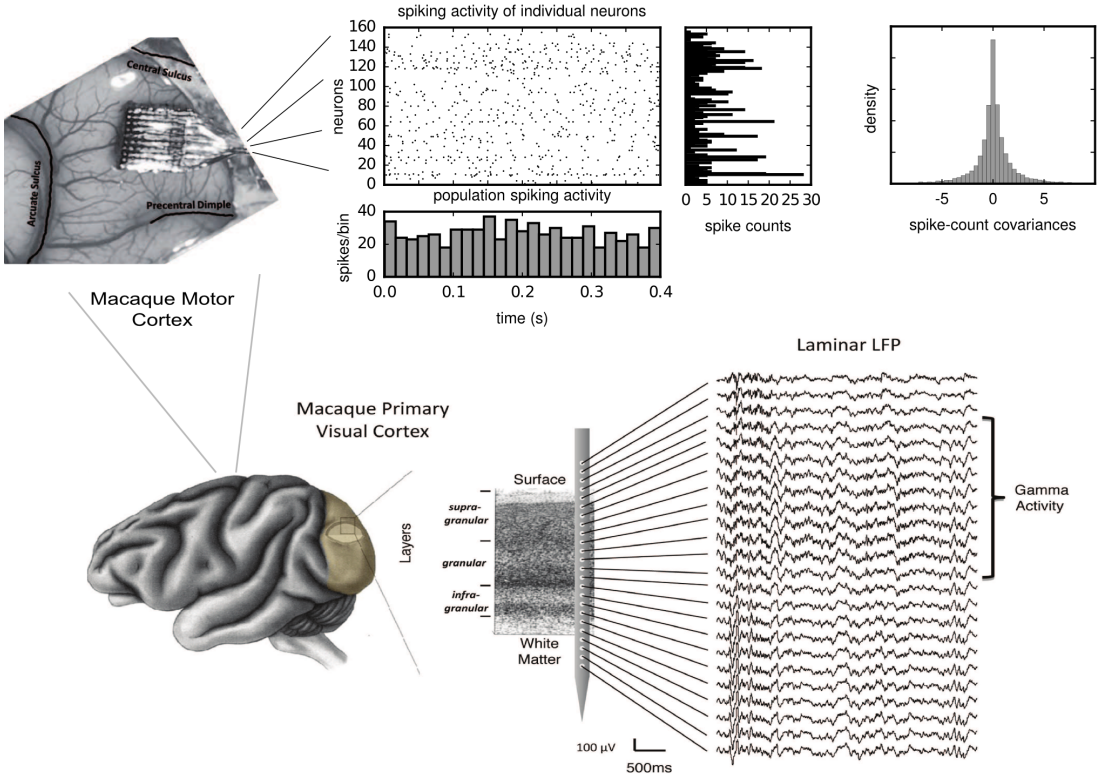


Figure 4: **Microscopic and mesoscopic brain activity.** *Top:* spiking activity (raster plot) of 155 neurons mostly located in layer 5 of macaque motor cortex (M1) and recorded with a 100-electrode "Utah" array (Blackrock Microsystems, Salt Lake City, UT, USA; left panel) with 400 μm interelectrode distance, covering an area of 4×4 mm^2. Data from $T = 400$ ms after trial start (TS) of a reach-to-grasp task (Riehle et al., 2013, subsession: i140703-001) with rather stationary spiking statistics as shown by the population activity (gray horizontal histogram), which is obtained as the summed and binned spike train of all neurons in 16ms bins. Spike counts within $T = 400$ ms show large variability across neurons (black vertical histogram) and trials (only one trial is shown). Spike-count covariances across trials (right panel) are widely distributed with mean close to zero. Picture of Utah array adapted with permission from Riehle et al. (2013). *Bottom:* local field potential across layers in macaque primary visual cortex (V1) recorded with laminar multielectrode. The depth-resolved LFP shows strong coherence across channels and varying spectral properties across layers. Bottom part of figure adapted with permission from Maier et al. (2010).

Mean-field theory for meso-scale dynamics

The interpretation of meso-scale activity measures relies on the understanding of the collective behavior of the large number of neurons contributing to these signals. Mean-field theory abstracts from the level of single-unit dynamics to a statistical description in terms of ensembles in order to simplify analytically intractable problems by approximating the statistics of inputs to neurons. Typically ensembles refer to neurons with similar properties defining neuronal populations such that the theory captures the population-averaged activity (see e.g. Sompolinsky et al., 1988; Brunel, 2000).

Mean-field theory originates from the theory of ferromagnets developed by Curie and Weiss. A ferromagnet is modeled as an infinite spin lattice with nearest-neighbor coupling (Fig. 5). All sites on the lattice therefore have the same properties and, hence, the same statistics of dynamics. The theory of Curie and Weiss is termed mean-field theory because the magnetic field that each spin experiences as a superposition of spins of neighboring sites is approximated by its mean, thus neglecting fluctuations. In the framework of statistical mechanics, this formally amounts to a saddle-point approximation of an auxiliary field which corresponds to the magnetic field discussed above (Negele & Orland, 1998).

Neuronal network models, much like spin glasses (Fischer & Hertz, 1991a), are generalizations of this model of a ferromagnet (Fig. 5). Despite disorder in connections, neurons in the same populations receive input with similar statistics. The variability of the latter can be formally captured by employing an ensemble average over realizations of the disordered connectivity (Sherrington & Kirkpatrick, 1975). In this ensemble description, a neuron receives input from presynaptic activities of connected neurons arising from the mean connectivity and its variability, respectively. For Gaussian connectivity statistics, this formally gives rise to definitions of two auxiliary fields, with the first one corresponding to the auxiliary field of the ordered system and the second one corresponding to fluctuations in the input induced by variability in connections (Sompolinsky & Zippelius, 1981, 1982). The saddle-point approximation of the latter therefore captures fluctuations in the input. In this approximation, the coupled network dynamics is reduced to the effective dynamics of a single unit that is exposed to a Gaussian input field whose statistics has to be determined self-consistently. The dynamics of this single unit then represents the population-averaged activity. This result can be intuitively understood using central-limit theorem arguments based on the superposition of a large number of weakly-correlated inputs (Ginzburg & Sompolinsky, 1994; van Vreeswijk & Sompolinsky, 1996; Renart et al., 2010).

Spin-glass mean-field techniques have been applied to networks of artificial neurons with activities characterized by rate variables which can be interpreted as the instantaneous firing probability of the neurons (Sompolinsky et al., 1988; Stern et al.,

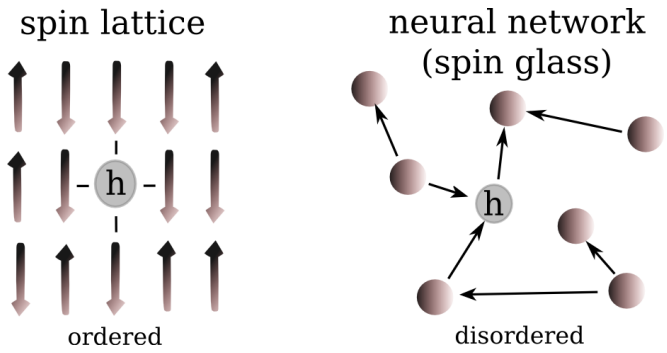


Figure 5: **Mean-field theory.** Left: ordered spin lattice with nearest neighbor coupling. Right: disordered spin glass or neuronal network with sparse directed connections. In both systems, the magnetic field or input h at a particular site is approximated by its expectation which is given by the average (nonlinearly transformed) activities of connected units.

2014; Kadmon & Sompolinsky, 2015; Aljadeff et al., 2015; Goedeke et al., 2016). Such rate-based models can be described using field-theoretic methods analogous to spin glasses (see Chap. 2 for a detailed review). More recent developments further derive consistent mean-field theories and generalizations including fluctuations for neural networks with Markovian dynamics (Buice & Cowan, 2007; Buice et al., 2010) and networks of phase oscillators (Hildebrand et al., 2007). While some spiking neuron models, such as the quadratic integrate-and-fire model, can be mapped to phase oscillators in specific dynamic regimes (Buice & Chow, 2013), to date there is no systematic field-theoretic framework for general spiking neuron models.

Heuristic mean-field approaches to spiking-neuron dynamics are based on rewriting dynamical equations in terms of the input variables instead of the membrane potential (see Bressloff, 2012, and references therein). A closure of equations is then achieved by a spatio-temporal coarse-graining, i.e. by considering the lumped activity of nearby neurons which receive temporally coarse-grained activity. Incoming spike trains are thereby replaced with the instantaneous firing rate of neurons and fluctuations arising from the discrete nature of inputs are neglected (see e.g. Ermentrout & Terman, 2010). This procedure works only well for slow synapses, i.e. for strong low-pass filtering of incoming spike trains and the resulting effective models are referred to as neural mass or neural field models (Wilson & Cowan, 1972a; Knight, 1972; Wilson & Cowan, 1973; Amari, 1972, 1977). In general, however, synaptic dynamics cannot be considered slow with respect to the membrane dynamics (Destexhe et al., 2003), and for fast synapses, exact timing of spikes becomes increasingly important.

A more systematic approach to approximate the statistics of inputs in spiking models is given by density methods (see e.g. Deco et al., 2008; Bressloff, 2012; Gerstner & Kistler, 2002, and references therein). The description of the system in terms of dynamical equations is given up for a statistical description for the states of neurons

(Abbott & van Vreeswijk, 1993; Brunel & Hakim, 1999). While this is possible for both deterministic and stochastic dynamics, closed equations for the temporal evolution of the density often rely on an ensemble description beyond single realizations of the system.

For spiking neurons with deterministic update, as e.g. the LIF model, and synaptic weights scaling as $J_{ij} \sim 1/\sqrt{N}$, the temporal evolution for the membrane-potential density is described by a partial differential equation, the Fokker-Planck equation (Risken, 1996; Amit & Brunel, 1997; Brunel & Hakim, 1999). While the intrinsic neural dynamics and the mean input to the neuron are captured by a 'drift' term, fluctuations in the input as arising from the spiking nature and the disordered connectivity give rise to a 'diffusion' term (Brunel & Hakim, 1999). The stochastic differential equation which corresponds to the Fokker-Planck equation is the Langevin equation (Risken, 1996). It describes the effective subthreshold dynamics where the input to the neuron is replaced by a Gaussian noise. Mean-field approximations, for the incoming spiking statistics are needed in order to determine the drift and diffusion term. Assuming the input to neurons is given by uncorrelated Poisson processes, both terms follow from the average firing rate of the presynaptic neuron populations (Brunel & Sergi, 1998). A self-consistent solution for the latter follows from the boundary condition of the Fokker-Planck equation: the probability flux at threshold contains the spiking nature of the interactions and determines the firing rate (Siegert, 1951; Fourcaud & Brunel, 2002).

The Fokker-Planck theory in combination with the mean-field approximation above goes beyond replacing the input to neurons by its mean and takes into account Gaussian fluctuations. It is thus conceptually similar to the mean-field approach employed in the context of rate-based models (Sompolinsky et al., 1988; Goedeke et al., 2016), where the fluctuations of the input are replaced by their expectation in the saddle-point approximation. For a scaling of synaptic weights $J_{ij} \sim 1/N$, the diffusion term in the Fokker-Planck equation vanishes in the thermodynamic limit $N \to \infty$ such that fluctuations become irrelevant for the macroscopic dynamics of the network. In this limit and for special choices of networks, such as fully-connected networks of quadratic integrate-and-fire neurons or phase oscillators, the dynamics are restricted to a subdomain of the phase space called the Otts-Antonsen manifold (Montbrió et al., 2015).

The Fokker-Planck theory is closely related to renewal theory, which is based on a refractory density instead of a membrane-potential density (Gerstner & van Hemmen, 1992; Gerstner & Kistler, 2002). Both formalisms have been the starting point for heuristic derivations of effective meso-scale dynamics. Stochastic rate-based models and generalizations to neural fields (Bressloff, 2012) have been motivated based on fluctuations around the stationary state firing rate. A decomposition of the Fokker-Planck operator into eigenmodes leads to decoupled equations for decaying oscillatory modes of the population activity (Mattia & Del Guidice, 2002). Neural mass models can be derived from the Fokker-Planck equation by replacing the probability density with a point mass (Deco et al., 2008).

Generalizations of this methodology have been used in the derivation of neural field models. The simulation of these effective rate models relies on the exchange of continuous variables instead of discrete spike events. A plethora of models furthermore considers instantaneous interactions between units. Time-discrete coupling between neurons decouples the neuronal dynamics of different neurons in between spike events which allows an efficient parallelization of simulations. An efficient framework for both rate-based and spike-based models therefore poses new challenges to simulation technology (see Chap. 4).

Despite the available density formalism, current mean-field approaches for spiking neurons rely on ad-hoc assumptions such as uncorrelated Poisson statistics of spike trains or homogeneity within populations. Effective models for meso-scale dynamics therefore lack a consistent treatment of microscopic fluctuations and correlations. Population signals such as the LFP, however, strongly depend on correlations between contributing neurons (Tetzlaff et al., 2008). The understanding of the local field potential therefore requires a two-step procedure, the characterization of correlations in the spiking activity and a translation of spikes to LFPs using a forward-modeling scheme.

The inclusion of correlations into mean-field approaches for spiking activity has so far been performed in a semi-systematic manner. The auto-correlation function of the population activity can be obtained by a weakly nonlinear stability analysis of the stationary state (Brunel & Hakim, 1999). It contains the population-averaged cross-correlations which has also been obtained using *linear-response theory* around the stationary state (Lindner et al., 2005; Schuecker et al., 2015). The resulting expressions have been used to determine the effect of network motifs on the overall correlation structure (Pernice et al., 2011; Trousdale et al., 2012), the origin of population-averaged correlations (Helias et al., 2014) , their suppression by inhibitory feedback (Tetzlaff et al., 2012; Helias et al., 2013), and the emergence of oscillations in complicated circuits (Bos et al., 2015).

On the conceptual level, the correlation structure between the linear-response modes can be interpreted as resulting from a set of coupled Ornstein-Uhlenbeck processes with output noise (Tetzlaff et al., 2012). On the population level, the same correlation structure with output noise replaced by input noise was shown to hold for networks of stochastic binary neurons (Grytskyy et al., 2013b). The mean-field theoretical Gaussian approximation of the input therefore provides a unified view on population-averaged correlations in large recurrent networks of different neuron models.

Focus of this study

This thesis contributes to the understanding of neural-network dynamics at the microscopic and mesoscopic scale in terms of theory, modeling and simulation (Fig. 6).

We extend current mean-field theoretical methods to capture the network dynamics and correlations on the level of individual cells, develop an efficient framework for simulations of rate-based single-neuron or population models in a spiking network simulator, and combine the established forward-modeling scheme for local field potentials with network dynamics of a cortical microcircuit model to study the relation between the LFP and the underlying spiking network dynamics.

Assuming that neurons solely communicate via spikes, the LFP follows as an epiphenomenon that can be derived using forward models. In order to understand the local field potential as measured for example with a laminar multielectrode in a cortical column of macaque primary visual cortex (Fig. 4) as well as its origin, one has to understand the concerted activity between individual neurons as arising in balanced networks such as the cortical microcircuit. While previous theoretical studies investigated the stationary state firing rates (Schuecker et al., 2016) and the origin of oscillations in this circuit (Bos et al., 2015), we here focus on the correlation structure between individual neurons in such balanced networks. Motivated by the unified description of population-averaged correlations (Grytskyy et al., 2013b), we employ a connectionist modeling approach and study the emergent dynamics on the level of single-unit activities and LFPs resulting from the complex structure in network connections between simplified neurons.

In Chap. 1, we study correlations in networks of binary neurons which constitute the classical model of collective phenomena in disordered systems (Dahmen et al., 2016a). Previous approaches are restricted to describe population-averaged correlations (Ginzburg & Sompolinsky, 1994; van Vreeswijk & Sompolinsky, 1996; Buice et al., 2010; Renart et al., 2010). The theoretical derivation furthermore relies on the assumption of weak couplings and heuristic central limit theorem arguments for replacing the input to neurons with a Gaussian field. We extend these existing mean-field approaches by deriving a systematic cumulant hierarchy which goes beyond the Gaussian approximation and recovers previous approaches as specific truncations. We derive self-consistency equations for pairwise-covariances between individual neurons in the Gaussian approximation for arbitrary network structures and coupling strengths. The solution of the equations reveals the equivalence of binary networks in the fluctuation driven regime to networks of Ornstein-Uhlenbeck processes. We thereby show that the networks of binary neurons are equivalent to coupled Ornstein-Uhlenbeck processes not only with respect to population-averaged correlations, but also on the level of correlations between individual neuron pairs.

In Chap. 2, we review mean-field methods for rate-based models including Ornstein-Uhlenbeck processes and nonlinear generalizations (Schuecker et al., 2016; Dahmen et al., 2016b) and present the derivation of the effective single-unit dynamics following from saddle-point approximations of auxiliary fields that was the basis for the seminal work of Sompolinsky et al. (1988) and more recent follow-up studies (Stern et al., 2014; Aljadeff et al., 2015; Kadmon & Sompolinsky, 2015) on chaotic dynamics in random networks. In the saddle-point approximation, neurons are decoupled

and cross-correlations therefore vanish. For the example of linear dynamics, we take into account finite-size effects that lead to fully symmetric systems instead of decoupled neurons after the ensemble average. These systems exhibit cross-correlations between neurons. We extend mean-field approximations to fluctuations of saddle-points through their source dependence. The latter dependence, which has been ignored in previous works, generates non-trivial fourth-order cumulants of activity, a second type of finite-size effect that is shown in the subsequent chapter to be crucial for capturing the heterogeneity across neurons.

In Chap. 3, Ornstein-Uhlenbeck processes are interpreted as linear-response modes of single-unit activity and their fourth-order cumulants are related to the experimentally observed dispersion of covariances (Dahmen et al., 2016b). Previous works could explain the low mean correlations in infinitely large networks (Renart et al., 2010) or in finite-size networks (Tetzlaff et al., 2012; Helias et al., 2014) as resulting from inhibitory feedback. The origin of the large dispersion of covariances (Fig. 4) was unclear, but mainly attributed to measurement noise (Ecker et al., 2010). By deriving relations between the statistics of connections and the statistics of covariances, we show that the latter are not an artifact, but provide insights into the properties of the underlying network structure and its operational regime. Applying the theory to experimental data of macaque motor cortex suggests an operation of the network close to a critical point. By relating the statistics of covariances to fourth moments of activity, we show that the extended mean-field theory goes beyond self-averaging and explains the meta-statistics of activity.

In Chap. 4, we describe a framework to integrate rate-based neuron dynamics in a simulator that is optimized for delayed spiking interactions (Hahne et al., 2016, 2017). The common simulation platform enables numerical validation of mean-field theories relating spike-based and rate-based dynamics and is the basis for multi-scale modeling approaches combining both levels of descriptions in a joint network model. The possibly instantaneous coupling between rate-based neurons is made compatible with the communication infrastructure of delayed spike events by employing the infrastructure for an iterative numerical solution scheme that was previously developed for gap-junction interactions in simulations of spiking networks (Hahne et al., 2015). This simulation framework preserves the decoupling of neuronal dynamics on a coarse time grid and thereby enhances parallelization and scaling behavior of the employed numerical solution schemes. Application of the methods is illustrated with a set of well-known network models from the neuroscientific literature.

In Chap. 5, we consider the mesoscopic reflection of large-scale network activity and describe a hybrid scheme for modeling local field potentials from point-neuron networks and its implementation in the hybridLFPy software package (Hagen et al., 2016). This methodology allows the validation of point-neuron network models in terms of experimentally accessible measures, such as the LFP. While previous attempts for conceptual reasons or limited performance of multicompartment-neuron simulations focused on the LFP as resulting from single neurons or small networks,

the current framework combines the efficiency of point-neuron network simulations with the biophysical principles of LFP generation and can therefore be used to investigate the origin of LFPs in terms of the spiking dynamics of the underlying large-scale networks.

In Chap. 6, we apply the hybrid scheme to a model of a cortical microcircuit to generate the spiking activity and LFP in a cortical column (Hagen et al., 2016). The microcircuit model has realistic cell density, but highly simplified neuron dynamics (LIF) and therefore focuses on the effect of the connectivity on the local network dynamics and LFP in such circuits. Despite this simplicity, the model displays features of spiking activity and LFP across populations and cortical layers in agreement with experimental observations. We further demonstrate the strong dependence of the LFP on the network state, which is determined by spike-train correlations on the microscopic scale. Furthermore, we derive the need for full-scale LFP simulations.

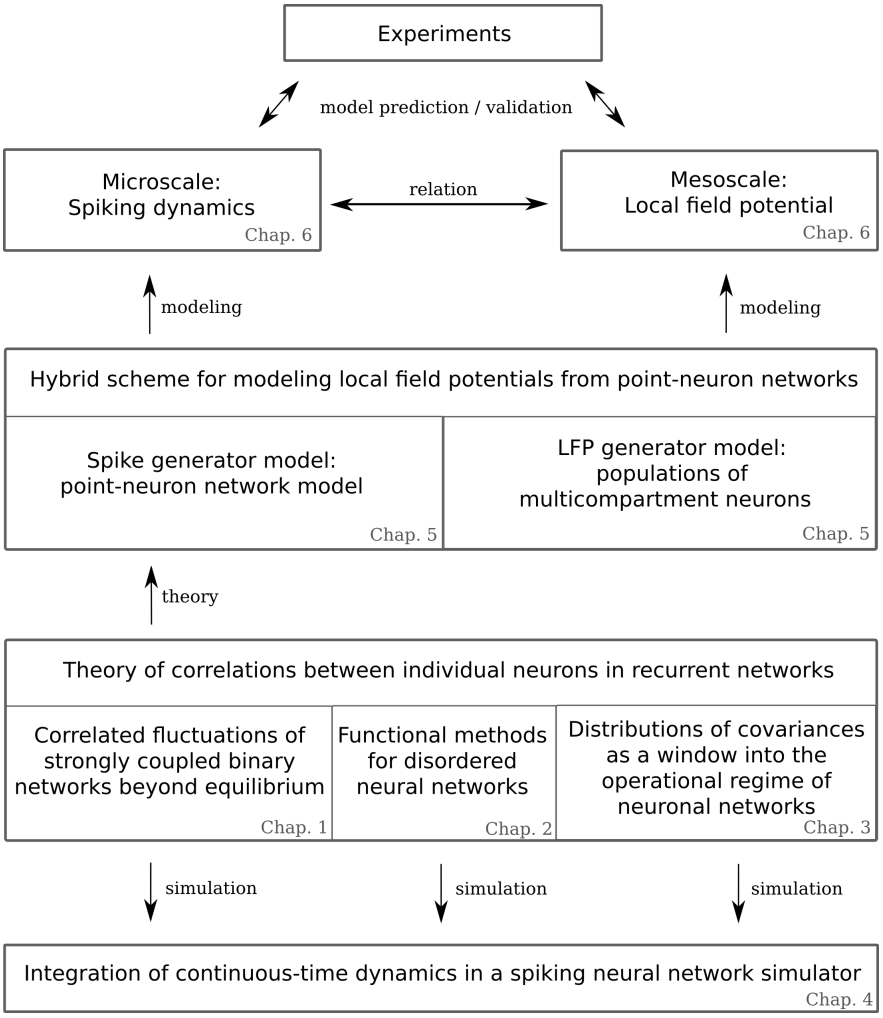


Figure 6: **Structure of this thesis.** This thesis focuses on the micro- and mesoscale dynamics of cortical networks. As an example, we develop a model for the spiking activity and LFPs in a cortical column (Chap. 6). The model builds on a hybrid scheme that translates spiking activity in point-neuron networks into extracellular potentials using populations of multicompartment-neuron models (Chap. 5). This scheme allows model predictions for experimental data and enables model validations on both spatial scales. The complex dynamics of point-neuron networks is amenable to mean-field theoretical methods. We assess correlated fluctuations in single realizations of networks composed of binary model neurons (Chap. 1), derive functional methods for ensembles of rate-based neural network models (Chap. 2), and combine these two works to explain the large dispersion of experimentally observed covariances across pairs of neurons (see Fig. 4) in terms of the operational regime of the network (Chap. 3). A framework for rate-based dynamics in a spiking neural network simulator (Chap. 4) allows numerical validation of the mapping between spiking and rate-based neuron models and forms the basis for future developments of multi-scale models combining spike- and rate-based descriptions.

Part II

Theory of correlations between individual neurons in recurrent networks

Chapter 1

Correlated fluctuations in strongly coupled binary networks beyond equilibrium

This chapter is based on the following publication:

Dahmen D, Bos H and Helias M (2016) *Correlated Fluctuations in Strongly Coupled Binary Networks Beyond Equilibrium*, Phys. Rev. X 6, 031024.

Preliminary results have been published in Dahmen et al. (2016b).

Author contributions:
Under the supervision of M Helias, the author performed all parts of the above publication that are included in this thesis. H Bos contributed section VII of the publication, which is not included in this thesis, and helped with the writing of the manuscript.

1.1 Introduction

Understanding collective phenomena arising from the interactions in a many-body system is the challenging subject of many-particle physics. The characterization of the emerging states typically rests on the quantification of correlations (Binney et al., 1992). Among the simplest classical models is the Ising (binary) model (Ising, 1925), or Glauber dynamics (Glauber, 1963) in its kinetic formulation. While for symmetric random couplings (Sherrington & Kirkpatrick, 1975) these systems are within the realm of equilibrium statistical mechanics (Thouless et al., 1977), the asymmetric kinetic Ising model (Hertz et al., 1986; Derrida, 1987; Crisanti & Sompolinsky, 1987), even in the stationary state, does not reach thermodynamic equilibrium. In their Markovian formulation (Kelly, 1979) these processes are studied in the field of nonequilibrium stochastic thermodynamics (for review see Seifert, 2012, chapter 6). The methods derived in these fields have proven useful in a variety of disciplines, including computer science, biology, artificial intelligence, social sciences, and economics (see e.g. Mézard et al., 1987; Nishimori, 2001; Barrat et al., 2008; Castellano et al., 2009; Stein & Newman, 2013; Sornette, 2014, and references therein).

Neuronal networks of the central nervous system are prominent examples of nonequilibrium systems due to Dale's principle (Eccles et al., 1954; Li & Dayan, 1999), which states that a neuron either excites or inhibits all its targets. The correlations between the activities of nerve cells are functionally important (Cohen & Kohn, 2011): They influence the representation of information in population signals depending on the readout in either a detrimental (Zohary et al., 1994) or beneficial (Shamir & Sompolinsky, 2001) manner, their time-dependent modulations are linked to behavior (Kilavik et al., 2009), and they determine the influence of neuronal activity on synaptic plasticity (Markram et al., 1997; Bi & Poo, 1998; Morrison et al., 2008), the biophysical mechanism underlying learning. In the case of binary units (Ising spins), knowledge of pairwise covariances, moreover, proves useful for sampling-based inference (reviewed in Fiser et al., 2010) as they constitute the next order in the systematic expansion of the joint probability distribution beyond independence (cf. Glauber, 1963, eq. (22)).

Dynamic mean-field theory (Sompolinsky & Zippelius, 1981, 1982; Sompolinsky et al., 1988; Cugliandolo & Kurchan, 1993; Aljadeff et al., 2015; Kadmon & Sompolinsky, 2015) in the $N \to \infty$ limit effectively reduces the many-body problem of a network comprised of a large number of units to a single particle interacting with a self-consistently determined field, but neglects the cross-covariances of activities between units. Ginzburg & Sompolinsky (1994) introduced pairwise correlations to the description of weakly coupled systems and Renart et al. (2010) extended the analysis to strongly coupled units in the large-N limit. Both approaches are, however, limited to averaged pairwise correlations. Similarly, approximate master equations for binary-state dynamics on complex networks, as recently discussed in context of the pair approximation by Gleeson (2011, 2013), are restricted to global dynamics in infinite

networks. For self-averaging observables, Coolen et al. (1996) presented a systematic approach to derive closed equations of motion, which, however, are only valid in the limit $N \to \infty$.

In this chapter, we derive a systematic cumulant expansion yielding an analytical description of correlations between the activities of individual pairs of units, which goes beyond averaged pairwise correlations (Ginzburg & Sompolinsky, 1994; Buice et al., 2010; Renart et al., 2010) and is applicable to a single realization of an asymmetric (directed) network. The framework links the coupling structure to the emerging correlated activity and describes the first- and second-order statistics for single units and pairs of units in a large, but finite, network of possibly strongly interacting elements. The analytical expressions predict a distribution of pairwise correlations with a small mean but large standard deviation, as observed experimentally (Ecker et al., 2010) in neural tissue and yet not explained theoretically. Moreover, the framework can be employed to infer the couplings between the units from the observed correlated activity, also termed the inverse problem (Aertsen & Preißl, 1990; Zeng et al., 2011; Mézard & Sakellariou, 2011; Tyrcha & Hertz, 2013).

In particular, we show that a Gaussian truncation of the presented cumulant hierarchy already yields good predictions for individual mean activities and pairwise correlations. Taking into account the subset of third-order cumulants that for binary variables can be expressed by lower-order ones, improves the prediction significantly. This finding demonstrates that the second-order statistics suffices to capture the major features of the collective dynamics arising in random binary networks, even when the coupling is strong. Our approach consistently takes into account the network-generated fluctuations in the marginal statistics of the input to each unit in a similar spirit as the seminal work by van Vreeswijk & Sompolinsky (1996). This approach exposes peculiar features of networks of binary threshold units. For a fixed number of incoming connections to each unit, mean-field theory predicts their averaged activities to be identical. Here, we show how distributed mean activities arise solely from the correlations emerging in the network. Units with a hard activation threshold render covariances independent of synaptic amplitudes J: If all incoming connections as well as the threshold of a unit are changed proportionally, the mean activity and correlations are maintained. Hence, scaling the threshold appropriately, it is impossible to increase the influence of one unit on another by larger coupling amplitudes. This finding questions the customary division into strong ($J \propto N^{-\frac{1}{2}}$) and weak coupling ($J \propto N^{-1}$) in such systems. Here, we show, in addition, that a network with hard threshold units implements the strongest possible coupling between stochastic binary units.

The independence of covariances with respect to coupling strengths implies that the latter cannot be reconstructed from the knowledge of the activity alone. However, together the amplitude and slope of covariance functions at zero time lag uniquely determine a linearized effective coupling strength between units. A linear approximation of the dynamics of fluctuations around the stationary state leads to a modified

Lyapunov equation. Decomposing the fluctuations into characteristic eigenmodes of the network shows that, to linear order, the kinetic binary network is equivalent to a system of coupled Ornstein-Uhlenbeck processes (Uhlenbeck & Ornstein, 1930; Risken, 1996).

1.2 First and second moments of the joint probability distribution

Here, we consider the classical network model of stochastic binary units, and denote the activity of unit k as $n_k(t)$ being either 0 or 1, where 1 indicates the active and 0 the inactive state (see e.g. Willshaw et al., 1969; Amari, 1975; Hertz et al., 1991; Ginzburg & Sompolinsky, 1994; van Vreeswijk & Sompolinsky, 1996; Buice et al., 2010; Renart et al., 2010; Gleeson, 2011, 2013). Since we may interpret a unit as representing an individual neuron, we use the terms "unit" and "neuron" interchangeably. The model neurons show transitions at random points in time between the two states 0 and 1 controlled by transition probabilities. The state of the network of N such units is described by a binary vector $\mathbf{n} = (n_1, \ldots, n_N) \in \{0,1\}^N$ that at each time point assumes one of the 2^N possible states. Using an asynchronous update (Rumelhart et al., 1986), in each infinitesimal interval $[t, t+\delta t)$ each unit in the network has the probability $\frac{1}{\tau}\delta t$ to be chosen for an update (Hopfield, 1982), where τ is the time constant of the dynamics. An equivalent implementation draws the time points of update independently for all units. For a particular unit, the sequence of update points has exponentially distributed intervals with mean duration τ, i.e. the update times form a Poisson process with rate τ^{-1}. We employ the latter implementation in the globally time-driven (Hanuschkin et al., 2010) spiking simulator NEST (Eppler et al., 2015) as described in Grytskyy et al. (2013b, Appendix), and use a discrete time resolution $\delta t = 0.1$ ms for the intervals. The stochastic update constitutes a source of noise in the system. Given that the kth neuron is selected for update, the probability to end in the active state ($n_k = 1$) is determined by the activation function $F_k(\mathbf{n})$ which depends on the activity $\mathbf{n}$ of all units i that are connected to neuron k via couplings J_{ki}. The probability to end in the inactive state ($n_k = 0$) is given by $1 - F_k(\mathbf{n})$.

The stochastic system at time t is completely determined by its probability distribution $p(\mathbf{n}, t)$. The time evolution of the two-point distribution $p(\mathbf{n}, t, \mathbf{q}, s)$ (describing the probability that the system was in state $\mathbf{q}$ at time s and is in state $\mathbf{n}$ at time t) obeys - because of the Markov property - the same master equation (see Eq. (A.1) in Sec. A.1) as $p(\mathbf{n}, t)$. Using the definitions of the first moment and the equal time as well as the two time point second moments,

$$\begin{aligned} \langle n_k(t)\rangle &:= \sum_{\mathbf{n}} p(\mathbf{n},t)\, n_k, \\ \langle n_k(t) n_l(t)\rangle &:= \sum_{\mathbf{n}} p(\mathbf{n},t)\, n_k n_l, \\ \langle n_k(t) n_l(s)\rangle &:= \sum_{\mathbf{n},\mathbf{q}} p(\mathbf{n},t,\mathbf{q},s)\, n_k q_l, \end{aligned} \tag{1.1}$$

one obtains a set of differential equations for the first and second moments (Ginzburg & Sompolinsky, 1994; Buice et al., 2010; Renart et al., 2010; Helias et al., 2014), which reads (for completeness, in the Appendix we included their derivation in Eq. (A.4))

$$\begin{aligned} & \tau \frac{\partial}{\partial t}\langle n_k(t)\rangle &&= -\langle n_k(t)\rangle + \langle F_k(\mathbf{n}(t))\rangle, \\ k \neq l \qquad & \tau \frac{\partial}{\partial t}\langle n_k(t) n_l(t)\rangle &&= -2\langle n_k(t)\, n_l(t)\rangle + \langle F_k(\mathbf{n}(t))\, n_l(t)\rangle + \langle F_l(\mathbf{n}(t))\, n_k(t)\rangle, \\ t > s \qquad & \tau \frac{\partial}{\partial t}\langle n_k(t) n_l(s)\rangle &&= -\langle n_k(t)\, n_l(s)\rangle + \langle F_k(\mathbf{n}(t))\, n_l(s)\rangle, \end{aligned} \tag{1.2}$$

where the expectation value $\langle\rangle$ defined in Eq. (1.1) can be interpreted as an average over realizations of the random dynamics. Note that the second line does not hold for $k = l$, but becomes the first line because of $\langle n_k^2\rangle = \langle n_k\rangle =: m_k$. The third line becomes the second line for $t \to s$. These equations are identical to eqs. (3.4)-(3.7) in Ginzburg & Sompolinsky (1994), to eqs. (3.12) and (3.13) in Buice et al. (2010), and to eqs. (18)-(22) in Renart et al. (2010, Supplementary material). The works by Ginzburg & Sompolinsky (1994) and Renart et al. (2010) considered the second-order statistics averaged over many pairs of neurons, i.e. $\langle n_\alpha n_\beta\rangle = \frac{1}{N_\alpha N_\beta}\sum_{i\in\alpha}\sum_{j\in\beta}\langle n_i n_j\rangle$. These averages are closely related to the population-averaged activities $n_\alpha(t) = \frac{1}{N_\alpha}\sum_{i\in\alpha} n_i(t)$ and can therefore be treated by population-averaging mean-field methods. While the methodology in Buice et al. (2010) is general, the authors consider interacting populations of neurons. Here, we go beyond the population-level description and consider second-order statistics of individual pairs for a particular coupling matrix $\mathbf{J}$, a setting which is beyond the disorder-averaged description of self-averaging observables (Coolen et al., 1996). In particular, we are interested in the stationary statistics of the mean activities of individual units and their zero time lag pairwise covariances $c_{kl} = \langle n_k(t) n_l(t)\rangle - \langle n_k(t)\rangle\langle n_l(t)\rangle$, which can be deduced from Eq. (1.2)

$$\begin{aligned} m_k &= \langle F_k(\mathbf{n})\rangle, \\ c_{kl} &= \begin{cases} \frac{1}{2}\langle F_k(\mathbf{n}) n_l\rangle + \frac{1}{2}\langle F_l(\mathbf{n}) n_k\rangle - m_k m_l & k \neq l \\ m_k(1-m_k) & k = l \end{cases}. \end{aligned} \tag{1.3}$$

We now assume a standard form for the activation function between neurons given by

$$\begin{aligned} F_k(\mathbf{n}) &= f(h_k(\mathbf{n})), \\ h_k &= \sum_{i=1}^{N} J_{ki} n_i, \end{aligned} \tag{1.4}$$

where the gain function $F_k(\mathbf{n}) = f(h_k(\mathbf{n}))$ depends on the state of the network only via the summed and weighted scalar activity h_k. In systems with symmetric couplings $J_{ki} = J_{ik}$ and activation function $f = \tanh$, this choice describes a system with a Hamiltonian, which is quadratic in the state variables and whose stationary distribution is of Boltzmann form. Instead, here we study general (nonsymmetric) couplings and arbitrary gain functions f. However, for concreteness, when comparing the analytical predictions to numerical results, we choose $f(h_k) = H(h_k - \theta)$, where $H(x) = 1 \quad \text{for } x \geq 0$ and 0 else is the Heaviside function. Intermediate results also hold for arbitrary gain functions f. If the distribution of h_k was known, the expectation value $\langle F_k \rangle$ could be directly calculated.

1.3 Gaussian approximation

We aim at a self-consistent equation for the N values of the first moments and the $N(N-1)$ values of the second moments of the activity variables (1.3). We note that van Vreeswijk & Sompolinsky (1996) and Renart et al. (2010) invoke the central-limit theorem to justify the assumption that the local field h_k typically follows a Gaussian distribution $\mathcal{N}(\mu_k, \sigma_k^2)$ with cumulants μ_k and σ_k^2. These cumulants are related to cumulants of the activity variables $\mathbf{n}$ via

$$\begin{aligned} \mu_k &= \langle h_k \rangle = \sum_i J_{ki} \langle n_i \rangle = (\mathbf{J}\,\mathbf{m})_k, \\ \sigma_k^2 &= \langle h_k^2 \rangle - \mu_k^2 = \sum_{i,j} J_{ki} J_{kj} \left(\langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle \right) \\ &= \left(\mathbf{J}\,\mathbf{C}\,\mathbf{J}^T \right)_{kk}, \end{aligned} \tag{1.5}$$

with the vector of mean activities $m_i = \langle n_i \rangle$ and the covariance matrix $c_{ij} = \langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle$, which contains $c_{ii} = \langle n_i \rangle (1 - \langle n_i \rangle)$ on the diagonal. In the seminal work by van Vreeswijk & Sompolinsky (1998), the influence of the cross-covariances on the variance in the input to a neuron was neglected. Instead, only the dominant contribution, given by the variances c_{ii} of the single neurons on the diagonal, was taken into account, leading to the expression $\sigma_k^2 = \sum_i J_{ki}^2 m_i (1 - m_i)$. The additional dependence of σ_k^2 on the off-diagonal elements (1.5) is important for the distribution of mean activities, as we will show in the following. We note that because of the marginal binary statistics of n_i, it follows that $\langle n_i^2 \rangle = \langle n_i \rangle$, illustrating that the second moment is uniquely determined by the first moment. We exploit this property in

the subsequent sections by expressing a subset of higher-order cumulants in terms of lower-order ones. Using the central-limit theorem, as outlined above, leads to

$$\begin{aligned} m_k = \langle F_k(\mathbf{n}) \rangle &\simeq \int_{-\infty}^{\infty} H(x-\theta)\, \mathcal{N}(\mu_k, \sigma_k^2, x)\, dx \\ &= \frac{1}{2} \mathrm{erfc}\left(-\frac{\mu_k - \theta}{\sqrt{2}\sigma_k} \right). \end{aligned} \tag{1.6}$$

To enable the extension to further corrections, we now aim at a more systematic calculation of expectation values containing the gain function. Using the Fourier representation $f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, \hat{f}(\omega) \exp(i\omega x)$ of the gain function f we obtain

$$\begin{aligned} \langle F_k(\mathbf{n}) \rangle_{\mathbf{n}} &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, \hat{f}(\omega) \left\langle \exp\left((i\omega) \sum_{j=1}^{N} J_{kj} n_j \right) \right\rangle_{\mathbf{n}} \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, \hat{f}(\omega) \left\langle \exp\left((i\omega)\, h_k(\mathbf{n}) \right) \right\rangle_{\mathbf{n}}, \end{aligned} \tag{1.7}$$

where we inserted the definition of h_k (1.4) in the second line. Defining the vector $\mathbf{t}$ with elements $t_j = i\omega J_{kj}$, the expectation value in the first line of the previous expression has the form $\langle \exp(\mathbf{t} \cdot \mathbf{n}) \rangle_{\mathbf{n}}$, which is the definition of the characteristic function or moment generating function

$$\begin{aligned} \varphi_n(\mathbf{t}) &\equiv \langle e^{\mathbf{t} \cdot \mathbf{n}} \rangle_{\mathbf{n}} \\ &= \sum_{\mathbf{n} \in \{0,1\}^N} p(\mathbf{n})\, e^{\mathbf{t} \cdot \mathbf{n}} \\ &= \exp\left(\Phi_{\mathbf{n}}(\mathbf{t}) \right) \end{aligned} \tag{1.8}$$

of the joint distribution of the binary variables $p(\mathbf{n})$ (Gardiner, 1985, p. 32), which can also be expressed by the cumulant generating function $\Phi_{\mathbf{n}}(\mathbf{t})$. Note that for compactness of notation we omitted the index k for $\mathbf{t}$. The Taylor expansion of $\Phi_{\mathbf{n}}(\mathbf{t})$ explicitly introduces a cumulant hierarchy showing that cumulants of activity variables n_j at all orders contribute to $\langle F_k(\mathbf{n}) \rangle$. If we neglect cumulants higher than order two, thus effectively approximating the binary states as correlated Gaussian variables, we use the corresponding quadratic cumulant generating function

$$\Phi_{\mathbf{n}}(\mathbf{t}) = \sum_i m_i t_i + \frac{1}{2} \sum_{i,j} c_{ij}\, t_i t_j \tag{1.9}$$

and obtain

$$\begin{aligned} \langle F_k(\mathbf{n}) \rangle_{\mathbf{n}} &\simeq \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, \hat{f}(\omega) \exp\left(\sum_i m_i t_i + \frac{1}{2} \sum_{i,j} c_{ij}\, t_i t_j \right) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\, \hat{f}(\omega) \exp\left(\mu_k i\omega + \frac{1}{2} \sigma_k^2 (i\omega)^2 \right). \end{aligned} \tag{1.10}$$

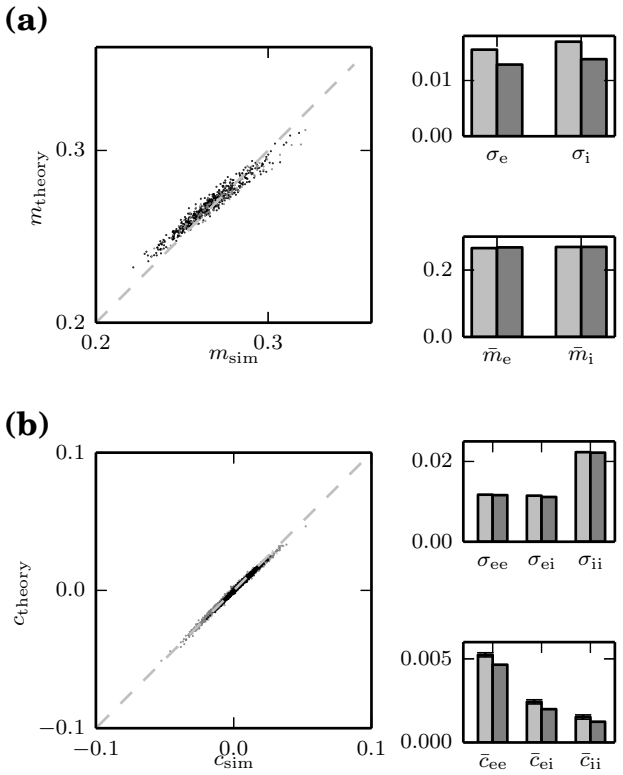


Figure 1.1: **Consistent Gaussian approximation for the mean activity and covariance matrix in a random network of binary neurons. (a)** Theoretical prediction of mean activities (1.15) versus simulation results (black: excitatory, gray: inhibitory). The diagonal is indicated by the dashed line. Lower inset: mean activity averaged over excitatory and inhibitory neurons (light gray: simulation, darker gray: theory). Black error bars show standard error of the mean obtained from 20 simulations (not visible, below line width). Upper inset: width of distribution of mean activities. **(b)** Theoretical prediction (1.15) versus simulated covariance. Only cross-covariances are shown, as auto-covariances follow trivially by Eq. (1.3) from the mean activities. Lower inset: mean covariances averaged over excitatory-excitatory, excitatory-inhibitory and over inhibitory-inhibitory neuron pairs (light gray: simulation, darker gray: theory). Error bars show standard error of the mean obtained from 20 simulations. Upper inset: width of the distributions of covariances. Network parameters: number of excitatory ($N_E = 500$) and inhibitory neurons ($N_I = 125$), threshold $\theta = -5.5$, excitatory coupling strength $J_{ij} = J = 1$, $\forall j \in E$, inhibitory coupling strength $J_{ij} = -6\,J$, $\forall j \in I$, fixed in-degree homogeneous random network with connection probability $p = 0.2$. Results averaged over 20 simulations with $T = 2,000,000$ ms each. Simulations were performed using the NEST simulator (Eppler et al., 2015). Theoretical predictions are obtained by damped fixed-point iteration Eq. (1.16).

Here we identified the sums in the penultimate line as the mean and variance of the input field (1.5). On the other hand in the second line of Eq. (1.7) we identify the moment generating function $\varphi_{h_k}(i\omega) \equiv \langle \exp\left((i\omega)\, h_k(\mathbf{n})\right)\rangle_{\mathbf{n}}$ of the input field h_k. From (1.10) we obtain its corresponding cumulant generating function as $\Phi_{h_k}(i\omega) = \mu_k i\omega + \frac{1}{2}\sigma_k^2 (i\omega)^2$. We therefore conclude that $h_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$ follows a Gaussian distribution. The mean activity is then written

$$m_k = \langle F_k(\mathbf{n})\rangle_{\mathbf{n}} \quad = \quad \langle f(h_k)\rangle_{h_k \sim \mathcal{N}(\mu_k, \sigma_k^2)}, \tag{1.11}$$

which for $f(x) = H(x-\theta)$ is identical to Eq. (1.6). The procedure hence reproduces the known result for the Heaviside gain function (van Vreeswijk & Sompolinsky, 1998; Renart et al., 2010, supplementary material, Sections 2.2, 2.3), but additionally yields a generalization for smooth gain functions f that takes into account the fluctuations of the synaptic input, which is in line with the treatment in (Grytskyy et al., 2013a, eq. (27)).

More importantly, this systematic calculation reveals the assumption that is underlying the Gaussian approximation for the input field, namely that cumulants higher than order two are ignored on the level of individual neuronal activities.

To obtain an expression for the zero time lag covariance, we start from Eq. (1.3). We hence need to evaluate expressions of the form $\langle F_k(\mathbf{n}) n_l\rangle$ which do not factorize trivially since the value of n_l may have an influence on neuron k through the gain function $F_k(\mathbf{n})$. However, along the same lines as above, we can derive $\langle F_k(\mathbf{n}) n_l\rangle$ using the identical approximation of $\langle \exp(\mathbf{n}\cdot\mathbf{t})\rangle$ by first noting

$$\begin{aligned}
\langle F_k(\mathbf{n}) n_l\rangle &= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, \hat{f}(\omega) \left\langle \exp\left(i\omega \sum_{j=1}^{N} J_{kj} n_j \right) n_l \right\rangle \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} d\omega\, \hat{f}(\omega)\, \frac{\partial}{\partial t_l} \langle \exp(\mathbf{n}\cdot\mathbf{t})\rangle .
\end{aligned}$$

Againg expressing the characteristic function by the cumulant generating function (1.9) yields

$$\begin{aligned}
\frac{\partial}{\partial t_l} \exp\left(\Phi_{\mathbf{n}}(\mathbf{t})\right) &= \underbrace{\frac{\partial \Phi_{\mathbf{n}}(\mathbf{t})}{\partial t_l}}_{= m_l + \sum_j J_{kj} c_{jl}(i\omega) \text{ by (1.9)}} \exp\left(\Phi_{\mathbf{n}}(\mathbf{t})\right) \\
&= \left(m_l + \sum_j J_{kj} c_{jl} \frac{\partial}{\partial \mu_k} \right) \exp\left(\Phi_{h_k}(i\omega)\right),
\end{aligned} \tag{1.12}$$

where we generated the factor $i\omega$ by a derivative ∂_{μ_k} in the last line. With Eq. (1.11)

this yields

$$\begin{aligned} \langle F_k(\mathbf{n}) n_l \rangle &= m_l \langle f(h_k) \rangle_{h_k \sim \mathcal{N}(\mu_k, \sigma_k^2)} + S_k \left(\mathbf{JC} \right)_{kl}, \\ S_k &= \frac{\partial}{\partial \mu_k} \langle f(h_k) \rangle_{h_k \sim \mathcal{N}(\mu_k, \sigma_k^2)} \\ &\overset{\text{i.b.p.}}{=} \langle f'(h_k) \rangle_{h_k \sim \mathcal{N}(\mu_k, \sigma_k^2)}, \end{aligned} \tag{1.13}$$

where we defined the local susceptibility S_k that determines the influence of an input to neuron k onto its output at the time point of its update. For a differentiable gain function f the susceptibility is equal to the slope f' averaged over the fluctuations of h_k (compare also Grytskyy et al., 2013a, eq. (27)), which follows from integrating the second line of Eq. (1.13) by parts (i.b.p.). Note that the second line in Eq. (1.13) also holds for nondifferentiable f, such as our particular choice $f(x) = H(x - \theta)$, for which the susceptibility has the form

$$S_k := \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(\mu_k - \theta)^2}{2\sigma_k^2}}, \tag{1.14}$$

and which, by Eq. (1.13), is the strongest possible coupling for a given input statistics. The self-consistent set of equations for the first and second cumulants thus reads

$$\begin{aligned} m_k &= \frac{1}{2} \mathrm{erfc}\left(-\frac{\mu_k - \theta}{\sqrt{2}\sigma_k} \right), \\ c_{kl} &= \begin{cases} \frac{1}{2} S_k \left(\mathbf{JC} \right)_{kl} + \frac{1}{2} S_l \left(\mathbf{JC} \right)_{lk} & k \neq l \\ m_k (1 - m_k) & k = l \end{cases}. \end{aligned} \tag{1.15}$$

To obtain the joint solution of equations Eq. (1.15), we use a damped fixed-point iteration with the nth iteration value denoted as $(m_k^{(n)}, c_{kl}^{(n)})$, which has the form

$$\begin{aligned} m_k^{(n+1)} &= \rho \frac{1}{2} \mathrm{erfc}\left(-\frac{\mu_k^{(n)} - \theta}{\sqrt{2}\sigma_k^{(n)}} \right) + (1 - \rho) m_k^{(n)}, \\ c_{kl}^{(n+1)} &\underset{k \neq l}{=} \rho \frac{1}{2} \left(S_k^{(n)} \left(\mathbf{JC}^{(n)} \right)_{kl} + S_l^{(n)} \left(\mathbf{JC}^{(n)} \right)_{lk} \right) \\ &\quad + (1 - \rho) c_{kl}^{(n)}, \\ c_{kk}^{(n+1)} &\underset{k = l}{=} m_k^{(n+1)} (1 - m_k^{(n+1)}), \end{aligned} \tag{1.16}$$

where we used the damping factor $\rho = 0.7$, which, for the current application, yields good convergence. Here μ_k and σ_k are determined by Eq. (1.5) in each step. The iteration continues until the summed absolute change of m_k and c_{kl} is smaller than the chosen tolerance of 10^{-14}, allowing a sufficient safety margin towards truncation errors induced by the employed double precision numerics. Fig. 1.1 shows that the theoretical prediction yields mean activities and covariances in line with simulations

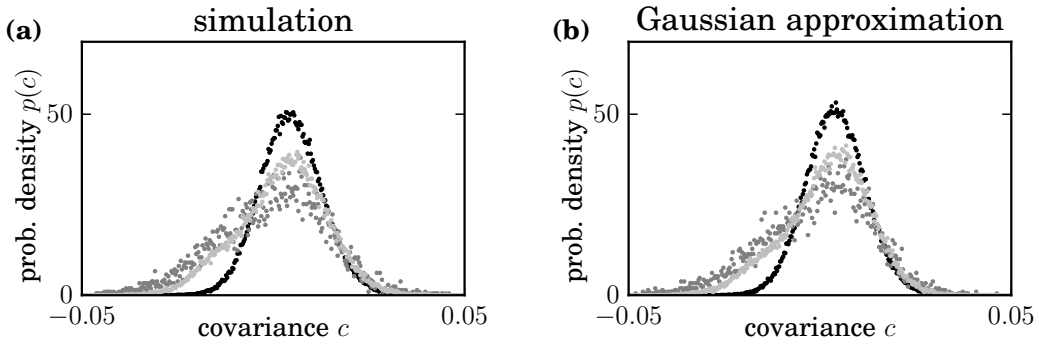


Figure 1.2: **Distribution of covariances in the Gaussian approximation for a random network of binary neurons. (a)** Distribution of cross-covariance between excitatory neurons (black), inhibitory neurons (gray) and between one inhibitory, one excitatory neuron (light gray) from simulation. **(b)** Same as (a), but showing the theoretical result for the Gaussian approximation (1.15). Parameters as in Fig. 1.1.

of a random fixed-in-degree network of excitatory and inhibitory neurons. While cross-covariances are explained with good accuracy, mean activities slightly, but systematically, deviate from simulated results: The scatter plot, showing the mean activities of individual neurons, has a slope below unity, indicating that the width of the distribution is underestimated by the theory. Moreover, we observe that the population-averaged covariances are slightly underestimated by the theory. However, in summary, the theory in the Gaussian approximation captures not only the mean and width of the covariance distribution to a large extent but also its general shape as shown in Fig. 1.2.

The systematic calculation presented here shows that the Gaussian assumption on the level of the individual statistics directly yields the linearized result of (Renart et al., 2010, supplementary material, Sections 2.3). Originally, the terms of the form $\langle F_k(\mathbf{n})n_l\rangle$ in Eq. (1.3) were obtained using relations between conditioned and unconditioned probability distributions of binary variables. After the Gaussian approximation, the resulting non-linear equation was linearized and second-order terms were discussed to be small in the large-N limit. In contrast, the derivation here shows that the linearization is not an additional assumption but appears naturally once the Gaussian approximation has been performed on the level of individual variables. Note also that no additional assumption regarding the strength of the coupling is necessary.

1.4 Third-order cumulants

The derivation in the previous section is systematic in the sense that all cumulants of order three and higher of the stochastic variables $\mathbf{n}$ are consistently neglected. The

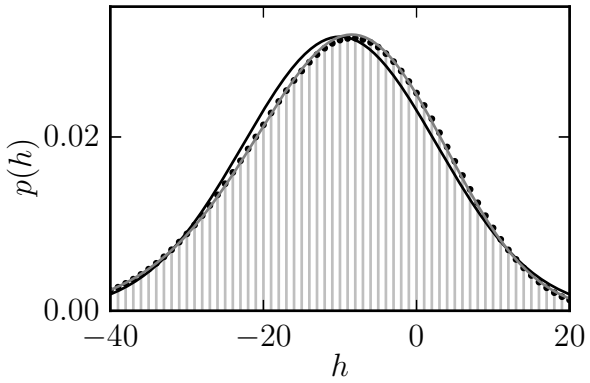


Figure 1.3: **Distribution of the summed input.** Circles show the distribution assuming uncorrelated activity of the input neurons with constant rate $m = 0.2$, i.e. $p(h) = \sum_{k,l} B(pN_E, m, k) B(pN_I, m, l)\, \delta_{h,\, J_E k + J_I l}$ with binomial probabilities $B(N, p, k)$. Gaussian approximation $\mathcal{N}(\mu, \sigma^2, h)$ (black) with the same moments as the summed binomial distribution $\mu = p\,(N_E J_E + N_I J_I)\; m$, $\sigma^2 = p\left(N_E J_E^2 + N_I J_I^2\right) m(1-m)$. Close-to-Gaussian approximation $p(h) \simeq \left(1 + \frac{\kappa_3}{6}\,\partial_\mu^3\right) \mathcal{N}(\mu, \sigma^2, h)$ (gray) taking into account the third cumulant $\kappa_3 = p\left(N_E J_E^3 + N_I J_I^3\right) (2m^3 - 3m^2 + m)$ of h. Other parameters as in Fig. 1.1.

original idea of a mean-field description, however, seeks an approximation of the local field h_k sensed by an individual unit k rather than a truncation of the cumulants of the individual variables $\mathbf{n}$ themselves. Since the local field is, by Eq. (1.4), a superposition of a large number of weakly correlated contributions, the distribution of h_k will be close to Gaussian by the central limit theorem. In the path-integral formulation, commonly employed to study disordered systems, the Gaussian approximation of h_k is equivalent to a saddle-point approximation to lowest order in the auxiliary field (see e.g. ref Sompolinsky & Zippelius, 1982, eq. (3.5) or ref Sompolinsky et al., 1988, eq. (3)). Even though one may expect the central-limit theorem to be applicable to the summed input h_k, it is easy to see that for networks of several hundreds of units higher order cumulants still have an effect. This is illustrated in Fig. 1.3, showing the distribution of the input to a neuron receiving a sum of binary, uncorrelated signals. The Gaussian approximation, with the same mean and variance as the exact distribution, has a peak that is slightly shifted to the left. Taking into account the third-order cumulant cures the displacement of the peak. The expansion in cumulants in Sec. 1.3 shows how the higher-order cumulants can be taken into account in the calculation of the expectation values of the gain function.

In the following, we expand the distribution of h_k up to third-order cumulants. To this end we use the relationship that the nth cumulant of a summed variable is the sum of nth cumulants of its constituents (A.9). We then use the property of binary variables that $n_k^K = n_k$ for $K \geq 1$, which relates cumulants of order $K + L$ to cumulants of order $L + 1$ in cases where neuron k appears K times in the cumulant (see Sec. A.4 for the detailed calculation). Note that similar relations also hold for classical spins

$n_i = \pm 1$, or generally for dichotomous units with arbitrary states $n_i \in \{\alpha, \beta\}$, as all these representations are related by an bijective affine-linear mapping. The first two cumulants of the summed input $h_k = \sum_l J_{kl} n_l$ to a neuron are therefore given as before by $\kappa_{1,k} = \mu_k$ and $\kappa_{2,k} = \sigma_k^2$ Eq. (1.5). The third cumulant $\kappa_{3,k}$ (by defining $x_l = J_{kl} n_l$ and $y = h_k$ and using Eq. (A.9)) reads

$$\kappa_{3,k} = \sum_{i,j,r} J_{ki} J_{kj} J_{kr} \langle\!\langle n_i n_j n_r \rangle\!\rangle, \tag{1.17}$$

where we use the notation $\langle\!\langle \circ \rangle\!\rangle$ to denote cumulants. For the mean activity (1.3), we need to evaluate the expectation value of a nonlinear function applied to h_k. We follow an analogous approach as in the previous section (see Sec. A.3 for details) and apply Eq. (A.11) to $F_k(\mathbf{n}) = H(\sum_j J_{kj} n_j - \theta)$, yielding a perturbative treatment for the effect of the third-order cumulant

$$\langle F_k(\mathbf{n}) \rangle = e^{\frac{1}{6}\kappa_{3,k}\left(\frac{\partial}{\partial \mu_k}\right)^3} \langle H(h_k - \theta) \rangle_{h_k \sim \mathcal{N}(\mu_k, \sigma_k^2)}. \tag{1.18}$$

To obtain a correction of the covariances, we determine the contribution of the third cumulant to the term of the form $\langle F_k(\mathbf{n})\, n_l \rangle = \langle H(h_k(\mathbf{n}) - \theta)\, n_l \rangle$. We apply the general result Eq. (A.12) with $x_l = J_{kl} n_l$, where we need to cancel a factor J_{kl} in the final result because of $n_l = x_l / J_{kl}$ to get

$$\begin{aligned} \langle F_k(\mathbf{n}) n_l \rangle &= \sum_{q=0}^{3} \Delta\kappa_{q,kl} \frac{1}{q!} \left(\frac{\partial}{\partial \mu_k}\right)^q \langle F_k(\mathbf{n}) \rangle \\ &\text{with} \\ \Delta\kappa_{q,kl} &= \sum_{i_1 \cdots i_q} J_{ki_1} \cdots J_{ki_q} \langle\!\langle n_{i_1} \cdots n_{i_q} n_l \rangle\!\rangle, \end{aligned} \tag{1.19}$$

where $\langle F_k(\mathbf{n}) \rangle$ is given by Eq. (1.18). The form of the terms $\Delta\kappa_{q,kl} \frac{1}{q!} \left(\frac{\partial}{\partial \mu_k}\right)^q = \Delta\kappa_{q,kl} \frac{\partial}{\partial \kappa_{q,k}}$ shows that, for $q \geq 1$, they correspond to an infinitesimal displacement of the qth cumulant $\kappa_{q,k}$ by $\Delta\kappa_{q,kl}$. These corrections come about by the presence of the variable n_l in the expectation value, which can alternatively be understood as a conditioned expectation value $\langle F_k(\mathbf{n}) n_l \rangle = m_l \langle F_k(\mathbf{n}) | n_l = 1 \rangle$, where the condition $n_l = 1$ changes the cumulants of h_k. The first two terms in the sum Eq. (1.19) are identical to the Gaussian approximation in Eq. (1.12). The difference in the approximation schemes on the level of $\mathbf{n}$ and h_k, respectively, is apparent in the term for $q = 2$, which contains a correction to the second-order cumulant due to the presence of n_l. This term is neglected in the Gaussian truncation of cumulants of $\mathbf{n}$ in Sec. 1.3. The consistent approximation of h_k up to third-order cumulants analogously requires the term for $q = 3$. This correction in turn depends on the fourth-order cumulant $\langle\!\langle n_{i_1} \cdots n_{i_q} n_l \rangle\!\rangle$.

We now use the properties of binary variables that allow us to express a subset of higher cumulants by lower-order ones because of the property $\langle n_i^K \rangle = \langle n_i \rangle$ (for all

integers $K \geq 1$). The kth raw moment is given by the product of all combinations of lower-order cumulants (van Kampen, 1992). For the third moment this yields

$$\begin{aligned} \langle n_l n_i n_j \rangle &= \langle\!\langle n_l n_i n_j \rangle\!\rangle \\ &+ c_{li} m_j + c_{ij} m_l + c_{jl} m_i \\ &+ m_l m_i m_j. \end{aligned} \tag{1.20}$$

If there are at least two identical indices, we can express the corresponding third-order cumulant by the two lower orders. Both cases $l = i \neq j$ and $l = i = j$ lead to the same expression Eq. (A.13)

$$\langle\!\langle n_l n_l n_j \rangle\!\rangle = c_{lj}(1 - 2m_l). \tag{1.21}$$

In the latter case, we get the third cumulant of a binary variable $\langle\!\langle n_l n_l n_l \rangle\!\rangle = m_l - 3m_l^2 + 2m_l^3$, which is uniquely determined by its mean. We can therefore take into account the contribution of all third-order cumulants (with at least two identical indices) to the statistics of h_k. Stated differently, we calculate the third-order cumulant of h_k using Eq. (1.17) and neglecting all cumulants of the binary variables where all indices are different ($\langle\!\langle n_i n_j n_l \rangle\!\rangle \simeq 0$ for $i \neq j \neq l$). A straightforward calculation (see Sec. A.4 Eq. (A.14) for details) yields

$$\begin{aligned} \kappa_{3,k} &\simeq \left[3(\mathbf{J} \circledast \mathbf{J})\, \mathrm{diag}(\{1 - 2m_i\})\, \mathbf{C}\mathbf{J}^T\right]_{kk} \\ &- \left[2\,(\mathbf{J} \circledast \mathbf{J} \circledast \mathbf{J})\ \mathrm{diag}\left(\{m_i - 3m_i^2 + 2m_i^3\}\right)\right]_k, \end{aligned} \tag{1.22}$$

where we use the symbol $\circledast$ for the element-wise (Hadamard) product of two matrices (Cichocki et al., 2009). For the third cumulant appearing explicitly in Eq. (1.19) we perform a similar reduction that yields the matrix Eq. (A.15)

$$\begin{aligned} \Delta\kappa_2 &\simeq \mathbf{J} \circledast \mathbf{J}\, \mathrm{diag}(\{1 - 2m_i\})\mathbf{C} \\ &+ 2\mathbf{J}\, \mathrm{diag}(\{1 - 2m_i\}) \circledast (\mathbf{J}\mathbf{C}) \\ &- 2(\mathbf{J} \circledast \mathbf{J})\, \mathrm{diag}\left(\{m_i - 3m_i^2 + 2m_i^3\}\right). \end{aligned} \tag{1.23}$$

The matrix $\Delta\kappa_3$ takes the form (see Sec. A.4 for details)

$$\begin{aligned} \Delta\kappa_3 &\simeq 3\mathbf{J} \circledast \left((\mathbf{J} \circledast \mathbf{J})\, \left\{\langle\!\langle n_i n_i n_j n_j \rangle\!\rangle_{ij}\right\}\right) \\ &+ 3(\mathbf{J} \circledast \mathbf{J}) \circledast \left(\mathbf{J}\left\{\langle\!\langle n_i n_j n_j n_j \rangle\!\rangle_{ij}\right\}\right) \\ &+ (\mathbf{J} \circledast \mathbf{J} \circledast \mathbf{J})\left(\left\{\langle\!\langle n_i n_i n_i n_j \rangle\!\rangle_{ij}\right\} + \mathrm{diag}(\{\langle\!\langle n_i n_i n_i n_i \rangle\!\rangle_i\})\right), \end{aligned} \tag{1.24}$$

where the fourth-order cumulants are given by Eq. (A.16)-Eq. (A.18).

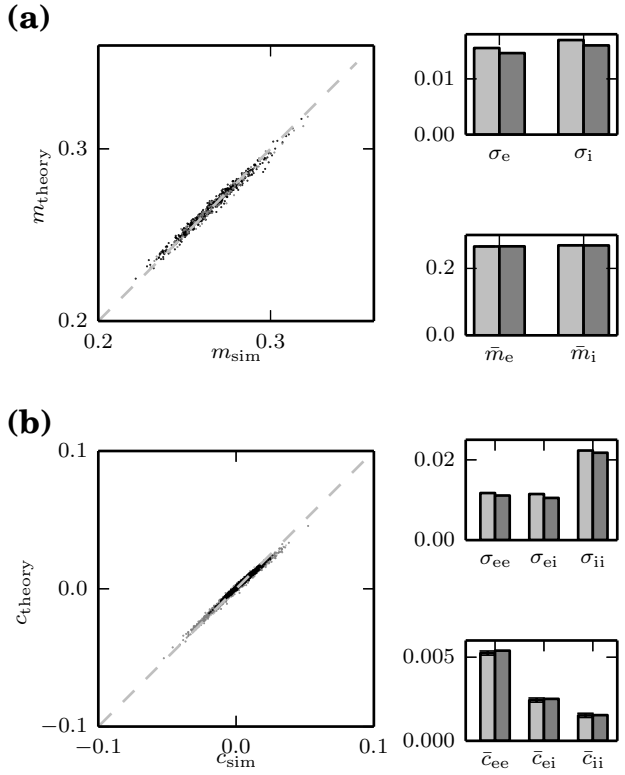


Figure 1.4: **Approximation for covariance matrix in random network of binary neurons including third order cumulants. (a)** Theoretical prediction of mean activities (1.25) versus simulation results (black: excitatory, gray: inhibitory). Lower inset: mean activity averaged over excitatory and inhibitory neurons (light gray: simulation, darker gray: theory). Black error bars (not visible, below line width) show standard error of the mean obtained from 20 simulations. Upper inset: standard deviation of distribution of mean activities of excitatory (σ_e) and inhibitory (σ_i) neurons. **(b)** Theoretical prediction Eq. (1.26) versus simulated covariance (black: excitation-excitation, gray: inhibition-inhibition, light gray: excitation-inhibition). Lower inset: mean covariances averaged over excitatory-excitatory, excitatory-inhibitory and over inhibitory-inhibitory neuron pairs (light gray: simulation, darker gray: theory). Error bars show standard error of the mean obtained from 20 simulations. Upper inset: standard deviations of the distributions of covariances. Network parameters and display as in Fig. 1.1. Theoretical predictions obtained by damped fixed-point iteration.

To evaluate Eq. (1.18) for the mean activity and Eq. (1.19) for the covariance, we need the nth derivative of the complementary error function. We use that $\frac{d}{dx}\frac{1}{2}\mathrm{erfc}(-x) = \frac{1}{\sqrt{\pi}}e^{-x^2}$ is a Gaussian and further that the nth derivative of a Gaussian $\left(\frac{d}{dx}\right)^n e^{-x^2} = (-1)^n H_n(x)\, e^{-x^2}$ can be expressed in terms of the n-th Hermite polynomial H_n (Olver et al., 2010, par. 18.3). Each differentiation with respect to μ_k yields an additional factor $\left(\sqrt{2}\sigma_k\right)^{-1}$. Hence, we get, for $n \geq 1$,

$$\begin{aligned} L_n(\mu_k,\sigma_k) &:= \left(\frac{d}{d\mu_k}\right)^n \frac{1}{2}\mathrm{erfc}\left(\frac{\theta-\mu_k}{\sqrt{2}\sigma_k}\right) \\ &= \frac{1}{\sqrt{\pi}}\left(\sqrt{2}\sigma_k\right)^{-n} H_{n-1}(x)\, e^{-x^2}\Big|_{x=\frac{\theta-\mu_k}{\sqrt{2}\sigma_k}}. \end{aligned}$$

Using this definition and the expansion $e^{\frac{1}{6}\kappa_{3,k}\left(\frac{\partial}{\partial\mu_k}\right)^3} \simeq 1 + \frac{1}{6}\kappa_{3,k}\left(\frac{\partial}{\partial\mu_k}\right)^3$, valid for small $\kappa_{3,k}$ compared to σ_k^3, we get the mean activity

$$\begin{aligned} m_k = \langle F_k(\mathbf{n})\rangle &\simeq \left(1 + \frac{1}{6}\kappa_{3,k}\left(\frac{\partial}{\partial\mu_k}\right)^3\right)\frac{1}{2}\mathrm{erfc}\left(\frac{\theta-\mu_k}{\sqrt{2}\sigma_k}\right) \\ &= L_0(\mu_k,\sigma_k) + \frac{1}{6}\kappa_{3,k}\, L_3(\mu_k,\sigma_k). \end{aligned} \tag{1.25}$$

For the term appearing in the covariance we get

$$\begin{aligned} \langle F_k(\mathbf{n})\, n_l\rangle &\simeq m_l m_k \\ &+ \sum_{j=1}^{N}\left(L_1(\mu_k,\sigma_k) + \frac{1}{6}\kappa_{3,k}\, L_4(\mu_k,\sigma_k)\right) J_{kj} c_{jl} \\ &+ \frac{1}{2}\left(L_2(\mu_k,\sigma_k) + \frac{1}{6}\kappa_{3,k}\, L_5(\mu_k,\sigma_k)\right)\Delta\kappa_{2,kl} \\ &+ \frac{1}{6}\left(L_3(\mu_k,\sigma_k) + \frac{1}{6}\kappa_{3,k}\, L_6(\mu_k,\sigma_k)\right)\Delta\kappa_{3,kl} \\ &+ o(\kappa_{3,k}^2). \end{aligned} \tag{1.26}$$

The latter expression shows, in particular, that the correction to the mean activity reappears (as the factor m_k) in the first line, indicating that the contribution of the third cumulant to lowest order ($q = 0$ in Eq. (1.19)) drops out of the covariance, as the term $m_l m_k$ is subtracted in Eq. (1.3). In a weakly correlated state of the network, the remaining terms are smaller than $m_l m_k$ as they give rise to the pairwise covariance (1.3). This explains why the Gaussian approximation is already fairly accurate.

Simultaneously solving Eqs. (1.25) and (1.26) by a damped fixed-point iteration, analogous to Eq. (1.16), yields an approximation of the mean activities and covariances shown in Fig. 1.4. The deviation of the mean activities from simulation results

(Fig. 1.4a) is reduced compared to the Gaussian approximation below the significance level. The width of the distributions is only slightly underestimated compared to simulations, as exhibited by the scatter plot aligned to the diagonal and the bar graph. The pairwise averaged cross-covariances (Fig. 1.4b) are within the error of the simulated results, in contrast to the Gaussian approximation (cf. Fig. 1.1b). A small contribution to the remaining difference in variance stems from the finite simulation time, naturally leading to a wider distribution in Fig. 1.4b compared to the theoretical prediction. The bigger contribution presumably comes from the nontrivial third-order cumulants $\langle\!\langle n_i n_j n_k \rangle\!\rangle$, where all indices are unequal. Neglecting the nontrivial covariances shows that the variability of σ_k from neuron to neuron is reduced, which in turn reduces the width of the distribution of the mean activities. For networks with fixed in-degree this even leads to a uniform mean activity across neurons, which, however, still matches the averaged mean activities from simulations indicating that averaged quantities are insensitive to variability in non-trivial higher-order cumulants (see Fig. 1.5 in Sec. 1.6). Analogously, we expect that the neglect of nontrivial third-order cumulants underlies the reduced width of the covariances.

1.5 Scale invariance of covariances

The equation for cross-covariances in Eq. (1.15) yields an additional insight: Assuming that mean activities and correlations were unchanged, a scaling of all incoming synapses to a neuron k by some factor $\alpha > 0$ amounts to a scaling of the strength of synaptic fluctuations in the same manner ($\sigma_k \propto \alpha$). The fixation of mean activities can be achieved for different choices of incoming synaptic amplitudes by adapting the threshold such that $(\mu_k(\{J_{kl}\}) - \theta_k)/\sigma_k$ remains invariant (cf. Eq. (1.15)). A uniform rescaling of all synapses by a factor $\alpha > 0$ therefore requires a change of the threshold by the same factor. The susceptibility (1.14) S_k then scales as $\sigma_k^{-1} \propto \alpha^{-1}$. Hence the term $S_k J_{kl}$ appearing in the equation for covariances in Eq. (1.15) is invariant under this scaling. This result implies that covariances are invariant with respect to the absolute value of synaptic amplitudes: The self-generated network noise causes a divisive normalization on the level of the synaptic input to each neuron. This also implies that scaling the synapses with $J \propto N^{-1}$, $J \propto N^{-\frac{1}{2}}$, or $J \propto 1$ as the network size tends to infinity yields the same covariances, given the thresholds are adapted so that the mean activity is preserved. On the level of population-averaged covariances this has been remarked earlier (Grytskyy et al., 2013a). The independence of the network dynamics on the absolute value of the synaptic weights even holds exactly, as noted earlier (van Albada et al., 2015). The reason for this is the absence of a length scale of a hard threshold; the only relevant length scale is the amplitude σ_k of the synaptic noise itself. Considering a single neuron k, the condition for the neuron to be activated is $h_k = \sum_l J_{kl} n_l > \theta_k$. Rescaling all incoming synapses as well as the threshold by the same factor $\alpha > 0$ multiplies both sides of this inequality; the neuron switches at the very same configuration $\mathbf{n}$ of incoming spins as in the original case.

The previous consideration is completely in line with the work of van Vreeswijk & Sompolinsky (1998). The latter work found that, when increasing the number of neurons N, a scaling of $1/\sqrt{N}$ is required to obtain a robust asynchronous state, if the system possesses variability of the thresholds with a standard deviation defined as unity: The width of the distribution of the thresholds induces a second length scale into the system. Invariant behavior can therefore only arise, if the synaptic noise σ_k and the standard deviation of the thresholds have a constant ratio. Scaling synapses as $J \propto 1/N$ in this setting leads to vanishing temporal fluctuations of h_k and hence spin-glass freezing, whereas for $J \propto 1$ the asynchronous state persists. Fixing J, while increasing the number of neurons N, results in more inputs per neuron and thus increased temporal fluctuations, which wash out the effect of distributed thresholds.

1.6 Mapping of fluctuations to Ornstein-Uhlenbeck processes

Here, we show an alternative interpretation of the Gaussian approximation in Sec. 1.3. Despite the fact that the covariances obey different equations on the diagonal $k = l$ and the off-diagonal $k \neq l$ (1.15), we can rewrite the equations as a single matrix equation

$$\begin{aligned} 2\mathbf{C} - \mathbf{D} &= \mathbf{W}\mathbf{C} + (\mathbf{W}\mathbf{C})^T \\ 0 &= (\mathbf{W} - \mathbf{1})\mathbf{C} + ((\mathbf{W} - \mathbf{1})\mathbf{C})^T + \mathbf{D}, \end{aligned} \tag{1.27}$$

where $\mathbf{D}$ is a diagonal matrix with elements constrained by the condition that the diagonal entries of the covariance matrix fulfill (1.3)

$$c_{kk} = m_k(1 - m_k), \tag{1.28}$$

and $\mathbf{W} = \mathbf{S}\mathbf{J}$ is the matrix of effective couplings which incorporates the susceptibility $\mathbf{S} = \mathrm{diag}(S_1, \ldots, S_N)$ of the individual units. We ensure this latter condition in the following way. We first solve Eq. (1.27) by multiplication with the left eigenvectors $\mathbf{v}_\alpha$ of the effective couplings, i.e. $\mathbf{v}_\alpha^T(\mathbf{W} - \mathbf{1}) = \lambda_\alpha \mathbf{v}_\alpha^T$ (Risken, 1996, chapter 6.5). The corresponding right eigenvectors are $\mathbf{u}_\alpha$, which fulfill the bi-orthogonality $\mathbf{v}_\alpha^T \mathbf{u}_\beta = \delta_{\alpha\beta}$ and completeness $\sum_\alpha \mathbf{u}_\alpha \mathbf{v}_\alpha^T = 1$ relation. With the notation $C^{\alpha\beta} := \mathbf{v}_\alpha^T \mathbf{C} \mathbf{v}_\beta$ (analogously for $\mathbf{D}$) we have

$$\begin{aligned} C^{\alpha\beta} &= -\frac{D^{\alpha\beta}}{\lambda_\alpha + \lambda_\beta}, \\ \mathbf{C} &= \sum_{\alpha,\beta} \mathbf{u}_\alpha \mathbf{u}_\beta^T C^{\alpha\beta}, \end{aligned} \tag{1.29}$$

where the latter expression follows from the completeness relation.

The expression reveals the connection between the eigenvalues λ_α of the linearized coupling matrix and the fluctuations in the system. In the case of a single eigenvalue

close to instability, i.e. $\Re(\lambda_\alpha) \simeq 0$, the fluctuations in the corresponding eigendirection constitute the dominant contribution to the covariance matrix $\mathbf{C} \simeq -\mathbf{u}_\alpha \mathbf{u}_\alpha^T \frac{\mathbf{v}_\alpha^T \mathbf{D} \mathbf{v}_\alpha}{2\lambda_\alpha}$. This scenario could experimentally be detected in a principle-component analysis, with the dominant principle component pointing in the direction of $\mathbf{u}_\alpha$.

Evaluating Eq. (1.29) on the diagonal and using $D^{\alpha\beta} = v_\alpha^T \mathbf{D} v_\beta = \sum_j v_{\alpha,j} v_{\beta,j} D_j$ we get

$$m_k(1-m_k) \stackrel{!}{=} c_{kk} = \sum_j \underbrace{-\left(\sum_{\alpha,\beta} \frac{u_{\alpha,k} v_{\alpha,j} \cdot u_{\beta,k} v_{\beta,j}}{\lambda_\alpha + \lambda_\beta} \right)}_{\equiv B_{kj}} D_j, \tag{1.30}$$

$$\mathbf{B} = -\sum_{\alpha,\beta} \frac{(\mathbf{u}_\alpha \mathbf{v}_\alpha^T) \circledast (\mathbf{u}_\beta \mathbf{v}_\beta^T)}{\lambda_\alpha + \lambda_\beta},$$

where the symbol $\circledast$ is to be understood as the element-wise multiplication of the two matrices (Hadamard product) in the numerator. The penultimate line is an ordinary matrix equation relating the (N-dimensional vector) D_k to the (N-dimensional vector) c_{kk}. To determine $\mathbf{D}$ as $\mathbf{D} = \mathbf{B}^{-1}\mathrm{diag}(\{c_{kk}\})$ we hence need to invert the matrix $\mathbf{B}$. The covariance matrix is then obtained by Eq. (1.29).

The result (1.27) can be further interpreted as a mapping of the fluctuations from the binary dynamics to an effective system of Ornstein-Uhlenbeck processes. Given the set of coupled Ornstein-Uhlenbeck processes

$$\tau \frac{\partial x_k(t)}{\partial t} = -x_k(t) + \sum_j w_{kj} x_j(t) + \xi_k(t) \tag{1.31}$$

with the Gaussian white noise $\langle \xi(t) \xi^T(s) \rangle = \tau \mathbf{D}\, \delta(t-s)$, the stationary equal-time covariance matrix fulfills the same continuous Lyapunov equation (Risken, 1996, chapter 6.5) as the binary network Eq. (1.27). Using the analogy of the continuous Lyapunov equations we see that the elements of the diagonal matrix $\mathbf{D}$ can be interpreted as the noise intensity injected into each neuron. The modified Lyapunov method Eq. (1.30) determines this intensity such that the variance of each continuous variable x_k agrees with that of the corresponding binary variable n_k given by Eq. (1.28), which, in turn, is fixed by the mean activity (1.15).

It is important to note that $\mathbf{W}$ and $\mathbf{D}$ in Eqs. (1.27) and (1.31) themselves depend on the covariances $\mathbf{C}$ via the susceptibility $\mathbf{S}$. This leads to Eq. (1.29) remaining an implicit equation for $\mathbf{C}$ that needs to be solved iteratively. However, by neglecting the contribution of cross-covariances in Eq. (1.5), $\mathbf{W}$ and $\mathbf{D}$ become independent of $\mathbf{C}$, rendering Eq. (1.27) a linear equation for the covariances and the above projection method an efficient algorithm to compute $\mathbf{C}$. This linear approximation consequently fails to predict the distribution of mean activities for networks with fixed in-degree, as shown in Fig. 1.5a. Nevertheless, the width of the distribution of the covariances

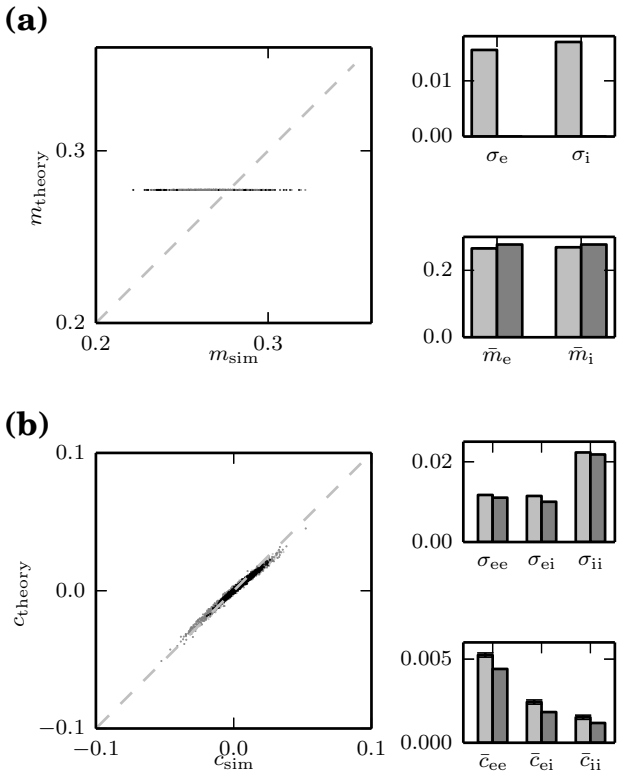


Figure 1.5: **Modified Lyapunov method.** Solution of the individual pairwise covariances by the projection method applied to the continuous Lyapunov equation in comparison to simulation results. **(a)** Mean activities predicted by the approximation that neglects cross-correlations (black: excitatory, gray: inhibitory). Lower inset: mean activity averaged over excitatory and inhibitory neurons (light gray: simulation, darker gray: theory). Black error bars show the standard error of the mean obtained from 20 simulations (barely visible, approximately at line width). Upper inset: width of distribution of mean activities. **(b)** Theoretical distribution of covariances (1.29) with $\mathbf{D}$ chosen according to Eq. (1.30) (same approximation as in (a) for $\mathbf{m}$ and $\mathbf{S}$, i.e. neglecting cross-correlations in the input to each neuron). Lower inset: mean covariances averaged over excitatory-excitatory, excitatory-inhibitory and over inhibitory-inhibitory neuron pairs (light gray: simulation, darker gray: theory). Error bars show the standard error of the mean obtained from 20 simulations. Upper inset: width of the distributions of covariances. Other parameters and display as in Fig. 1.1.

shown in Fig. 1.5b is only slightly underestimated, showing that the distribution of mean activities contributes only marginally to the distribution of covariances.

The viability of the linear approximation for the covariances shows that fluctuations in the binary model are practically equivalent to those in linear Ornstein-Uhlenbeck processes. This equivalence has been reported earlier for effective equations of the population-averaged pairwise correlations (Grytskyy et al., 2013b). The latter work used the approximate result $c_{kk} \simeq \frac{D_k}{2}$, which ignores the effect of the covariances onto the autocovariances in Eq. (1.27), i.e. assuming that $(\mathbf{WC})_{kk}$ is smaller than c_{kk} itself and hence neglecting the right-hand side of Eq. (1.27) when determining the diagonal elements c_{kk}. In the absence of self-connections, this amounts to neglecting the off-diagonal cross-covariances in $\mathbf{C}$. For weakly correlated network states, this is a good approximation. However, in the present network setting, it slightly overestimates the population averages of the covariances as well as the width of their distribution (data not shown).

1.7 Reconstruction of couplings

In the current section, we investigate the inverse problem, i.e., the inference of the couplings J_{kl} of the network from the observed activity. The time-lagged cross-covariance function in a network of Ornstein-Uhlenbeck processes (for $t > s$) fulfills the differential equation (Risken, 1996, chapter 6.5)

$$q_{kl} := \tau \frac{\partial}{\partial t} c_{kl}(t,s) \quad + c_{kl}(t,s) \overset{t \geq s}{=} \sum_j w_{kj} c_{jl}(t,s). \tag{1.32}$$

This means we can uniquely reconstruct the effective couplings w_{kl} from the knowledge of the two matrices, the covariance and its slope at time lag zero as

$$\begin{aligned} \mathbf{W} &= \mathbf{Q}\mathbf{C}^{-1} \\ &= \mathbf{1} + \tau \frac{\partial}{\partial t} \mathbf{C}(t,s)|_{t=s} \, \mathbf{C}^{-1}. \end{aligned} \tag{1.33}$$

For an Ornstein-Uhlenbeck process, obeying Eq. (1.27) and Eq. (1.32), the reconstruction of the couplings w_{kl} is exact, as shown in Fig. 1.7b.

We will now demonstrate that a similar relationship approximately also holds in binary networks. For $t > s$, the time-lagged cross-covariance function $c_{kl}(t,s) = \langle n_k(t) n_l(s) \rangle - \langle n_k(t) \rangle \langle n_l(t) \rangle$ for the binary network fulfills Eq. (A.5)

$$\begin{aligned} q_{kl}(t,s) \quad &:= \quad \tau \frac{\partial}{\partial t} c_{kl}(t,s) + c_{kl}(t,s) \\ &\overset{t \geq s}{=} \quad \langle F_k(\mathbf{n}(t)) \, n_l(s) \rangle - \langle n_k(t) \rangle \langle n_l(s) \rangle. \end{aligned} \tag{1.34}$$

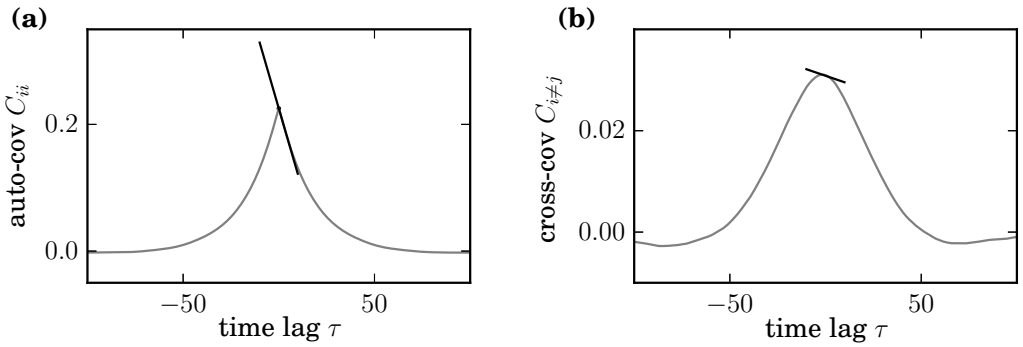


Figure 1.6: **Autocovariance and cross-covariance functions.** Example of an auto- (**a**) and cross-covariance function (**b**) in the network described in Fig. 1.1. The respective slopes of the functions at time lag 0+ are indicated by the black tangent lines.

An example of an auto- and cross-covariance function together with the slope at zero time lag is shown in Fig. 1.6.

In the limit of vanishing time lag the form of $q_{kl} = \lim_{t\to s} q_{kl}(t,s) = \langle F_k(\mathbf{n}(t))\, n_l(t)\rangle - \langle n_k(t)\rangle\langle n_l(t)\rangle$ shows that q_{kl} measures the direct influence of fluctuations of neuron l on the gain function of neuron k. In the Gaussian approximation (see Sec. 1.3) we get for $k \neq l$

$$q_{kl} \quad = \quad \sum_j S_k J_{kj} c_{jl}. \tag{1.35}$$

For $k = l$, however, we have to evaluate $\langle H(h_k(\mathbf{n}(t)-\theta)\, n_k(t)\rangle$ on the right-hand side. Since the term n_k appears in the expectation value, the statistics of h_k is effectively conditioned on the state $n_k = 1$. Since the transitions of neuron k directly depend on the value of h_k, this conditioning violates the close-to Gaussian approximation of $h_k|n_k = 1$. An approximate treatment on the diagonal is possible under the assumption that the autocovariance function of h_k is dominated by contributions of the autocovariances of the binary variables n, yielding a nonlinear differential equation for the temporal shape of the autocovariance function (van Vreeswijk & Sompolinsky, 1998, eq. (5.17)). Approximating the temporal profile of the autocovariance function $a(s) := \langle H(h_k(t+s)-\theta)\, H(h_k(t)-\theta)\rangle \simeq \langle m_k\rangle + \langle m_k\rangle(1-\langle m_k\rangle)\, e^{-\frac{|s|}{\tau}}$, which has the right asymptotic behaviors ($a(0) = m_k$ and $a(s\to\infty) = m_k^2$) and assumes fluctuations to be correlated on the time scale τ of the update, leads to the approximate value $\langle H(h_k(t)-\theta)\, n_k(t)\rangle \simeq \frac{1}{2} m_k(1-m_k)$ (by using van Vreeswijk & Sompolinsky, 1998, eq. (A.10)). The corresponding approximation of the slope following from the former approximation with Eq. (1.34) is shown in Fig. 1.7a. However, the resulting expression for the slope at zero time lag cannot be written in the form Eq. (1.35). Therefore, we assume that Eq. (1.35) also holds on the diagonal, although the resulting predicted slopes have a slight negative bias compared to simulation results.

Given the two covariance matrices entering Eq. (1.34), as well as the average activities m_i for each neuron, we can derive a procedure for reconstructing the couplings J_{kl}/σ_k

relative to the noise. The mean and variance of the input to a neuron are determined by Eq. (1.5). Inverting the relation Eq. (1.6) as $y(m_k) := \frac{\mu_k - \theta}{\sqrt{2}\sigma_k} = -\mathrm{erfc}^{-1}(2m_k)$ allows us to determine the susceptibility Eq. (1.14) as $S(m_k, \sigma_k) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{-y(m_k)^2}$. Hence we obtain an expression for the ratio of the incoming synaptic weight and the total synaptic noise of neuron k

$$\frac{J_{kl}}{\sigma_k} = \sqrt{2\pi} w_{kl} e^{y(m_k)^2} \quad . \tag{1.36}$$

This again shows that correlations are only controlled by the ratio J_{kl}/σ_k rather than by J_{kl} and σ_k alone, in line with the invariance found in Sec. 1.5. Fig. 1.7b shows the reconstruction of couplings from the simulated covariance functions. Due to the previously discussed approximations, the exact reconstruction of the relative couplings J_{kl}/σ_k is not possible for a binary network. The observed deviations from the identity line are predominantly caused by the approximation of the slope of the autocovariance functions. The reconstructed coupling matrix correctly infers all excitatory and all inhibitory connections, but additionally yields a considerable number of false-positive excitatory and inhibitory connections.

The described procedure, moreover, allows us to determine the remaining parameters of the binary network. With regard to correlations, we are free to choose an arbitrary σ_k, e.g. $\sigma_k = 1$ to determine J_{kl} by Eq. (1.36). As the mean activity and correlations are known, the actual magnitude of fluctuations σ_k^{loc} caused by the local inputs from the network follows from Eq. (1.5). If these fluctuations are smaller than or equal to our arbitrary choice ($\sigma_k^{\mathrm{loc}} \leq \sigma_k = 1$), it is possible to supply each neuron with additional noise of variance $1 - \left(\sigma_k^{\mathrm{loc.}}\right)^2$. Only in this case can we construct a binary network satisfying the given constraints. Having fixed J_{kl}, and given the mean activities by Eq. (1.5), determines the mean input μ_k to each cell. Finally the threshold θ must be chosen such that $\theta_k = \mu_k - \sqrt{2}y(m_k)$.

Different routes to inverse problems have been proposed earlier. For equilibrium systems, a statistical-mechanics formulation allows the use of principal component analysis (Cocco et al., 2011) or maximum-entropy approaches (Roudi et al., 2009; Mora et al., 2010). For off-equilibrium systems with a sequential Glauber update in discrete time steps, methods related to our approach have previously been proposed by Mézard & Sakellariou (2011). Their results can be related to our continuous-time method by time discretization of the slope appearing on the left-hand side of (38), leading to similar expressions as their eqs. (8) and (18). An alternative approach, based on the Thouless-Anderson-Palmer (TAP) approximation (Thouless et al., 1977), is described in Zeng et al. (2011, see their eq. (14)).

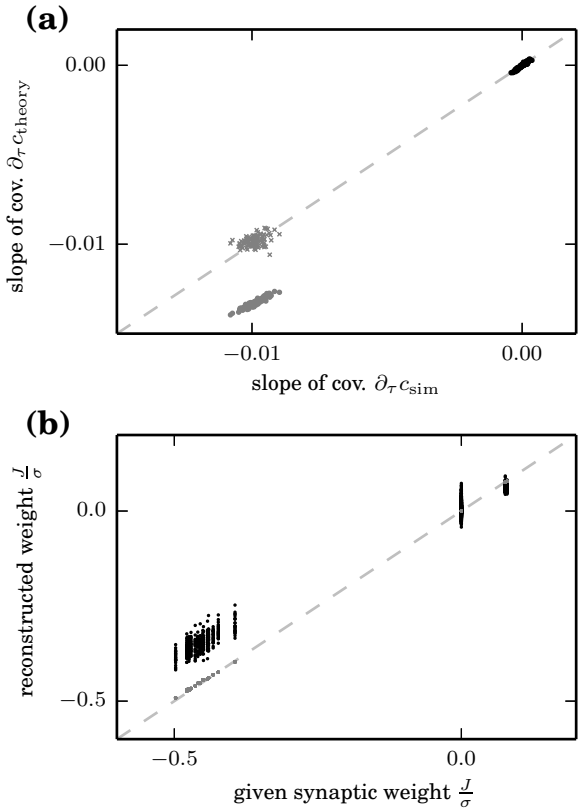


Figure 1.7: **Reconstruction of couplings from the matrix of covariances and slopes of the covariance functions at zero time lag. (a)** Slope of auto- (gray) and cross-covariance functions (black) at zero time lag from the Gaussian approximation versus slope taken from simulation results. Dots show the prediction of the linear theory Eq. (1.32), crosses the approximation $-\frac{1}{2}m_k(1-m_k)$ for the slope of auto-covariances. **(b)** Reconstructed synaptic amplitude w_{kl} in a network of Ornstein-Uhlenbeck processes Eq. (1.33) (gray) and reconstructed weights J_{kl}/σ_k Eq. (1.36) (black) in a binary network versus original couplings used in simulation. Covariance and slope averaged over 20 repetitions, each simulated for $T = 2,000,000$ ms. Other network parameters as in Fig. 1.1.

1.8 Conclusion

In this chapter, we presented a theoretical description of fluctuations in strongly-coupled networks of binary (Ising) units that, for the first time, faithfully captures the statistics of individual units as well as pairs of units. The fluctuations are characterized by self-consistent equations for the mean activity and pairwise correlations, including finite-size effects down to hundreds of units. The method considers a single realization of the random couplings and can be applied to a wide range of networks, in particular networks with asymmetric and strong couplings. The systematic cumulant expansion of single unit activities goes beyond mean-field theory and is applicable outside thermodynamic equilibrium. The linearized theory yields an expansion of the correlation functions in collective eigenmodes, leads to an efficient algorithm solving the inverse problem, and shows that correlations are invariant under scaling of the interaction strengths. The exposed equivalence to systems of Ornstein-Uhlenbeck processes on the single unit level builds the corner stone for further analytical investigations of such disordered systems and their correlation structure.

Chapter 2

Functional methods for disordered networks

This chapter is based in parts on the following publications:

[1] Schuecker J, Goedeke S, Dahmen D and Helias M (2016) *Functional methods for disordered neural networks*, arXiv 1605.06758
[2] Dahmen D, Diesmann M and Helias M (2016) *Distributions of covariances as a window into the operational regime of neuronal networks*, arXiv 1605.04153
Preliminary results have been published in Dahmen et al. (2015, 2016a).

Author contributions:
[1] Under the supervision of M Helias, the author performed all parts of the above publication that are included in this thesis. Supervised by M Helias, J Schuecker and S Goedeke developed sections II C,D,E,F of the publication, which are not included in this thesis, and contributed to the derivations presented in section 2.3.2. All authors contributed to the writing of the manuscript.
[2] Under the supervision of M Helias, the author performed all parts of the above publication. All authors contributed to the design of the numerical experiments and the writing of the manuscript.

2.1 Introduction

In the previous chapter, we established the link between networks of binary neurons and coupled Ornstein-Uhlenbeck processes. A similar mapping has been derived from linear response theory for other neuron models (Grytskyy et al., 2013b), such as leaky integrate-and-fire (Lindner et al., 2005; Pernice et al., 2012; Trousdale et al., 2012) and Poisson model neurons (Hawkes, 1971; Pernice et al., 2011). Therefore, many qualitative features of emerging collective phenomena, such as correlated activity (e.g. in Trousdale et al., 2012; Tetzlaff et al., 2012; Helias et al., 2014), stability, response to inputs and oscillations (Bos et al., 2015), as well as chaotic and regular behavior (Sompolinsky et al., 1988; Goedeke et al., 2016) can be understood in such simple linear rate models and their nonlinear generalizations. These are accessible to a treatment in statistical mechanics, or, more precisely, classical statistical field theory.

This chapter presents the fundamentals behind contemporary developments in the theory of neural networks of rate units (e.g. Aljadeff et al., 2015; Stern et al., 2014; Hermann & Touboul, 2012) that are based on methods from statistical mechanics of classical systems with a large number of interacting degrees of freedom. In particular, we focus on a relevant class of systems that have quenched (time independent) disorder. In neural networks, the main source of disorder arises from random synaptic couplings between neurons. These systems are in many respects similar to spin glasses (Fischer & Hertz, 1991a). The chapter therefore also explains the methods for these disordered systems as far as they are applied in neuroscience. The disorder average that can be performed in this field-theoretic formulation allows to go beyond the description level of a single realization of the disordered connectivity towards a statistical description of the ensemble of networks. The latter is much simpler to treat analytically as it only depends on the statistical properties of the disorder and therefore exposes more symmetries than the single realizations.

The presentation consists of two parts. In the first part, we introduce stochastic differential equations in the Ito formulation and present their treatment in the Martin–Siggia–Rose-De Dominicis path integral formalism (Martin et al., 1973; De Dominicis & Peliti, 1978), reviewed in Altland & B. (2010); Chow & Buice (2015); Hertz et al. (2016). In the second part, we will employ this language to derive the dynamic mean-field theory for random networks. In deriving the formalism, we will follow the De Dominicis approach (De Dominicis & Peliti, 1978), that was also employed to obtain the dynamic mean-field theory of spin glasses (Sompolinsky & Zippelius, 1981, 1982; Crisanti & Sompolinsky, 1987). Furthermore, we extend mean-field theory to capture finite-size effects arising from mean connections and fluctuations of auxiliary fields (Dahmen et al., 2016b). In the subsequent Chap. 3, this extension is shown to go beyond self-averaging as it enables the assessment of the meta-statistics of activity, a setting beyond previously developed mean-field techniques that employ the thermodynamic limit $N \to \infty$.

2.2 Functional formulation of stochastic differential equations

We here follow Chow & Buice (2010) to derive the Martin-Siggia-Rose (Martin et al., 1973; Janssen, 1976; De Dominicis, 1976; De Dominicis & Peliti, 1978; Altland & B., 2010; Chow & Buice, 2015) path integral representation of a stochastic differential equation and Wio et al. (1989) to obtain the Onsager-Machlup path integral. Hertz et al. (2016) also provide a pedagogical survey of the Martin-Siggia-Rose path integral formalism for the dynamics of stochastic and disordered systems.

The presented functional formulation of dynamics is advantageous in several respects. First, it recasts the dynamical equations into a path integral, where the dynamic equations give rise to the definition of an "action". In this way, the known tools from theoretical physics, such as perturbation expansions with the help of Feynman diagrams or the loopwise expansions (Zinn-Justin, 1996) to obtain a systematic treatment of fluctuations can be applied. Within neuroscience, the recent review by Chow & Buice (2015) illustrates the first, the work by Buice & Cowan (2007) the latter approach. Moreover, this formulation will be essential for the treatment of disordered systems in Sec. 2.3, following the spirit of the work by De Dominicis & Peliti (1978), to obtain a generating functional that describes an average system belonging to an ensemble of systems with random parameters.

Many dynamic phenomena can be described by differential equations. Often, the presence of fluctuations is represented by an additional stochastic forcing. We therefore consider the stochastic differential equation (SDE)

$$\begin{aligned} dx(t) &= f(x)\,dt + d\mathcal{W}(t) \\ x(0+) &= a, \end{aligned} \tag{2.1}$$

where a is the initial value and $d\mathcal{W}$ a stochastic increment. Stochastic differential equations are defined as the limit $h \to 0$ of a dynamics on a discrete time lattice of spacing h. For discrete time $t_l = lh$, $l = 0, \ldots, M$, the solution of the SDE consists of the discrete set of points $x_l = x(t_l)$. For the discretization there are mainly two conventions used, the Ito and the Stratonovich convention (Gardiner, 2009). Since we only consider additive noise, i.e. the stochastic increment in Eq. (2.1) does not depend on the state, both conventions yield the same continuous-time limit. In the limit $h \to 0$, the symbolic notation of Eq. (2.1) can be interpreted as

$$x_{i+1} - x_i = f(x_i)\,h + a\delta_{i0} + \mathcal{W}_i, \tag{2.2}$$

where $\mathcal{W}_i$ is a stochastic increment that follows a probabilistic law ("Euler-Maruyama" scheme, see Chap. 4). A common choice for $\mathcal{W}_i$ is a normal distribution $\rho(\mathcal{W}_i) = \mathcal{N}(0, hD)$, called a Wiener increment. Here the parameter D controls the variance of the noise. The term $a\delta_{i0}$ ensures that the solution obeys the stated initial condition, assuming that $x_{i\leq 0} = 0$ in the absence of noise $\mathcal{W}_0 = 0$. Another widely-used differential

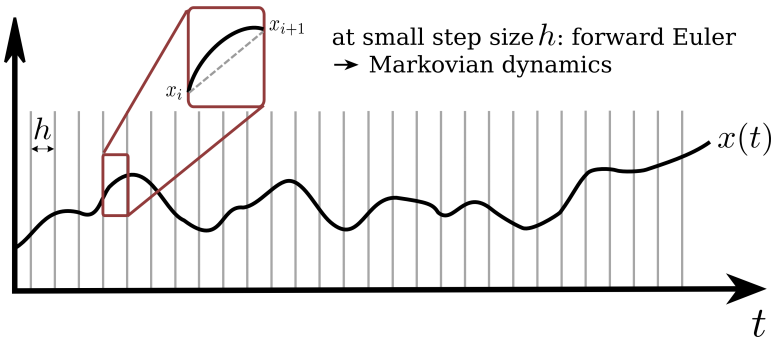


Figure 2.1: **Illustration of the path integral formalism for stochastic differential equations.** The discretized temporal dynamics is approximated using Euler steps such that the probability $p(x(t)) = p(x_1, \ldots, x_M|a)$ for the path $x(t)$ factorizes into Markovian update probabilities $p(x_{i+1}|x_i)$. The product of the latter can be expressed as a path integral in the limit $M \to \infty$.

notation for Eq. (2.1) with the Wiener process $\mathcal{W}(t)$ is the so-called Langevin-form

$$\frac{dx(t)}{dt} = f(x) + \xi(t) \tag{2.3}$$

with Gaussian white noise $\xi(t)$ defined via $d\mathcal{W}(t) = \xi(t)dt$, such that $\langle \xi(t) \rangle = 0$ and $\langle \xi(t)\xi(t') \rangle = D\delta(t - t')$. The moments follow from the fact that increments of the Wiener process are independent and that the variance of the increment $\mathcal{W}_i$ is proportional to the time step h. More details on stochastic differential equations will be presented in Chap. 4. Due to independent noise at each time step, the probability density of the path $x(t)$, i.e. a distribution in the points $x_1, \ldots, x_M$, can be written as

$$p(x_1, \ldots, x_M|a) \equiv \int \Pi_{i=0}^{M-1} d\mathcal{W}_i \, \rho(\mathcal{W}_i) \, \delta(x_{i+1} - y_{i+1}(\mathcal{W}_i, x_i)), \tag{2.4}$$

where, by Eq. (2.2), $y_{i+1}(\mathcal{W}_i, x_i) = x_i + f(x_i)\, h + a\delta_{i0} + \mathcal{W}_i$ is understood as the solution of Eq. (2.2) at time point $i+1$ given the noise realization $\mathcal{W}_i$ and the solution until the previous time point x_i (Fig. 2.1).

The notation $y_{i+1}(\mathcal{W}_i, x_i)$ indicates that the solution only depends on the last time point x_i, but not on the history longer ago, which is called the Markov property of the process. This form also shows that the density is correctly normalized, because integrating over all paths

$$\int dx_1 \cdots \int dx_M \, p(x_1, \ldots, x_M|a) = \int \Pi_{i=0}^{M-1} d\mathcal{W}_i \rho(\mathcal{W}_i) \underbrace{\int dx_{i+1} \, \delta(x_{i+1} - y_{i+1}(\mathcal{W}_i, x_i))}_{=1}$$

$$= \Pi_{i=0}^{M-1} \int d\mathcal{W}_i \rho(\mathcal{W}_i) = 1$$

yields the normalization condition of $\rho(\mathcal{W}_i)$, $i = 0, \ldots, M-1$, the distribution of the stochastic increments. In the limit $M \to \infty$, we therefore define the probability functional as $p[x|a] := \lim_{M\to\infty} p(x_1, \ldots, x_M|a)$.

Using Eq. (2.4) and the substitution $\delta(y)\,dy = \delta(\phi(x_{i+1}))\phi' dx_{i+1}$ with $y = \phi(x_{i+1}) = \mathcal{W}_i(x_{i+1})$ obtained by solving Eq. (2.2) for $\mathcal{W}_i$

$$\begin{aligned} \mathcal{W}_i(x_{i+1}) &= x_{i+1} - x_i - f(x_i)\,h - a\delta_{i0} \\ \frac{\partial \mathcal{W}_i}{\partial x_{i+1}} &= \phi' = 1 \end{aligned} \tag{2.5}$$

we obtain

$$\begin{aligned} p(x_1, \ldots, x_M|a) &= \int \Pi_{i=0}^{M-1} d\mathcal{W}_i\, \rho(\mathcal{W}_i) \times \\ &\times \delta(\mathcal{W}_i - \underbrace{x_{i+1} - x_i - f(x_i)\,h - a\delta_{i0}}_{\mathcal{W}_i(x_{i+1})}). \\ &= \Pi_{i=0}^{M-1} \rho(x_{i+1} - x_i - f(x_i)\,h - a\delta_{i0}). \end{aligned} \tag{2.6}$$

In Sec. 2.2.1 we will look at the special case of Gaussian noise and derive the so called Onsager-Machlup path integral (Onsager & Machlup, 1953). This path integral has a square in the action, originating from the Gaussian noise. For many applications, this square complicates the analysis of the system. The formulation presented in Sec. 2.2.2 removes this square on the expense of the introduction of an additional field, the so called response field.

2.2.1 Onsager-Machlup path integral

For the case of a Gaussian noise $\rho(\mathcal{W}_i) = \mathcal{N}(0, Dh) = \frac{1}{\sqrt{2\pi Dh}} e^{-\frac{\mathcal{W}_i^2}{2Dh}}$ the variance of the increment is

$$\begin{aligned} \langle \mathcal{W}_i \mathcal{W}_j \rangle &= \begin{cases} Dh & i = j \\ 0 & i \neq j \end{cases} \\ &= \delta_{ij}\, Dh. \end{aligned} \tag{2.7}$$

Using the Gaussian noise and then taking the limit $M \to \infty$ of Eq. (2.6) we obtain

$$\begin{aligned} p(x_1, \ldots, x_M|a) &= \Pi_{i=0}^{M-1} \rho(x_{i+1} - x_i - f(x_i)\,h - a\delta_{i0}) \\ &= \Pi_{i=0}^{M-1} \frac{1}{\sqrt{2\pi Dh}} \exp\left[-\frac{1}{2Dh}(x_{i+1} - x_i - f(x_i)\,h - a\delta_{i0})^2\right] \\ &= \left(\frac{1}{\sqrt{2\pi Dh}}\right)^M \exp\left[-\frac{1}{2D}\sum_{i=0}^{M-1}\left[(\frac{x_{i+1} - x_i}{h} - f(x_i) - a\frac{\delta_{i0}}{h})^2\right]h\right]. \end{aligned}$$

We will now define a symbolic notation by recognizing $\lim_{h\to 0} \frac{x_{i+1}-x_i}{h} = \partial_t x(t)$ as well as $\lim_{h\to 0} \frac{\delta_{i0}}{h} = \delta(t)$ and $\lim_{h\to 0} \sum_i f(hi)\, h = \int f(t)\, dt$

$$p[x|x(0+)=a]\, \mathcal{D}_{\sqrt{2\pi Dh}} x = \exp\left(-\frac{1}{2D}\int_0^T (\partial_t x - f(x) - a\delta(t))^2\, dt\right) \mathcal{D}_{\sqrt{2\pi Dh}} x \qquad (2.8)$$
$$:= \lim_{M\to\infty} p(x_1,\ldots,x_M|a)\, dx_1 \ldots dx_M,$$

where we defined the integral measure $\mathcal{D}_{\sqrt{2\pi Dh}} x := \Pi_{i=1}^{M} \frac{dx_i}{\sqrt{2\pi Dh}}$ to obtain a normalized density $1 = \int \mathcal{D}_{\sqrt{2\pi Dh}} x\, p[x|x(0+)=a]$.

2.2.2 Martin-Siggia-Rose-De Dominicis-Janssen (MSRDJ) path integral

The square in the action Eq. (2.8) sometimes has disadvantages for analytical reasons, for example if quenched averages are to be calculated, as we will do in Sec. 2.3. To avoid the square we here introduce an auxiliary field, the response field $\tilde{x}$. This field enters the probability functional Eq. (2.6) by representing the δ-distribution by its Fourier integral

$$\delta(x) \;=\; \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} d\tilde{x}\, e^{\tilde{x}x}. \qquad (2.9)$$

Replacing the δ-distribution at each time slice by an integral over $\tilde{x}_i$ at the corresponding slice, Eq. (2.6) takes the form

$$\begin{aligned} p(x_1,\ldots,x_M|a) &= \Pi_{i=0}^{M-1}\left\{\int d\mathcal{W}_i \rho(\mathcal{W}_i) \int_{-i\infty}^{i\infty} \frac{d\tilde{x}_i}{2\pi i}\, e^{\tilde{x}_i(x_{i+1}-x_i-f(x_i)h-\mathcal{W}_i-a\delta_{i0})}\right\} \\ &= \Pi_{i=0}^{M-1}\left\{\int_{-i\infty}^{i\infty} \frac{d\tilde{x}_i}{2\pi i}\, e^{\tilde{x}_i(x_{i+1}-x_i-f(x_i)h-a\delta_{i0})}\, Z_{\mathcal{W}}(-\tilde{x}_i)\right\} \qquad (2.10) \\ Z_{\mathcal{W}}(-\tilde{x}) &\equiv \int d\mathcal{W}_i \rho(\mathcal{W}_i)\, e^{-\tilde{x}\mathcal{W}_i} = \langle e^{-\tilde{x}\mathcal{W}_i}\rangle_{\mathcal{W}_i}. \end{aligned}$$

Here $Z_{\mathcal{W}}(-\tilde{x})$ is the moment generating function (Gardiner, 1985) also known as the characteristic function of the noise process, which is identical to the Fourier transform of the density (with $i\omega = -\tilde{x}$). Note the index i of the field $\tilde{x}_i$ is the same as the index of the noise variable $\mathcal{W}_i$, which allows the definition of the characteristic function $Z_{\mathcal{W}}$. Hence the distribution of the noise only appears in the probability functional in the form of $Z_{\mathcal{W}}(-\tilde{x})$. For Gaussian noise Eq. (2.7) the characteristic function is

$$\begin{aligned} Z_{\mathcal{W}}(-\tilde{x}) &= \frac{1}{\sqrt{2\pi Dh}}\int d\mathcal{W}\, e^{-\frac{\mathcal{W}^2}{2Dh}}\, e^{-\tilde{x}\mathcal{W}} = \frac{1}{\sqrt{2\pi Dh}}\int d\mathcal{W}\, e^{-\frac{1}{2Dh}(\mathcal{W}+Dh\tilde{x})^2}\, e^{\frac{Dh}{2}\tilde{x}^2} \\ &= e^{\frac{Dh}{2}\tilde{x}^2}. \qquad (2.11) \end{aligned}$$

2.2.3 Moment generating functional

The probability distribution (2.10) is a distribution for the random variables $x_1, \ldots, x_M$. We can alternatively describe the probability distribution by the moment-generating functional by adding the terms $\sum_{l=1}^{M} j_l x_l h$ to the action and integrating over all paths

$$Z(j_1, \ldots, j_M) \;:=\; \Pi_{l=1}^{M} \left\{ \int_{-\infty}^{\infty} dx_l \, \exp\left(j_l x_l h\right) \right\} p(x_1, \ldots, x_M | a). \tag{2.12}$$

Moments of the path can be obtained by taking derivatives (writing $\mathbf{j} = (j_1, \ldots, j_M)$)

$$\begin{aligned} \left. \frac{\partial}{\partial(h\, j_k)} Z(\mathbf{j}) \right|_{\mathbf{j}=0} &= \Pi_{l=1}^{M} \left\{ \int_{-\infty}^{\infty} dx_l \right\} p(x_1, \ldots, x_M | a)\, x_k \\ &\equiv \langle x_k \rangle. \end{aligned} \tag{2.13}$$

For $M \to \infty$ and $h \to 0$ the additional term $\exp\left(\sum_{l=1}^{M} j_l\, x_l h\right) \overset{h\to 0}{\to} \exp\left(\int j(t) x(t)\, dt\right)$. So the derivative on the left hand side of Eq. (2.13) turns into the functional derivative

$$\frac{\partial}{\partial(h j_k)} Z(\mathbf{j}) \equiv \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left(Z(j_1, \ldots, j_k + \frac{\epsilon}{h}, j_{k+1}, \ldots, j_M] - Z(j_1, \ldots, j_k, \ldots, j_M) \right) \overset{h\to 0}{\to} \frac{\delta}{\delta j(t)} Z[j],$$

and the moment becomes $\langle x(t) \rangle$ at time point $t = hk$. The generating functional takes the explicit form

$$\begin{aligned} Z(\mathbf{j}) \;=\; & \Pi_{l=1}^{M} \left\{ \int_{-\infty}^{\infty} dx_l \exp\left(j_l x_l h\right) \right\} \Pi_{k=0}^{M-1} \left\{ \int_{-i\infty}^{i\infty} \frac{d\tilde{x}_k}{2\pi i} \, Z_{\mathcal{W}}(-\tilde{x}_k) \right\} \times \\ & \times \exp\left(\sum_{l=0}^{M-1} \tilde{x}_l (x_{l+1} - x_l - f(x_l) h - a \delta_{l0}) \right). \end{aligned} \tag{2.14}$$

Note the different index ranges for the path coordinates $x_1, \ldots, x_M$ and the response field $\tilde{x}_0, \ldots, \tilde{x}_{M-1}$. Letting $h \to 0$ we now define the path integral as the generating functional (2.14) and introduce the notations $\Pi_{i=1}^{M} \int_{-\infty}^{\infty} dx_i \overset{h\to 0}{\to} \int \mathcal{D}x$ as well as $\Pi_{i=0}^{M-1} \int_{-i\infty}^{i\infty} \frac{d\tilde{x}_i}{2\pi i} \overset{h\to 0}{\to} \int \mathcal{D}\tilde{x}$. Note that the different index ranges and the different integral boundaries are implicit in this notation, depending on whether we integrate over $x(t)$ or $\tilde{x}(t)$. We hence write symbolically for the probability distribution (2.10)

$$\begin{aligned} p[x(t)|x(0+) = a] &= \int \mathcal{D}\tilde{x} \, \exp\left(\int_{-\infty}^{\infty} \tilde{x}(t) (\partial_t x - f(x) - a\delta(t))\, dt \right) Z_{\mathcal{W}}[-\tilde{x}] \\ &= \int \mathcal{D}\tilde{x} \, \exp\left(\tilde{x}^{\mathrm{T}} (\partial_t x - f(x) - a\delta(t)) \right) Z_{\mathcal{W}}[-\tilde{x}] \end{aligned} \tag{2.15}$$

$$\begin{aligned} Z_{\mathcal{W}}[-\tilde{x}_k] &= \left\langle \exp\left(- \int_{-\infty}^{\infty} \tilde{x}(t)\, d\mathcal{W}(t) \right) \right\rangle_{\mathcal{W}} \\ &= \left\langle \exp\left(-\tilde{x}^{\mathrm{T}} d\mathcal{W} \right) \right\rangle_{\mathcal{W}} \end{aligned}$$

where the respective second lines use the definition of the inner product on the space of functions

$$x^{\mathrm{T}} y := \int_{-\infty}^{\infty} x(t) y(t)\, dt. \tag{2.16}$$

This vectorial notation also reminds us of the discrete origin of the path integral. Note that the lattice derivative appearing in Eq. (2.15) follows the definition $\partial_t x = \lim_{h\to 0} \frac{1}{h}\left(x_{t/h+1} - x_{t/h}\right)$. We compactly denote the generating functional (2.14) as

$$Z[j] = \int \mathcal{D}x \int \mathcal{D}\tilde{x}\, \exp\left(\int \tilde{x}(t)(\partial_t x - f(x) - a\delta(t)) + j(t)x(t)\, dt \right) Z_{\mathcal{W}}[-\tilde{x}]. \tag{2.17}$$

For Gaussian white noise we have with Eq. (2.11) the moment generating functional $Z_{\mathcal{W}}[-\tilde{x}] = \exp\left(\frac{D}{2}\tilde{x}^{\mathrm{T}}\tilde{x}\right)$ and we get

$$Z[j] = \int \mathcal{D}x \int \mathcal{D}\tilde{x}\, \exp\left(\tilde{x}^{\mathrm{T}}(\partial_t x - f(x) - a\delta(t)) + \frac{D}{2}\tilde{x}^{\mathrm{T}}\tilde{x} + j^{\mathrm{T}} x \right). \tag{2.18}$$

2.3 Dynamic mean-field theory for random networks

Systems with many interacting degrees of freedom present a central quest in physics. While disordered equilibrium systems show fascinating properties such as the spin-glass transition (Parisi, 1980; Sompolinsky & Zippelius, 1981), new collective phenomena arise in non-equilibrium systems: Large random networks of neuron-like units can exhibit complex (chaotic) dynamics (Sompolinsky et al., 1988; van Vreeswijk & Sompolinsky, 1996; Monteforte & Wolf, 2010) with important functional consequences (Legenstein & Maass, 2007a; Sussillo & Abbott, 2009; Toyoizumi & Abbott, 2011). In the following, we characterize the dynamics by studying the emergent correlation structure in these systems.

2.3.1 Definition of the model and generating functional

We study the coupled set of first order stochastic differential equations

$$d\mathbf{x}(t) + \mathbf{x}(t)\, dt = \boldsymbol{W}\phi(\mathbf{x}(t))\, dt + d\boldsymbol{\mathcal{W}}(t), \tag{2.19}$$

where

$$W_{ij} \sim \begin{cases} \mathcal{N}(0, \frac{\sigma^2}{N}) \text{ i.i.d.} & \text{for } i \neq j \\ 0 & \text{for } i = j \end{cases} \tag{2.20}$$

are i.i.d. Gaussian random couplings, ϕ is a (non-linear) gain function applied element-wise, the $d\mathcal{W}_i$ are pairwise uncorrelated Wiener processes with $\langle d\mathcal{W}_i^2(t)\rangle = D\,dt$.

We formulate the problem in terms of a generating functional from which we can derive all moments of the activity. Introducing the notation $\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{x} = \sum_i \int \tilde{x}_i(t)x_i(t)\,dt$, we obtain the moment-generating functional

$$\begin{aligned} Z[\mathbf{j}] &= \int \mathcal{D}\mathbf{x} \int \mathcal{D}\tilde{\mathbf{x}}\, \exp\left(S_0[\mathbf{x},\tilde{\mathbf{x}}] - \tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{W}\boldsymbol{\phi}\left(\mathbf{x}\right) + \mathbf{j}^{\mathrm{T}}\mathbf{x}\right) \\ \text{with } S_0[\mathbf{x},\tilde{\mathbf{x}}] &= \tilde{\mathbf{x}}^{\mathrm{T}}\left(\partial_t + 1\right)\mathbf{x} + \frac{D}{2}\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}, \end{aligned} \tag{2.21}$$

where the measures are defined as $\int \mathcal{D}\mathbf{x} = \lim_{M\to\infty} \Pi_{i=1}^{N}\Pi_{k=1}^{M} \int_{-\infty}^{\infty} dx_i^k$ and $\int \mathcal{D}\tilde{\mathbf{x}} = \lim_{M\to\infty} \Pi_{i=1}^{N}\Pi_{k=0}^{M-1} \int_{-i\infty}^{i\infty} \frac{d\tilde{x}_i^k}{2\pi i}$. Here the superscript k denotes the kth time slice. The action S_0 is defined to contain all single unit properties, therefore excluding the coupling term $-\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{W}\boldsymbol{\phi}\left(\mathbf{x}\right)$, which is written explicitly.

2.3.2 Average over the quenched disorder

The dynamics of Eq. (2.19) shows invariant features independent of the actual realization of the couplings, only dependent on their statistics, here parameterized by σ. To capture these properties that are generic to the ensemble of the models, we introduce the averaged functional

$$\langle Z[\mathbf{j}]\rangle \;\coloneqq\; \int \Pi_{ij} dW_{ij}\, \mathcal{N}(0, \frac{\sigma^2}{N}, W_{ij})\, Z[\mathbf{j}]. \tag{2.22}$$

We use that the coupling term $\exp(-\sum_{i\neq j} W_{ij}\int \tilde{x}_i(t)\phi(x_j(t))\,dt)$ in Eq. (2.21) factorizes into $\Pi_{i\neq j}\exp(-W_{ij}\int \tilde{x}_i(t)\phi(x_j(t))\,dt)$ as does the distribution over the couplings (due to W_{ij} being independently distributed). We make use of the couplings appearing linearly in the action and complete the square (in each W_{ij} separately) to obtain for $i \neq j$

$$\begin{aligned} &\int dW_{ij}\mathcal{N}(0,\frac{\sigma^2}{N},W_{ij})\, \exp\left(-W_{ij}\int \tilde{x}_i(t)\phi(x_j(t))\,dt\right) \\ =\;& \exp\left(\frac{\sigma^2}{2N}\left(\int \tilde{x}_i(t)\phi(x_j(t))\,dt\right)^2\right). \end{aligned} \tag{2.23}$$

We reorganize the last term including the sum $\sum_{i\neq j}$ as

$$\begin{aligned}
&\exp\left(\frac{\sigma^2}{2N}\sum_{i\neq j}\left(\int \tilde{x}_i(t)\phi(x_j(t))\,dt\right)^2\right)\\
=\;&\exp\left(\frac{\sigma^2}{2N}\sum_{i\neq j}\int\int \tilde{x}_i(t)\phi(x_j(t))\,\tilde{x}_i(t')\phi(x_j(t'))\,dt\,dt'\right)\\
=\;&\exp\left(\frac{1}{2}\int\int\left(\sum_i \tilde{x}_i(t)\tilde{x}_i(t')\right)\left(\frac{\sigma^2}{N}\sum_j \phi(x_j(t))\phi(x_j(t'))\right)dt\,dt'\right)\times\\
&\exp\left(-\frac{\sigma^2}{2N}\int\int\sum_i \tilde{x}_i(t)\tilde{x}_i(t')\phi(x_i(t))\phi(x_i(t'))\,dt\,dt'\right),
\end{aligned}$$

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 last term is the diagonal element that is to be taken out of the double sum. It is a correction of order N^{-1} and will be neglected in the following. The disorder-averaged generating functional Eq. (2.22) therefore takes the form

$$\begin{aligned}
\langle Z[\mathbf{j}]\rangle \;=\;& \int\mathcal{D}\mathbf{x}\int\mathcal{D}\tilde{\mathbf{x}}\,\exp\left(S_0[\mathbf{x},\tilde{\mathbf{x}}]+\mathbf{j}^{\mathrm{T}}\mathbf{x}\right)\times \\
&\times\exp\Big(\frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\sum_i \tilde{x}_i(t)\tilde{x}_i(t')\right)\underbrace{\left(\frac{\sigma^2}{N}\sum_j \phi(x_j(t))\phi(x_j(t'))\right)}_{=:Q_1(t,t')}dt\,dt'\Big).
\end{aligned}\tag{2.24}$$

The coupling term in the last line contains quantities that depend on four fields. We now aim to decouple these terms into terms of products of pairs of fields. The aim is to make use of the central limit theorem, namely that the quantity Q_1 indicated by the curly braces in Eq. (2.24) is a superposition of a large (N) number of (weakly correlated) contributions, which will hence approach a Gaussian distribution. The outcome of the saddle point approximation to lowest order will be the replacement of Q_1 by its expectation value, as we will see in the following steps. We define

$$Q_1(t,s) := \frac{\sigma^2}{N}\sum_j \phi(x_j(t))\phi(x_j(s)) \tag{2.25}$$

and enforce this condition by inserting the Dirac-δ functional

$$\begin{aligned}
&\delta[-\frac{N}{\sigma^2}Q_1(s,t)+\sum_j \phi(x_j(s))\,\phi(x_j(t))] \\
&=\int\mathcal{D}Q_2\,\exp\left(\iint\; Q_2(s,t)\left[-\frac{N}{\sigma^2}Q_1(s,t)+\sum_j \phi(x_j(s))\,\phi(x_j(t))\right]ds\,dt\right).
\end{aligned}\tag{2.26}$$

We here note that as for the response field, the field $Q_2 \in i\mathbb{R}$ is purely imaginary due to the Fourier representation Eq. (2.9) of the δ.

We aim at a set of self-consistent equations for the auxiliary fields. We therefore introduce one source term for each of the fields to be determined. Extending our notation by defining $Q_1^\mathrm{T} Q_2 := \iint Q_1(s,t)\, Q_2(s,t)\, ds\, dt$ and $\tilde{x}^\mathrm{T} Q_1 \tilde{x} := \iint \tilde{x}(s)\, Q_1(s,t)\, \tilde{x}(t)\, ds\, dt$ we hence rewrite Eq. (2.24) as

$$\begin{aligned} \langle Z \rangle &= \int \mathcal{D}Q_1 \int \mathcal{D}Q_2 \exp\left(-\frac{N}{\sigma^2} Q_1^\mathrm{T} Q_2 + N \ln Z[Q_1, Q_2]\right) \\ Z[Q_1, Q_2] &= \int \mathcal{D}x \int \mathcal{D}\tilde{x} \exp\left(S_0[x,\tilde{x}] + \frac{1}{2}\tilde{x}^\mathrm{T} Q_1 \tilde{x} + \phi(x)^\mathrm{T} Q_2 \phi(x)\right), \end{aligned} \tag{2.27}$$

where integral measures $\mathcal{D}Q_{1,2}$ 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 latter only contains functional integrals over the two scalar fields x, $\tilde{x}$. This shows that we have reduced the problem of N interacting units to that of a single unit exposed to a set of external fields Q_1 and Q_2.

The remaining problem can be considered a field theory for the auxiliary fields Q_1 and Q_2. The form of Eq. (2.27) clearly exposes the N dependence of the action for these latter fields in Eq. (2.27): It is of the form $\int dQ \exp(Nf(Q))\, dQ$, which, for large N, suggests a saddle point approximation.

In the saddle point approximation (Sompolinsky & Zippelius, 1982) we seek the stationary point of the action determined by

$$0 = \frac{\delta S[Q_1, Q_2]}{\delta Q_\alpha} = \frac{\delta}{\delta Q_\alpha}\left(-\frac{N}{\sigma^2} Q_1^\mathrm{T} Q_2 + N \ln Z[Q_1, Q_2]\right) = 0 \qquad \alpha \in \{1, 2\}. \tag{2.28}$$

We here set the value for the source fields $j = \tilde{j} = 0$ to zero. This corresponds to finding the point in the space (Q_1, Q_2) which provides the dominant contribution to the probability mass. This can be seen by writing the probability functional as $p[\mathbf{x}] = \iint DQ_1 DQ_2\, p[\mathbf{x}; Q_1, Q_2]$ with

$$\begin{aligned} p[\mathbf{x}; Q_1, Q_2] &= \exp\left(-\frac{N}{\sigma^2} Q_1^\mathrm{T} Q_2 + \sum_i \ln \int D\tilde{x} \exp\left(S_0[x_i, \tilde{x}] + \frac{1}{2}\tilde{x}^\mathrm{T} Q_1 \tilde{x} + \phi(x_i)^\mathrm{T} Q_2 \phi(x_i)\right)\right) \\ b[Q_1, Q_2] &:= \int \mathcal{D}\mathbf{x}\, p[\mathbf{x}; Q_1, Q_2], \end{aligned} \tag{2.29}$$

where we defined $b[Q_1, Q_2]$ as the contribution to the entire probability mass for a given value of the auxiliary fields Q_1, Q_2. Maximizing b therefore amounts to the condition Eq. (2.28), illustrated in Fig. 2.2. We here used the convexity of the exponential function.

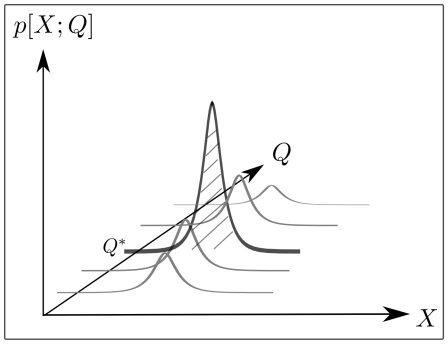


Figure 2.2: **Finding saddle point by maximizing contribution to probability:** The contribution to the overall probability mass depends on the value of the parameter Q, i.e. we seek to maximize $b[Q] := \int \mathcal{D}x\, p[\mathbf{x}; Q]$ (2.29). The point at which the maximum is attained is denoted as Q^*, the value $b[Q^*]$ is indicated by the hatched area.

The functional derivative in the stationarity condition (2.28) applied to $\ln Z[Q_1, Q_2]$ produces an expectation value with respect to the distribution (2.29): the fields Q_1 and Q_2 here act as sources. This yields the set of two equations

$$\begin{aligned} 0 &= -\frac{N}{\sigma^2} Q_1^*(t,t') + \frac{N}{Z} \left.\frac{\delta Z[Q_1,Q_2]}{\delta Q_2(s,t)}\right|_{Q^*} \\ \leftrightarrow \quad Q_1^*(s,t) &= \sigma^2 \langle \phi(x(s))\phi(x(t))\rangle_{Q^*} =: \sigma^2 C_{\phi(x)\phi(x)}(t,t') \\ 0 &= -\frac{N}{\sigma^2} Q_2^*(t,t') + \frac{N}{Z} \left.\frac{\delta Z[Q_1,Q_2]}{\delta Q_1(s,t)}\right|_{Q^*} \\ \leftrightarrow \quad Q_2^*(s,t) &= \frac{\sigma^2}{2}\langle \tilde{x}(s)\tilde{x}(t)\rangle_{Q^*} = 0, \end{aligned} \tag{2.30}$$

where we defined the average autocorrelation function $C_{\phi(x)\phi(x)}(t,t')$ of the non-linearly transformed activity of the units. The second saddle point $Q_2^* = 0$ vanishes, as it would otherwise alter the normalization of the generating functional through mixing of retarded and non-retarded time derivatives which then yield acausal response functions (Sompolinsky & Zippelius, 1982).

The expectation values $\langle\rangle_{Q^*}$ appearing in Eq. (2.30) must be computed self-consistently, since the values of the saddle points, by Eq. (2.27), influence the statistics of the fields $\mathbf{x}$ and $\tilde{\mathbf{x}}$, which in turn determines the functions Q_1^* and Q_2^* by (2.30).

Inserting the saddle point solution into the generating functional (2.27) we get

$$\langle Z\rangle \quad \propto \quad \int \mathcal{D}x \int \mathcal{D}x \exp\left(S_0[x,\tilde{x}] + \frac{\sigma^2}{2}\tilde{x}^{\mathrm{T}} C_{\phi(x)\phi(x)}\tilde{x}\right). \tag{2.31}$$

As the saddle points only couple to the sums of fields, the action has the important property that it decomposes into a sum of actions for individual, non-interacting

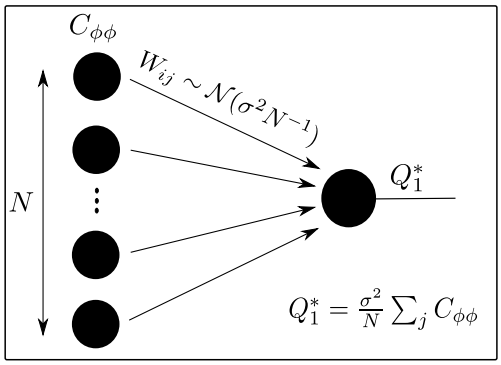


Figure 2.3: **Interpretation of the saddle point value Q_1^* given by Eq. (2.30):** The summed covariances $C_{\phi\phi}$ received by a neuron in the network, weighted by the synaptic couplings W_{ij}, which have Gaussian statistics with variance $\sigma^2 N^{-1}$.

units that feel a common field with self-consistently determined statistics, characterized by its second cumulant $C_{\phi(x)\phi(x)}$. Hence the saddle-point approximation reduces the network to N non-interacting units, or, equivalently, a single unit system. The second term in Eq. (2.31) is a Gaussian noise with a two point correlation function $C_{\phi(x)\phi(x)}(t,t')$. 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 $\sigma^2 N^{-1}$, which incorporates the Gaussian statistics of the couplings. This intuitive picture is shown in Fig. 2.3.

The interpretation of the noise can be appreciated by explicitly considering the moment generating functional of a Gaussian noise with a given autocorrelation function $C(t,t')$, which leads to the cumulant generating functional $\ln Z_\zeta[\tilde{x}]$ that appears in the exponent of Eq. (2.31) and has the form

$$\begin{aligned} \ln Z_\zeta[\tilde{x}] &= \ln\langle \exp\left(\int \tilde{x}(t)\, \zeta(t)\, dt \right) \rangle \\ &= \frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tilde{x}(t)\, C(t,t')\, \tilde{x}(t')\, dt\, dt' \\ &= \frac{1}{2} \tilde{x}^{\mathrm{T}}\, C\, \tilde{x}. \end{aligned}$$

Note that the effective noise term only has a non-vanishing second cumulant. This means the effective noise is Gaussian, as the cumulant generating function is quadratic. It couples pairs of time points that are correlated.

This is the starting point in Sompolinsky et al. (1988, eq. (3)), stating that the effective mean-field dynamics of the network is given by that of a single unit

$$(\partial_t + 1)\, x(t) = \eta(t) \tag{2.32}$$

driven by a Gaussian noise $\eta = \zeta + \frac{dW}{dt}$ with autocorrelation

$$\langle \eta(t)\eta(s)\rangle = \sigma^2 C_{\phi(x)\phi(x)}(t,s) + D\delta(t-s).$$

In the cited paper the white noise term $\propto D$ is absent, though. The generalization is presented in Goedeke et al. (2016). Using this effective mean-field dynamics the previously mentioned works derived self-consistent equations for the autocorrelation function of the neurons to determine a transition between regular and chaotic dynamics.

The above methods, however, can not explain cross-correlations between neurons in finite-size networks: The collective dynamics of the N interacting units are reduced to N pairwise independent units, each subject to a self-consistently determined auxiliary field in the thermodynamic limit $N \to \infty$. Covariances of individual neurons are self-averaging in this limit, i.e. identical over unit index and independent of the realization of the randomness in the network connectivity. In particular, cross-covariances vanish. Interactions between neurons can be included on average as a finite-size correction within this self-averaging framework by considering nonzero average connection weights (Ginzburg & Sompolinsky, 1994; Renart et al., 2010; Trousdale et al., 2012; Tetzlaff et al., 2012; Helias et al., 2014). However, additional finite-size effects arise from fluctuations in the saddle points that were neglected so far. In the next section, we will consider both features in a simpler network, the set of coupled Ornstein-Uhlenbeck processes as derived from the binary network in Chap. 1. This network is easier to treat analytically as it allows for a reduction from path integrals to ordinary integrals. Furthermore, it captures the linear-response dynamics of nonlinear network models.

2.3.3 Mean-field theory for finite-size networks

Let us now consider linear dynamics $\phi(x) = x$, i.e. a set of Ornstein-Uhlenbeck processes (cf. Eq. (1.31) in Sec. 1.6)

$$\frac{d\mathbf{x}(t)}{dt} = -\mathbf{x}(t) + \mathbf{W}\mathbf{x}(t) + \boldsymbol{\xi}(t). \tag{2.33}$$

with generating functional

$$\begin{aligned} Z[\mathbf{j}] &= \int \mathcal{D}\mathbf{x} \int \mathcal{D}\tilde{\mathbf{x}} \exp\left(S_0[\mathbf{x},\tilde{\mathbf{x}}] + \mathbf{j}^{\mathrm{T}}\mathbf{x}\right) \\ \text{with } S_0[\mathbf{x},\tilde{\mathbf{x}}] &= \tilde{\mathbf{x}}^{\mathrm{T}}\left(\partial_t + 1 - \mathbf{W}\right)\mathbf{x} + \frac{D}{2}\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}. \end{aligned} \tag{2.34}$$

The latter can easily be interpreted in Fourier domain due to the linearity of Eq. (2.33) and the invariance of scalar products under unitary transforms

$$Z[\mathbf{J}] = \int \mathcal{D}\mathbf{X} \int \mathcal{D}\tilde{\mathbf{X}} \exp\left(S_0[\mathbf{X},\tilde{\mathbf{X}}] + \mathbf{J}^{\mathrm{T}}\mathbf{X}\right)$$
$$\text{with } S_0[\mathbf{X},\tilde{\mathbf{X}}] = \tilde{\mathbf{X}}^{\mathrm{T}}\left(i\omega + 1 - \mathbf{W}\right)\mathbf{X} + \frac{D}{2}\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}, \tag{2.35}$$

with Fourier transformed variables denoted by capital letters. The scalar product in frequency domain reads $\tilde{\mathbf{X}}^{\mathrm{T}}\mathbf{X} = \sum_i \int \tilde{X}_i(-\omega)X_i(\omega)\,d\omega$. The generating functional factorizes into generating functions for each frequency ω. Since ω couples to $-\omega$ in the scalar product, it is easiest to focus on the component for $\omega = 0$. In the following, we will only discuss zero frequency and therefore omit the frequency argument, i.e. we write $\mathbf{X} \equiv \mathbf{X}(0)$ and correspondingly for sources $\mathbf{J}$. After integration over all non-zero frequencies one obtains the generating function for zero frequency

$$Z[\mathbf{J}] = \det(1-\mathbf{W}) \int \mathcal{D}\mathbf{X} \int \mathcal{D}\tilde{\mathbf{X}} \exp\left(S_0[\mathbf{X},\tilde{\mathbf{X}}] + \mathbf{J}^{\mathrm{T}}\mathbf{X}\right) \tag{2.36}$$
$$= \exp\left(\frac{1}{2}\mathbf{J}^{\mathrm{T}}\left(\mathbf{1}-\mathbf{W}\right)^{-1}\mathbf{D}\left(\mathbf{1}-\mathbf{W}^{\mathrm{T}}\right)^{-1}\mathbf{J}\right)$$
$$\text{with } S_0[\mathbf{X},\tilde{\mathbf{X}}] = \tilde{\mathbf{X}}^{\mathrm{T}}\left(\mathbf{1}-\mathbf{W}\right)\mathbf{X} + \frac{D}{2}\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}, \tag{2.37}$$

with the single-frequency scalar product defined as $\tilde{\mathbf{X}}^{\mathrm{T}}\mathbf{X} = \sum_i \tilde{X}_i X_i$, and integration measures $\int \mathcal{D}\mathbf{X} = \prod_j \int_{-\infty}^{\infty} dX_j$ and $\int \mathcal{D}\tilde{\mathbf{X}} = \prod_j \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} d\tilde{X}_j$. The determinant in Eq. (2.36) follows from the normalization condition $Z[\mathbf{J}=0] = 1$.

Similar to the general case discussed in Sec. 2.3.2, the generating function formalism allows an algorithmic integration of the statistics of $\mathbf{W}$. Ignoring insignificant variations in the normalization of $Z[\mathbf{J}]$ [1], the disorder average only affects the coupling term in Eq. (2.36)

$$\left\langle \exp\left(\tilde{\mathbf{X}}^{\mathrm{T}}\mathbf{W}\mathbf{X}\right)\right\rangle \quad = \quad \prod_{i,j}\phi(\tilde{X}_i X_j) = \prod_{i,j}\exp\left(\sum_{k=1}^{\infty}\frac{\kappa_k}{k!}(\tilde{X}_i X_j)^k\right).$$

For clarity, we here assume independent and identically distributed weights W_{ij}. In the resulting cumulant expansion (Sherrington & Kirkpatrick, 1975; De Dominicis & Peliti, 1978; Nishimori, 2001; Fischer & Hertz, 1991b), ϕ is the characteristic function for a single connection W_{ij} and κ_k its k-th cumulant (Gardiner, 1985). Independence of network size can only be expected, if the fluctuations of the input to a neuron are independent of N, requiring synaptic weights to scale with $1/\sqrt{N}$ (van Vreeswijk & Sompolinsky, 1996, 1998), such that the cumulant expansion is an expansion in $1/\sqrt{N}$. A truncation at the second cumulant ($\propto N^{-1}$) maps $\mathbf{W}$ to a Gaussian connec-

[1] Variability in the determinant can be accounted for perturbatively, but only yields subleading contributions which are suppressed for large network size.

tivity $\mathcal{N}(\mu, \sigma^2/N)$ so that

$$\begin{aligned}\langle Z[\mathbf{J}]\rangle &\sim \int \mathcal{D}\mathbf{X} \int \mathcal{D}\tilde{\mathbf{X}} \exp\left(S_0(\mathbf{X},\tilde{\mathbf{X}}) + S_{\text{int}}(\mathbf{X},\tilde{\mathbf{X}}) + \mathbf{J}^{\mathrm{T}}\mathbf{X}\right),\\ S_0(\mathbf{X},\tilde{\mathbf{X}}) &= \tilde{\mathbf{X}}^{\mathrm{T}}\left(-\mathbf{1}+\boldsymbol{\mu}\right)\mathbf{X} + \frac{D}{2}\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}},\\ S_{\text{int}}(\mathbf{X},\tilde{\mathbf{X}}) &= \frac{\sigma^2}{2N}\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\mathbf{X}^{\mathrm{T}}\mathbf{X},\end{aligned}$$

with a homogeneous mean connection weight $\mu_{ij} = \mu = \mathcal{O}(1/\sqrt{N})$. The second cumulant ($\sigma^2/N$) is the first non-trivial contribution to the second moment of covariances. While higher cumulants of the connectivity have an impact on higher moments of the distribution of covariances, their effect on the first two moments is suppressed by the large network size.

The interaction term S_{int} prevents an exact calculation of the disorder-averaged generating function. A converging perturbation series can be obtained in the auxiliary field formulation (Moshe & Zinn-Justin, 2003), where a field $Q_1 = \frac{\sigma^2}{N}\mathbf{X}^{\mathrm{T}}\mathbf{X}$ is introduced for the sum of a large number of statistically equivalent activity variables. Using the Hubbard-Stratonovich transformation

$$\exp\left(S_{\text{int}}(\mathbf{X},\tilde{\mathbf{X}})\right) \sim \int \mathcal{D}\mathbf{Q}\, \exp\left(-\frac{N}{\sigma^2}Q_1Q_2 + \frac{1}{2}Q_1\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}} + Q_2\mathbf{X}^{\mathrm{T}}\mathbf{X}\right),$$

one obtains a free theory, i.e. quadratic action in the activity ($\mathbf{X}$) and response variables ($\tilde{\mathbf{X}}$), on the background of fluctuating fields $\mathbf{Q}$

$$\begin{aligned}\langle Z[\mathbf{J}]\rangle &\sim \int \mathcal{D}\mathbf{Q}\, \exp\left(-\frac{N}{\sigma^2}Q_1Q_2 + \ln\left(Z_{\mathbf{Q}}[\mathbf{J}]\right)\right), \qquad (2.38)\\ Z_{\mathbf{Q}}[\mathbf{J}] &= \int \mathcal{D}\mathbf{X} \int \mathcal{D}\tilde{\mathbf{X}}\, \exp\left(S_0(\mathbf{X},\tilde{\mathbf{X}}) + \frac{1}{2}Q_1\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}} + Q_2\mathbf{X}^{\mathrm{T}}\mathbf{X} + \mathbf{J}^{\mathrm{T}}\mathbf{X}\right).\end{aligned}$$

The high dimensional integrals of the free theory $Z_{\mathbf{Q}}[\mathbf{J}]$ can be solved analytically yielding a two-dimensional interacting theory in the auxiliary fields Q_1 and Q_2. The auxiliary field formalism translates the high-dimensional ensemble average over $\mathbf{W}$ to a low-dimensional average over $\mathbf{Q}$, i.e. a mapping of the local disorder in the connections to fluctuations of global fields $\mathbf{Q}$ interacting with a highly symmetric all-to-all connected network, illustrated in Fig. 2.4. Only in the special case of vanishing mean connection strength $\mu = 0$, the system factorizes into N unconnected units, each interacting with the same set of fields $\mathbf{Q}$ (see Sec. 2.3.2). The all-to-all network not only captures the autocovariance of a single neuron, but also the cross-covariance with any other neuron.

Performing a change of variables $\mathbf{Q} = \mathbf{Q}^*[\mathbf{J}] + \frac{\delta\mathbf{Q}}{\sqrt{N}}$ in Eq. (2.38), where $\mathbf{Q}^*[\mathbf{J}]$ are the saddle points of the exponent $-\frac{N}{\sigma^2}Q_1Q_2 + \ln\left(Z_{\mathbf{Q}}[\mathbf{J}]\right) =: Y[Q_1, Q_2, \mathbf{J}]$ determined by

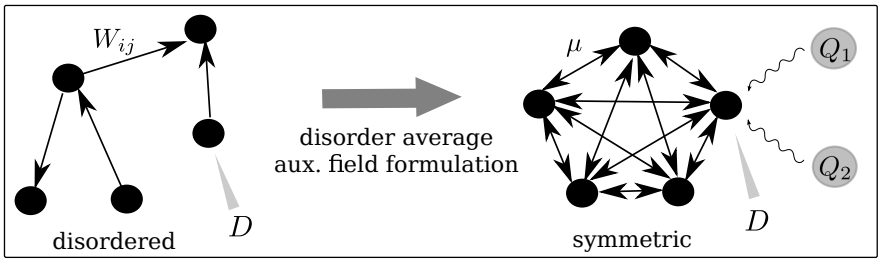


Figure 2.4: Disorder average maps network with frozen variability in connections to highly symmetric network on the background of fluctuating auxiliary fields $\mathbf{Q}$, globally contributing to the covariance that drives the network fluctuations (illustrated by wavy arrows).

saddle-point equations

$$0 \quad = \quad \frac{\partial}{\partial Q_\alpha} Y[Q_1, Q_2, \mathbf{J}]\Big|_{\mathbf{Q}=\mathbf{Q}^*[\mathbf{J}]} \qquad \alpha \in \{1,2\}, \tag{2.39}$$

and expanding $Y[Q_1, Q_2, \mathbf{J}]$ around $\mathbf{Q}^*[\mathbf{J}]$, yields a $1/N$ expansion (Moshe & Zinn-Justin, 2003; Amit, 2005). Note that in contrast to Sec. 2.3.2, we here preserve the dependence of $\mathbf{Q}^*[\mathbf{J}]$ on external sources $\mathbf{J}$. This dependence was neglected in prior work (Sompolinsky & Zippelius, 1982) since it scales to leading order as $1/N$ (see below). We focus only on the leading-order contribution $Y\left[Q_1^*[\mathbf{J}], Q_2^*[\mathbf{J}], \mathbf{J}\right]$ which describes tree-level diagrams in the $\mathbf{Q}$-theory and drop all source dependence in higher Taylor coefficients

$$\langle Z(\mathbf{J}) \rangle \sim \exp\left(Y\left[Q_1^*[\mathbf{J}], Q_2^*[\mathbf{J}], \mathbf{J}\right]\right), \tag{2.40}$$

with corresponding action

$$S(\mathbf{X}, \tilde{\mathbf{X}}) = \tilde{\mathbf{X}}^{\mathrm{T}} \left(-\mathbf{1} + \boldsymbol{\mu}\right) \mathbf{X} + \frac{D + Q_1^*[\mathbf{J}]}{2} \tilde{\mathbf{X}}^{\mathrm{T}} \tilde{\mathbf{X}} + Q_2^*[\mathbf{J}]\, \mathbf{X}^{\mathrm{T}} \mathbf{X}. \tag{2.41}$$

Note that this Gaussian theory in $\mathbf{X}, \tilde{\mathbf{X}}$ still contains in the latter two terms effects arising from the quartic interaction term S_{int} of the original theory. Variability in $\mathbf{Q}^*[\mathbf{J}]$ through their source dependence therefore gives rise to cumulants of the activity variables beyond second order.

The action (2.41) shows that $Q_1^*[\mathbf{J}]$ acts as a global contribution to the Gaussian noise whereas $Q_2^*[\mathbf{J}]$ directly contributes to the inverse of the covariance matrix. By integrating out the response variables $\tilde{\mathbf{X}}$, one can alternatively interpret both ${Q_\alpha}^*[\mathbf{J}]$ as contributions to the covariance matrix of the noise.

Saddle points $\mathbf{Q}^*[\mathbf{J}]$ are obtained self-consistently from Eq. (2.39) which reduces to

the set of equations

$$\begin{aligned} Q_1^*[\mathbf{J}] &= \frac{\sigma^2}{N}\frac{\int \mathcal{D}\mathbf{X}\int \mathcal{D}\tilde{\mathbf{X}}\,\mathbf{X}^{\mathrm{T}}\mathbf{X}\exp\left(S(\mathbf{X},\tilde{\mathbf{X}})+\mathbf{J}^{\mathrm{T}}\mathbf{X}\right)}{\int \mathcal{D}\mathbf{X}\int \mathcal{D}\tilde{\mathbf{X}}\exp\left(S(\mathbf{X},\tilde{\mathbf{X}})+\mathbf{J}^{\mathrm{T}}\mathbf{X}\right)}, \\ Q_2^*[\mathbf{J}] &= \frac{\sigma^2}{2N}\frac{\int \mathcal{D}\mathbf{X}\int \mathcal{D}\tilde{\mathbf{X}}\,\tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\exp\left(S(\mathbf{X},\tilde{\mathbf{X}})+\mathbf{J}^{\mathrm{T}}\mathbf{X}\right)}{\int \mathcal{D}\mathbf{X}\int \mathcal{D}\tilde{\mathbf{X}}\exp\left(S(\mathbf{X},\tilde{\mathbf{X}})+\mathbf{J}^{\mathrm{T}}\mathbf{X}\right)}. \end{aligned} \tag{2.42}$$

Moments of activity are determined by saddle points and their derivatives evaluated at zero source. Setting $\mathbf{J} = 0$ yields

$$\begin{aligned} Q_1^*[0] &= \frac{\sigma^2}{N}\left\langle \mathbf{X}^{\mathrm{T}}\mathbf{X}\right\rangle_\Omega, \\ Q_2^*[0] &= \frac{\sigma^2}{2N}\left\langle \tilde{\mathbf{X}}^{\mathrm{T}}\tilde{\mathbf{X}}\right\rangle_\Omega, \end{aligned} \tag{2.43}$$

with correlators $\left\langle X_i X_j\right\rangle_\Omega := \frac{d^2}{dJ_i dJ_j}\Omega[\mathbf{J}]\Big|_{\mathbf{J}=0}$ and the cumulant generating function $\Omega[\mathbf{J}] = \ln\left(Z_{\mathbf{Q}^*[0]}[\mathbf{J}]\right)$. Analogously, correlators including $\tilde{X}_i$ can be obtained by introducing a source term $\tilde{\mathbf{J}}^{\mathrm{T}}\tilde{\mathbf{X}}$ into $Z_{\mathbf{Q}^*[0]}$. A self-consistent result reads

$$\left\langle \tilde{X}_i\tilde{X}_j\right\rangle_\Omega = 0, \tag{2.44}$$

$$\left\langle X_i X_j\right\rangle_\Omega = \left[(\mathbf{1}-\boldsymbol{\mu})^{-1}\left(D+Q_1^*[0]\right)\left(\mathbf{1}-\boldsymbol{\mu}^{\mathrm{T}}\right)^{-1}\right]_{ij}, \tag{2.45}$$

$$\left\langle \tilde{X}_i X_j\right\rangle_\Omega = \left[(\mathbf{1}-\boldsymbol{\mu})^{-1}\right]_{ij}, \tag{2.46}$$

with

$$\begin{aligned} Q_1^*[0] &= \frac{R^2}{1-R^2}D, \\ Q_2^*[0] &= 0, \end{aligned}$$

$R = \sqrt{1+\gamma}\,\sigma$ and $\gamma = \mathcal{O}(1/N)$ resulting from subleading terms in Eq. (2.45) containing μ (see Sec. B.2). Eq. (2.45) shows that $Q_1^*[0]$ leads to a renormalization of the noise D. The renormalized noise reads $D_{\mathrm{r}} = D + Q_1^*[0] = D/(1-R^2)$.

Analogously, second derivatives of saddle points with respect to sources can be calculated from Eq. (2.42) with the final result

$$\begin{aligned} \left.\frac{d^2Q_1^*}{dJ_k dJ_l}\right|_{\mathbf{J}=0} &= \frac{\sigma^2}{\beta N}\sum_i \left\langle X_i X_k\right\rangle_\Omega\left\langle X_i X_l\right\rangle_\Omega + \frac{\sigma^2}{\beta N}\sum_{i,a}\left\langle X_i X_a\right\rangle_\Omega^2 \left.\frac{d^2}{dJ_k dJ_l}Q_2^*[\mathbf{J}]\right|_{\mathbf{J}=0}, \\ \left.\frac{d^2Q_2^*}{dJ_k dJ_l}\right|_{\mathbf{J}=0} &= \frac{\sigma^2}{\beta N}\sum_i \left\langle \tilde{X}_i X_k\right\rangle_\Omega\left\langle \tilde{X}_i X_l\right\rangle_\Omega, \qquad \text{with } \beta = 1-\frac{\sigma^2}{N}\sum_{i,a}\left\langle X_i\tilde{X}_a\right\rangle_\Omega^2, \end{aligned}$$

where we used that $\frac{dQ_\alpha^*}{dJ_k} = 0$ since three-point correlators of Ω vanish.

Using partial derivative calculus $\frac{d}{dJ_i}\langle Z(\mathbf{J})\rangle = \left(\frac{\partial Q_1^*}{\partial J_i}\frac{\partial}{\partial Q_1^*} + \frac{\partial Q_2^*}{\partial J_i}\frac{\partial}{\partial Q_2^*} + \frac{\partial}{\partial J_i}\right)\left\langle Z(Q_1^*, Q_2^*, \mathbf{J})\right\rangle$ and the saddle-point condition $\frac{\partial}{\partial Q_\alpha^*}\left\langle Z(Q_1^*, Q_2^*, \mathbf{J})\right\rangle = 0$ which follows from Eq. (2.39) and Eq. (2.40), we find that second moments of activity, i.e. average covariances, are not influenced by the source dependence of saddle points

$$\left\langle \left\langle X_i X_j \right\rangle_x \right\rangle = \frac{d^2}{dJ_i dJ_j}\langle Z(\mathbf{J})\rangle\Big|_{\mathbf{J}=0} = \left\langle X_i X_j \right\rangle_\Omega = \left[(\mathbf{1}-\boldsymbol{\mu})^{-1} D_\mathrm{r}\left(\mathbf{1}-\boldsymbol{\mu}^\mathrm{T}\right)^{-1}\right]_{ij}. \tag{2.47}$$

The non-zero mean connection strength $\mu = \mathcal{O}(1/\sqrt{N})$ yields cross-covariances as finite-size effects. Additional finite-size effects arise when considering fourth moments of activity. Formally, two of the four derivatives act on $\mathbf{Q}^*[\mathbf{J}]$ while the other two act on the source term $\mathbf{J}^\mathrm{T}\mathbf{X}$ to yield

$$\begin{aligned}
\left\langle \left\langle X_i X_j X_i X_j \right\rangle_x \right\rangle &= \frac{d^4}{dJ_i dJ_j dJ_i dJ_j}\langle Z(\mathbf{J})\rangle\Big|_{\mathbf{J}=0} \\
&= 4\,\frac{d^2 Q_1^*}{dJ_i dJ_j}\Big|_{\mathbf{J}=0}\sum_a \left\langle X_i \tilde{X}_a\right\rangle_\Omega \left\langle X_j \tilde{X}_a\right\rangle_\Omega + 2\,\frac{d^2 Q_1^*}{dJ_i dJ_i}\Big|_{\mathbf{J}=0}\sum_a \left\langle X_j \tilde{X}_a\right\rangle_\Omega \left\langle X_j \tilde{X}_a\right\rangle_\Omega \\
&+ 4\,\frac{d^2 Q_2^*}{dJ_i dJ_j}\Big|_{\mathbf{J}=0}\sum_a \left\langle X_i X_a\right\rangle_\Omega \left\langle X_j X_a\right\rangle_\Omega + 2\,\frac{d^2 Q_2^*}{dJ_i dJ_i}\Big|_{\mathbf{J}=0}\sum_a \left\langle X_j X_a\right\rangle_\Omega \left\langle X_j X_a\right\rangle_\Omega \\
&+ 2\left\langle X_i X_j\right\rangle_\Omega \left\langle X_i X_j\right\rangle_\Omega + \left\langle X_i X_i\right\rangle_\Omega \left\langle X_j X_j\right\rangle_\Omega .
\end{aligned} \tag{2.48}$$

The latter three terms in Eq. (2.48) correspond to the trivial Wick decomposition from the Gaussian part of the theory i.e. $\mathbf{Q}^*[\mathbf{J}=0]$, whereas the second derivatives of saddle points $\frac{d^2 Q_\alpha^*}{dJ_i dJ_j}\Big|_{\mathbf{J}=0} = \mathcal{O}(1/N)$ evaluated at zero source determine the non-vanishing fourth cumulants. Note that no quartic derivatives of saddle points appear in Eq. (2.48) due to the saddle-point condition $\frac{\partial}{\partial Q_\alpha^*}\left\langle Z(Q_1^*, Q_2^*, \mathbf{J})\right\rangle = 0$. We will show an application of this result in the next chapter where we relate the meta-statistics of activity, i.e. the dispersion of covariances, to the fourth cumulants calculated above.

2.4 Conclusion

We reviewed the functional formalism for classical stochastic systems by De Dominicis & Peliti (1978) and, by a combination of tools from spin glasses (Crisanti & Sompolinsky, 1987) and large-N field theory (Moshe & Zinn-Justin, 2003), derived and extended the mean-field theory for the dynamics of large, disordered networks underlying the seminal work of Sompolinsky et al. (1988). This mean-field theory reduces the disordered to highly symmetric networks by performing ensemble averages which allow descriptions of the dynamics in terms of the statistics of connections rather than single realizations. In the thermodynamic limit, $N \to \infty$, the ensemble-averaged system is decoupled. A saddle-point approximation of the fluctuating auxiliary fields yields effective equations of motion for isolated single units. Taking into

account finite-size effects from mean connection strengths and fluctuations of auxiliary fields yields a non-Gaussian theory, which exhibits non-trivial cross-correlations between units.

Chapter 3

Distributions of covariances as a window into the operational regime of neuronal networks

This chapter is based on the following publication:

Dahmen D, Diesmann M and Helias M (2016) *Distributions of covariances as a window into the operational regime of neuronal networks*, arXiv 1605.04153
Preliminary results have been published in Dahmen et al. (2015, 2016a).

Author contributions:
Under the supervision of M Helias, the author performed all parts of the above publication. All authors contributed to the design of the numerical experiments and the writing of the manuscript.

3.1 Introduction

A network of neurons constitutes a many-particle system with interactions mediated by directed (asymmetric) and random synaptic connections. This quenched randomness is the defining feature of a disordered system. The large number and divergence of outgoing connections implies that each pair of neurons receives a substantial amount of common inputs, leading to positively correlated activity on average (De la Rocha et al., 2007; Shea-Brown et al., 2008). The dominant negative feedback, which stabilizes the strongly fluctuating activity in recurrent networks (van Vreeswijk & Sompolinsky, 1996), gives rise to active decorrelation (Tetzlaff et al., 2012). As a consequence, correlations are positive on average, but close to zero (Ecker et al., 2010; Cohen & Kohn, 2011) and their mean is predicted to vanish in inverse proportion to the number of neurons in the network (Renart et al., 2010). On the level of individual pairs of neurons, however, experimentally a wide distribution of correlations is observed, shown in Fig. 3.1 for recordings in macaque motor cortex (see also Fig. 4). The mechanism responsible for the large width is beyond available theories, which are restricted to population averages.

Functionally, correlations are important, because they influence the information contained in the activity of neural populations (Shadlen & Newsome, 1998; Sompolinsky et al., 2001; Moreno-Bote et al., 2014). The recent availability of massively parallel recordings of neural activity (Stevenson & Kording, 2011) poses the question whether the joint statistics allows conclusions on the structure and the operational regime of the network. For example, a fundamental transition from regular to chaotic dynamics is known to occur in large networks at a precisely defined interaction strength, the point at which the regular state looses linear stability (Sompolinsky et al., 1988). Highest computational performance (Crutchfield & Young, 1989) is expected at the edge of this chaotic state (Toyoizumi & Abbott, 2011).

Whether and how this critical point is reflected in pairwise correlations is yet unknown, since the employed mean-field theories (Sompolinsky et al., 1988; Molgedey et al., 1992; van Vreeswijk & Sompolinsky, 1996; Aljadeff et al., 2015; Kadmon & Sompolinsky, 2015) are restricted to self-averaging quantities in the $N \to \infty$. In this limit, cross-covariances vanish, such that the standard mean-field theory (Sommers et al., 1988; Le Cun et al., 1991) as described in Sec. 2.3.2 cannot explain their dispersion. Connections in neuronal networks have, however, limited range, so that the effective network size is bounded well below the size of the entire brain. Understanding correlations therefore requires us to preserve finite-size effects.

The mean-field theory derived in Sec. 2.3.3 captures finite-size fluctuations of external auxiliary fields which derive from the quenched disorder of the connections and explain the neuron-to-neuron variability. Employing this formalism, we obtain closed-form expressions for the mean and width of the distribution of covariances and we explain why their ratio is $\propto \sqrt{N}$. The dependence on the structural parameters further allows us to infer network parameters from the observed activity. The

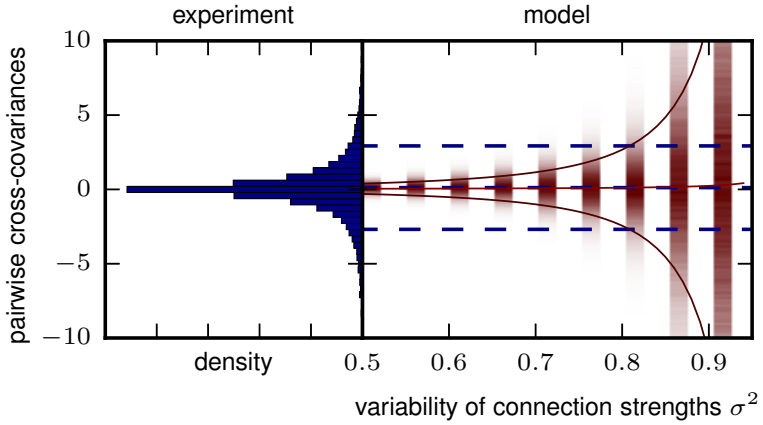


Figure 3.1: Distribution of cross-covariances $c_{ij} = \frac{1}{T}\left(\langle n_i n_j\rangle_t - \langle n_i\rangle_t\langle n_j\rangle_t\right)$ between spike counts n_i of neurons in macaque motor cortex (blue histogram, $\langle\rangle_t$ is the average across trials) as compared to integral cross-covariances given by Eq. (3.1) with different variability σ^2 of connection strengths $W_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2/N)$ (red shading indicates density of histogram, red curves the analytical prediction for mean and ±1 standard deviation). The low mean and large standard deviation (blue dashed horizontal lines) of experimentally observed cross-covariances (blue) are explained by a model network (red) with high variability of connections ($\sigma^2 \approx 0.8$). The experimental data from 155 neurons mostly located in layer 5 of macaque motor cortex (M1) are recorded with a 100-electrode "Utah" array (Blackrock Microsystems, Salt Lake City, UT, USA) with 400 μm interelectrode distance, covering an area of 4×4 mm^2. Spike counts n_i of activity are obtained within $T = 400$ ms after trial start (TS) of a reach-to-grasp task (Riehle et al., 2013) in 141 trials (subsession: i140703-001, same data as in Fig. 4). The data set is also available with all details on the recording and annotations in Brochier et al. (to be submitted to Scientific Data). Note that the standard deviation of cross-covariances was corrected for bias due to limited number of trials (see Sec. B.1). Numerical solution of Eq. (3.1) and analytical predictions for the mean (3.6) and standard deviation (3.7) are performed with network size $N = 1000$, uniform covariance $D = 3.36$, Gaussian connectivity with mean connection strength $\mu = 6.6 \cdot 10^{-4}$, and varying σ^2 (right horizontal axis). Data courtesy of A. Riehle and T. Brochier.

equations establish a link between the distance to criticality, i.e. the spectral radius of the connectivity matrix, and the width of the distribution of covariances. The experimental data introduced in Fig. 3.1 strongly supports the operation of the brain close to this critical point.

3.2 Meta-statistics of activity and criticality

In the asynchronous irregular regime (Brunel, 2000) resembling cortical activity in the absence of external stimuli, linear response theory applied to a variety of neuron models (Grytskyy et al., 2013b), such as binary (Ginzburg & Sompolinsky, 1994; Dahmen et al., 2016a), leaky integrate-and-fire (Pernice et al., 2012; Trousdale et al., 2012), and Poisson model neurons (Hawkes, 1971; Pernice et al., 2011) yields equations for the covariance matrix between network fluctuations $\mathbf{x}(t)$ around the stationary state (see Fig. 4). Considering the time-lag integral of the covariance functions, these different expressions reduce to a unique form

$$\begin{aligned} c_{ij} &= \int_{-\infty}^{\infty} \left\langle x_i(t+t')x_j(t)\right\rangle_x dt' \\ &= \left[(\mathbf{1}-\mathbf{W})^{-1}\,\mathbf{D}\left(\mathbf{1}-\mathbf{W}^{\mathrm{T}}\right)^{-1}\right]_{ij}, \end{aligned} \tag{3.1}$$

which, in particular, is independent of the connection delays and time constants of the system. Here $\mathbf{W}$ is the connectivity matrix and $\mathbf{D}$ is a matrix of covariances related to single unit fluctuations and correlated external input (see Sec. B.3,Sec. B.4). In the absence of the latter and for spiking models, the entries D_{ii} are the stationary variances of spike counts in the absence of coupling determined by the mean activities (Pernice et al., 2012; Helias et al., 2014), and c_{ij} is the covariance between spike counts n_i and n_j (Fig. 3.1).

Distributions of the covariances c_{ij} (3.1) over different pairs of neurons arise from distributed mean activities and variability of shared and correlated external input across neurons (entering $\mathbf{D}$), and from the structural variability in the disordered connectivity $\mathbf{W}$. We here focus on the latter variability which is generated within the network. The found results are robust with respect to variability in $\mathbf{D}$ (see Sec. B.3,Sec. B.4).

In nature, the disordered connectivity between neurons is cell-type specific and distance dependent. However, already an isolated homogeneous random network, i.e. a network with independent and identically distributed weights and uniform covariance $\mathbf{D} = D \cdot \mathbf{1}$, exhibits widely distributed cross-covariances with a small mean (Fig. 3.1). To expose the fundamental mechanism, we here neglect the distance dependence and cell-type specificity and focus on the effect of the disordered connectivity alone.

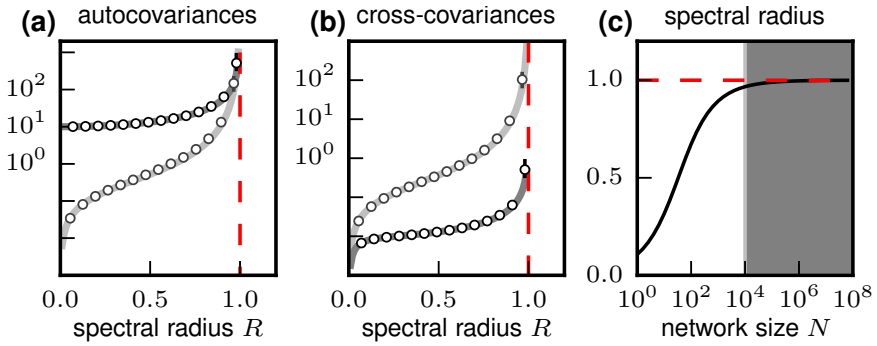


Figure 3.2: Mean (absolute value; dark gray) and standard deviation (light gray) of integral auto- (a) and cross-covariances (b) for different spectral radii R. Solid curves indicate analytical predictions, symbols show numerical results of Eq. (3.1) for 100 network realizations with Erdős-Rényi connectivity and $N = 1000$. Circles denote the ensemble averages, bars the average upward/downward deviation of the non-Gaussian statistics. Bars are below marker size except for rightmost data points. Spectral radius numerically estimated by the maximal real part of the eigenvalues. (c) Predicted spectral radius of the effective connectivity of the macaque motor cortex network as a function of network size for given moments of experimentally observed and bias-corrected parameters of the covariance distribution (Fig. 3.1, $\overline{c_{ij}} = 0.12$, $\overline{\delta c_{ij}^2} = 6.11$, $\overline{c_{ii}} = 16.16$, see Sec. B.1). The shaded area marks the range of biologically plausible effective network sizes corresponding to the spatial scale of the recordings. The red dashed line indicates the critical point $R = 1$.

Determining the moments of covariances from Eq. (3.1) is hindered by the occurrence of the inverse of the random connectivity matrix $(\mathbf{1} - \mathbf{W})^{-1}$. A systematic calculation of moments, which does not rely on ad-hoc approximations can, however, be performed in the equivalent generating function formulation by noting that the correlation structure Eq. (3.1) can be regarded as originating from the time evolution of a set of coupled Ornstein-Uhlenbeck processes (Risken, 1996).

We seek to determine the dispersion of covariances and therefore must not expect individual covariances to be self-averaging. We may, however, expect their meta-statistics, the moments of the distribution of covariances, to be self-averaging (Fischer & Hertz, 1991b). We denote with $\bar{}$ the empirical average, with $\langle\rangle_x$ the expectation over realizations of the processes $\boldsymbol{x}$, and with $\langle\rangle$ the ensemble average over the disordered connectivity $\mathbf{W}$. Exchanging the order of differentiation and averaging allows expressing second moments of covariances as derivatives of a single disorder-averaged generating function: We first assume the empirical average to be self-averaging

$$\overline{c_{ii}^2} = \left\langle \overline{c_{ii}^2} \right\rangle, \tag{3.2}$$

use its definition $\overline{c_{ii}^2} = \frac{1}{N}\sum_{i=1}^{N} c_{ii}^2$ and the definition of $c_{ij} = \int_{-\infty}^{\infty} \left\langle x_i(t+t')x_j(t)\right\rangle_x dt' = \left\langle X_i X_j \right\rangle_x$ as the second cumulant of the zero frequency components $\mathbf{X} = \mathbf{X}(\omega = 0) =$

$\int_{-\infty}^{\infty} \mathbf{x}(t)\,dt$ of Ornstein-Uhlenbeck processes $\mathbf{x}$ (Wiener-Khinchin theorem, see Gardiner, 1985) to obtain

$$\overline{c_{ii}^2} = \frac{1}{N}\sum_{i=1}^{N}\left\langle \langle X_i X_i\rangle_x^2 \right\rangle . \tag{3.3}$$

As the action Eq. (2.36) for a single realization of $\mathbf{W}$ is quadratic, Wick's theorem applies, $\left\langle X_i X_i X_j X_j\right\rangle_x = 2\left\langle X_i X_j\right\rangle_x^2 + \langle X_i X_i\rangle_x \left\langle X_j X_j\right\rangle_x = 2c_{ij}^2 + c_{ii}c_{jj}$ for the special case $i = j$, such that we identify the squared second moment as a fourth moment of the generating functional

$$\begin{aligned}\overline{c_{ii}^2} &= \frac{1}{N}\sum_{i=1}^{N}\left\langle \langle X_i X_i\rangle_x^2 \right\rangle \overset{\text{Wick's th.}}{=} \frac{1}{N}\sum_{i=1}^{N}\frac{1}{3}\left\langle \langle X_i X_i X_i X_i\rangle_x \right\rangle \\ &= \frac{1}{N}\sum_{i=1}^{N}\frac{1}{3}\left\langle \frac{d^4}{dJ_i^4} Z[\mathbf{J}]\right\rangle\Big|_{\mathbf{J}=0} = \frac{1}{3}\,\frac{d^4}{dJ_i^4}\langle Z[\mathbf{J}]\rangle\Big|_{\mathbf{J}=0} .\end{aligned} \tag{3.4}$$

In the last step, we used the symmetry of the disorder-averaged network. Analogously follows for $i \neq j$ and $N-1 \approx N$,

$$\overline{c_{ij}^2} = \left\langle \overline{c_{ij}^2}\right\rangle = \frac{1}{2N(N-1)}\sum_{i\neq j=1}^{N}\frac{d^4}{dJ_i^2 dJ_j^2}\langle Z[\mathbf{J}]\rangle\Big|_{\mathbf{J}=0} - \frac{1}{2}\langle \overline{c_{ii}}\rangle^2 = \frac{1}{2}\,\frac{d^4}{dJ_i^2 dJ_j^2}\langle Z[\mathbf{J}]\rangle\Big|_{\mathbf{J}=0} - \frac{1}{2}\langle \overline{c_{ii}}\rangle^2 . \tag{3.5}$$

For large networks, the required two- and four-point correlators can be calculated from the zeroth order term in the saddle-point expansion of the fields $\mathbf{Q}$. With Eq. (2.47) and Eq. (2.48) in Sec. 2.3.3 we obtain the mean integral covariances

$$\overline{c_{ij}} = \left[(\mathbf{1}-\boldsymbol{\mu})^{-1}\,\frac{D}{1-R^2}\left(\mathbf{1}-\boldsymbol{\mu}^{\mathrm{T}}\right)^{-1}\right]_{ij} = D_{\mathrm{r}}\gamma_{ij} \tag{3.6}$$

and the variance of integral covariances

$$\overline{\delta c_{ij}^2} = R^2\left[\frac{1}{(1-R^2)^2} + \frac{1}{1-R^2}\right] D_{\mathrm{r}}^2\chi_{ij}, \tag{3.7}$$

with the spectral radius $R = \sqrt{1+\gamma}\,\sigma$ of the connectivity matrix $\mathbf{W}$ (Rajan & Abbott, 2006) with $W_{ij} \sim \mathcal{N}(\mu, \frac{\sigma^2}{N})$, and the covariance $D_{\mathrm{r}} = D + Q_1^*[0] = D/(1-R^2)$, which is renormalized by the structural variability contained in Q_1^* (see Fig. 2.4). The mean connectivity μ enters the coefficients $\gamma_{ij} = \delta_{ij} + \gamma$ (with δ_{ij} the Kronecker symbol), $\chi_{ij} = \frac{1}{N}\left(1 + \delta_{ij} + \mathcal{O}(1/N)\right)$ and $\gamma = \mathcal{O}(1/N)$ (see Sec. B.2), and acts as a negative feedback in inhibitory or inhibition-dominated networks (Tetzlaff et al., 2012). While this feedback suppresses mean cross-covariances (3.6), it only yields a subleading contribution to the dispersion (3.7). The spread of individual cross-covariances is determined by fluctuations in connection weights. These fluctuations, which are formally reflected in fluctuations of auxiliary fields $\mathbf{Q}$, cause broad distributions of

cross-covariances of both signs even in a homogeneous network [1].

The theoretical predictions (Eqs. (3.6) and (3.7)) correctly reproduce numerical results for networks (Fig. 3.2a,b) that differ in higher-order cumulants neglected by the theory. The small spread in numerical results across realizations of the random connections furthermore justifies the self-averaging assumption of the meta-statistics in terms of mean and variance of the covariance distributions.

At a critical point $R = 1$ (Fig. 3.2), where the linearized system becomes unstable due to the largest eigenvalue of the connectivity matrix $\mathbf{W}$ exceeding unity, we observe a divergence of auto- and cross-covariances. In the slightly sub-critical regime, slowly decaying fluctuations generate individual covariances in the network much larger than the average across neurons. The same correlation structure exists in a nonlinear network (Sompolinsky et al., 1988) with infinitesimal additive noise, leading to fluctuating activity also in the regular regime. In this model, the point of linear instability coincides with a transition to chaos and a divergence in topological complexity (Wainrib & Touboul, 2013).

The analytic expressions for the moments of covariances Eqs. (3.6) and (3.7) can be used to infer network parameters from experimentally observed covariance distributions. For the highly symmetric network considered here, the application of Wick's theorem yields at leading order a trivial factor two between the variance of integral auto- and cross-covariances (see definition χ_{ij} above) and thus requires one free parameter in the inversion of Eqs. (3.6) and (3.7), here chosen as the network size N since its order of magnitude is most easily accessible in experiments. For given network size N, the spectral radius of the effective connectivity $\mathbf{W}$ is predominantly determined by $\Delta := \overline{\delta c_{ij}^2}/\overline{c_{ii}}^2$, which is the width of the distribution of cross-variances normalized by the mean autocovariances squared,

$$R^2 \quad = \quad 1 - \sqrt{\frac{1}{1 + N\Delta}}, \tag{3.8}$$

with $\mathcal{O}(1/N)$ corrections due to subleading terms in κ and χ_{ij}. Distributions of covariances can therefore be used to infer the operational regime of the network, i.e. the distance to criticality. Fig. 3.2c shows that for the distribution of covariances measured in macaque motor cortex (Fig. 3.1) and biologically plausible network sizes, the linear network model must operate close to criticality to explain the data. The small change of the spectral radius in the biologically relevant range further illustrates the robustness of the result with respect to a potential bias of the experimental estimates

[1] A decomposition of fluctuations into the population-averaged fluctuation and orthogonal modes shows that the population-averaged fluctuation, which exclusively determines average covariances, is suppressed by the average inhibitory connection weight $\mu < 0$, while the second moment of covariances results from the fluctuations of all remaining modes, which are not suppressed by negative feedback (see Sec. B.2).

due to limited number of recorded neurons or minor deviations of Δ, arising from distributed mean activities, external inputs or correlated connections (see Sec. B.3-Sec. B.5).

Mean integral autocovariances (3.6) and the variance of integral cross-covariances (3.7) are predominantly determined by the average stationary mean activities and the spectral radius. They are rather insensitive to specific deterministic features of the connectivity and distributions of noise amplitudes across neurons, which arise from variability in stationary mean activities or external input. This suggests that Eq. (3.8) also holds qualitatively for more complicated network topologies.

One can use Eqs. (3.6) and (3.7) to uniquely determine the parameters of a homogeneous network of arbitrary size that generates distributions of covariances matching the first two moments of experimental data (Fig. 3.1). An exception is the variance of integral autocovariances that is sensitive to the inter-neuron variability of the noise level (see Sec. B.3). A better agreement between the higher-order moments of the experimental distributions for auto- and cross-covariances and the model results requires more realistic network topologies, such as excitatory and inhibitory populations with heterogeneous mean activities driven by external input and spatially dependent connectivity.

3.3 Conclusion

Massively parallel recordings of spiking activity in macaque motor cortex (M1) show that spike count covariances vary widely across pairs of neurons. Their low average is well understood, but an explanation for the wide distribution in relation to the quenched disorder of the connectivity in recurrent random networks was so far elusive. We interpret the correlation structure of linear response modes of network activity as arising from a set of Ornstein-Uhlenbeck processes and applied the mean-field theory beyond self-averaging developed in Sec. 2.3.3. Using Wick's theorem, we relate fourth-order cumulants of activity arising from fluctuations of auxiliary fields to the dispersion of covariances. The exposed analytical relation between the statistics of connections and the statistics of pairwise covariances shows that both, average and dispersion of the latter, diverge at a critical coupling. At this point, a network of nonlinear units transits from regular to chaotic dynamics. Applying these results to the recordings from macaque motor cortex suggests its operation close to this edge of criticality.

Part III

Simulation of stochastic rate-based neuron models

Chapter 4

Integration of continuous-time dynamics in a spiking network simulator

This chapter is based on the following publications:

Hahne J*, Dahmen D*, Schuecker J*, Frommer A, Bolten M, Helias M and Diesmann M (2016) *Integration of continuous-time dynamics in a spiking neural network simulator*, arXiv 1610.09990
Hahne J*, Dahmen D*, Schuecker J*, Frommer A, Bolten M, Helias M and Diesmann M (2017) *Integration of continuous-time dynamics in a spiking neural network simulator*, Front. Neuroinform. 11, 34.
*These authors contributed equally to this publication.

Author contributions:
Under the supervision of M Helias and M Diesmann, the author jointly worked with J Hahne and J Schuecker on all parts of the above publication. Together with J Schuecker, the author thereby 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.

4.1 Introduction

Over the past decades, multiple strategies of neural network modeling have emerged in computational neuroscience. Functionally inspired top-down approaches that aim to understand the computation in neural networks 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 (1972b) 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), perceptual rivalry (Laing & Chow, 2002; Wilson, 2003), geometric visual hallucination patterns (Ermentrout & Cowan, 1979; Bressloff et al., 2001), ocular dominance and orientation selectivity (Ben-Yishai et al., 1995; Bressloff & Cowan, 2003; Stevens et al., 2013), spatial navigation (Samsonovich & McNaughton, 1997), and movement preparation (Erlhagen & Schöner, 2002). 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). Ideas from functional network models have further inspired the field of artificial neuronal networks in the domain of engineering (Haykin, 2009).

In contrast, bottom-up approaches explicitly model individual neurons, employing biophysically grounded spiking neuron models that simulate the time points of action potentials. These 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, 1997; Shadlen & Newsome, 1998), membrane-potential fluctuations (Destexhe & Paré, 1999), asynchronous firing (Brunel, 2000; Ecker et al., 2010; Renart et al., 2010; Ostojic, 2014), correlations in neural activity (Gentet et al., 2010; Okun & Lampl, 2008; Helias et al., 2013), self-sustained activity (Ohbayashi et al., 2003; Kriener et al., 2014), rate distributions across neurons (Griffith & Horn, 1966; Koch & Fuster, 1989; Roxin et al., 2011) and across laminar populations (Potjans & Diesmann, 2014), as well as resting state activity (Deco & Jirsa, 2012).

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. The situation underlines that bottom-up and top-down 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 recent advances in simulation (Djurfeldt et al., 2008; 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) enable full-density bottom-up models of complete circuits (Potjans & Diesmann, 2014; Markram et al., 2015). In particular, it has become feasible to build spiking models (Schmidt et al., 2016b) that describe the same macroscopic system as the rate-based descriptions (Chaudhuri et al., 2015).

Mean-field theories pave the way to relate spiking and rate-based descriptions of neuronal dynamics (Deco et al., 2008). Population density methods using Fokker-Planck theory can be used to determine stationary state activities of spiking networks (Siegert, 1951; Brunel, 2000). The dynamics of rate fluctuations around this background activity can be obtained using linear response theory (Brunel & Hakim, 1999; Lindner et al., 2005; Ostojic & Brunel, 2011; Trousdale et al., 2012; Grytskyy et al., 2013b; Schuecker et al., 2015), moment expansions for mode decompositions of the Fokker-Planck operator (Mattia & Del Guidice, 2002, 2004; Deco et al., 2008), or other specific ansatzes (Montbrió et al., 2015). An alternative derivation of rate-based dynamics aims at a closure of equations for synaptic currents of spiking networks in a coarse-graining limit by replacing spiking input with the instantaneous firing rate (Bressloff, 2012). Using field-theoretical methods (Buice & Chow, 2013) that were originally developed for Markovian network dynamics (Buice & Cowan, 2007; Buice et al., 2010) allows a generalization of this approach to fluctuations in the input (Bressloff, 2015). In any case, 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. A unified code base that features both models would highly simplify these benchmarks rendering duplication of code obsolete.

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 benchmarks, 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, can not 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.

This chapter 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). The technology described here 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 the long-term collaborative project to provide the technology for neural systems simulations (Gewaltig & Diesmann, 2007).

In the following, we briefly review numerical solution schemes for ordinary and stochastic (delay) differential equations in Sec. 4.2 and their application to neural networks in Sec. 4.2.2. 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. 4.2.3. In Sec. 4.3, 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., 2013b), 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., 2016), and briefly mention straight-forward generalizations.

4.2 Material and 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. 4.2.3) 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

$$X(t) = X_0 + \int_{t_0}^{t} a(s, X(s))\, ds + \int_{t_0}^{t} b(s, X(s))\, dW(s)\,, \tag{4.1}$$

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. (4.1) reads

$$dX(t) = a(t, X(t))\, dt + b(t, X(t))\, dW(t) \tag{4.2}$$

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

$$\frac{dX(t)}{dt} = a(t, X(t)) + b(t, X(t))\, \xi(t)\,, \tag{4.3}$$

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. (4.3) 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. (4.1) as $dW(t) \equiv \xi(t)dt$.

4.2.1 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 Δt 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

$$dX(t) = a(t, X(t))\,dt + b(t, X(t))\,dW(t) \tag{4.4}$$

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. (4.1) with their left-sided values. The update formula reads

$$X_{k+1} = X_k + a(t_k, X_k) \cdot \Delta t + b(t_k, X_k) \cdot \Delta W_k \tag{4.5}$$

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

$$X_{k+1} = X_k + a(t_{k+1}, X_{k+1}) \cdot \Delta t + b(t_k, X_k) \cdot \Delta W_k \tag{4.6}$$

with $\Delta W_k \sim \mathcal{N}(0, \Delta t)$ for $k = 0, \ldots, n-1$. 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. (4.4) 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

$$\int_{t_0}^{t} e^{A(t-s)} b(s, X(s))\, dW(s) \approx \int_{t_0}^{t} e^{A(t-s)} b(t_0, X(t_0))\, dW(s)\,. \tag{4.7}$$

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. (4.7), amounts to (Adamu, 2011)

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

The corresponding scheme reads

$$X_{k+1} = e^{A\Delta t} X_k + A^{-1}(e^{A\Delta t} - I) \cdot f(t_k, X(t_k)) + \sqrt{A^{-1}\left(\frac{e^{2A\Delta t} - I}{2}\right)} \cdot b(t_k, X_k) \cdot \eta_k \quad (4.8)$$

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. (4.8) 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).

4.2.2 Network of rate models

We here consider networks of N rate-based model neurons where each neuron receives recurrent input from the network. The system fulfills the Itô-SDEs

$$\tau^i dX^i(t) = \left[-X^i(t) + \mu^i + \phi\left(\sum_{j=1}^{N} w^{ij}\psi\left(X^j(t - d^{ij})\right)\right)\right] dt + \sqrt{\tau^i}\sigma^i\, dW^i(t) \qquad i = 1, \ldots, N \quad (4.9)$$

with possibly nonlinear input-functions $\phi(x)$ and $\psi(x)$, connection weights w^{ij}, mean input μ^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 σ, τ, μ, and d. The considered class of rate models only contains additive noise. Therefore, as noted above, the system Eq. (4.9) 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. (4.9) with $d = 0$. The Euler-Maruyama update step reads

$$X^i_{k+1} = X^i_k + \left[-X^i_k + \mu + \phi\left(\sum_{j=1}^{N} w^{ij}\psi\left(X^j_k\right)\right)\right] \frac{1}{\tau}\Delta t + \frac{1}{\sqrt{\tau}}\sigma\Delta W^i_k\,. \quad (4.10)$$

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

$$X^i_{k+1} = e^{-\Delta t/\tau} X^i_k + \left(1 - e^{-\Delta t/\tau}\right)\left[\mu + \phi\left(\sum_{j=1}^{N} w^{ij}\psi\left(X^j_k\right)\right)\right] + \sqrt{\frac{1}{2}(1 - e^{-2\Delta t/\tau})}\,\sigma\eta^i_k \quad (4.11)$$

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

$$\tau dX(t) = \left[A \cdot X(t) + \mu\right] dt + \sqrt{\tau}\sigma dW(t) \qquad i = 1, \ldots, N \tag{4.12}$$

with $A = -I + W$ and $W = (w^{ij})_{N\times N}$. Here the stochastic exponential Euler scheme Eq. (4.8) yields the exact solution of the system.

The numerical schemes presented in Sec. 4.2.1 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 Δt. 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.

4.2.3 Implementation in spiking network simulation code

This section describes the embedding of rate-based models (Sec. 4.2.2) in a simulation code for spiking neuronal networks. Examples how to create, connect and record activity from rate models will be made available with the release of our reference implementation of the continuous-time dynamics in the simulation code NEST.

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 `SecondaryEvents`, 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.

4.2.3.1 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. (4.6) 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. 4.3.1 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 not able to add the influence from all other rate neurons to the matrix A in Eq. (4.12), 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. (4.11). This also has the benefit of avoiding the need to numerically evaluate a general matrix exponential as A is a diagonal matrix (see Sec. 4.2.2 for details).

4.2.3.2 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 analysis of the numerical schemes in Sec. 4.3 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 (4.9). In this implementation neurons can handle both events, i.e. `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 (4.9). 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 (4.9) 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. 4.1 gives an overview of 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`.

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

Table 4.1: **Template-derived rate-based neuron models.** The table shows the 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`.

4.2.3.3 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. 4.2). 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. 4.1B 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 communication interval and converges to the solution obtained with the standard approach (Fig. 4.1A). The iteration terminates when a user chosen convergence tolerance `wfr_tol` (see Tab. 4.2) 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

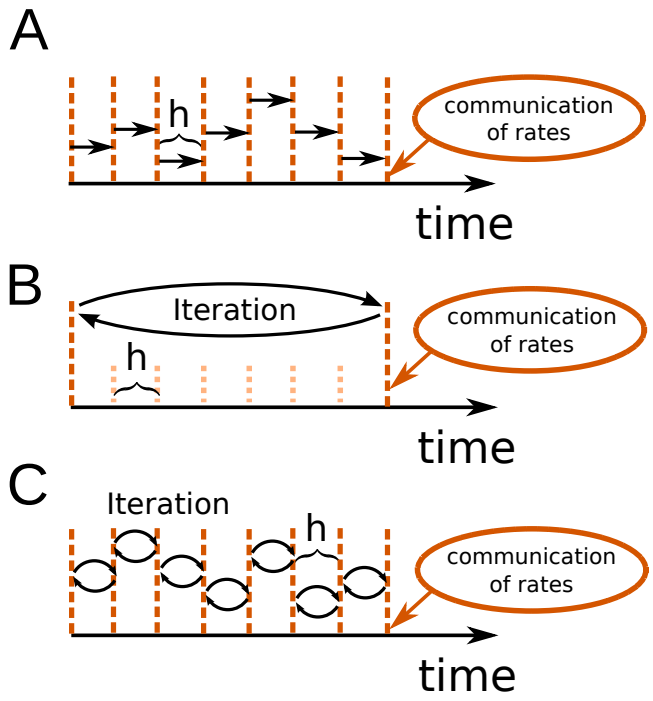


Figure 4.1: **Different communication strategies for distributed simulations.** Distance between neighboring dotted orange lines indicates computation time step of size h. Distance between neighboring dashed red lines symbolizes one communication interval where rates (and other events like spike events) 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.

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. 4.3.2 we investigate the performance of the iterative (Fig. 4.1B) and the standard approach (Fig. 4.1A) 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. 4.1C 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. 4.3.1 demonstrates that this is not an efficient option for the integration of rate-based models in distributed simulations.

Tab. 4.2 summarizes the parameters of our reference implementation of the waveform-relaxation technique. A subset was previously introduced by Hahne et al. (2015) (`wfr_interpolation_order`, `wfr_max_iterations`, `wfr_tol`), 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.

4.3 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.

4.3.1 Stability and accuracy of integration methods

To investigate the accuracy and stability of the different numerical methods (see Sec. 4.2.1) we consider an exactly solvable network of linear rate neurons with $\mu = 0$ (see also Sec. 4.2.2)

$$\tau dX(t) = A \cdot X(t)\, dt + \sqrt{\tau}\sigma\, dW(t)\,. \tag{4.13}$$

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. (4.8)). 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 4.2.3.1). 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.

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* ($d_{\min}$,`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.

Table 4.2: **Parameters of the waveform relaxation algorithm.** The table shows the different parameters of the waveform relaxation algorithm together with their C++ data-type, default value, and a brief description.

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

$$\tau dZ^i(t) = \lambda_i \cdot Z^i(t)\,dt + \sum_{j=1}^{N} \sqrt{\tau}\sigma t^{ij}\,dW^j(t) \qquad i = 1, \ldots, n \tag{4.14}$$

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. (4.13) 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 Δt 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.

We investigate the accuracy of the numerical methods for an all-to-all connected network with inhibitory connections of weight $w^{ij} = \frac{-1}{\sqrt{N}}$. The eigenvalues of $A = -I + \frac{-1}{\sqrt{N}} \cdot J$ with $J^{ij} = 1$ are $\lambda_1 = -1 - \sqrt{N}$ 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. 4.2 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 $f(X) = \frac{-1}{\sqrt{N}} \cdot J \cdot X$, 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

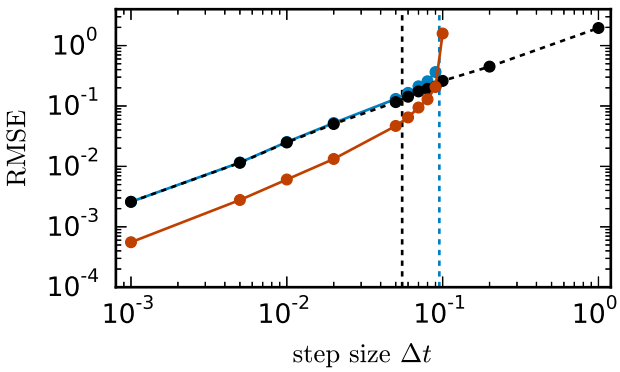


Figure 4.2: **Comparison of numerical methods for an all-to-all connected inhibitory network.** RMSE $= \sqrt{\frac{1}{N(t_n - t_0)} \sum_{i=1}^{N} \sum_{j=1}^{n} (X_j^i - \widehat{X}_j^i)^2}$ of the solution X obtained by the approximate solvers (black dashed curve: implicit Euler method solved with 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 $\tau = 1\,\text{ms}$.

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 4.2.3.1). 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 much 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. (4.9), 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 Δt 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 τ the step size Δt needs to be reduced.

The fully connected network constitutes the worst case test for the class of rate-based models Eq. (4.9), 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

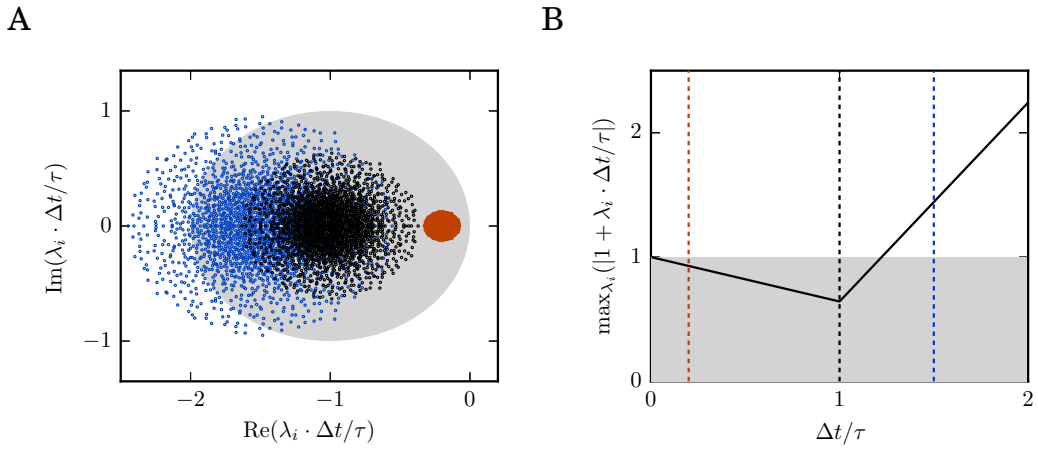


Figure 4.3: **Stability analysis for a balanced excitatory-inhibitory network.** The network contains in total N neurons where the number of excitatory neurons is four times larger then the number of inhibitory neurons. Each neuron receives input from a fixed number of $0.8 \cdot p \cdot N$ excitatory 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 λ_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 λ_i dependent on $\Delta t/\tau$. The gray area again indicates the region where the stability criterion of the Euler Maruyama method is meet. Colored vertical lines correspond to the rescaled eigenvalues displayed in panel (A).

is a perfectly balanced network of excitatory and inhibitory neurons (Rajan & Abbott, 2006). For these models the conditions on the step size Δt 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. 4.3B 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 τ 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.

4.3.2 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 4.2.3.3 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. 4.3.1, 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. 4.4A) 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 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 4.2.3.3) 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. 4.4B 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

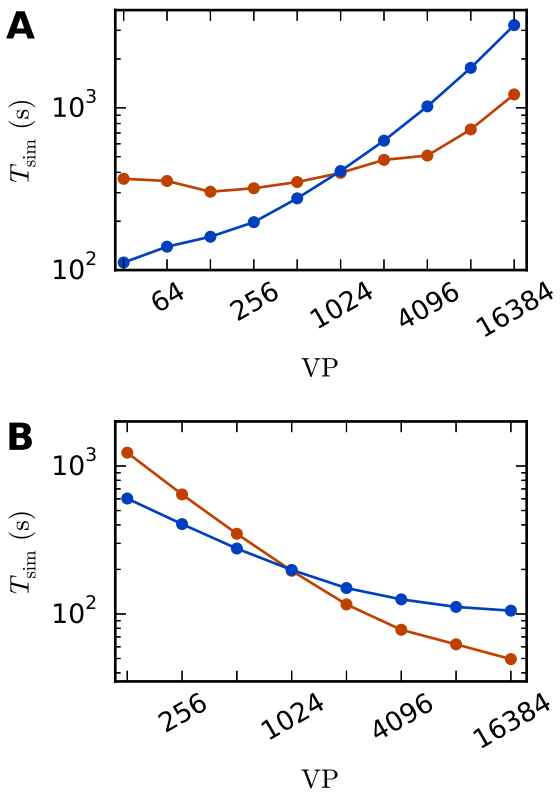


Figure 4.4: **Scaling behavior of an excitatory-inhibitory network.** Simulation time with waveform relaxation (red curves, `wfr_comm_interval`: 1.0 ms, `wfr_tol`: 10^{-4}) and without waveform relaxation (blue curves) as a function of the number of virtual processes in double logarithmic representation. The simulations span $T = 100$ ms of biological time at a computation step size of $h = 0.1$ ms. **(A)** Weak scaling with 100 neurons per virtual process VP. **(B)** Strong scaling with a total number of $N = 51,200$ neurons. Other network parameters as in Sec. 4.3.1 with $\mu = 0$, $\sigma = 1$ and $\tau = 10$ ms.

(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.

4.3.3 Applications

This section presents several applications of the simulation framework. 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.

4.3.3.1 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., 2013b; Trousdale et al., 2012; Pernice et al., 2011; Dahmen et al., 2016a). Networks of such noisy linear rate models have been investigated to explain features such as oscillations (Bos et al., 2015) 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 (4.9) with $\phi(x) = \psi(x) = x$, $\mu = 0$, and noise amplitude σ,

$$\tau dX^i(t) = \left(-X^i + \sum_{j=1}^{N} w^{ij} X^j(t) \right) dt + \sqrt{\tau}\sigma dW^i(t) \,. \tag{4.15}$$

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., 2016a)

$$c(t) = \sum_{i,j} \frac{v^{i\mathrm{T}} \sigma^2 v^j}{\lambda_i + \lambda_j} u^i u^{j\mathrm{T}} \left(\theta(t) \frac{1}{\tau} e^{-\lambda_i \frac{t}{\tau}} + \theta(-t) \frac{1}{\tau} e^{\lambda_j \frac{t}{\tau}} \right), \tag{4.16}$$

where θ denotes the Heaviside function. The λ_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., 2013b). Fig. 4.5 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.

4.3.3.2 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

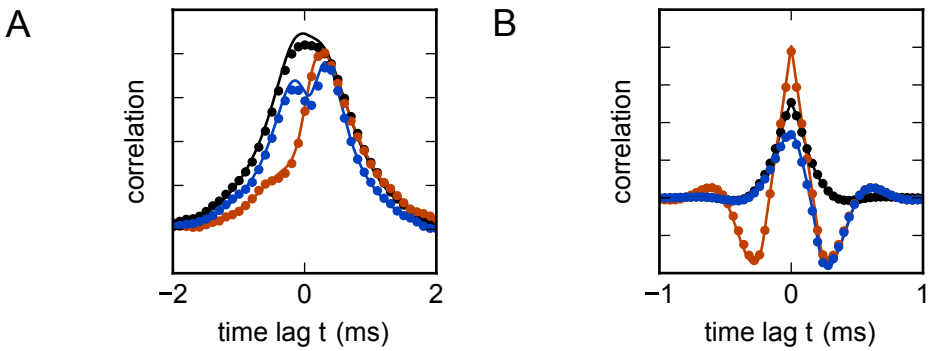


Figure 4.5: **Linear rate model of a random excitatory-inhibitory network. (A)** Cross-correlation functions of two pairs of excitatory neurons (black, red) and an excitatory-inhibitory neuron pair (blue) in a network without delay. The variability across correlation functions arises from heterogeneity 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 $w_I = -6\,w_E$, $\tau = 1\,\text{ms}$, $\mu = 0$, $\sigma = 1$. Step size: $h = 0.1\,\text{ms}$.

network of non-linearly coupled rate units. The original model is deterministic and has been recently extended to stochastic dynamics (Goedeke et al., 2016). The model definition follows from Eq. (4.9) with $\mu = 0$, $\phi(x) = x$, $\psi(x) = \tanh(x)$, i.e.

$$\tau dX^i(t) = \left(-X^i(t) + \sum_{j=1}^{N} w^{ij} \tanh\left(X^j(t)\right) \right) dt + \sqrt{\tau}\sigma\, dW^i(t)\,, \tag{4.17}$$

where $w^{ij} \approx \mathcal{N}(0, g^2/N)$ are Gaussian random couplings. In the thermodynamic limit $N \to \infty$, the population averaged autocorrelation function $c(t)$ can be determined within dynamic mean-field theory (Sompolinsky et al., 1988; Goedeke et al., 2016; Schuecker et al., 2016). Comparing $c(t)$ obtained by simulation of a network (Eq. (4.17)) with the analytical result (Goedeke et al., 2016, eqs. 6 and 8) demonstrates excellent agreement (Fig. 4.6). The simulated network is two orders of magnitude smaller than the cortical microcircuit, illustrating that in this context finite-size effects are already negligible at this scale.

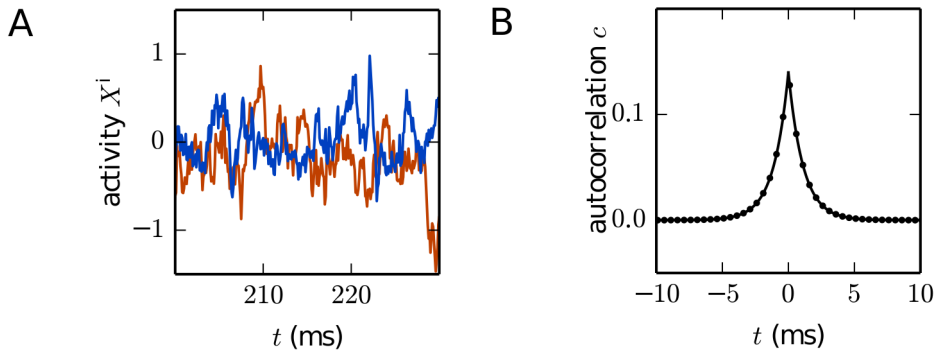


Figure 4.6: **Nonlinear network model.** Simulation of the network specified by (4.17) with $N = 1000$ neurons. **(A)** Noisy example trajectories of two neurons. **(B)** Autocorrelation function obtained by simulation averaged over all neurons (dots) and theory (solid curve). Other parameters: $\tau = 1$ ms, $\sigma = 0.5$, $g = 0.5$. Step size: $h = 0.1$ ms.

4.3.3.3 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

$$\tau dX(\varphi,t) = \left(-X(\varphi,t) + \phi\left[I_{\text{ext}} + \int_{-\pi}^{\pi} d\varphi'\, w(|\varphi-\varphi'|)X(\varphi',t-d)\right]\right) dt \tag{4.18}$$

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

$$w(|\varphi-\varphi'|) = w_0 + w_1 \cos(\varphi-\varphi'). \tag{4.19}$$

The spatial variable φ can also be interpreted as the preferred orientation of a set of neurons, thus rendering Eq. (4.18) a model in feature space (Hansel & Sompolinsky,

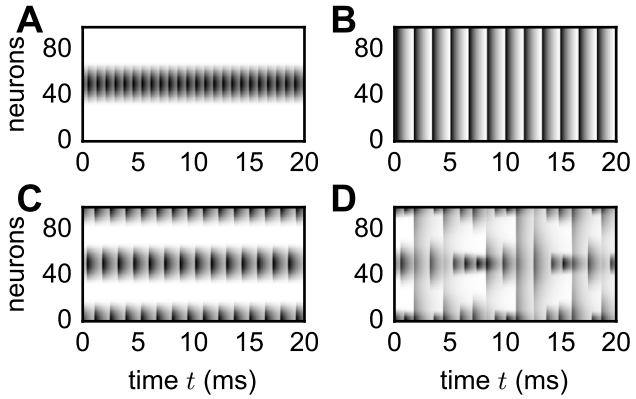


Figure 4.7: **Spatial patterns in functional neural-field model.** Vertical axis shows neuron indices organized according to ascending angle $\varphi \in [-\pi, \pi)$. Neuronal activity $X_i(t) = X(\varphi_i, t)$ encoded by gray scale with white denoting no activity. Initial transients not shown. Patterns reproduce the analytically derived phase diagram in the original study by Roxin et al. (2005). Parameters: $N = 100$, $d = 0.1$ ms, $\tau = 1$ 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: $X_i(0) = X(\varphi_i, 0) = \pi^2 - \varphi_i^2$. Step size: $h = 0.01$ ms.

1998). Discretizing space into N segments yields the following set of coupled ODEs

$$\tau dX^i = \left(-X^i + \phi \left[I_{\text{ext}} + \sum_{j=1}^{N} w^{ij} X^j(t-d) \right] \right) dt \tag{4.20}$$

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. To illustrate applicability of the simulation framework to neural-field models, we reproduce various patterns (Fig. 4.7) 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. 4.7 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.

4.3.3.4 Mean-field analysis of complex networks

A network of spiking neurons constitutes a high dimensional and complex system. To investigate its stationary state, one can describe the activity in terms of averages across neurons and time, leading to population averaged stationary firing rates (Brunel, 2000). Here, the spatial average collapses a large number of neurons into a single population, which interpreted as a single rate unit. The ability to represent

spiking as well as rate dynamics by the same simulation framework allows a straightforward analysis of the spiking network by replacing the spiking neuron populations by single rate-based model neurons.

In more formal terms, we now consider networks of neurons structured into N interconnected populations. A neuron in population α receives $K_{\alpha\beta}$ incoming connections from neurons in population β, each with synaptic efficacy $w_{\alpha\beta}$. Additionally, each neuron in population α is driven by $K_{\alpha,\text{ext}}$ Poisson sources with rate X_{ext} and synaptic efficacy w_{ext}. We assume leaky integrate-and-fire model neurons with exponentially decaying post-synaptic currents. The dynamics of membrane potential V and synaptic current I_s is (Fourcaud & Brunel, 2002)

$$\begin{aligned} \tau_m \frac{dV^i}{dt} &= -V^i + I_s^i \\ \tau_s \frac{dI_s^i}{dt} &= -I_s^i + \tau_m \sum_{j=1}^{N} w^{ij} \sum_k \delta(t - t_k^j - d)\,, \end{aligned} \tag{4.21}$$

where t_k^j denotes the k-th spike-time of neuron j, and τ_m and τ_s are the time constants of membrane and synapse, respectively. The membrane resistance has been absorbed in the definition of the current. Whenever the membrane potential V crosses the threshold θ, the neuron emits a spike and V is reset to the potential V_r, where it is clamped for a period of length τ_r. If we assume that all neurons have identical parameters, a diffusion approximation leads to the population-averaged firing rates X_α (Fourcaud & Brunel, 2002)

$$\begin{aligned} \frac{1}{X_\alpha} &= \tau_\text{r} + \tau_\text{m}\sqrt{\pi} \int_{(V_\text{r}-\mu_\alpha)/\sigma_\alpha+\gamma\sqrt{\tau_\text{s}/\tau_\text{m}}}^{(\theta-\mu_\alpha)/\sigma_\alpha+\gamma\sqrt{\tau_\text{s}/\tau_\text{m}}} e^{u^2}\,(1+\text{erf}(u))\,\text{d}u \\ &=: 1/\Phi_\alpha(X) \qquad (4.22) \\ \mu_\alpha &= \tau_\text{m} \sum_\beta K_{\alpha\beta} w_{\alpha\beta} X_\beta + \tau_\text{m} K_{\text{ff,ext}} w_\text{ext} X_\text{ext} \qquad (4.23) \\ \sigma_\alpha^2 &= \tau_\text{m} \sum_\beta K_{\alpha\beta} w_{\alpha\beta}^2 X_\beta + \tau_\text{m} K_{\text{ff,ext}} w_\text{ext}^2 X_\text{ext}. \qquad (4.24) \end{aligned}$$

Here, $\gamma = |\zeta(1/2)|/\sqrt{2}$, with ζ denoting the Riemann zeta function (Abramowitz & Stegun, 1974). We find the fixed points of Eq. (4.22) by solving the first-order differential equation (Wong & Wang, 2006; Schuecker et al., 2016)

$$\tau \text{d}X_\alpha = (-X_\alpha + \Phi(X))\,dt, \tag{4.25}$$

which constitutes a network of rate neurons with the dimension equal to the number of populations N.

Next we apply this framework to a cortical microcircuit model (Potjans & Diesmann, 2014) constituting roughly 80000 spiking neurons structured into 8 populations across 4 layers $[L23, L4, L5, L6]$, with one excitatory and one inhibitory cell

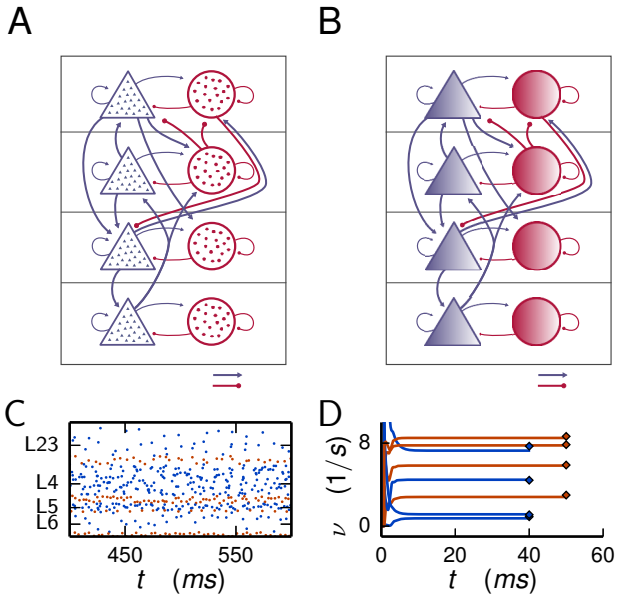


Figure 4.8: **Reduction of spiking microcircuit model to rate dynamics**. **(A)** Sketch of a microcircuit model (Potjans & Diesmann, 2014) with excitatory (blue triangles) and inhibitory (red circles) neuron populations, each consisting of a large number of neurons indicated by the small triangles and 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. **(D)** Dynamics of the eight neurons of the rate network Eq. (4.25) (curves) and comparison to population averaged firing rates obtained from direct simulations of the spiking network (diamonds).

type each (Fig. 4.8). The model exhibits irregular and stationary spiking activity (Fig. 4.8C). Replacing each population by a single rate neuron (4.8B) results in an eight-dimensional rate network Eq. (4.25) which converges to a fixed point corresponding to the population-averaged firing rates obtained from direct simulation of the spiking model (Fig. 4.8D).

4.3.3.5 Further neuron models

Nonlinear neuron dynamics A common feature of various rate neurons considered so far is the leaky neuron dynamics, i.e the linear term $-X^i(t)$ in Eq. (4.9). However the presented framework can be easily extended to nonlinear neuron dynamics as used for example by Stern et al. (2014). In a more general form Eq. (4.9) reads

$$\tau dX^i(t) = \left[a(X^i(t)) + \phi\left(\sum_{j=1}^{N} w^{ij} \psi\left(X^j(t-d) \right) \right) \right] dt + \sqrt{\tau}\sigma dW^i(t) \tag{4.26}$$

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,

$$X^i_{k+1} = X^i_k + \left[a(X^i_k) + \mu + \phi\left(\sum_{j=1}^{N} w^{ij} \psi\left(X^j_{k-\frac{d}{\Delta t}} \right) \right) \right] \frac{1}{\tau} \Delta t + \frac{1}{\sqrt{\tau}} \sigma \Delta W^i_k. \tag{4.27}$$

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

$$X^i_{k+1} = e^{-\Delta t/\tau} X^i_k + \left(1 - e^{-\Delta t/\tau}\right) \left[f(X^i_k) + \phi\left(\sum_{j=1}^{N} w^{ij} \psi\left(X^j_{k-\frac{d}{\Delta t}} \right) \right) \right] + \sqrt{\frac{1}{2}(1 - e^{-2\Delta t/\tau})}\, \sigma \eta^i_k, \tag{4.28}$$

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 (1972b, 1973). In the most general form, this amounts to

$$\tau dX^i(t) = \left[-X^i(t) + H(X^i) \cdot \phi\left(\sum_{j=1}^{N} w^{ij} \psi\left(X^j(t-d) \right) \right) \right] dt + \sqrt{\tau} \sigma dW^i(t)\,, \tag{4.29}$$

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

$$X^i_{k+1} = e^{-\Delta t/\tau} X^i_k + \left(1 - e^{-\Delta t/\tau}\right) \left[H(X^i_k) \cdot \phi\left(\sum_{j=1}^{N} w^{ij} \psi\left(X^j_{k-\frac{d}{\Delta t}} \right) \right) \right] + \sqrt{\frac{1}{2}(1 - e^{-2\Delta t/\tau})}\, \sigma \eta^i_k, \tag{4.30}$$

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

Multiplicative noise So far, we have considered rate models subject to additive noise corresponding to $b(t, x(t)) = b(t)$ in Eq. (4.4). The linear rate model considered in Sec. 4.3.3.1 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. (4.8).

Output noise Grytskyy et al. (2013b) 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. Here the noise is added to the output rate of the afferent neurons

$$\tau \frac{dX^i(t)}{dt} = -X^i(t) + \mu + \phi\left(\sum_{j=1}^{N} w^{ij}\psi\left(X^j(t-d) + \sqrt{\tau}\sigma\xi^j(t)\right)\right) \qquad i = 1,\ldots,N \tag{4.31}$$

and we cannot write the system as a SDE of type Eq. (4.2), as the nonlinearities $\phi(x)$ and $\psi(x)$ are also applied to the white noise ξ^j. In addition to the implementation `rate_neuron_ipn` for the rate-based models Eq. (4.9) 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 ξ^j to be constant over the update interval which leads to the update formula

$$X^i_{k+1} = e^{-\Delta t/\tau} X^i_k + \left(1 - e^{-\Delta t/\tau}\right)\left[\mu + \phi\left(\sum_{j=1}^{N} w^{ij}\psi\left(X^j_k + \sqrt{\frac{\tau}{\Delta t}}\sigma\eta^j_k\right)\right)\right]. \tag{4.32}$$

The term $X^j_k + \sqrt{\frac{\tau}{\Delta t}}\sigma\eta^j_k$ with $\eta^j_k \sim \mathcal{N}(0,1)$ 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.

4.4 Conclusion

Contemporary modeling approaches to the dynamics of neural networks consider two main classes of models: biologically grounded spiking neurons and functionally inspired rate-based units. The unified simulation framework presented here supports the combination of the two for multi-scale modeling approaches, the quantitative validation of mean-field approaches by spiking network simulations, and an increase in reliability by usage of the same simulation code and the same network model specifications for both model classes. While most efficient spiking simulations rely on the communication of discrete events, rate models require time-continuous interactions between neurons. Exploiting the conceptual similarity to the inclusion of gap junctions in spiking network simulations, we arrived at a reference implementation of instantaneous and delayed interactions between rate-based models in a spiking network simulator. These rate-based models are described by ordinary and stochastic (delay) differential equations whose solution can be best obtained in a distributed simulator using an exponential Euler scheme. The iterative waveform-relaxtion scheme improves scalability on supercomputers by reducing communication at the cost of additional computations. Extensibility of the framework is achieved by using an implementation of neuron models in terms of `C++` templates, and the

framework's broad applicability is demonstrated by various examples from the literature ranging from random networks to neural-field models.

Part IV

Modeling of mesoscale fluctuations based on microscopic dynamics

Chapter 5

Hybrid scheme for modeling local field potentials from spiking point-neuron networks

This chapter is based on the following publication:

Hagen E*, Dahmen D*, Stavrinou ML, Lindén H, Tetzlaff T, van Albada SJ, Grün S, Diesmann M and Einevoll GT (2016) *Hybrid Scheme for Modeling Local Field Potentials from Point-Neuron Networks,* Cerebral Cortex, bhw237.
*These authors contributed equally to this publication.
Preliminary results have been published in Hagen et al. (2013); Dahmen et al. (2013); Stavrinou et al. (2014); Dahmen et al. (2014); Hagen et al. (2015a,b).

Author contributions:
Under the supervision of T Tetzlaff, S Grün and M Diesmann, the author contributed to all parts of the above publication. Considering the equal contribution of the author and E Hagen, the author focused more on the point-neuron-network component and the analysis of the example model of a cortical microcircuit, while E Hagen focused more on the multicompartment-model component and the development of the software package. SJ van Albada and M Stavrinou helped with the derivation of the spatial connectivity profiles. All authors contributed to the design of the numerical experiments, the data analysis and the writing of the manuscript.

5.1 Introduction

With rapidly advancing multielectrode recording technology, the local field potential (LFP), the low-frequency component ($\lesssim$ 500 Hz) of the extracellular potential recorded in the brain, has again become a popular measure of neuronal activity in both research and clinical applications. Proper understanding of the LFP, which originates from transmembrane currents across neuronal dendrites (Nicholson & Freeman, 1975), requires detailed mathematical modeling incorporating the anatomical and electrophysiological features of the thousands or even millions of neurons near the recording electrode, as well as synaptic inputs from the entire large-scale neuronal circuitry (Kajikawa & Schroeder, 2011; Lindén et al., 2011; Łęski et al., 2013; Einevoll et al., 2013). Large-scale models are therefore necessary to include contributions to the LFP from local and distant populations.

Modeling large-scale neural-network dynamics with individual spiking neurons is challenging due to the memory required to represent the large number of synapses. Currently, the largest simulations of neural networks comprise up to 10^9 model neurons with highly simplified dynamics (Diesmann, 2013; Kunkel et al., 2014). Typically, these models neglect the spatial aspects of neuronal morphologies and describe neurons as points in space (point-neuron models). Point-neuron networks are amenable to mathematical analysis (see previous chapters) and can be efficiently evaluated numerically (Plesser et al., 2007; Brette et al., 2007; Helias et al., 2012; Kunkel et al., 2014).

Although point-neuron networks capture many features of *in vivo* spiking activity, they fail to predict extracellular potentials that result from transmembrane currents distributed across the cell surface. According to Kirchhoff's law of current conservation, the sum of all transmembrane currents, including all ionic and capacitive currents, must be zero for each neuron. In a point-neuron model, all transmembrane currents are collapsed in a single point in space. The net transmembrane current, and hence the extracellular potential, therefore vanishes. Only the spatial separation between current sinks and sources leads to a non-zero extracellular potential (Pettersen et al., 2012; Einevoll et al., 2013). *A priori*, the prediction of extracellular potentials therefore requires spatially extended neuron models accounting for the spatial distribution of transmembrane currents, commonly handled using multicompartment-neuron models (De Schutter & Van Geit, 2009). Existing multicompartment-neuron-network models accounting for realistic cell morphologies are, however, restricted to sizes of $\sim 10^4$–10^5 neurons (Hines et al., 2008; Reimann et al., 2013; Migliore et al., 2014; Markram et al., 2015).

In several previous studies (Bazhenov et al., 2001; Hill & Tononi, 2005; Ursino & Cara, 2006; Mazzoni et al., 2008, 2010, 2011), the activity of point-neuron networks (e.g. population firing rates, synaptic currents, membrane potentials) has nevertheless been used as a proxy for the LFP when comparing with experiments. In a recent

study comparing different candidate proxies, it was found that a suitably chosen sum of synaptic currents could provide a good LFP proxy, but only for the case when the LFP is generated from transmembrane currents of a single population of pyramidal neurons (Mazzoni et al., 2015). In cortex, however, several populations in general contribute to the LFP, and there are spatial cancellation effects when positive LFP contributions from one population overlap in space with negative LFP contributions from other populations. This effect cannot be accounted for by a simple LFP proxy.

In this chapter, we present a hybrid modeling scheme which combines the simplicity and efficiency of point-neuron network models and the biophysical principles underlying LFP generation captured by multicompartment neuron models with anatomically reconstructed morphologies. The scheme allows for arbitrary numbers of LFP-contributing populations, and directly incorporates spatial cancellation effects. Further, the spatially extended LFP-generating neurons assure that effects from intrinsic dendritic filtering of synaptic inputs are included in the predicted LFP (Lindén et al., 2010). The scheme assumes that the spiking activity of the neural network (Fig. 5.1B) generating the synaptic input reflected in the LFP is well described by a point-neuron network model (Fig. 5.1A). The network spiking activity serves as synaptic input to a population of mutually unconnected multicompartment model neurons with realistic morphologies positioned in 3D space (Fig. 5.1C) and is thereby translated into a distribution of transmembrane currents and, hence, an LFP (Fig. 5.1D). Thus each multicompartment model neuron has its equivalent in the point-neuron network and receives input spikes from the same presynaptic neurons as this point-neuron equivalent.

In the proposed hybrid modeling scheme, the LFP stems from the presynaptic spiking activity, but does not affect the spike-generation dynamics. Thus, the modeling of the spike trains and the LFP generation are separated so that the effects of the spatial and electrophysiological properties of the postsynaptic (multicompartment) neurons on the LFP can be investigated independently of the spike-generation dynamics. Due to the linearity of Maxwell's equations and volume conduction theory linking transmembrane currents to extracellular potentials (Pettersen et al., 2012; Einevoll et al., 2013), the compound LFP results from the linear superposition of all single-cell LFPs generated by the collection of neurons in the multicompartment model neuron population (Einevoll et al., 2013). Note that this linear superposition principle applies even for nonlinear cell dynamics (e.g., nonlinear synaptic integration, action-potential generation, active conductances) as in Reimann et al. (2013). As ephaptic interactions (Anastassiou et al., 2011) are neglected, the LFP contribution from each multicompartment model neuron can be treated independently from the others. The computational hybrid LFP scheme proposed here exploits the methodological and conceptual advantages due to the independence of the contributions to the LFP from each multicompartment model neuron: The evaluation of the LFPs becomes 'embarrassingly parallel' (see Foster, 1995) and simulations of the multicompartment model neuron dynamics can be easily distributed in parallel across many compute

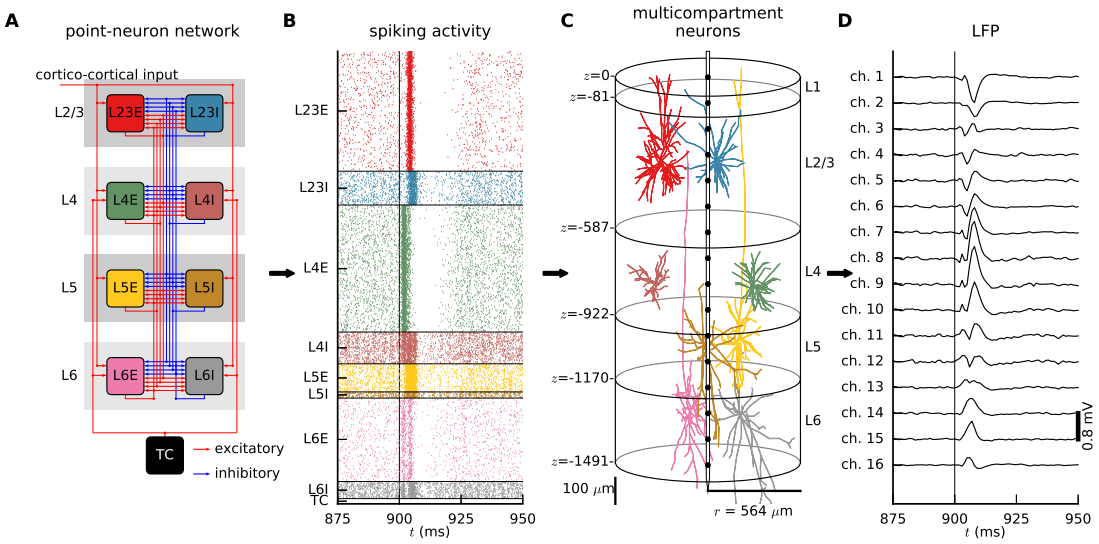


Figure 5.1: **Overview of the hybrid LFP modeling scheme for a cortical microcircuit model.** **A** Sketch of the point-neuron network representing a 1 mm^2 patch of early sensory cortex (adapted from Potjans & Diesmann (2014)). The network consists of 8 populations of leaky integrate-and-fire (LIF) neurons, representing excitatory (E) and inhibitory neurons (I) in cortical layers 2/3, 4, 5 and 6. External input is provided by a population of thalamo-cortical (TC) neurons and cortico-cortical afferents. The color coding of neuron populations is used consistently throughout this chapter. Red arrows: excitatory connections. Blue arrows: inhibitory connections. See Tab. 5.1-5.2, D.1-D.2 for details on the network model. **B** Spontaneous ($t < 900$ ms) and stimulus-evoked spiking activity (synchronous firing of TC neurons at time $t = 900$ ms, denoted by thin vertical line) generated by the point-neuron network model shown in panel A, sampled from all neurons in each population. Each dot represents the spike time of a particular neuron. **C** Populations of LFP-generating multicompartment model neurons with reconstructed, layer- and cell-type specific morphologies. Cells are distributed within a cylinder spanning the cortex. Layer boundaries are marked by horizontal black lines (at depths z relative to cortex surface $z = 0$). Only one representative neuron for each population is shown (see Fig. 5.2 for a detailed overview of cell types and morphologies). Sketch of a laminar recording electrode (gray) with 16 contacts separated by 100 μm (black dots). **D** Depth-resolved LFP traces predicted by the model (cf. Tab. 5.3-5.4). Note that channel 1 is at the pial surface, so that channel 2 corresponds to a cortical depth of 100 μm and so forth.

units (i.e., CPUs). Although tailored towards use on high-performance computing facilities, the hybrid simulation can in principle be run on a single laptop.

The hybrid scheme predicts spatially and temporally resolved neural activity at various scales: spikes, synaptic currents, membrane potentials, current-source densities (CSD, see e.g., Nicholson & Freeman (1975); Pettersen et al. (2006, 2008)), and LFPs. It therefore allows for investigation of relationships between different measures of neural activity. Thus, although point-neuron networks until now only have connected to

in vivo experiments via measurement of spikes, single-neuron membrane potentials and currents, the present hybrid scheme allows for comparison of model predictions also with measured LFPs (and associated CSDs).

As an illustration, we apply the hybrid scheme to a multi-layered point-neuron network model of an early sensory cortical microcircuit (Potjans & Diesmann, 2014). We thereby demonstrate how to obtain LFP predictions from point-neuron network models using additional spatial connectivity information from anatomical data (Binzegger et al., 2004; Izhikevich & Edelman, 2008).

The network model of Potjans & Diesmann (2014) is chosen here since it has a minimum level of detail in the sense that individual neurons have simplified leaky integrate-and-fire (LIF) dynamics, but still represents a cortical column with full density of neurons and connections. The connectivity in such a full-scale circuit alone suffices to explain realistic firing rates across populations as well as propagation of activity through layers (Potjans & Diesmann, 2014). Applicability of the scheme is, however, not restricted to this model as it in principle can be used for all network models generating spikes.

5.2 Point-neuron network model

The point-neuron network model is a key component of the hybrid scheme. The hybrid scheme enables LFP predictions from network models with an arbitrary number of populations and thus permits application to a large class of networks with arbitrarily complex single-neuron and synapse dynamics. The example network of 'spike-generators' used here is, except for some minor adjustments (see below), the multi-layered model of a cortical microcircuit published by Potjans & Diesmann (2014). The model is implemented and included in `NEST` (http://www.nest-simulator.org, Eppler et al. (2015)) and was recently made freely available online (http://www.opensourcebrain.org/projects/potjansdiesmann2014).

The network model describes 1 mm^2 of primary sensory cortex and consists of four layers with one excitatory (E) and one inhibitory (I) neuron population each, as illustrated in Fig. 5.1A. The network receives modulated thalamic input in addition to stationary external input. Whereas the neuron (leaky-integrate-and-fire) and synapse (static, exponential-current-based) model are intentionally left simple, the focus of this network implementation is on the complex connectivity which integrates multiple sources of anatomical and electrophysiological data (Potjans & Diesmann, 2014, and references therein). Apart from the layer identity, the model does not explicitly account for cell positions. For the full network description, see Tab. 5.1,5.2 and D.1. The microcircuit model reproduced experimentally observed distributions of firing rates across populations and propagation of activity across layers (Potjans & Diesmann, 2014). It thus forms a suitable starting point for LFP predictions in a cortical column.

A	**Model summary**
Structure	Multi-layered excitatory-inhibitory (E-I) network
Populations	8 cortical in 4 layers, 1 thalamic (TC)
Connectivity	Random, independent, population-specific, fixed number of connections
External input	Cortico-cortical: constant current with population-specific strength
Neuron model	Cortex: leaky integrate-and-fire (LIF); TC: point process
Synapse model	Exponential postsynaptic currents, static weights, population-specific weight distributions
Measurements	Spike activity, input currents, membrane potential of each neuron

B	**Network model**
Connectivity	Connection probability C_{YX} ($X, Y \in \{\text{L2/3, L4, L5, L6}\} \times \{\text{E,I}\} \cup \text{TC}$, $C_{YX} = 0$ for $Y = \text{TC}$) - Fixed number of synapses K_{YX} between populations X and Y - Binomial in-/outdegrees
Input	Cortico-cortical direct current I_Y^{ext}

C	**Neuron model**
	Cortex
Type	Leaky integrate-and-fire neuron
Description	Dynamics of membrane potential $V_i(t)$ (neuron $i \in [1, N]$): - Spike emission at times t_l^i with $V_i(t_l^i) \geq \theta$ - Subthreshold dynamics: $\tau_{\text{m}} \dot{V}_i = -V_i + R_{\text{m}} I_i(t)$ if $\forall l : t \notin (t_l^i, t_l^i + \tau_{\text{ref}}]$ - Reset + refractoriness: $V_i(t) = V_{\text{reset}}$ if $\forall l : t \in (t_l^i, t_l^i + \tau_{\text{ref}}]$ Exact integration with temporal resolution dt (Rotter & Diesmann, 1999) Uniform distribution of membrane potentials at $t = 0$
	Thalamus
Type	- DC current for constant background input - Nonstationary Poisson process for modulation
Description	DC current included in external DC input Types of thalamic input modulation: - Spontaneous activity: no modulation in activation of thalamic neurons - Thalamic pulses: fixed-interval coherent activation of all N_{TC} thalamic neurons - AC modulation: Poisson spike trains with sinusoidally modulated rate profile (discretized with time resolution dt): $\nu_{\text{th}}(t) = \overline{\nu}_{\text{TC}} + \Delta\nu_{\text{TC}} \sin(2\pi t f_{\text{TC}})$ (5.1)

Table 5.1: Description of point-neuron network for the cortical microcircuit model (continued in Tab. 5.2) following the guidelines of Nordlie et al. (2009).

D	**Synapse model**
Type	Exponential postsynaptic currents, static weights
Description	Input current of neuron j of synapses formed with presynaptic neurons i: $$I_j(t) = \sum_i J_{ji} \sum_l \exp\left(-(t - t_l^i - d_i)/\tau_s\right) H(t - t_l^i - d_i) + I_j^{ext}$$ - Static synaptic weights $J_{ji} = \mathrm{sgn}(X)\,\lvert J_{YX}\rvert$ ($i \in X$, $j \in Y$); $\mathrm{sgn}(X) = 1$ for $X \in \{\text{L2/3E, L4E, L5E, L6E, TC}\}$, -1 otherwise - Absolute weights $\lvert J_{YX}\rvert$ drawn from lognormal distribution $$p(\lvert J_{YX}\rvert) = \frac{1}{\sqrt{2\pi}\sigma_{YX}\lvert J_{YX}\rvert} \exp\left(-\frac{(\ln\lvert J_{YX}\rvert - \mu_{YX})^2}{2\sigma_{YX}^2}\right) \quad (5.2)$$ or normal distribution $$p(\lvert J_{YX}\rvert) = \frac{1}{\sqrt{2\pi}\sigma_{YX}} \exp\left(-\frac{(\lvert J_{YX}\rvert - \mu_{YX})^2}{2\sigma_{YX}^2}\right) \quad (5.3)$$ with $\mu_{YX} = g_{YX} J$ and $\sigma_{YX} = \sigma_{J,\mathrm{rel}}\,\mu_{YX}$ - Delays $d_i = d_X$ ($i \in X$) drawn from (left-clipped) Gaussian distribution $$p(d_X) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(-\frac{(d_X - \mu_X)^2}{2\sigma_X^2}\right) \quad (5.4)$$ with mean $\mu_X = d_E, d_I$ for X exc., inh., stdev $\sigma_X = \sigma_{d,\mathrm{rel}}\,\mu_X$ and $d_X \in [dt, \infty)$ - $H(t) = 1$ for $t \geq 0$, and 0 elsewhere. - External DC input $I_j^{ext} = I_Y^{ext} = I^{ext} k_Y^{ext}$ ($j \in Y$)

Table 5.2: Description of point-neuron network for the cortical microcircuit model (continuation of Tab. 5.1).

The stationary thalamic Poisson input and cortico-cortical input to the microcircuit present in the original model of Potjans & Diesmann (2014), are here replaced by DC currents. DC input slightly increases the degree of synchrony (see, e.g., Brunel (2000, Fig. 5)), but retains network dynamics and firing rate distributions across populations as in Potjans & Diesmann (2014). The network of Potjans & Diesmann (2014) shows slightly synchronous behavior due to the E-I network of layer 4 being close to the synchronous irregular (SI) regime (Brunel, 2000; Bos et al., 2015). In order to reduce synchrony, we here increased the average synaptic weight from neurons in population L4I (inhibitory) to L4E (excitatory) neurons by 12.5%, resulting in attenuated oscillations in layer 4. Taking advantage of the fact that point-neuron networks are amenable for theoretical analysis, we derived modified weights based on predictions from dynamical mean-field theory applied to the microcircuit model (Bos et al., 2015). Moreover, we found that high-frequency network oscillations seen for Gaussian synaptic weight distributions are reduced when using lognormally distributed synaptic weights (Sarid et al., 2007; Iyer et al., 2013; Teramae & Fukai, 2014). This made the dynamics more similar to experimental observations (Song et al., 2005; Buzsaki & Mizuseki, 2014), and we thus also chose this for our network. Henceforth, we refer to our modified network as the 'reference network'. Modulated activity of each thalamo-cortical (TC) neuron in the external thalamic population was modeled as synchronous spikes or as independent non-stationary Poisson processes with sinusoidally oscillating rate profiles (cf. Eq. (5.1)).

5.3 Populations of multicompartment model neurons

Cancellation effects from positive and negative contributions to extracellular potentials and effects of intrinsic dendritic filtering can only be captured with spatially extended multicompartment neuron models (Einevoll et al., 2013). In the hybrid scheme, extracellular potentials are estimated from the spiking activity in the point-neuron network through synaptic activation of populations of multicompartment model neurons ('LFP generators'). In principle, these mutually unconnected model neurons mirror their network counterparts and receive inputs from exactly the same point neurons. In addition to the description of the point-neuron network model, different types of spatial information are thus needed to predict LFPs. For one, detailed dendritic morphologies are required for each individual network population (Fig. 5.2). Further, the positions of neurons and synaptic connections must also be specified, as well as the separation of network populations into morphologically distinct cell types (Fig. 5.2 and 5.4).

Availability of detailed cell-type specific connectivity of neural circuits, especially including information about synapse positions, is limited due to the substantial experimental effort involved. However, several ongoing large-scale neuroscience projects (Kandel et al., 2013) address this issue and detailed connectomes are beginning to

become publicly available (Jiang et al., 2015; Reimann et al., 2015; Markram et al., 2015). In the present example application we used the connection probabilities as given by Izhikevich & Edelman (2008) derived from Binzegger et al. (2004). Note that the point-neuron network connectivity was partially derived from the same data (Potjans & Diesmann, 2014). Quantitative data was provided for the number of connections in five cortical layers (layer 1 (L1), layers 2 and 3 grouped into a joint layer 2/3 (L2/3), and layers 4 (L4), 5 (L5) and 6(L6)) between 17 cortical cell types, cortico-cortical connections from other areas, and two thalamo-cortical relay cell types. We follow the nomenclature of Izhikevich & Edelman (2008), where y = p23 denotes pyramidal cell types in layer 2/3, y = b23 and y = nb23 basket interneurons and non-basket interneurons within the same layer, y = ss4(L23) spiny stellate cells in layer 4 with targets mainly within layer 2/3, y = p4 layer 4 pyramidal cells and so forth. Out of the 17 covered intracortical cell types only the y = nb1 cell type is not associated with any point-neuron network population in our scheme. To account for the lack of layer 1 neurons in our model, we renormalized the connection probabilities for the remaining 16 cortical cell types including the two thalamo-cortical (TC) relay-cell types, such that the occurrences F_y of all cell types y summed to 100% as given in Tab. D.4. Further, we assumed that the excitatory point-neuron network populations within one layer are composed of pyramidal cells and spiny stellate cells if both are present in the layer, and that inhibitory network populations encompass both types of interneurons. This results in the grouping of cell types y into postsynaptic populations Y illustrated in Fig. 5.2. The neuron count N_y of each cell type is then trivially computed from the frequency of occurrence F_y as given in Tab. D.4 and Fig. 5.2.

Inclusion of cell-type and layer-specific connections in the present hybrid scheme has some implications for how we proceed with setting up equivalent populations consisting of morphologically detailed model neurons. Different cell types belonging to a particular population may have different spatial distributions of synapses, or the populations may consist of different morphological classes of neurons (Kisvarday & Eysel, 1992; Nowak et al., 2003; Stepanyants et al., 2008). An example is layer 4 in which spiny stellate cells lack apical dendrites, while pyramidal cells have apical dendrites extending into layer 1. To incorporate some of this morphological diversity we considered altogether 16 cell types for the 8 cortical network populations.

For each of the 16 cell types, we acquired representative morphological reconstructions of predominantly cat visual cortex neurons from several sources (Contreras et al., 1997; Kisvarday & Eysel, 1992; Mainen & Sejnowski, 1996; Stepanyants et al., 2008) (cf. Fig. 5.2, Tab. D.3)). Morphology files were obtained either from NeuroMorpho.org (Ascoli et al., 2007) or through personal communication with the authors. Constrained by layer boundary depths (Stepanyants et al., 2008, Fig. 3) and laminar connectivities (see below) we applied an intermediate preprocessing step to our pyramidal cell morphologies. Assuming that the soma compartments of each cell type were centered in their corresponding layer, and noting that the layer-specific connectivity (cf. Tab. D.4) implies connections to layer 1, we stretched the apical dendrites along the axis perpendicular to the cortical surface such that they reached

A	**Model summary**
Topology	Cortical column under $1\,\mathrm{mm}^2$ of cortical surface
Populations	8 excitatory and 8 inhibitory cell types
Input	Spiking activity of thalamic and cortical populations as modeled by point-neuron network
Neuron model	Multicompartment, passive cable formalism
Synapse model	Exponential postsynaptic current, static weights
Measurements	Current source density (CSD), local field potential (LFP)

B	**Topology**
Type	Cylindrical volume with layer-specific distribution of cell types and synapses
Description	Cylinder radius r Laminar, defining upper/lower boundaries of layers 1, 2/3, 4, 5, 6

C	**Populations**
Type	Each cell type y assigned to population Y, $y \in Y$
Description	Populations $Y \in \{\text{L2/3, L4, L5, L6}\} \times \{\text{E,I}\}$ (population size N_Y, cell types $y \in Y$) (e.g., L4E = {p4, ss4(L23), ss4(L4)}, cf. Fig. 5.2). Cell types y: - Size $N_y = F_y \sum_Y N_Y$, F_y is the occurrence of cell type y in the full model - Morphology M_y - Extrapolated according to spatial connectivity data (Tab. D.3) Somatic placement, population Y: - Random soma placement in cylindrical volumes with radius r, thickness h - Volumes centered between boundaries of layers 2/3–6
Morphologies	
Type	3D histological reconstructions from slice preparations (see Jacobs et al. (2009); De Schutter & Van Geit (2009)) of cat visual and somatosensory cortices
Description	One morphology M_y per cell type: - Excitatory and inhibitory cells in layers 2/3–6 - For all cells $j \in y$: $M_j = M_y$ - For some cell types y, y': $M_y = M_{y'}$ (limited availability) Orientations: - Pyramidal cells: apical dendrites oriented along depth axis with random depth-axis rotation - Interneurons, stellate cells: random rotation around all axes Corrections: - Apical dendrites of pyramidal cells elongated to accommodate spatial connectivity - Axons removed if present Reconstructed morphologies (cf. Fig. 5.2): - Cat visual cortex (Kisvarday & Eysel, 1992; Mainen & Sejnowski, 1996; Contreras et al., 1997; Stepanyants et al., 2008) - Cat somatosensory cortex from NeuroMorpho.org (Contreras et al., 1997; Ascoli et al., 2007).

Table 5.3: Description of multicompartment-neuron populations for the cortical microcircuit model (continued in Tab. 5.4).

D	**Neuron models**
Type	Passive, multicompartment, reconstructed morphologies
Description	Compartment n membrane potential $V_{\mathrm{m}jn}$ of cell j having length l_{jn}, diameter d_{jn} and surface area A_{jn}: $C_{\mathrm{m}jn}\frac{dV_{\mathrm{m}jn}}{dt} = \sum_{k=1}^{m} I_{\mathrm{a}jkn} - G_{\mathrm{L}jn}(V_{\mathrm{m}jn} - E_{\mathrm{L}}) - \sum_i I_{jin}$, (5.5) $C_{\mathrm{m}jn} = c_{\mathrm{m}} A_{jn}$, (5.6) $I_{\mathrm{a}jkn} = G_{\mathrm{a}jkn}\left(V_{\mathrm{m}jk} - V_{\mathrm{m}jn}\right)$, (5.7) $G_{\mathrm{a}jkn} = \pi(d_{jk}^2 + d_{jn}^2)/(4r_{\mathrm{a}}(l_{jk} + l_{jn}))$, (5.8) $G_{\mathrm{L}jn} = A_{jn}/r_{\mathrm{m}}$, (5.9) $I_{\mathrm{m}jn} = C_{\mathrm{m}jn}\frac{dV_{\mathrm{m}jn}}{dt} + G_{\mathrm{L}jn}(V_{\mathrm{m}jn} - E_{\mathrm{L}}) + \sum_i I_{jin}$. (5.10) $C_{\mathrm{m}jn}$ is compartment capacitance, $G_{\mathrm{L}jn}$ its passive leak conductance, E_{L} the passive leak reversal potential, $I_{\mathrm{a}jkn}$ axial current between compartment n and neighboring compartment k (out of m compartments), $G_{\mathrm{a}jkn}$ axial conductance between n and k, I_{jin} synaptic currents, and $I_{\mathrm{m}jn}$ transmembrane current of compartment n. For specific parameter values, see Tab. D.2. Membrane potentials and transmembrane currents are computed using `NEURON` through `LFPy` (Carnevale & Hines, 2006; Lindén et al., 2014), assuming the extracellular potential to be zero everywhere on the outside of the neuron (infinite extracellular conductivity).

E	**Synapse model**
Type	Exponential postsynaptic current, static weights
Description	Neuron j input current of synapse formed with presynaptic neuron i: $I_{ji}(t) = I_{ji,\max} \sum_l \exp\left(-(t - t_l^i - d_i)/\tau_{\mathrm{s}}\right) H(t - t_l^i - d_i)$, (5.11) $I_{ji,\max} = -C_{\mathrm{m}}\,\mu_{YX}$ of point-neuron network, (5.12) $H(t) = 1$ for $t \geq 0$, and 0 elsewhere. (5.13) - Static synaptic weights $J_{ji} = \mu_{YX}$ ($j \in Y$, $i \in X$) (see Tab. 5.2) - Delays d_i from Gaussian distribution with mean d_X ($i \in X$), relative standard deviation $\sigma_{d,\mathrm{rel}}$ - Synapse activation times: network spike trains plus delay - No cortico-cortical connections: $I^{\mathrm{ext}} = 0$ (cf. Tab. D.1)

F	**Input**
Type	Spike times t_l^i of spiking neuron network (including thalamic input spikes), no cortico-cortical input
Description	Synapse placement, postsynaptic cell $j \in y$, $y \in Y$ (see Methods): - Number of synapses from presynaptic population X in layer L: k_{yXL} - Compartment specificity of connections: $A_{jn}/\sum_{n \in L} A_{jn}$, comp. $n \in L$ - Synapse locations within layers are chosen randomly among dendritic compartments only

G	**Measurements**
Type	Local field potential (LFP) and current source density (CSD)
Description	Laminar multielectrode, see parameter values in Tab. D.2: - Axis perpendicular to pial surface - n_{contacts}: number of contacts - h_{contacts}: intercontact distance - r_{contact}: contact surface radius

Table 5.4: Description of multicompartment-neuron populations for the cortical microcircuit model (continuation of Tab. 5.3).

the pial surface. The only exception was the p6(L4) morphology, which we extended to reach the center of layer 2/3 in accordance with Stepanyants et al. (2008) and the observation that Tab. D.4 predicts zero connections within layer 1 and very few connections in layer 2/3 to the p6(L4) cell type. Due to lack of available morphologies of sufficient reconstruction quality, certain cell types were represented by the same neuron morphology. Interneuron types and spiny stellate cells in a given layer shared morphologies, the same interneuron morphology was reused in both layer 5 and 6, and finally the p5(L23) and p6(L56) cell morphology were similar except for the stretching of the apical dendrites.

Preserving the laminar cell density under 1 mm^2 surface area of the point-neuron network model, we created for each postsynaptic cell type y model populations where somas were assigned random locations in 3D within cylindrical slabs with radius $r = 564$ μm and thickness $h = 50$ μm, each centered in their respective layer (illustrated in Fig. 5.1D, see also Fig. 6.1E). Regardless of the vertical offset of the soma of pyramidal cells, postsynaptic target dendrites were therefore present within the ~ 80 μm thick (Stepanyants et al., 2008) uppermost layer 1 except for cell type p6(L4). For simplicity, each cell type was represented by a single reconstructed morphology in the present application. The full specification of the populations is given in Tab. 5.3.

Each neuron is modeled using the multicompartmental, passive cable formalism (Rall, 1964, 2009; De Schutter & Van Geit, 2009), describing the changes in membrane voltage and the associated transmembrane currents throughout all parts of the neuron geometry (Tab. 5.4). We used (non-plastic) exponential current-based synapses as in the point-neuron network model (Tab. 5.2). Synapse locations were randomly assigned onto cell compartments assuming a probability proportional to the compartment's surface area divided by the total surface area of the same cell within the target layer. Tab. 5.4-D.2 summarize parameters relevant for the synapse models, synapse locations and passive parameters of the multicompartment models.

5.4 Spatial synaptic connectivity

A full description of the connectivity in networks of multicompartment model neurons requires a 3-dimensional (3D) representation, for example in the form of sparse $N_X \times N_Y \times n_{\text{comp}}$ matrices of synaptic weights and spike-transmission delays between presynaptic neurons $i \in [1, N_X]$ and compartments $n \in [1, n_{\text{comp}}]$ of postsynaptic cells $j \in [1, N_Y]$. Here, N_X and N_Y denote the number of pre- and postsynaptic neurons in population X and Y, respectively, and n_{comp} the number of compartments of the postsynaptic cell. In point-neuron networks, in contrast, connectivity is by definition only 2-dimensional (2D) as the cell morphology is collapsed into a single point and, consequently, the specificity of synapse locations on the postsynaptic morphology is ignored. In the proposed hybrid modeling scheme, the connectivity within the

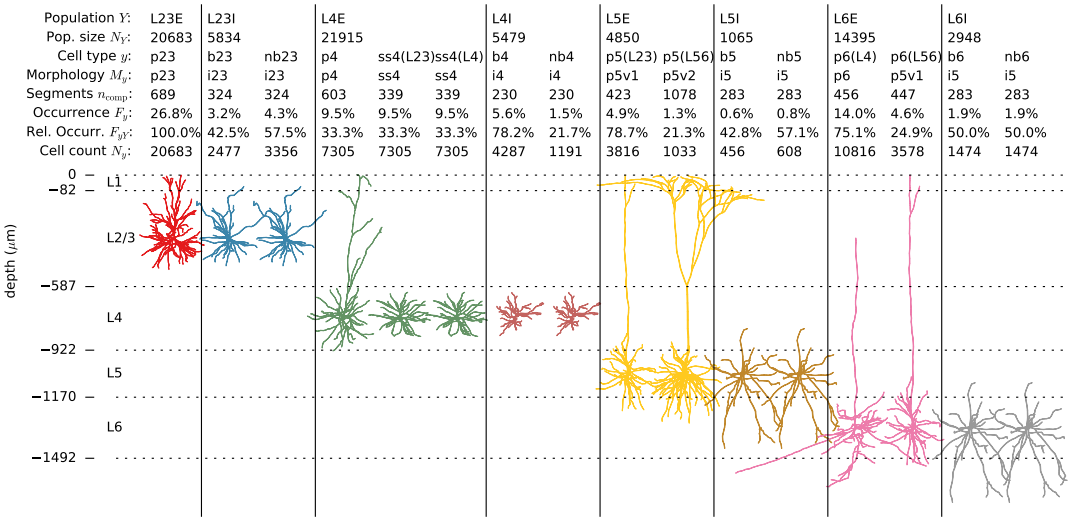


Figure 5.2: **Cell types and morphologies of the multicompartment-neuron populations.** The 8 cortical populations Y of size N_Y in the microcircuit network model are represented by 16 subpopulations of cell type y with detailed morphologies M_y (Binzegger et al., 2004; Izhikevich & Edelman, 2008). Neuron reconstructions are obtained from cat visual cortex and cat somatosensory cortex (source: NeuroMorpho.org (Ascoli et al., 2007), Contreras et al. (1997); Mainen & Sejnowski (1996); Kisvarday & Eysel (1992); Stepanyants et al. (2008), cf. Tab. D.3). Each morphology M_y is here shown in relation to the layer boundaries (horizontal lines). Colors distinguish between network populations as in Fig. 5.1. The number of compartments (n_{comp}), frequencies of occurrence (F_y), relative occurrence (F_{yY}) and cell count (N_y) is given for each cell type $y \in Y$.

point-neuron network is consistent with the connectivity between point neurons and multicompartment model neurons. Ideally, each multicompartment model neuron has its equivalent in the point-neuron network and receives inputs from exactly the same presynaptic sources as its point-neuron counterpart. Synapses should be positioned on the dendritic tree according to anatomical data, and synaptic weights and time constants should be adapted such that the somatic membrane potential or somatic current match the point-neuron counterparts. Such mapping between point neurons and passive multicompartment neurons is feasible (Koch & Poggio, 1985; Wybo et al., 2013, 2015).

In the current application of the hybrid scheme to the cortical microcircuit model, we make the simplest approximation to the mapping problem and fixed the current amplitudes $I_{ji,\text{max}}$ and synaptic time constants as in the network model, with compartment specificity of connections dependent on compartment surface area (see Tab. 5.4). We further preserve only the statistics of connections (average number of inputs, distribution of spike-transmission delays) for each pair of pre- and postsynaptic neuron populations, exploiting that connections between network populations are drawn randomly with fixed probabilities. Finally, we simplify the positioning of

synapses to a layer-specificity of connections. The activation times of each synapse are then given by the spike train of a randomly drawn point neuron in the network model, with random delays consistent with the delay distribution in the network (Tab. 5.2 and 5.4).

We first show how to derive a 2D point-neuron connectivity from a given 3D multicompartment-neuron connectivity and describe the case where the complexity of the point-neuron network is further reduced by pooling cell types. Then we describe the opposite procedure, connecting an existing (published) point-neuron network with a predefined 2D connectivity to a population of multicompartment model neurons such that the resulting 3D connectivity is as consistent as possible with anatomical datasets accounting for the compartment specificity of connections (for example, the layer specificity of connections as in the anatomical data published by Binzegger et al. (2004)). The procedures outlined below, allow a reduction of complexity within the point-neuron network while accounting for the full diversity in cell types and synapse locations for multicompartment-neuron populations which is essential for predicting extracellular potentials (Fig. 5.3).

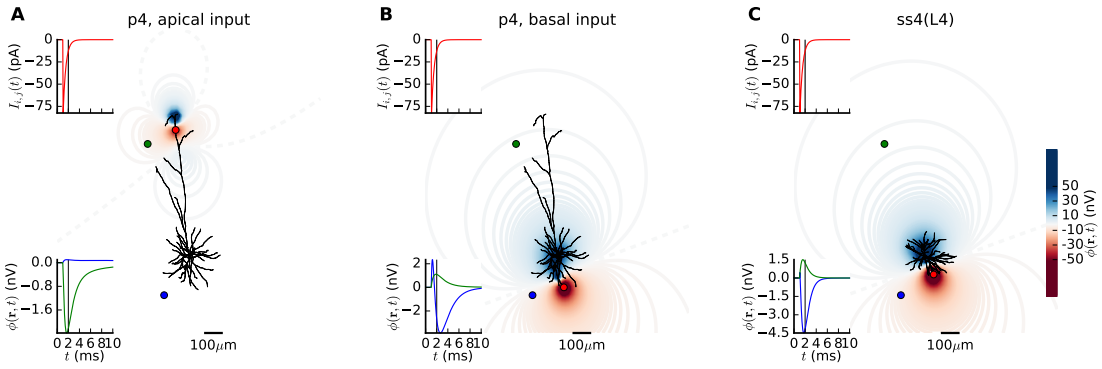


Figure 5.3: **Example LFP responses from single-synapse activations of layer 4 neurons. A** Illustration of the non-trivial relationship between apical synaptic input (red circle) onto a reconstructed morphology (black) of a pyramidal cell in layer 4 and the corresponding extracellular potential. The exponential synaptic input current $I_{i,j}(t)$ (upper inset) results in deflections in the extracellular potential $\phi(\mathbf{r},t)$ here shown as time courses at two locations in proximity to the input site and the basal dendrites (green and blue circle, respectively; lower inset). The color-coded isolines show the magnitude of the scalar extracellular potential at t = 2 ms (vertical black line in insets) in the vicinity of the cell. **B** Same as in panel A, however with the synaptic input current relocated to a basal dendrite, resulting in an extracellular potential with a different spatiotemporal signature less dependent on the geometry of the apical dendritic tree. At the location denoted by the blue circle, the extracellular potential changes sign with time due to interactions between signal propagation in the passive model neuron and volume conduction. **C** Same as panels B and C for a spiny stellate cell in layer 4 receiving an excitatory synaptic input on a basal dendrite.

5.4.1 Construction of point-neuron network connectivity

For our example point-neuron network model, the cortical microcircuit model by Potjans & Diesmann (2014), the connectivity is to a large extent based on anatomical data from cat visual cortex (Binzegger et al., 2004; Izhikevich & Edelman, 2008) (cf. Tab. D.4). From Tab. D.4 we obtain (i) the number N_y of neurons belonging to cell type y, (ii) the average total number k_{yL} of synapses on all compartments in layer L (input layer) of a single postsynaptic neuron of type y, and (3) the fraction p_{yxL} of the k_{yL} synapses formed with presynaptic neurons of cell type x. The quantity

$$k_{yxL} = p_{yxL}\, k_{yL} \tag{5.14}$$

defines the number of synapses between all presynaptic cells of type x and a single postsynaptic cell of type y in input layer L (cf. network connectivity in Izhikevich & Edelman (2008)). The number of synapses between all neurons in x and all neurons in y, irrespective of the input layer L, is given by

$$K_{yx} = N_y \sum_L k_{yxL} \ . \tag{5.15}$$

The number K_{yx} of connections in combination with a chosen connectivity model (e.g., random graphs with binomially distributed (Erdős & Rényi, 1959), fixed in-/out-degree (Newman, 2003) or random graphs with defined higher-order statistics (Song et al., 2005; Zhao et al., 2011)) is sufficient for setting up the point-neuron network. Assuming independently drawn synapses (allowing multiple connections between neurons), the probability C_{yx} of at least one connection between a neuron of type x and a neuron of type y can be obtained from K_{yx} as (Potjans & Diesmann, 2014)

$$C_{yx} = 1 - \left(1 - \frac{1}{N_x N_y}\right)^{K_{yx}} \ . \tag{5.16}$$

In our case, the point-neuron microcircuit model consists of excitatory and inhibitory populations X, Y (see Tab. 5.1-5.2) pooling different pre- and postsynaptic cell types $x \in X$ and $y \in Y$ (cf. Fig. 5.4). Given a single multicompartment model neuron of type y we compute the number k_{yXL} of incoming connections (in-degree) from cell types x in each presynaptic population X in a given layer L by pooling all connections as illustrated in Fig. 5.4A as

$$k_{yXL} = \sum_{x \in X} k_{yxL} \ . \tag{5.17}$$

The total number of connections onto postsynaptic cells y from cells in X is then

$$K_{yXL} = N_y k_{yXL} \ . \tag{5.18}$$

The layer-specific connection probability C_{yXL} (Fig. 5.5B) can be derived from Eq. (5.18) analogously to Eq. (5.16) for a presynaptic population size N_X (here, $N_X = \sum_{x \in X} N_x$). In order to obtain the connectivity within the point-neuron network, i.e., between

populations X, Y, we also need to pool over all synapses of input layers L and cell types y within the postsynaptic population Y (dashed/dotted lines in Fig. 5.4B). Thus

$$K_{YX} = \sum_{y \in Y} K_{yX} = \sum_{y \in Y} \sum_{L} K_{yXL} \,, \tag{5.19}$$

which yields the connectivity of the simplified network structure C_{YX} (cf. Eq. (5.16), Fig. 5.5A).

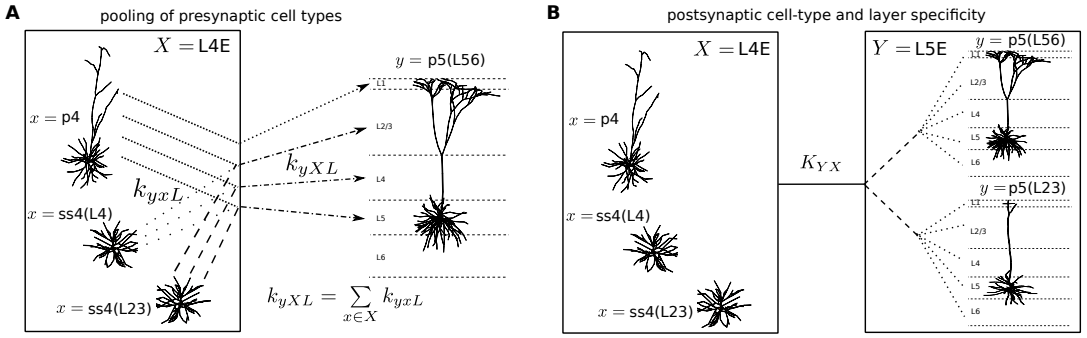


Figure 5.4: **Constructing spatial synaptic connectivity for the cortical microcircuit model. A** Illustration of pooling of presynaptic cell types. Presynaptic populations X in the point-neuron model (left box; here X = L4E) consist of multiple cell types x (here $x \in$ {p4, ss4(L4), ss4(L23)}). The layer-specific number of synapses k_{yXL} (dash-dotted lines) formed between one cell of postsynaptic cell type y (right part of panel A: morphology projected onto cortical layers 1–6; here y = p5(L56)) and a presynaptic population X is given by the sum of all individual cell-type resolved synapse counts k_{yxL} (dotted or dash-dotted lines). **B** Bi-directional cell- and layer-specific pooling and dispersing of synapses between pre- and postsynaptic cell types. Postsynaptic populations Y (right box; here Y = L5E) in the point-neuron model consist of multiple cell types y (here $y \in$ {p5(L56), p5(L23)}). A given presynaptic population X (left box; here X = L4E) containing cell types x (here $x \in$ {p4, ss4(L4), ss4(L23)}) forms cell-type and layer-specific connections within Y (black connection tree). For the number of synapses K_{yXL} between population X and cells of type y in layer L (right-most branching of connection tree) the synapse count K_{YX} between all cells in X and Y can be obtained by pooling all synapses onto cell types $y \in Y$ and input layers L. Conversely, for a given total number of synapses K_{YX} between all cells in X and Y, the number of synapses K_{yXL} onto a specific cell type y and layer L can, as described by Eq. (5.22), be obtained by calculating the cell-type and layer specificity of connections $\mathcal{T}_{yX}$ and $\mathcal{L}_{yXL}$ (see Fig. 5.5) from anatomical data (Tab. D.3).

5.4.2 From pooled to specific network connectivity

In case of an already existing point-neuron network model, the reverse task of creating a spatial connectivity C_{yXL} from a given point-neuron network connectivity C_{YX}

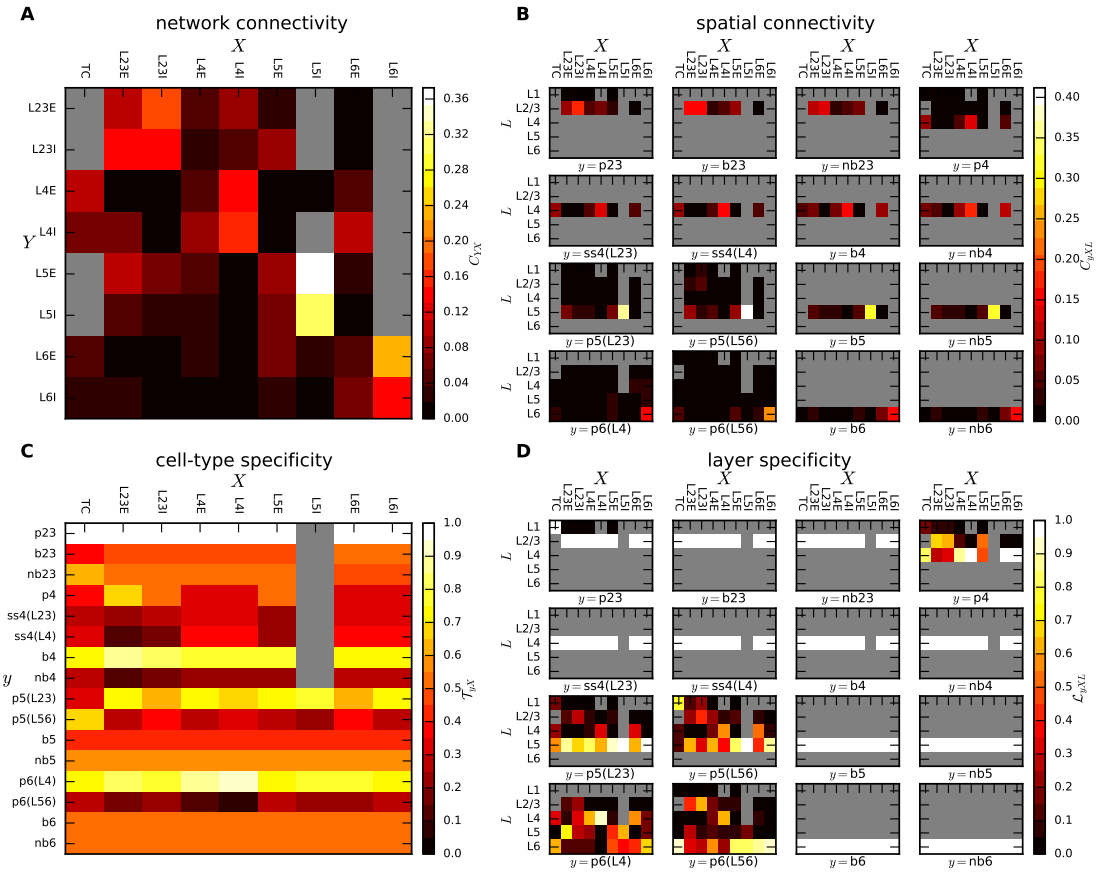


Figure 5.5: **Connectivity of the cortical microcircuit model**. **A** Connection probability C_{YX} between presynaptic population X and postsynaptic population Y of the cortical microcircuit model by Potjans & Diesmann (2014) given in Tab. D.1. Zero values are shown as gray here and in subsequent panels. **B** Layer- and cell-type specific connectivity map C_{yXL}, where X, y and L denote presynaptic populations, postsynaptic cell types and the synapse location (layer), respectively. This map is computed from the connectivity of the point-neuron network (panel A) and cell-type (panel C) and layer specificity (panel D) of connections. **C** Cell-type specificity $\mathcal{T}_{yX}$ of connections quantified as the fraction of synapses between pre- and postsynaptic populations X and Y formed with a specific postsynaptic cell type y. **D** Layer specificity $\mathcal{L}_{yXL}$ of connections denoting the fraction of synapses between population X and cell type y formed in a particular layer L. Both $\mathcal{T}_{yX}$ and $\mathcal{L}_{yXL}$ in panels C and D respectively are calculated from anatomical data (Binzegger et al., 2004; Izhikevich & Edelman, 2008), cf. Tab. D.4.

is necessary. This inverse procedure compared to pooling over cell types and input layers entails introducing the cell-type specificity

$$\mathcal{T}_{yX} = \frac{K_{yX}}{K_{YX}}\,, \tag{5.20}$$

which describes the fraction of synapses between populations X and Y that are formed with a specific postsynaptic cell type y (Fig. 5.5C), and the layer specificity of connections

$$\mathcal{L}_{yXL} = \frac{K_{yXL}}{K_{yX}}, \tag{5.21}$$

denoting the fraction of synapses between population X and all cells of cell type y formed in a particular layer L (Fig. 5.5D). The product $\mathcal{T}_{yX}\mathcal{L}_{yXL}$ defines the probability of a synapse between populations X and Y formed with a specific postsynaptic cell type y in a particular layer L (Fig. 5.4B). Thus, if K_{YX} is given, the total number of connections in layer L onto postsynaptic cells y from cells in X is

$$K_{yXL} = K_{YX}\mathcal{T}_{yX}\mathcal{L}_{yXL}\ . \tag{5.22}$$

If K_{YX} is constructed from the same data as $\mathcal{T}_{yX}$ and $\mathcal{L}_{yXL}$, Eqs. (5.20)-(5.22) are fully consistent. However, K_{YX} can also be computed from any given point-neuron network connectivity C_{YX}. This is particularly relevant for the network connectivity C_{YX} (Fig. 5.5A) of Potjans & Diesmann (2014) that includes additional data sets for which spatial information on synapse locations is not available. Here the number of synapses k_{yXL} from population X established in layer L on each multicompartment model neuron of type y is obtained from Eq. (9) as $k_{yXL} = K_{yXL}/N_y$.

5.5 Forward modeling of extracellular potentials

The LFP signal reflects transmembrane currents weighted according to the distance from the source to the measurement location (Einevoll et al., 2013), and here we compute the LFP from the model neurons using a now well established forward-modeling scheme combining multicompartment neuron modeling and electrostatic (volume-conduction) theory (Holt & Koch, 1999; Gold et al., 2006; Pettersen et al., 2008; Lindén et al., 2010, 2011; Reimann et al., 2013; Lindén et al., 2014; Tomsett et al., 2014).

The cable formalism for multicompartment-neuron models (see e.g., De Schutter & Van Geit (2009)) describes the time course of the membrane potential as well as axial and transmembrane currents across neuronal morphologies as resulting from spike-triggered synaptic input. In volume conduction theory (Rall & Shepherd, 1968; Holt & Koch, 1999; Nunez & Ramesh, 2006), the extracellular medium is the assumed to be ohmic such that the electric field E is linearly related to the current density field $J = \sigma_e E$, with electric conductivity σ_e. In the present application where the signal frequencies are well below 1,000 Hz, we can employ the quasi-static approximation of Maxwell's equations (in particular $\partial B/\partial t = 0$), which states that changes in the magnetic field B do not influence the electric field (Hämäläinen et al., 1993). As a consequence, E is conservative ($\nabla \times E = 0$) and thus defined by a scalar electric potential $E = -\nabla\phi$. Ohm's law can be written in differential form $\nabla \cdot (\sigma_e \nabla\phi) = -I$ to relate

the electric potential ϕ to the current-source density $I = \nabla \cdot J$ given by transmembrane currents. Hereby, one employs Gauss's theorem which relates the outward flux of the current density field through a closed surface, the cell membrane, to the volume integral of the divergence of the current field over the intracellular region. For constant conductivity σ_e, this equation reduces to Poisson's equation which can be solved with the Green's function formalism. Due to linearity of the above equations, the extracellular potential generated by all cells results as a linear superposition of the potentials arising from transmembrane currents of individual cells, or even individual compartments. In the following, we further assume the extracellular medium to be linear, isotropic, homogeneous (Pettersen et al., 2012; Einevoll et al., 2013) and represented by a scalar extracellular conductivity.

For multicomparment neurons, each morphology is spatially discretized into compartments using the `d_lambda` rule (Hines & Carnevale, 2001) with electrotonic length constants computed at $f = 100$ Hz. The cable-equation description is summarized in box D in Tab. 5.4. Eq. (5.10) relates synaptic input currents I_{jin} onto compartment n in a neuron j from presynaptic neurons i, the membrane voltages V_{mjn} and transmembrane currents I_{mjn}, and is derived from the assumption (Kirchoff's law) of current balance in the intracellular node of the equivalent electrical circuit of a cylindrical compartment n with m neighboring compartments. We use the standard convention that a positive membrane current is a positive current from the intracellular to the extracellular space across the membrane. I_{mjn} is assumed to be homogeneously distributed across the outer surface of the cylindrical compartment, and the calculation assumes that the electrical potential on the outside boundary of the membrane is zero at all times. Hence, there are no mutual interactions (i.e., ephaptic coupling (Anastassiou et al., 2011)) between the extracellular potential estimated using volume conduction theory, the transmembrane currents, and intracellular potentials.

Given a time-varying point current source with magnitude $I(t)$ at position $\mathbf{r}'$, the scalar extracellular potential $\phi(\mathbf{r}, t)$ at position $\mathbf{r}$ and time t is then given by (Nunez & Ramesh, 2006; Lindén et al., 2014)

$$\phi(\mathbf{r}, t) = \frac{I(t)}{4\pi\sigma_e|\mathbf{r} - \mathbf{r}'|} \,. \tag{5.23}$$

Contributions to the extracellular potential from multiple current sources, i.e., transmembrane currents of all individual compartments n from all cells j in a population of N cells sum linearly. In accordance with the assumed homogeneous current distribution along each cylindrical compartment, the *line-source* approximation is used for dendritic compartments (Holt & Koch, 1999). The line-source forward-modeling formula is obtained by integrating Eq. (5.23) along the cylindrical axis of each compartment n, and summing the contributions from all n_{comp} compartments (Holt & Koch, 1999; Pettersen et al., 2008; Lindén et al., 2014):

$$\phi(\mathbf{r}, t) = \sum_{j=1}^{N} \sum_{n=1}^{n_{\text{comp}}} \frac{I_{mjn}(t)}{4\pi\sigma_e} \int \frac{1}{|\mathbf{r} - \mathbf{r}_{jn}|} d\mathbf{r}_{jn} \,. \tag{5.24}$$

Presently, we approximate the thick soma compartments as spherical current sources, and thus combine the point-source equation (Eq. (5.23)) with the line-source formula (Eq. (5.24)) for dendrite compartments, obtaining (Lindén et al., 2014):

$$\begin{aligned}\phi(\mathbf{r},t) &= \sum_{j=1}^{N}\frac{1}{4\pi\sigma_{\mathrm{e}}}\left(\frac{I_{\mathrm{m}j,\mathrm{soma}}(t)}{|\mathbf{r}-\mathbf{r}_{j,\mathrm{soma}}|}+\sum_{n=2}^{n_{\mathrm{comp}}}\int\frac{I_{\mathrm{m}jn}(t)}{|\mathbf{r}-\mathbf{r}_{jn}|}d\mathbf{r}_{jn}\right)\\ &= \sum_{j=1}^{N}\frac{1}{4\pi\sigma_{\mathrm{e}}}\left(\frac{I_{\mathrm{m}j,\mathrm{soma}}(t)}{|\mathbf{r}-\mathbf{r}_{j,\mathrm{soma}}|}+\sum_{n=2}^{n_{\mathrm{comp}}}\frac{I_{\mathrm{m}jn}(t)}{\Delta s_{jn}}\ln\left|\frac{\sqrt{h_{jn}^2+r_{\perp jn}^2}-h_{jn}}{\sqrt{l_{jn}^2+r_{\perp jn}^2}-l_{jn}}\right|\right).\end{aligned} \tag{5.25}$$

Here, Δs_{jn} denotes compartment length, $r_{\perp jn}$ the perpendicular distance from the electrode point contact to the axis of the line compartment, h_{jn} the longitudinal distance measured from the start of the compartment, and $l_{jn} = \Delta s_{jn} + h_{jn}$ the longitudinal distance from the other end of the compartment. If the distance between electrode contacts and dendritic current sources becomes smaller than the radius of the dendritic segment, an unphysical singularity in our extracellular potential may occur. In these cases singularities are avoided by setting $|\mathbf{r}-\mathbf{r}_{j,\mathrm{soma}}|$ or $r_{\perp jn}$ equal to the compartment radius. Electrode contacts of real recording devices have finite spatial extent and are not point contacts as assumed above. However, the recorded signal can be well approximated as the mean of the potential averaged across the uninsulated surface (Robinson, 1968; Nelson et al., 2008; Nelson & Pouget, 2010; Ness et al., 2015), at least for current sources positioned further away than an electrode radius or so (Ness et al., 2015). Here we employ the *disc-electrode* approximation to the potential (Camuñas Mesa & Quiroga, 2013; Lindén et al., 2014; Ness et al., 2015):

$$\phi_{\mathrm{disc}}(\mathbf{r},t) = \frac{1}{A_S}\iint_S \phi(\mathbf{r}',t)\,d^2\mathbf{r}' \approx \frac{1}{m}\sum_{h=1}^{m}\phi(\mathbf{r}_h,t)\,. \tag{5.26}$$

We further consider circular electrode contacts with a radius of $r_{\mathrm{contact}} = 7.5$ μm, and we average the point-contact potential in Eq. (5.25) over $m = 50$ random locations $\mathbf{r}_h$ across the contact surface S, A_S being the surface area. The chosen locations are distributed with uniform probability on circular discs representing each contact surface, with surface vectors oriented perpendicular to the electrode axis (Lindén et al., 2014). Calculations of extracellular potentials are facilitated by `LFPy` (http://LFPy.github.io) (Lindén et al., 2014), in which `NEURON` simulation software is used for calculations of transmembrane currents (i.e., solving Eq. (5.10)) (Hines & Carnevale, 2001; Hines et al., 2009; Carnevale & Hines, 2006).

5.6 The `hybridLFPy` `Python` package

To facilitate usage of the hybrid scheme by other users, a novel `Python` software package, `hybridLFPy`, has been made publicly available under the General Public License version 3 (GPLv3, http://www.gnu.org/licenses/gpl-3.0.html) on GitHub (http:

//github.com/INM-6/hybridLFPy). Compatibility with a host of different machine architectures and operating systems (∗nix, OSX, Windows) is ensured with the freely available, object-oriented programming language `Python` (http://www.python.org). `Python` adds tremendous flexibility in terms of interfacing a large number of packages and libraries for, e.g., performing numerical analysis and data visualization, such as `numpy` (http://www.numpy.org) and `matplotlib` (http://www.matplotlib.org), while several other neural simulation softwares also come with their own `Python` interfaces, such as `NEST` (http://www.nest-initiative.org) (Eppler et al., 2015) and `NEURON` (http://www.neuron.yale.edu) (Hines et al., 2009).

The source code release of `hybridLFPy` provides a set of classes implementing the hybrid scheme, as well as example network simulation codes implemented with `NEST` (Eppler et al., 2015) (at present a simplified two-population network (Brunel, 2000) and the full cortical microcircuit model of Potjans & Diesmann (2014) adapted from public codes). The class `hybridLFPy.CachedNetwork` uses an efficient `sqlite3` database implementation for reading in all point-neuron network spike events and interfacing network spike events with the main simulation in which the LFP and CSD are calculated. The class `hybridLFPy.Population` defines populations of multicompartment model neurons representing each cell type, assigns synapse locations across the laminae, selects spike trains for each synapse location from the appropriate presynaptic population, and calculates the LFP and CSD. The single-cell calculations are handled using `LFPy` (http://LFPy.github.io) (Lindén et al., 2014) which builds on `NEURON` (Carnevale & Hines, 2006; Hines et al., 2009). As there are no mutual interactions between the multicompartment model neurons in the calculation of LFPs, these calculations remain embarrassingly parallel operations. Finally, the class `hybridLFPy.PostProcess` constructs the full compound signals in terms of LFPs and CSDs created by multiple instances of the `Population` class, and performs the main analysis steps as described above. Reproducible simulation and data analysis are assured by tracking code revisions using `git` and by fixing random number generation seeds and the number of parallel processes. Further documentation and information on installing and using `hybridLFPy` is provided online, see http://github.com/INM-6/hybridLFPy.

The present implementation of `hybridLFPy` and corresponding simulations were made possible by `Open MPI` (v.1.6.2), `HDF5` (v.1.8.13), `sqlite3` (v.3.6.20), as well as `Python` (v.2.7.3) with modules `Cython` (v.0.23dev), `NeuroTools` (v.0.2.0dev), `h5py` (v.2.5.0a0), `SpikeSort` (v.0.13), `ipython` (v.0.13), `matplotlib` (v.1.5.x), `mpi4py` (v.1.3), `numpy` (v.1.10.0.dev-c63e1f4), `pysqlite` (v.2.6.3) and `scipy` (v.0.17.0.dev0-357a1a0). Point-neuron network simulations were performed using `NEST` (v.2.8.0 ff71a29), and simulations of multicompartment neurons using `NEURON` (v.7.4 1186:541994f8f27f) through `LFPy` (dev. v.3761c4). All software was compiled using `GCC` (v.4.4.6). Simulations were performed in parallel (256 threads) on the Stallo high-performance computing facilities (NOTUR, the Norwegian Metacenter for Computational Science) consisting of 2.6 GHz Intel E5-2670 CPUs running the Rocks Cluster Distribution (Linux) operating system (v.6.0).

5.7 Conclusion

We have here described a hybrid modeling scheme for computing the local field potential (LFP) incorporating both large-scale neural network dynamics and the biophysics underlying LFP generation on the single-neuron level. The hybrid modeling scheme was illustrated with a full-scale network model of a cortical column in early sensory cortex (Potjans & Diesmann, 2014).

The hybrid scheme combines the simplicity and efficiency of point-neuron network models with the biophysics-based modeling of LFP by means of multicompartment model neurons with detailed dendritic morphologies. The neuronal network dynamics are governed by the point-neuron network model independent of LFP predictions. The spikes of the point-neuron network are distributed to the synapses of the multicompartment model neurons with realistic cell-type and layer-specific connectivity. Synapse activation results in spatially distributed transmembrane currents, which are mapped to an LFP signal according to well-established volume-conduction theory.

Chapter 6

Local field potential of a cortical column

This chapter is based on the following publication:

Hagen E*, Dahmen D*, Stavrinou ML, Lindén H, Tetzlaff T, van Albada SJ, Grün S, Diesmann M and Einevoll GT (2016) *Hybrid Scheme for Modeling Local Field Potentials from Point-Neuron Networks,* Cerebral Cortex, bhw237.
*These authors contributed equally to this publication.
Preliminary results have been published in Hagen et al. (2013); Dahmen et al. (2013); Stavrinou et al. (2014); Dahmen et al. (2014); Hagen et al. (2015a,b).

Author contributions:
Under the supervision of T Tetzlaff, S Grün and M Diesmann, the author contributed to all parts of the above publication. Considering the equal contribution of the author and E Hagen, the author focused more on the point-neuron-network component and the analysis of the example model of a cortical microcircuit, while E Hagen focused more on the multicompartment-model component and the development of the software package. SJ van Albada and M Stavrinou helped with the derivation of the spatial connectivity profiles. All authors contributed to the design of the numerical experiments, the data analysis and the writing of the manuscript.

6.1 Introduction

To illustrate the application of the hybrid scheme for predictions of LFPs from point-neuron networks (Chap. 5), we here present results for a modified version of the point-neuron network model of Potjans & Diesmann (2014) with ~ 78,000 neurons mimicking a 1 mm^2 patch of cat primary visual cortex (see Sec. 5.2). The microcircuit model has realistic cell density and deliberately neglects many biological details on the single-cell level, focusing on the effect of the connectivity on the local network dynamics in such circuits. Despite this simplicity, the model displayed firing rates across populations in agreement with experimental observations (Potjans & Diesmann (2014)), as well as propagation of spiking activity across layers. Likewise, the microcircuit model in conjunction with simplified, passive multicompartment populations is used here to study the effect of the (spatial) connectivity on the laminar pattern of spontaneous and stimulus-evoked CSD and LFP signals. For this, CSD and LFP signals for a laminar multielectrode recording at different cortical depths are computed. The example illustrates the effect of network dynamics on the LFP and the interpretation of the LFP in terms of underlying laminar neuron populations. The large network size of the model is further used to explain the effect of correlations and neuron density on CSD and LFP predictions. To this end, we show that correct LFP predictions can only be obtained by accounting for realistic neuron densities.

6.2 Model measurements

The main simulation output of the hybrid scheme consists of spike trains of each neuron in the point-neuron network, 'ground-truth' current-source density (CSD) and local field potentials (LFP) of each neuron in the morphologically detailed post-synaptic model populations (see Tab. 5.3 and 5.4). Here the term 'ground-truth' refers to the fact that the CSD is computed from transmembrane currents rather than estimated from the LFP.

As the transmembrane current of each compartment (Eq. (5.10)) is known at each simulation time step, we follow the procedure of Pettersen et al. (2008) to compute the 'ground-truth' CSD in addition to the LFP. From N model neurons with n_{comp} compartments having membrane currents $I_{\text{m}jn}(t)$ and lengths Δs_{jn}, we calculate the CSD signals $\rho(\mathbf{r},t)$ inside cylinder elements $V_\rho(\mathbf{r})$ centered around each electrode contact as:

$$\rho(\mathbf{r},t) = \frac{1}{\pi r^2 h_{\text{elec}}} \sum_{j=1}^{N} \sum_{n=1}^{n_{\text{comp}}} I_{\text{m}jn}(t) \frac{\Delta s_{jn,\text{inside}}(\mathbf{r})}{\Delta s_{jn}} \,. \tag{6.1}$$

$\Delta s_{jn,\text{inside}}(\mathbf{r})$ denotes the length of the line source contained within $V_\rho(\mathbf{r})$. In contrast to Pettersen et al. (2008), we do not apply a spatial filter to the CSD. The volumes have

radii equal to the population radius r and heights equal to the electrode separation h_{elec} (cf. Tab. D.2).

In the present example application, the extracellular potential is computed at locations corresponding to a laminar multi-electrode array with 16 recording electrodes with an inter-electrode distance of $h_{\text{elec}} = 100$ μm, positioned at the cylindrical axis of the model column with the topmost contact at the pial surface (cf. Fig. 5.1D, see Tab. D.2 for details). Each electrode contact is set to have a radius of 7.5 μm (cf. Eq. (5.26)). In the network we also record membrane voltages and input currents from a subset of cells in each of the eight cortical populations (see Tab. D.5).

We ran our simulations for a total duration of $T = 5{,}200$ ms using a temporal resolution of $dt = 0.1$ ms (cf. Tab. D.1 and D.7). However, LFP and CSD signals were resampled prior to file storage to a temporal resolution of $dt_\psi = 1$ ms by (i) applying a 4th-order Chebyshev type I low-pass filter with critical frequency $f_c = 400$ Hz and 0.05 dB ripple in the passband using a forward-backward linear filter operation, and (ii) then selecting every 10th time sample.

6.3 Post-processing and data analysis

As the contributions to the CSD and LFP of the different cells sum linearly (cf. Eq. (5.25)), we compute population-resolved signals as the sum over contributions from all cells in a population, and the full compound signals as the sum over all population signals (see Tab. D.5). In Sec. 6.8 we derive a rescaled 'low-density predictor' $\phi^{\gamma\xi}(\mathbf{r}, t)$ of the LFP from random subsets of neurons in all populations. Thereby, we make a downscaled LFP-generating model setup with the same column volume, but with neuron density reduced to a factor $\gamma \in (0, 1)$ of the original density, while preserving the in-degrees, i.e., the number of synaptic connections onto individual neurons. The LFP from the downscaled setup is multiplied by an overall scaling factor ξ chosen to roughly preserve the LFP from the full-scale model. For analysis and plotting, the initial 200 ms of results after simulation onset was removed, and the signal mean was subtracted from LFP and CSD traces emulating DC filtering during experimental data acquisition. The contribution from each population to the overall LFP signal was assessed by computing and comparing the signal variances (see Tab. D.6).

Cross-correlations between single-cell LFPs $\phi_i(\mathbf{r}, t)$ were quantified by analyzing the power spectrum of the compound extracellular signals. Given that the compound LFP/CSD is a linear superposition of single-cell LFP/CSD contributions, the power spectrum of the compound signal $P_\phi(\mathbf{r}, f)$ can be obtained as the sum of all single-cell power spectra $P_{\phi_i}(\mathbf{r}, f)$ and all pairwise cross-spectra $C_{\phi_i\phi_j}(\mathbf{r}, f)$, or equivalently from the average single-cell power spectrum $\overline{P_\phi}(\mathbf{r}, f)$, the average pairwise cross-spectrum $\overline{C_\phi}(\mathbf{r}, f)$, and the total cell count N, as shown in Tab. D.6. Note that $\overline{C_\phi}(\mathbf{r}, f)$ and

hence the average pairwise single-cell LFP coherence $\overline{\kappa_\phi}(\mathbf{r}, f)$ (Eq. (D.5)) are real, while $C_{\phi_i\phi_j}(\mathbf{r}, f)$ is complex. Note that this definition allows for negative values of $\overline{\kappa_\phi}(\mathbf{r}, f)$. In the sum over all i and j in Eq. (D.4), the imaginary parts of $C_{\phi_i\phi_j}(\mathbf{r}, f)$ and $C_{\phi_j\phi_i}(\mathbf{r}, f)$ cancel because $C_{\phi_i\phi_j}(\mathbf{r}, f) = C_{\phi_j\phi_i}(\mathbf{r}, f)^*$ (* denotes the complex conjugate). The power spectrum $P_{\phi^{\gamma\xi}}(\mathbf{r}, f)$ of the compound signals of the 'downscaled' network (i.e., low-density LFP predictor, see above) is given by the reduced cell count γN, the average single-cell power spectrum $\overline{P_{\phi^{\gamma\xi}}}(\mathbf{r}, f)$, and the average pairwise single-cell cross-spectrum $\overline{C_{\phi^{\gamma\xi}}}(\mathbf{r}, f)$ calculated from that subset of neurons (see Tab. D.6). These single-cell averages are the same as the respective single-cell averages of the full-scale model setup, apart from variability due to subsampling.

Throughout this chapter, signal power spectra are estimated using Welch's average periodogram method (Welch, 1967) (with the `matplotlib.mlab.psd` implementation in `Python`, see Tab. D.7). Temporal cross-correlations are quantified as the zero time-lag correlation coefficient (Eq. (D.2) in Tab. D.6). Distributions of chance correlation coefficients were determined using surrogate data that preserved the overall spectra. We generated this data by transforming one input vector to the Fourier domain, assigned a random phase $\in [0, 2\pi)$ to each frequency $f > 0$, and transformed the data back to the time domain prior to computing the correlation coefficients (Schreiber & Schmitz, 2000). We show 1% significance levels obtained by computing the 99th percentile of chance correlation coefficients across 1000 trials. Spike-trigger-averaged LFP (stLFP) signals are computed as the cross-covariance between the time-resolved population spike rate $\nu_X(t)$ and the compound LFP $\phi(\mathbf{r}, t)$, divided by the total number of spikes (i.e., $\int_0^T \nu_X(t)dt$).

6.4 Spontaneous vs. stimulus-evoked LFP

We first consider the LFP generated by spontaneous network activity. The output of our hybrid scheme covers various scales and measurement modalities, from spikes of each neuron (Fig. 6.1A), population-averaged firing rates (Fig. 6.1B), excitatory, inhibitory, and total synaptic input currents (Fig. 6.1C), and membrane voltages (Fig. 6.1D), to the compound CSD and LFP stemming from all populations of different cell types (Fig. 6.1E-G). For spontaneous activity, i.e., no modulated thalamic input (cf. Tab. 5.1), we observe asynchronous irregular spiking in all populations (Fig. 6.1A) and firing rates similar to the original model (Potjans & Diesmann, 2014) (Fig. 6.1B). In particular, layer 2/3 exhibits low firing rates, and generally inhibitory neurons fire with higher rates than excitatory neurons of the respective layers. The network is in a balanced regime (Brunel, 2000), reflected by the substantial cancellation of population-averaged excitatory and inhibitory input currents (Fig. 6.1C). The population-averaged membrane potential fluctuates below the fixed firing threshold $\theta = -50$ mV down to ~ -80 mV (Fig. 6.1D).

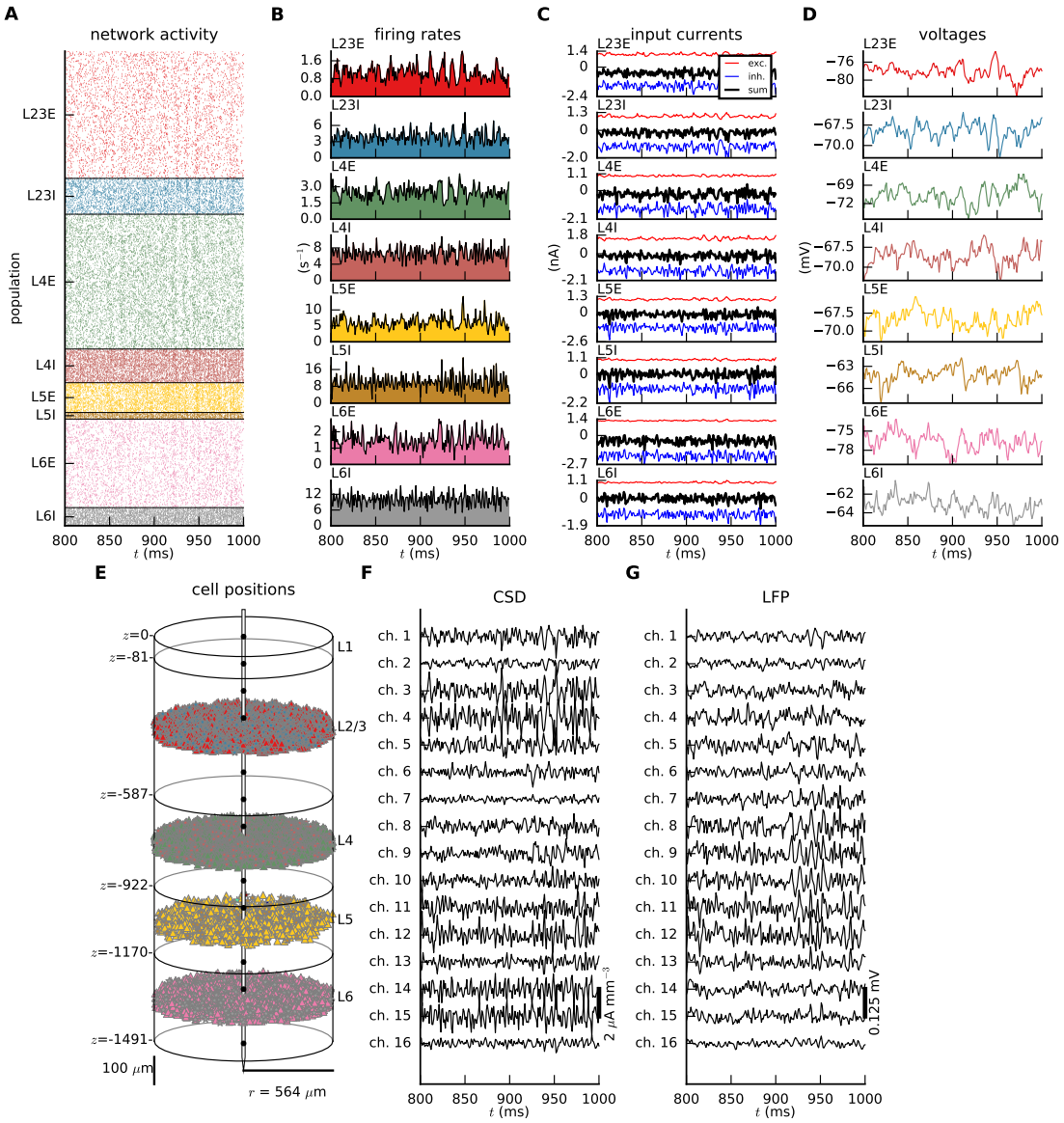


Figure 6.1: **Overview of output signals obtained from application of the hybrid scheme to a cortical microcircuit (spontaneous activity).** *Point-neuron network:* **A** Spiking activity. Each dot represents the spike time of a point neuron (color coding as in Fig. 5.1). **B** Population-averaged firing rates for each population. **C** Population-averaged somatic input currents (red: excitatory, blue: inhibitory, black lines: total). **D** Population-averaged somatic voltages. Averaged somatic input currents and voltages are obtained from 100 neurons in each population. *Multicompartment model neurons:* **E** Somas of excitatory (triangles) and inhibitory (stars) multicompartment cells and layer boundaries (gray/black ellipses). Illustration of a laminar electrode (gray) with 16 recording channels (black circles). **F** Depth-resolved current-source density (CSD) obtained from summed transmembrane currents in cylindrical volumes centered at each contact. **G** Depth-resolved local field potential (LFP) calculated at each electrode contact from transmembrane currents of all neurons in the column. Channel 1 is at pial surface, channel 2 at 100 μm depth, etc.

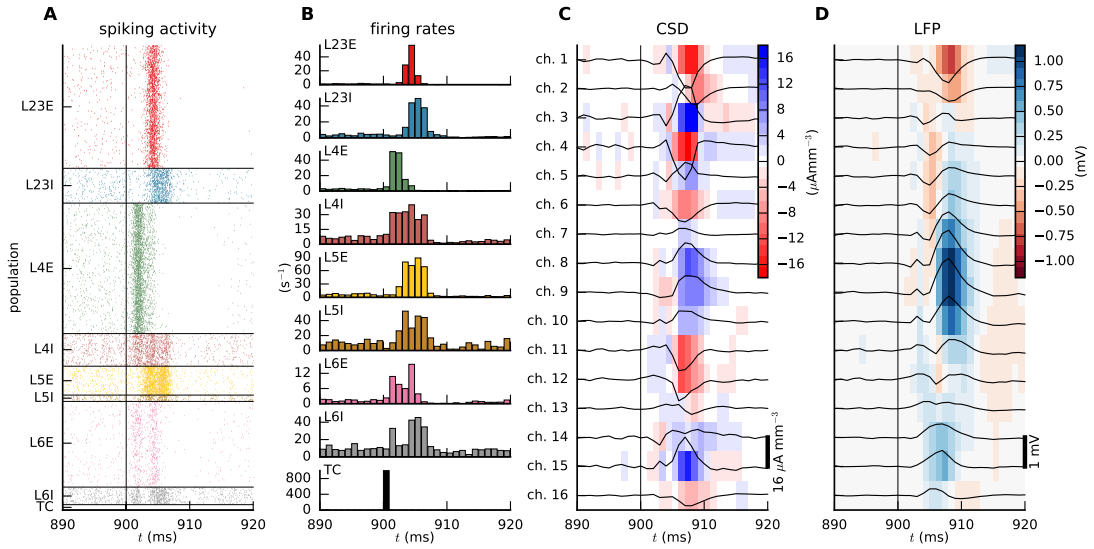


Figure 6.2: **Network activity following transient activation of thalamocortical afferents**. **A** Raster plot of spiking activity before and after δ-shaped thalamic stimulus presented at t = 900 ms (vertical black line in panels A, C and D). **B** Population-averaged firing rate histogram for each population (color coding as in panel A). **C** Depth-resolved compound current-source density (CSD) of all populations (shown both in color and by the black traces). **D** Depth-resolved compound local field potential (LFP, shown both in color and by the black traces) at each electrode channel as generated by all populations. Channel 1 is at pial surface, channel 2 at 100 μm depth, etc.

The corresponding compound CSD and LFP signals with contributions from all cortical populations are shown in Fig. 6.1F and G, respectively. As expected for spontaneous cortical network activity, the LFP signal amplitudes are small, $\simeq$ 0.1 mV (intriguingly close to what has been seen in experiments, see, e.g., Fig. 1 in Maier et al. (2010) for macaque visual cortex, Fig. 7 in Hagen et al. (2015c) for mouse visual cortex), and exhibit strong cross-channel covariance and coherence in line with experimental observations, e.g., Einevoll et al. (2007); Maier et al. (2010); Riehle et al. (2013); Hagen et al. (2015c). In the model these correlations stem from dendritic cable properties and volume conduction effects (Pettersen et al., 2008; Lindén et al., 2011; Łęski et al., 2013). The correlations across channels are, as expected, generally less visible in the more localized CSD signal where volume conduction effects are absent (Nicholson & Freeman, 1975; Pettersen et al., 2006, 2008).

LFP and associated CSD studies have commonly been used to investigate stimulus-evoked responses in sensory cortices, see, e.g., Mitzdorf & Singer (1979); Mitzdorf (1985); Di et al. (1990); Swadlow et al. (2002); Einevoll et al. (2007); Sakata & Harris (2009); Szymanski et al. (2009); Maier et al. (2010); Jin et al. (2011); Szymanski et al. (2011), as well as Einevoll et al. (2013) and references therein. To model the situation with a sharp onset of a visual stimulus (or direct electrical stimulation of

the thalamocortical pathway (Mitzdorf & Singer, 1979)) we drive the network with a short thalamic pulse mimicking a volley of incoming spikes onto primary visual cortex from the visual thalamus (lateral geniculate nucleus, LGN). The activation targets populations in layers 4 and 6 (see Fig. 5.1A or Fig. 5.5A,B) and propagates in the network to populations in layers 2/3 and 5 (Fig. 6.2A,B). At the level of spiking activity, the results match the behavior of the original model (Potjans & Diesmann, 2014, Fig. 10A,B) and even agree qualitatively with experimental findings in rodents from stimulus-evoked activation of auditory cortex (Sakata & Harris, 2009) and somatosensory cortex (Armstrong-James et al., 1992; Einevoll et al., 2007; Reyes-Puerta et al., 2014). This points to a general biological plausibility of this generic network model, based largely on data from cat V1.

The corresponding CSD and LFP profiles across depth associated with this spiking activity are determined by the synapse locations and dendritic filtering by cell-type specific morphologies (see Fig. 5.3, Fig. 5.5D). For the CSD (Fig. 6.2C) a complex alternating spatiotemporal pattern of current sinks (negative CSD) and sources (positive CSD) is observed. Due to volume conduction this detailed spatial pattern is largely smeared out in the LFP profile (Fig. 6.2D), which displays a strong positivity across the middle layer around $t = 910$ ms. We also note the different spatiotemporal profiles of the spiking activity (Fig. 6.2A,B) compared to laminar CSD and LFP profiles. Not only do the CSD and LFP signals typically fade out 5–10 ms later than the spiking, the spatial profiles are also very different. For example, the LFP signal is very weak between channels 11 and 12 (i.e., between $1,100$ and $1,200$ μm depth), even if the firing rate of layer 5 positioned between these channels is very high. We also note that the predicted LFP magnitudes are an order of magnitude larger compared to the LFP predicted for spontaneous activity, that is, ~ 1 mV for stimulus-evoked versus ~ 0.1 mV for spontaneous activity.

Although the present example has not been tuned to address specific experiments, we nevertheless observe that the model predictions display several features seen in experiments. For example, stimulus-evoked LFP amplitudes on the order of 1 mV are similar to the maximal amplitudes ($\sim 1-3$ mV) observed in cat V1 following electric stimulation of thalamocortical axons (optical radiation) (Mitzdorf, 1985, Fig. 3A) and ventroposterior lateral (VPL) pathway axons projecting to rat somatosensory cortex (Castro-Alamancos & Connors, 1996, Fig. 3), and in rat somatosensory (barrel) cortex following whisker flicks (Di et al. (1990, Fig. 1), Einevoll et al. (2007, Fig. 1A), Reyes-Puerta et al. (2014, Fig. 1D-F)). Also some qualitative features of the spatiotemporal LFP and CSD patterns from the cat visual cortex experiments of Mitzdorf (1985, Fig. 3A,B) can be recognized. One example is the early CSD sink around layer 4 (i.e., at channel 6 close to the boundary between layers 2/3 and 4), another is the large positivity in LFP extending through most of the layers following the initial response to the thalamic input volley.

The time courses of the LFP and CSD in our simulations with δ-pulse thalamic activations are not as directly comparable with experimental stimulus-evoked activity,

where the input is temporally filtered by several cell populations before reaching thalamus on the way to cortex. However, we note that very swift stimulus-evoked responses lasting not much longer than the $\sim$ 10 ms response volleys seen in our simulations also are observed in the somatosensory system (Di et al., 1990; Einevoll et al., 2007; Reyes-Puerta et al., 2014).

6.5 Effect of network dynamics on LFP

The spiking activity in the microcircuit model is highly sensitive to modification of intrinsic model parameters and external input (Bos et al., 2015). In general LFPs reflect synaptic input both from local and distant neurons (Herreras et al., 2015), and also depend on network state (see, e.g., Kelly et al., 2010; Gawne, 2010). With the hybrid scheme, we first illustrate as an example the dependence of the spontaneous LFP, i.e., the LFP without thalamic input, on local network dynamics as determined by intrinsic network parameters. In particular we compare our reference network model with the original model proposed by Potjans & Diesmann (2014) where the strength of connections from L4I to L4E neurons is weaker and the synaptic weights are drawn from a Gaussian distribution. We next investigate the LFP when the network is stimulated by sinusoidally modulated thalamic input.

For spontaneous activity (Fig. 6.3A-E), the spiking (Fig. 6.3A) is asynchronous irregular in all populations. The firing-rate power spectra (Fig. 6.3B) vary from relatively flat to more band-pass-like with a maximum power around 80 Hz. The suppressed power at lower frequencies arises from active decorrelation due to inhibitory feedback (Tetzlaff et al., 2012). In combination with the low-pass filtering involved in the generation of LFPs from spiking activity (Lindén et al., 2010; Łęski et al., 2013), the firing-rate spectra translate into an LFP power spectrum that, depending on recording depth, has either low-pass or band-pass filter characteristics (Fig. 6.3D). The notably sharper attenuation of the LFP power spectra compared to the firing-rate power spectra above $\gtrsim$ 100 Hz is expectedly due to the intrinsic dendritic filtering effect (Lindén et al., 2010). As the effect of this filtering depends on the position of the electrode compared to the neuronal morphology (Lindén et al., 2010), the result is a variable LFP power spectral density (PSD) profile across cortical depths, even in the absence of any structured external input (Fig. 6.3E).

There is experimental evidence that cortical microcircuits receive oscillating input at various frequencies from remote areas and subcortical structures (Bastos et al., 2015; van Kerkoerle et al., 2014; Ito et al., 2014). To illustrate the effect of an oscillatory external input in our model, we model thalamic input as independent realizations of non-stationary Poisson processes with a sinusoidal rate profile (see Tab. D.1, Fig. 6.3F-J). The spiking activity of all populations as well as the CSD and LFP across depth are sensitive to thalamic input, showing that the network response goes beyond the layers receiving thalamic input, i.e., layers 4 and 6. The stimulus-evoked

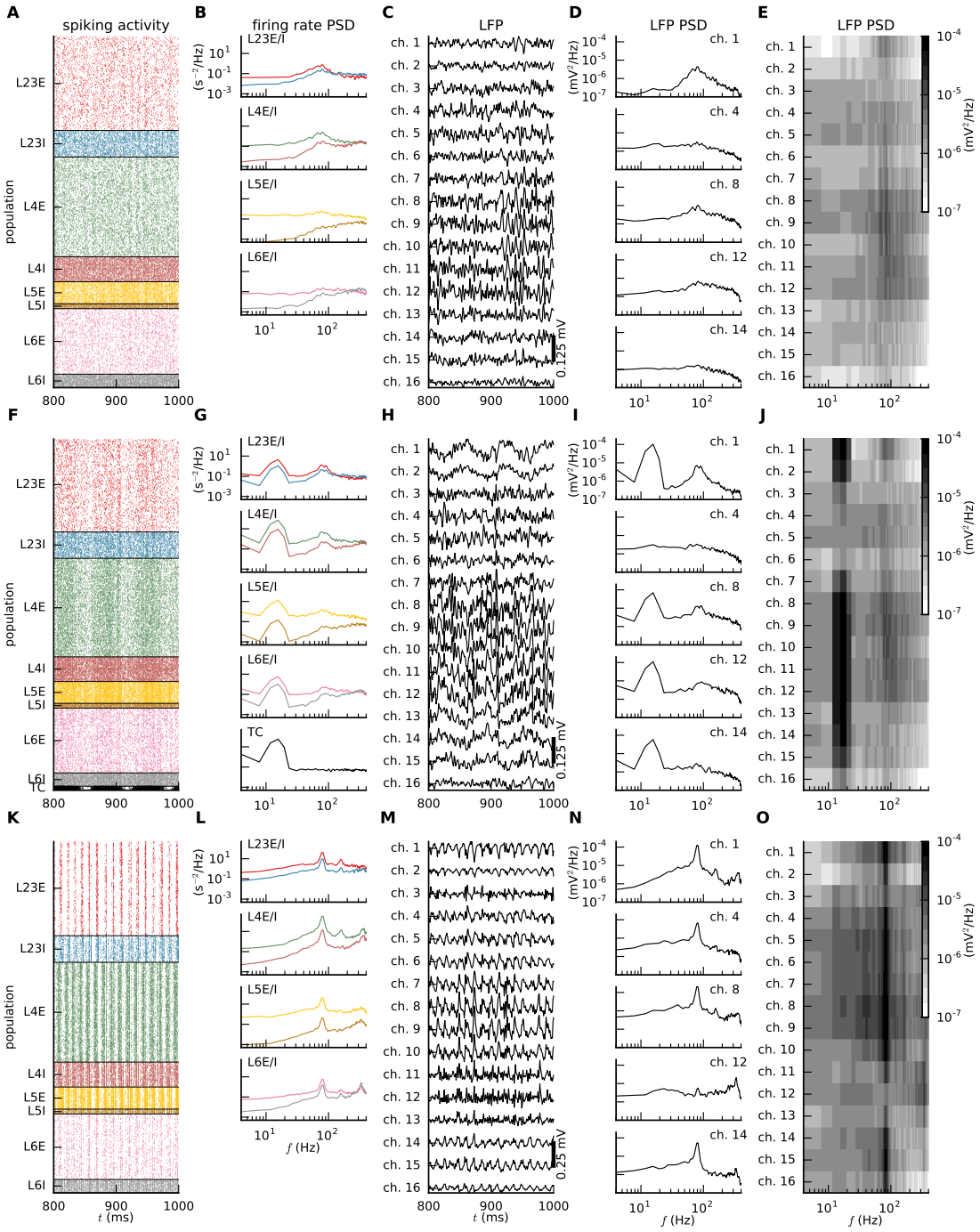


Figure 6.3: **Effect of network dynamics on LFP.** Comparison of two different thalamic input scenarios and two different networks. **Top:** Reference network, spontaneous activity. **Center:** Reference network, oscillatory thalamic activation. **Bottom:** Original model by Potjans & Diesmann (2014), spontaneous activity. **A,F,K** Population-resolved spiking activity. **B,G,L** Population-averaged firing rate spectra. **C,H,M** Depth-resolved LFP. **D,I,N** LFP power spectra in layer 1 and at typical somatic depths of network populations. **E,J,O** LFP power spectra across all channels. Channel 1 is at pial surface, channel 2 at 100 μm depth, etc.

spiking activity follows the 15 Hz modulated rate of the thalamic input (Fig. 6.3F). The 15 Hz oscillation is reflected in the firing-rate spectrum as a peak around 15 Hz (Fig. 6.3G) and is robustly transferred to the LFP (Fig. 6.3H-J). The LFP oscillation strength varies with depth and is greatest in channels 1–2 and channels 8–14, while the oscillation is barely seen in channels 3–6. Populations L4E/I and L6E/I receive thalamic inputs around the depths of channels 8–10 and channels 14–15, and these channels are strongly affected by the stimulus. However, recurrent connections between populations, dendritic propagation of currents, and volume conduction produce strong LFP oscillations also in other channels.

The LFP amplitudes are not only influenced by the temporal structure of the external input, but also by synaptic weights in two ways: first via their influence on the spiking dynamics in the point-network simulation and second via the influence on the size of the synaptic currents setting up the transmembrane currents in the multicompartment models in the LFP-computing step. In order to illustrate the weight dependence, we compare the LFP under spontaneous activity in our reference network model (Fig. 6.3A-E) with the corresponding spontaneous LFP in the original model by Potjans & Diesmann (2014) (Fig. 6.3K-O). The lower inhibition from population L4I onto L4E and the narrow weight distribution compared to our model gives a higher degree of spike synchrony in all populations. Although our model exhibits asynchronous irregular activity (Fig. 6.3A), the original model is closer to a synchronous irregular regime (Brunel, 2000) (Fig. 6.3K), resulting in high-frequency oscillations around 80 Hz and in the 300 – 400 Hz band (Fig. 6.3L), both associated with delay loops in the multi-layered network (Bos et al., 2015). The 80 Hz oscillation also appears in the LFP and its corresponding power spectra (Fig. 6.3M-O), but the magnitude of the peak is not constant across depth. The lowest magnitudes are located in the vicinity of layers 2/3 and 5 (in channels 3–4 and channels 11–13, compare also Fig. 4).

6.6 Contributions from individual populations to CSD and LFP

The direct interpretation of CSD and LFP signals in terms of the underlying activity of different populations or input pathways is inherently ambiguous and thus difficult: For example, a CSD sink observed in cortical layer 2/3 can alternatively stem from excitatory synaptic inputs to the basal dendrites of layer 2/3 cells, similar inputs into the apical dendrites of layer 5 cells, or even return currents from appropriately placed inhibitory inputs onto the same cells (Lindén et al., 2010; Einevoll et al., 2013). Several schemes for decomposition of CSD and LFP data into contributions from cortical populations have thus been proposed: principal component analysis (PCA) (Di et al., 1990), laminar population analysis (Einevoll et al., 2007), and independent component analysis (ICA) (Łęski et al., 2010; Makarov et al., 2010; Głąbska et al., 2014;

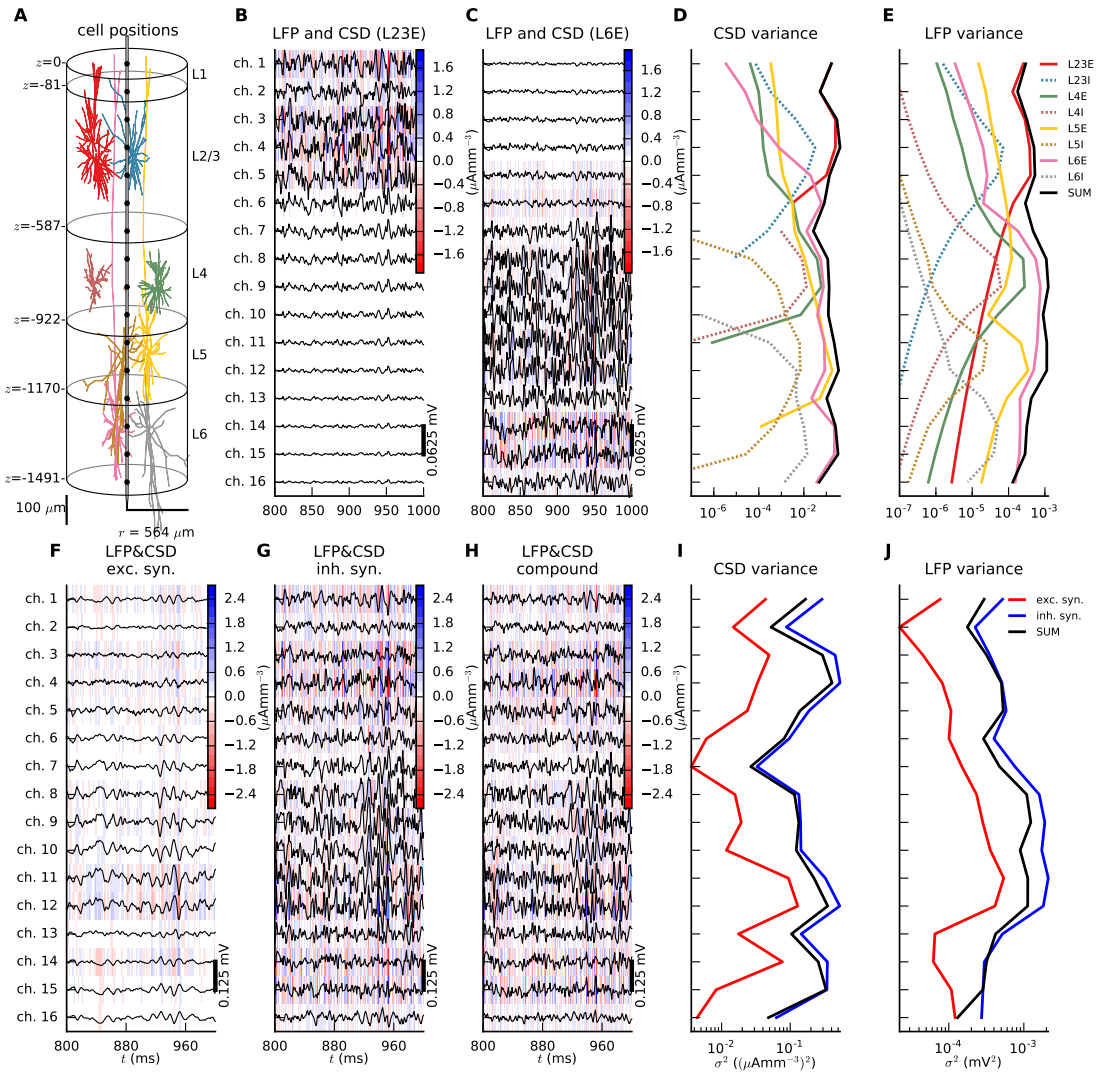


Figure 6.4: **Composition of CSD and LFP during spontaneous activity.** **A** Representative morphologies of each population Y illustrating dendritic extent. **B** LFP (black traces) and CSD (color plot) produced by the superficial population L23E for spontaneous activity in the reference network. **C** Similar to panel B for population L6E (summing over contributions of $y \in$ {p6(L4), p6(L56)}). **D** CSD variance as function of depth for each individual subpopulation (colored lines) and for the full compound signal (thick black line). **E** Same as in panel D, but for LFPs. Variances $\sigma^2 < 10^{-7}$ mV^2 not shown. **F** Compound LFP (black traces) and CSD (color plot) resulting from only excitatory input to the LFP-generating multicompartment model neurons. **G** Conversely, LFP (black traces) and CSD (color plot) resulting from only inhibitory input to the neurons. **H** Full compound LFP (black traces) and CSD (color plot) resulting from both excitatory and inhibitory synaptic currents. **I** Compound CSD variance as a function of depth with all synapses intact (thick black line), or having only excitatory (red) or inhibitory synapse input (blue). **J** Same as in panel I, but for the LFP signal.

Herreras et al., 2015). In our model we have the benefit of having the contributions from the various populations, connections or different synapse types to the CSD and LFP signals directly accessible.

Here, we focus on the LFP and CSD contributions from individual populations and different synapse types. To quantify the contributions from each postsynaptic population, i.e., the CSD and LFP stemming from the transmembrane currents of a population, we simply summed all single-cell CSD and LFP contributions from all neurons in the population. Results for spontaneous network activity are shown in Fig. 6.4. As different cell types are assigned appropriate morphologies and cortical depths based on available anatomical data, a trivial consequence is that the neurons in each population make their main contributions to the CSD and LFP at depths spanned by their dendrites (Fig. 6.4B,C for post-synaptic populations L23E and L6E, respectively). For example, L6E neurons have a high density of afferent synapses on apical dendrites in layer 4, and as a consequence, the amplitude of CSDs and LFPs generated by population L6E (Fig. 6.4C) is large in the vicinity of layer 4 and not just in layer 6.

To quantify the relative contributions from the various populations we show in Fig. 6.4D,E the CSD and LFP variances across time (corresponding to power spectral densities summed over all non-zero frequencies) for all depths (Lindén et al., 2011; Łęski et al., 2013). For the compound CSD, only the L23E neurons contribute substantially in the superficial channels (ch. 1–4) (Fig. 6.4D). At deeper contacts the main contributing population is L6E. L4E and L5E populations make sizable contributions only in the channels closest to their somatic location reflecting that the net associated return currents of their distributed synaptic inputs are largely restricted to somatic regions (Lindén et al., 2010). The bulk of the variance of the CSD in layer 5 arises in about equal parts from the L6E and relatively sparse L5E populations. The magnitude of the depth-resolved CSD variances of each layer's inhibitory population (L23I, L4I, ...) is consistently one order of magnitude or more smaller than that of the corresponding excitatory cell populations. Moreover, they span comparatively small depth ranges as determined by the maximum extent of the dendrites.

The depth-resolved LFP variance (Fig. 6.4E) has similar features as the CSD variance, as expected from their common biophysical origin. However, volume conduction has some qualitative effects, such as the reduced relative contribution from the L6E population in layer 2/3. As correlated synaptic inputs are known to amplify and increase the spread of the LFP generated from a cortical population (Lindén et al., 2011; Łęski et al., 2013), this may reflect that the synaptic inputs to the L23E population are more correlated than to the L6E population. Overall, we conclude that for the spontaneous activity in our network, the L23E and L6E populations dominate the compound LFP and CSD with only smaller contributions from the other excitatory populations. The signal variances from the inhibitory populations are typically much smaller than the contributions from the excitatory populations, suggesting that they can be safely neglected, in line with recent findings of Mazzoni et al. (2015).

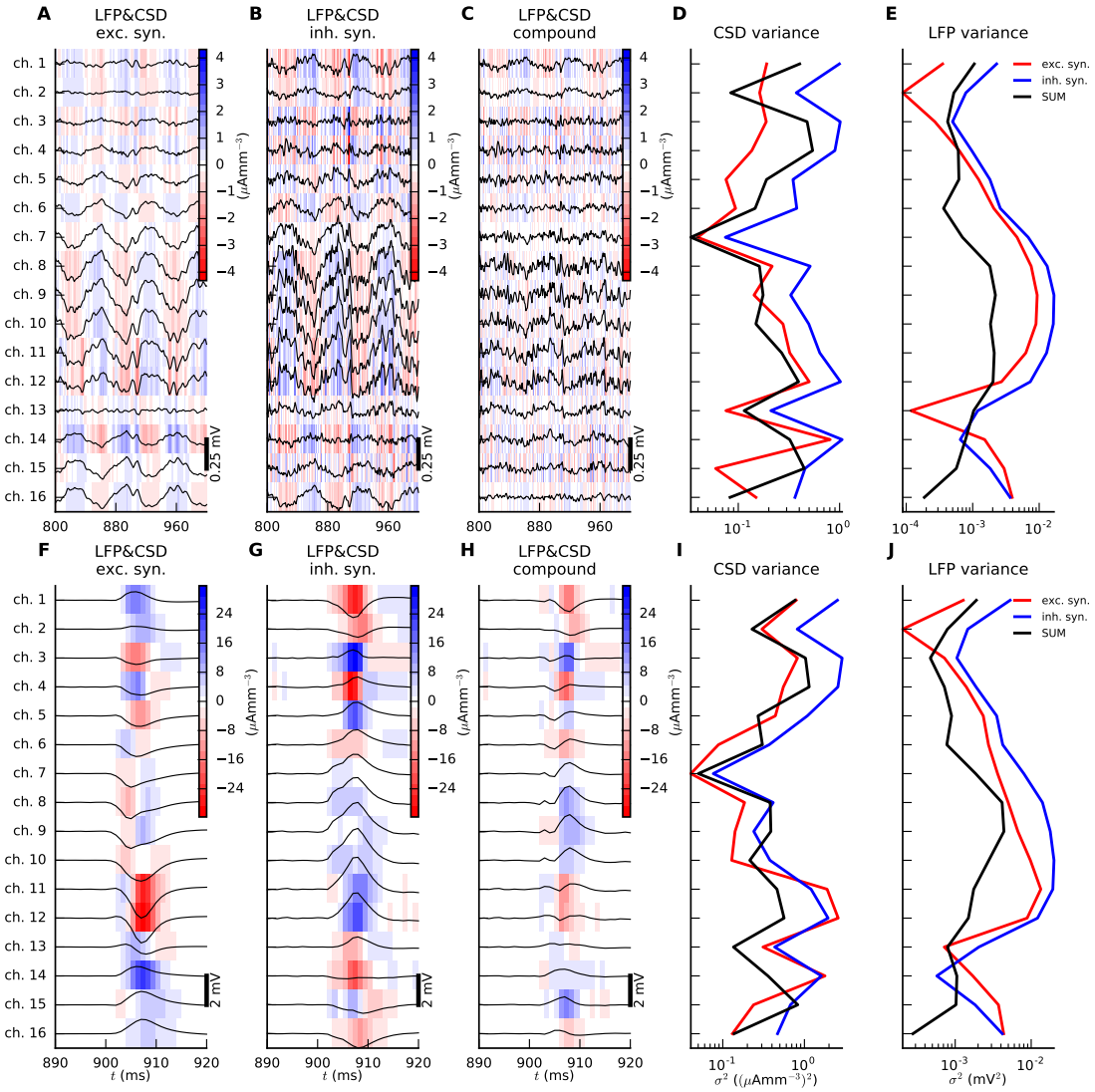


Figure 6.5: **Decomposition of CSD and LFP into contributions due to excitatory and inhibitory inputs for thalamic activation. A–E** Oscillatory thalamic activation (f = 15 Hz). **F–J** Transient thalamic activations at $t = 900 + n \cdot 1000$ ms for $n \in \{0, 1, 2, 3, 4\}$. Same row-wise figure arrangement as in Fig. 6.4F-J.

Although transmembrane currents in inhibitory neurons provide little of the observed CSD and LFP, the inhibitory synaptic inputs onto excitatory neurons provide a substantial contribution. In the present network, inhibitory synaptic currents have a four-fold larger amplitude compared to most excitatory synapses (Tab. D.1), inhibitory neurons have higher overall firing rates compared to excitatory neurons (Potjans & Diesmann, 2014), and inhibition specifically targets soma-proximal sections (Markram et al., 2004). Since our LFP-generating model is linear (passive cable formalism, linear synapse model), we can decompose the compound signal into contributions from each synapse type. For example, selective removal of either inhibitory or excitatory synaptic currents in the CSD and LFP modeling (Fig. 6.4F,G) shows that the CSD and LFP signals (Fig. 6.4H) are dominated by inhibitory synaptic currents and their associated return currents (Fig. 6.4I,J) when the network operates in the spontaneous asynchronous irregular firing regime. Further, at most depths the variance of the signal arising from inhibitory input exceeds the compound variance, implying that the inhibitory component is generally negatively correlated with the excitatory component (Fig. 6.4I,J). Visual inspection of Fig. 6.4F-H also reveals that the inhibitory dominance appears particularly strong at high frequencies, in accordance with the firing-rate PSDs in Fig. 6.3B showing less power of inhibitory spiking, and thus inhibitory synaptic input currents, at low frequencies.

The relative contribution from excitation and inhibition to the CSD and LFP depends on thalamic input and network state, however. For oscillatory thalamic input, the CSDs and LFPs from excitatory (Fig. 6.5A) or inhibitory synapses (Fig. 6.5B) alone show much stronger oscillations than the compound signals (Fig. 6.5C). This follows from the observations that the contributions to the CSD and LFP from excitatory and inhibitory synapses are anticorrelated, which, in turn, is a consequence of the dynamical balance between excitation and inhibition in asynchronous-irregular states of balanced random networks (Hertz, 2010; Renart et al., 2010; Tetzlaff et al., 2012).

For transient thalamic activation, the same cancellation of excitation and inhibition can be observed (Fig. 6.5F-H), although it is more pronounced at some depths (layers 5 and 6, i.e., channels 11-13) than at others, depending on whether excitation and inhibition are in phase or not (Fig. 6.5I,J). For such strong and transient thalamic input, the network activity is briefly imbalanced as inhibition cannot keep up with thalamic excitation on the short time scales. Since thalamic input is most prominent in layer 4 (channels 7–10), fewer cancellation effects are present here: in channel 9 in Fig. 6.5I the total CSD variance is, e.g., seen to be larger than the individual contributions from inhibitory and excitatory synapses.

6.7 Effect of input correlations

Synaptic inputs to two neighboring cells are typically correlated because (i) they receive, to some extent, inputs from the same presynaptic sources ('shared-input

correlation'), and (ii) the spike trains of the presynaptic neurons may be correlated ('spike-train correlation'). The net synaptic-input correlation is determined by the interplay between these two contributions, shared-input correlations and spike-train correlations (Renart et al., 2010; Tetzlaff et al., 2012). As the LFP is largely generated by synaptic inputs, synaptic-input correlations result in correlated single-cell LFP contributions $\phi_i(\mathbf{r}, t)$ (for details, see Lindén et al., 2011; Łęski et al., 2013). As outlined in the following, these single-cell-LFP correlations play a dominating role for the spectrum of the compound LFP.

The power spectrum $P_\phi(\mathbf{r}, f)$ (Eq. (D.6)) of the compound LFP $\phi(\mathbf{r}, t) = \sum_{i=1}^{N} \phi_i(\mathbf{r}, t)$ of a population of N neurons is given by

$$P_\phi(\mathbf{r}, f) = N\overline{P_\phi}(\mathbf{r}, f) + N(N-1)\overline{P_\phi}(\mathbf{r}, f)\overline{\kappa_\phi}(\mathbf{r}, f) \ . \tag{6.2}$$

Here $\overline{P_\phi}(\mathbf{r}, f)$ is the average single-cell LFP power spectrum (Eq. (D.3)) and $\overline{\kappa_\phi}(\mathbf{r}, f)$ the average pairwise single-cell LFP coherence (Eq. (D.5)), a measure for cross-correlations, across all cells. Note that, while the first term in Eq. (6.2) scales linearly with the number of neurons N, the second term is proportional to $N(N-1) \approx N^2$ for large N. Hence, for large N, even small cross-correlations may dominate the spectrum of the compound LFP. Here, we investigate this situation by calculating the power spectrum $P^0_\phi(\mathbf{r}, f)$ of the compound LFP under the assumption of zero cross-correlation (where it simply reduces to a sum over single-cell spectra $P_{\phi_i}(\mathbf{r}, f)$), and compare to the true spectrum $P_\phi(\mathbf{r}, f)$. The ratio between these quantities is given by

$$\frac{P_\phi(\mathbf{r}, f)}{P^0_\phi(\mathbf{r}, f)} = 1 + (N-1)\overline{\kappa_\phi}(\mathbf{r}, f) \ . \tag{6.3}$$

With weak or no cross-correlations, i.e., $(N-1)\overline{\kappa_\phi}(\mathbf{r}, f) \ll 1$, the ratio approaches unity, and the power of the compound LFP is essentially the sum of the power of the single-cell LFPs. For $N\overline{\kappa_\phi}(\mathbf{r}, f) \gg 1$, i.e., in the correlation-dominated regime, this ratio is instead proportional to the number of neurons N. Note also that anti-correlated signals ($\overline{\kappa_\phi}(\mathbf{r}, f) < 0$) may lead to a ratio $P_\phi(\mathbf{r}, f)/P^0_\phi(\mathbf{r}, f) < 1$.

In the example application, both for spontaneous (Fig. 6.6A,B) and for evoked activity (Fig. 6.6C,D) the compound power spectra $P_\phi(\mathbf{r}, f)$ are systematically (across channels and frequencies) larger than $P^0_\phi(\mathbf{r}, f)$, demonstrating the importance of cross-correlations in the present network. Depending on the recording depth and frequency, the ratio varies from ~ 1 to 10^3 (see Fig. 6.6B,D). For spontaneous activity (see Fig. 6.3A-E), the largest effects of cross-correlations are typically found at higher frequencies (Fig. 6.6A,B). At low frequencies, cross-correlations are suppressed by inhibitory feedback (cf. Fig. 6.3B, Tetzlaff et al. (2012)). The thalamic sinusoidally modulated input to the network (Fig. 6.3F-J) synchronizes single-cell CSDs and LFPs at the stimulus frequency and gives a large boost of the LFP power at this frequency (see peak in Fig. 6.6C,D). Close inspection reveals that there is also in fact a boost of the power at around 80 Hz, but much less so than around 15 Hz. Note that the ex-

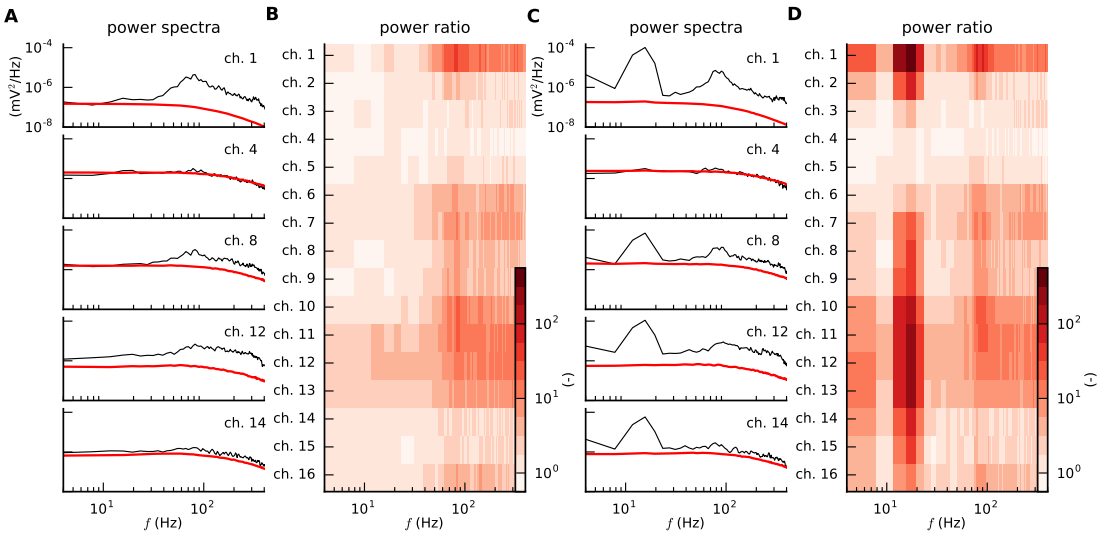


Figure 6.6: **Effect of single-cell-LFP cross-correlations on compound-LFP power spectra during spontaneous activity (A,B) and for oscillatory thalamic input (C,D).** **A,C** Compound-LFP power spectra $P_\phi(\mathbf{r}, f)$ (black traces) and compound spectra $P^0_\phi(\mathbf{r}, f)$ (red traces) obtained when omitting cross-correlations between single-cell LFPs (red traces; computed for 10 % of the cells and multiplied by a factor 10) at recording channels corresponding approximately to the centers of layers 1, 2/3, 4, 5 and 6. **B,D** Depth and frequency-resolved ratio $P_\phi(\mathbf{r}, f)/P^0_\phi(\mathbf{r}, f)$ of LFP power spectra, cf. Eq. (6.3).

ternal activation hardly affects the single-cell spectra (see red curves in Fig. 6.6A,C). LFP synchronization is thus mainly encoded in the phase of $\phi_i(\mathbf{r}, t)$.

In conclusion, cross-correlations between single-cell LFP contributions play a pivotal role in shaping the compound LFP spectra (similar for CSD spectra, results not shown). To account for the dominant features of the LFP (CSD) in such models, it is therefore essential to include the main factors determining the synaptic-input correlations, i.e., realistic correlations in presynaptic spike trains and shared-input structure. The findings presented here for the cortical microcircuit model hold in general in the presence of correlated activity. Only the details of the spectra depend on the specific underlying network dynamics.

6.8 Network downscaling

Due to the computational cost associated with modeling LFPs, it would be desirable to downsize the postsynaptic populations of multicompartment model neurons to a fraction γN ($\gamma \in (0, 1)$) while leaving the point-neuron network at full size (N) and at the same time preserve the in-degrees, i.e., the number of inputs, of each postsynaptic cell. The power spectrum of the full-scale LFP can indeed be estimated

from the population-averaged single-cell power spectra $\overline{P_{\phi^\gamma}}(\mathbf{r}, f) \approx \overline{P_\phi}(\mathbf{r}, f)$ and coherences $\overline{\kappa_{\phi^\gamma}}(\mathbf{r}, f) \approx \overline{\kappa_\phi}(\mathbf{r}, f)$ computed for downsized networks by means of Eq. (6.2). These quantities are preserved except for deviations due to smaller sampling size γN (Eq. (D.6) in Tab. D.6). However, due to lack of phase information in the power spectra, one cannot estimate the LFP time course.

One could attempt to obtain a time-course estimate $\phi^{\gamma\xi}(\mathbf{r}, t)$, i.e., a 'low-density LFP prediction', of the full-scale signal $\phi(\mathbf{r}, t)$ by upscaling single-cell LFPs $\phi_i(\mathbf{r}, t)$ computed in the downsized setup by a scalar factor ξ (cf. Eq. (D.1)). Such a naive upscaling can grossly recover the amplitude of the full-scale LFP $\phi(\mathbf{r}, t)$, but it still only partially reconstructs its detailed time course (Fig. 6.7A,E). Also, this approach does not generally give accurate power spectra as the two terms in Eq. (6.2) scale differently with ξ: The rescaling introduces a prefactor ξ^2 in the population-averaged single-cell power spectra $\overline{P_{\phi^{\gamma\xi}}}(\mathbf{r}, f) \approx \xi^2 \overline{P_\phi}(\mathbf{r}, f)$, while the coherences $\overline{\kappa_{\phi^{\gamma\xi}}}(\mathbf{r}, f) \approx \overline{\kappa_\phi}(\mathbf{r}, f)$ are unchanged. Thus the compound spectra $P_\phi(\mathbf{r}, f)$ and $P_{\phi^{\gamma\xi}}(\mathbf{r}, t)$ of the full-size LFP and the low-density LFP predictor, respectively, differ. Their ratio

$$\frac{P_\phi(\mathbf{r}, f)}{P_{\phi^{\gamma\xi}}(\mathbf{r}, f)} = \frac{1 + (N-1)\overline{\kappa_\phi}(\mathbf{r}, f)}{\gamma\xi^2 + \gamma\xi^2(\gamma N - 1)\overline{\kappa_\phi}(\mathbf{r}, f)} = \begin{cases} 1/(\gamma\xi^2) & \text{for } \overline{\kappa_\phi}(\mathbf{r}, f) = 0 \\ 1/(\gamma^2\xi^2) & \text{for } \overline{\kappa_\phi}(\mathbf{r}, f) = 1 \end{cases} \tag{6.4}$$

demonstrates that in the general case there is no scaling factor ξ which allows for the recovery of the full-size compound LFP power, i.e., makes the ratio in Eq. (6.4) equal to one for all spatial positions $\mathbf{r}$ and frequencies f. This can only be done in the special case where $\overline{\kappa_\phi}(\mathbf{r}, f)$ is a constant c. Here the two extreme cases correspond to no correlation ($\overline{\kappa_\phi}(\mathbf{r}, f) = 0$ with $\xi = 1/\sqrt{\gamma}$) and full correlation between all single-cell signals ($\overline{\kappa_\phi}(\mathbf{r}, f) = 1$ with $\xi = 1/\gamma$).

The substantial scaling effects observed for our microcircuit model in the asynchronous state (Fig. 6.7A–D) suggest that correlations cannot be neglected even when modeling the LFP for spontaneous network activity. Choosing the scaling factor $\xi = 1/\sqrt{\gamma}$ corresponding to $\overline{\kappa_\phi}(\mathbf{r}, f) = 0$ (red lines in Fig. 6.7A,C) leads to a severe underestimation of the full-size compound power spectrum (Fig. 6.7C,D). Even though the correlation (i.e., Pearson's correlation coefficients) between the full-size full-size LFP signals and low-density LFP predictions are quite high (Fig. 6.7B), the power ratios (Fig. 6.7D) reveal that the rescaled signals are systematically wrong in frequency bands where single-cell LFPs are most strongly correlated (i.e., the frequencies for which the compound spectra are much larger than the predictions when omitting cross-correlations (Fig. 6.6A,B)). Assuming the full-correlation scaling factor $\xi = 1/\gamma$, on the other hand, typically overestimates the full-size compound power spectrum (cf. gray spectra in Fig. 6.7C), particularly at low frequencies.

The results for the sinusoidally stimulated network (Fig. 6.7E–H) are quite similar to the spontaneous-activity results, except around the stimulation frequency 15 Hz where the modulated input leads to strongly correlated single-cell LFP contributions and a strong boost of the compound LFP. The approximate downscaling procedure

assuming the full-correlation scaling factor $\xi = 1/\gamma$ thus essentially agrees with the full-size compound spectrum for 15 Hz (while giving a strong overestimation for frequencies other than ~ 15 and ~ 80Hz).

We note in passing that in contrast to the power spectra, the computation of the spike-trigger-averaged LFP (stLFP) (Swadlow et al., 2002; Nauhaus et al., 2009; Jin et al., 2011; Denker et al., 2011) in downsized networks do not have a principled problem due to cross-correlations between single-cell LFPs. As stLFPs are linearly dependent on the single-neuron LFP contributions, the only principled problem with downsizing is increased noise in the estimates due to sampling over fewer postsynaptic neurons.

6.9 LFP prediction from population firing rates

An important question in systems neuroscience is to what extent the dynamics of networks of thousands or millions of neurons can be described by much simpler mathematical descriptions in terms of neural *populations* (Deco et al., 2008; Blomquist et al., 2009). Likewise, we here ask the question of whether LFPs can be predicted from knowledge of the population firing rate (Einevoll et al., 2007; Moran et al., 2008; Einevoll et al., 2013). The hybrid scheme is excellently suited for testing and development of simplified numerical schemes for LFP prediction as the ground truth, i.e., the LFP from the full network, is available as benchmarking data.

The use of current-based synapses and passive dendrites in the present application of the hybrid scheme, renders synaptic events independent of each other in the LFP prediction. This inherent linearity results in a unique spatio-temporal relation $H_X^i(\mathbf{r}, \tau)$ for $\tau \in [-\infty, \infty]$ between a spike event of a point neuron i in population X and its contribution to the compound LFP $\phi(\mathrm{r}, t)$ from all its postsynaptic multicompartment model neurons. In this scheme the link is causal, i.e., the spikes drive the LFP, so that $H_X^i(\mathbf{r}, \tau) = 0$ for $\tau < 0$ (as in laminar population analysis (LPA) (Einevoll et al., 2007)). $H_X^i(\mathbf{r}, \tau)$ encompasses connectivity, spike transmission delays and all postsynaptic responses including effects of synaptic input currents and passive return currents. With a linear, current-based model such as our example cortical column, it is in principle possible by linear superposition to fully reconstruct the compound LFP if $H_X^i(\mathbf{r}, \tau)$ and the spike times t_l^i are known for all neurons i in each population of the network. It is, however, in the case of large networks impractical to assess each $H_X^i(\mathbf{r}, \tau)$, as the LFP response needs to be determined for every neuron separately. In contrast, a large reduction in dimensionality can be achieved by determining the population-averaged LFP responses $\overline{H}_X(\mathbf{r}, \tau)$ of a spike within each population X. We thereby ignore heterogeneity in kernels $H_X^i(\mathbf{r}, \tau)$ due to the variability in the connections from neurons in population X. An approximate compound LFP $\phi^*(\mathbf{r}, t)$ based on population firing rates (Einevoll et al., 2007) can be computed from these extracted

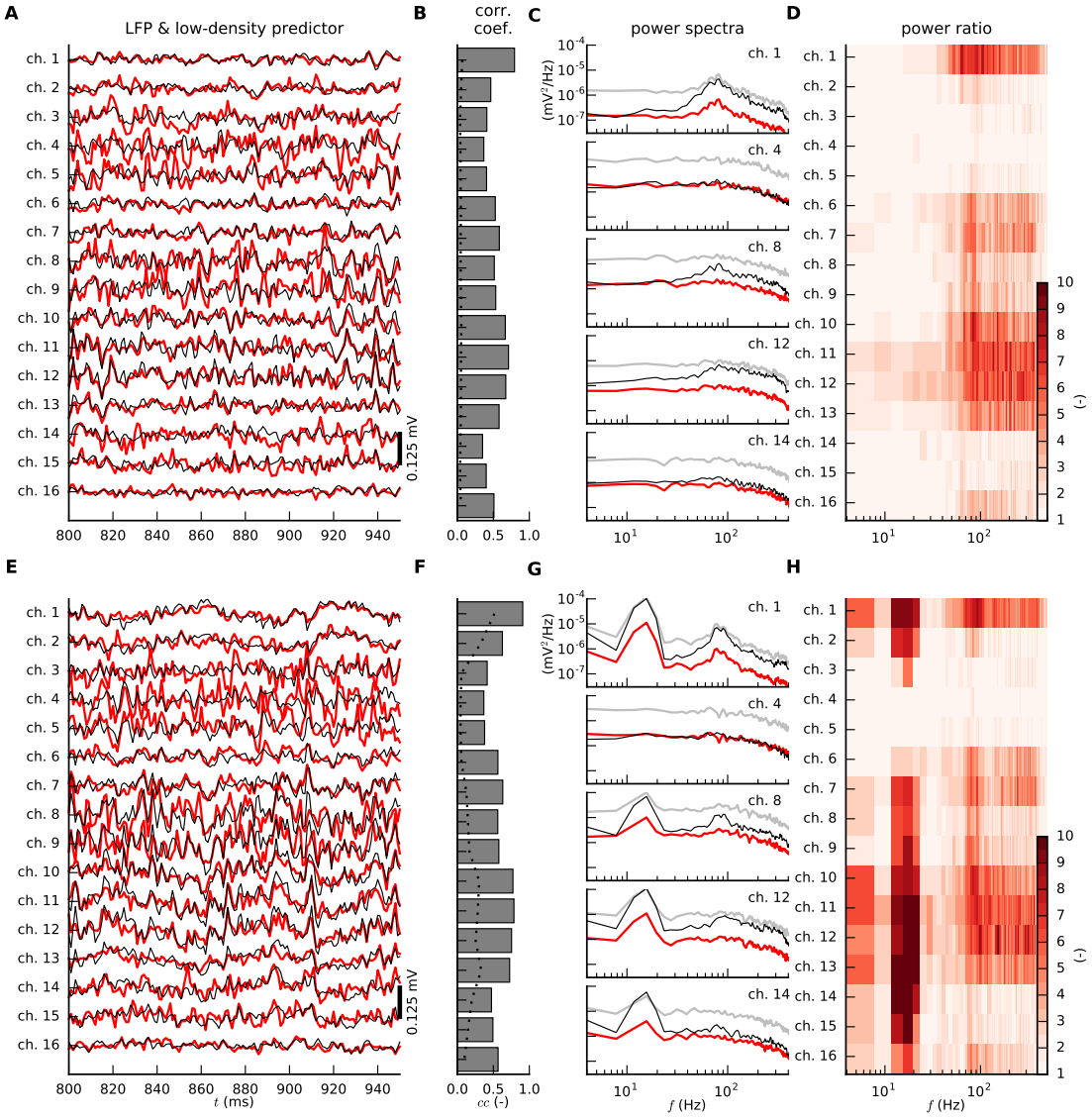


Figure 6.7: **Prediction of LFPs from downsized networks.** Top row: Spontaneous activity. Bottom row: Oscillatory thalamic activation. **A,E** Full-scale LFP traces $\phi(\mathbf{r},t)$ (black) and low-density predictors $\phi^{\gamma\xi}(\mathbf{r},t)$ (red) obtained from a fraction $\gamma = 0.1$ of neurons in all populations and upscaling by a factor $\xi = \gamma^{-1/2}$. **B,F** Correlation coefficients (gray bars) between full-scale LFP and low-density predictor shown in panels A and E, respectively. The dashed lines denote 1% significance levels obtained after computing the chance correlation coefficients for 1000 trials. **C,G** Power spectra $P_\phi(\mathbf{r},f)$ and $P_{\phi^{\gamma\xi}}(\mathbf{r},f)$ of full-scale LFPs (black) and low-density predictors with $\gamma = 0.1$ and $\xi = \gamma^{-1/2}$ (red) or $\xi = \gamma^{-1}$ (gray). **D,H** Ratio $P_\phi(\mathbf{r},f)/P_{\phi^{\gamma\xi}}(\mathbf{r},f)$ between power spectra of full-scale LFP and low-density predictor with $\gamma = 0.1$ and $\xi = \gamma^{-1/2}$ (cf. Eq. (6.4)).

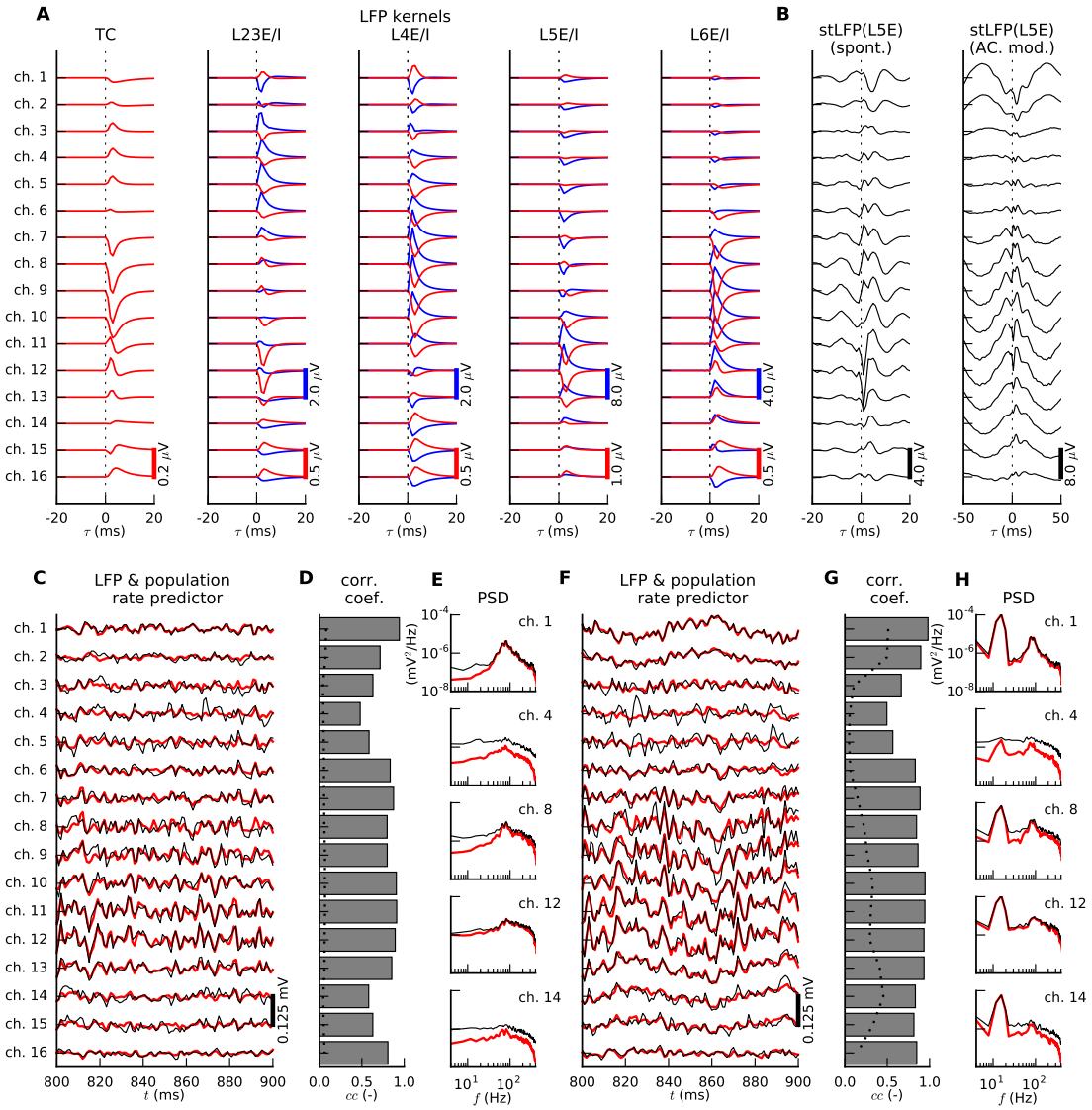


Figure 6.8: **Linear prediction of LFPs from population firing rates. A** LFP responses $\overline{H}_X(\mathbf{r},\tau)$ (kernels) to simultaneous firing of all neurons in a single presynaptic population X (see subpanel titles) at time $\tau = 0$ ms, normalized by size N_X of the presynaptic population (red/blue: responses to firing of excitatory/inhibitory presynaptic populations). **B** Spike-trigger-averaged LFPs (stLFP) triggered on spikes of L5E neurons during spontaneous activity (left) and oscillatory thalamic network activation (right), averaged across all L5E spikes (T = 5,000 ms simulation time). **C,F** LFP traces of the full model (black) compared to predictions (red) obtained from superposition of linear convolutions of population firing rates ν_X with LFP kernels $\overline{H}_X(\mathbf{r},\tau)$ shown in panel A. **D,G** Correlation coefficients (gray bars) between LFPs and population-rate predictors shown in panels C and F. The dashed lines denote 1% significance levels obtained after computing the chance correlation coefficients for 1000 trials. **E,H** Power spectra of LFPs (black) and the population-rate predictors (red) for different recording channels. Panels C-E and F-H show results for spontaneous activity and oscillatory thalamic activation, respectively.

population kernels by means of the convolution $\phi^*(\mathbf{r},t) = \sum_X \left(\nu_X * \overline{H}_X\right)(\mathbf{r},t)$, where $\nu_X(t)$ are the instantaneous population firing rates.

Here we estimate the population LFP kernels by computing the response to synchronous activation of all neurons in a population (Fig. 6.8). The spatio-temporal kernels $\overline{H}_X(\mathbf{r},\tau)$ are extracted from time slices $[t_X - 20\text{ ms}, t_X + 20\text{ ms}]$ of the compound LFP response $\phi(\mathbf{r},t)/N_X$, where N_X is the number of neurons in a presynaptic population. The procedure results in unique kernels $\overline{H}_X(\mathbf{r},\tau)$ for each excitatory and inhibitory population in the network (Fig. 6.8A).

In the example application, the population kernels $\overline{H}_X(\mathbf{r},\tau)$ differ significantly between populations. Excitatory spike events result in prominent LFP negativities at depths where most connections are made, such as in layer 4 (channels 7–10) for thalamocortical connections ($\overline{H}_{\mathrm{TC}}(\mathbf{r},\tau)$, column 1 in Fig. 6.8A, cf. Fig. 5.5C). In contrast, spikes of inhibitory point neurons on average produce prominent LFP positivities in their corresponding layer, such as in layer 2/3 (channels 3–6) for population L23I (column 2 in Fig. 6.8A). In all cases, the signatures of opposite-sign return currents and also other, weakly connected populations are seen across depth.

As seen in Fig. 6.8C,F the population-rate predictions are in good qualitative agreement with the ground-truth LFP for both spontaneous and sinusoidally modulated network activity with correlation coefficients (*cc*) between 0.48 and 0.94 for spontaneous activity (Fig. 6.8D) and 0.52 and 0.98 for thalamically evoked oscillations (Fig. 6.8G). Overall, the population-rate predictions appear to be best for the lower frequencies (Fig. 6.8E,H), while the inherent variability in the individual point-neuron kernels $H_X^i(\mathbf{r},\tau)$ (which is not accounted for in the population approximation) has a larger effect on the higher frequencies. This can be understood on biophysical grounds, as the lower frequencies are expected to mainly reflect the gross anatomical features of the postsynaptic populations and their presynaptic connections patterns where the individual variability plays a lesser role (Lindén et al., 2010; Pettersen et al., 2012). The correlation coefficients and power spectra of the population-rate prediction thus show that the population-rate LFP predictor is more accurate than the low-density LFP predictors (Fig. 6.7C,F) in case of substantial downscaling.

Although the spike-trigger-averaged LFP (stLFP) (Swadlow et al., 2002; Nauhaus et al., 2009; Jin et al., 2011; Denker et al., 2011), calculated as the cross-covariance between the population spike rate $\nu_X(t)$ and the compound LFP $\phi(\mathbf{r},t)$ divided by the total number of spikes (i.e., $\int_0^T \nu_X(t)dt$), is related to our LFP population kernels, it measures very different aspects of cortical dynamics. The population kernels $\overline{H}_X(\mathbf{r},\tau)$ are causal and independent of effects of spike-train correlations. The stLFP, on the other hand, is non-causal and strongly depends on spike-train correlations, and thus also network state (Einevoll et al., 2013). The stLFP is thus not only very different from $\overline{H}_X(\mathbf{r},\tau)$, it also varies strongly between the spontaneous and sinusoidally modulated network state (see example for L5E neurons in Fig. 6.8B).

6.10 Conclusion

We described the local field potential (LFP) of a cortical column as simulated using the hybrid scheme and investigated the impact of individual populations, network dynamics and cell density on the mesoscopic LFP signal. Even though the example application was not tuned to address specific experiments, the predicted LFP amplitudes for ongoing and evoked activity as well as layer-specific activation patterns are generally consistent with experimental results. LFPs at different cortical depth are highly coherent due to passive volume conduction and power spectra differ across depth due to contributions of specific neuron populations. The superposition of LFP contributions from different neuron populations and synapse types is non-trivial, and may, for example, result in local cancellations of the signal. In general, inhibitory currents onto excitatory neuron populations dominate the LFP, which is strongly influenced by synaptic-input correlations, and thus highly sensitive to local and remote network dynamics. The effect of correlations on the LFP cannot be fully accounted for in downscaled LFP models even if the synaptic input correlations are correctly accounted for. However, LFPs can be derived to a good approximation from population firing rates rather than from spikes of individual neurons.

Part V

Discussion

The organization of the cerebral cortex extends over a wide range of spatial scales, from fine-scale specificity of the connectivity between small groups of neurons to hierarchically organized networks of entire cortical areas. Cortical activity exhibits correlations on a multitude of spatial and temporal scales. Linking the microscopic scale, on which the precise form of the interaction among neurons via synapses is known, to the mesoscopic scale, on which experimentally observable activity and network function arise, is a fundamental requirement to understand the brain.

A bridge from microscopic to mesoscopic dynamics is moreover required in neuroscience to compare model predictions to experimental signals that are indicative of collective network dynamics: Routinely accessible signals in cortex only arise on mesoscopic scales from the superposition of the activity of large numbers of cells. The local field potential (LFP), the low frequency components of the electrical potential in the extracellular space, is such a readily available mass signal with contributions from hundreds of thousands to millions of neurons (Kajikawa & Schroeder, 2011; Lindén et al., 2011; Łęski et al., 2013). Forward-modeling studies have shown that the LFP can be approximated as a linearly weighted sum of transmembrane currents across the dendritic tree of neurons (Einevoll et al., 2013). These currents are triggered primarily by synaptic events that arise from spiking activity, the main type of interactions between neurons. Recent advances in electrode technology and spike-sorting techniques further allow the assessment of single-unit activity of up to hundreds of neurons in parallel extracellular recordings, showing a rich correlation structure on the level of individual pairs of neurons. Neuronal networks therefore share many features with interacting many-body systems in physics that exhibit collective behavior on multiple scales.

This thesis contributes to the understanding of the microscopic and mesoscopic brain activity by deriving a theory of correlations between individual neurons in recurrent networks, building simulation software for rate-based models of neuronal activity, and by developing and implementing a modeling scheme for local field potentials based on spiking activity of point-neuron networks. The following sections discuss individual aspects related to the chapters in this thesis, which are put into context in the outlook that illustrates future directions.

Correlated fluctuations in strongly coupled binary networks beyond equilibrium

In Chap. 1, we address the characterization of fluctuations in strongly interacting, non-equilibrium, disordered systems. For the prototypical model of such systems, dynamic binary units introduced by Glauber (1963), we derive for the first time a closed hierarchy of cumulant equations that jointly describe local mean activities (“magnetizations”) and pairwise correlations including finite-size effects.

Standard approaches describing systems with symmetric pairwise interactions in terms of a partition function cannot be extended to nonsymmetric coupling, because these systems do not reach thermodynamic equilibrium. Asymmetries arising from disordered couplings or deterministic constraints, such as Dale's law, are treated by using ensemble averages over the random couplings (Fischer & Hertz, 1991a; Hermann & Touboul, 2012) or the random dilution of the network (Shinomoto, 1987), respectively. By construction, these approaches are, however, constrained to describe self-averaging properties of systems. Here, we are interested in the dynamics of individual units, which is not self-averaging, and we set out to capture the statistics of a particular realization of the system in terms of the cumulant hierarchy for the activity variables. Truncating the hierarchy after second order yields a closed set of equations that already provides a good approximation for the zero time lag covariances between individual units. We show the equivalence of this truncation to the Gaussian approximation of the input field (van Vreeswijk & Sompolinsky, 1998; Renart et al., 2010), which follows from the central-limit theorem. Non-Gaussian, finite-size effects in networks of only several hundreds of units are effectively taken into account by incorporating a subset of third-order cumulants that can be expressed in terms of lower-order cumulants due to the binary nature of the activity variables. This significantly improves the prediction of mean activities, which are strongly affected by the nonlinearity of the gain function. The resulting set of nonlinear coupled equations can efficiently be solved numerically by damped fixed-point iteration.

We demonstrate that a Heaviside activation function - independent of the choice of coupling amplitudes - constitutes the strongest possible interaction for binary networks. This finding generalizes the invariance of population-averaged pairwise correlations under proportional scaling of couplings and threshold of all units (Grytskyy et al., 2013a; van Albada et al., 2015) to the invariance of individual pairwise correlations under the scaling of the corresponding units' parameters. Weaker coupling in binary networks can only be achieved by a smoother activation function, but not by different scalings of coupling amplitudes, e.g., $1/N$ versus $1/\sqrt{N}$, contradictory to frequently employed arguments in the literature (as discussed in van Albada et al., 2015). In networks with plastic synapses or networks receiving time-varying input, dynamically changing thresholds could similarly act as a homeostasis mechanism on the level of mean activities and correlations and prevent systems from freezing in a local minimum (Roach et al., 2016). While the scaling invariance is generic, the presented Gaussian and close-to-Gaussian approximations of the input field hold for the commonly considered narrowly distributed couplings. For wide distributions, the statistics of the input field is likely to depart from the close-to Gaussian assumption.

The contribution of cross-covariances to the marginal statistics of the input to each neuron causes a distribution of the mean activities, even in networks composed of neurons each receiving an identical number of inputs. The classical treatment of neurons with a Heaviside nonlinearity neglects this effect (van Vreeswijk & Sompolinsky, 1998). Still, distributed numbers of synaptic input typically dominate the

distribution of mean activities in sparsely (van Vreeswijk & Sompolinsky, 1998) and densely connected random networks (Renart et al., 2010; Helias et al., 2014).

The Gaussian closure, that is, the truncation of the cumulant hierarchy after second order, yields a set of equations that accurately predicts the second-order statistics. This result is in line with experimental evidence showing that pairwise correlations sufficiently constrain maximum-entropy models of collective activity (Schneidman et al., 2006). Analogously, the truncation after first order, the Curie-Weiss mean-field theory, neglects fluctuations of the local order parameter. Formally, this theory can be obtained as the leading order of a saddle-point approximation in the auxiliary fields (e.g. Negele & Orland, 1998, section 4.3). Still it yields a good estimate of the first moments. In the context of balanced networks, this description is sometimes referred to as "balance equations" (van Vreeswijk & Sompolinsky, 1998, eqs. (4.1) and (4.2)). Taking into account fluctuations in the input, due to the variance of individual units (van Vreeswijk & Sompolinsky, 1998), constitutes an intermediate step in the cumulant hierarchy. While for continuous random variables the higher-order cumulants are, by definition, independent of all lower ones, for discrete variables this is not necessarily the case: All cumulants of a random binary variable are completely determined by its mean. The constraint on only two possible values causes this dependence. The variance can therefore be determined by exploiting the binary character of the variables. A consequence of the lower-order moments depending only weakly on the statistics of higher-order moments is that cross-covariances can be determined from linear fluctuations around the steady state, which itself is determined by neglecting correlations altogether. We may therefore speculate that an approximation of third- or higher-order correlations can be obtained from the fluctuations around a state that itself is determined self-consistently by taking into account only up to second-order correlations.

The consistent truncation of the cumulant hierarchy further provides deeper insights: The expression for covariances between individual neuron pairs follows naturally from a consistent Gaussian closure, without further approximation. In previous studies, it was obtained as a linear approximation for weak correlations (Renart et al., 2010, supplementary eqs. (31)-(33)) and population-averaged covariances (Ginzburg & Sompolinsky, 1994, eq. (6.8)). Here, we show that solving the modified Lyapunov equation naturally leads to a decomposition of fluctuations into eigenmodes of the system. The presented method solves the problem that the modified Lyapunov equation does not hold on the diagonal, in contrast to the case of population averages (Ginzburg & Sompolinsky, 1994). The approach moreover exposes that fluctuations in the Gaussian approximation are described by a set of coupled Ornstein-Uhlenbeck processes. It is well known that the fluctuations obtained from a systematic system-size expansion (van Kampen, 1992, chapter X), to lowest order, obey a Langevin equation. This result has been applied to networks of homogeneous populations (Ohira & Cowan, 1997; Bressloff, 2009; Grytskyy et al., 2013b). The interesting point here is the feasibility of such a reduction directly on the level of individual binary variables. In contrast to the population-averaged activity, the jumps of a binary variable

cannot be considered small compared to its value; hence the precondition to apply the system-size expansion is not valid. The approach taken here to expose the correspondence between binary dynamics and Ornstein-Uhlenbeck processes is hence complementary. In particular, the result shows that the effective noises appearing in the equivalent set of Ornstein-Uhlenbeck processes are not uniquely determined by the activities of the neurons alone, in contrast to the population-averaged case (Grytskyy et al., 2013b, sec. 3.1).

We show that the effective interaction strength, that is, the product of a unit's susceptibility and the incoming coupling amplitude, can uniquely be reconstructed from the covariance matrix and the slope of the covariance functions at zero time lag. This relation builds on the regression theorem (Risken, 1996), which states that the fluctuations in the system, to linear order, follow the same differential equation as the time-delayed covariance functions. Coupling amplitudes, for principle reasons, cannot be inferred unambiguously: Correlations only depend on J/σ, where σ measures the strength of the overall incoming fluctuations to the node. The inference of the effective interaction strength J/σ still allows the classification of connections into excitatory and inhibitory ones, but leads to a considerable rate of false-positive connections. The method presented here requires the measurement of covariances between all pairs of units in the network. In the case of severe subsampling, its application is restricted to small subnetworks. Thus, in general settings, more sophisticated approaches that take into account the influence of hidden units on the observed covariances (Tyrcha & Hertz, 2013) or on the inference of the global network state (Buice & Chow, 2011) are more promising. The presented expressions that relate the covariance matrix and its slope at zero time lag to the coupling matrix may still prove useful to construct similar inference methods for the investigated class of networks. Moreover, the algorithm constructing a network that generates activity with desired mean activities and covariances is a useful tool to generate surrogate data.

The invariance of effective interactions with respect to scaling of couplings implies that in Erdős-Rényi random binary networks with fixed weights, the spectral radius of the matrix of effective couplings is bounded by $\sqrt{2(1-p)/\pi} < 0.8 < 1$ (Grytskyy et al., 2013a), so that all modes of the network are stable. An instability of the mean activities can only be achieved with coupling statistics or motifs that differ from an Erdős-Rényi network. This finding is in qualitative contrast to spiking networks that indeed show a rate instability at a critical coupling weight (Ostojic, 2014; Harish & Hansel, 2015). This insight has consequences for the use of random networks with fixed weights in the framework of reservoir computing (Jaeger, 2001; Maass et al., 2002), where highest computational performance (Crutchfield & Young, 1989) is achieved at the edge of chaos (for a review, see Legenstein & Maass, 2007b).

Previous works have addressed several aspects of pairwise correlations. The smallness of the average magnitude of covariances in strongly coupled balanced networks (van Vreeswijk & Sompolinsky, 1996) has been explained by the influential work

of Renart et al. (2010) in the large-N limit. In general, decorrelation follows from the dominant negative feedback in balanced networks at any network size (Tetzlaff et al., 2012; Helias et al., 2014). These global properties are determined by collective fluctuations and can hence be described by population-averaged mean-field theory (Ginzburg & Sompolinsky, 1994). While such self-averaging observables in disordered systems can be obtained in ensembles of system realizations, here we discuss the statistics of individual units in a particular realization of the couplings, which goes beyond the existing approaches.

Our results are complementary to the method presented by Buice et al. (2010), who consider Markov systems of interacting populations of neurons that consequently have integer numbers of active states, as opposed to the Glauber dynamics (Glauber, 1963) of individual neurons considered here. In consequence, the truncation in the former work is performed on the level of normal-ordered cumulants, which represent an approximation around Poisson statistics, while here we derive an expansion in terms of ordinary cumulants (Gardiner, 2004). Moreover, Buice et al. (2010) and Ginzburg & Sompolinsky (1994) consider all fluctuations beyond the mean activity perturbatively and therefore rely on a smooth activation function, while our results are also applicable to deterministic single units with hard thresholds.

In the path-integral formulation of Markovian neuronal network dynamics, introduced by Buice & Cowan (2007), a systematic treatment of fluctuations without resorting to *ad hoc* approximations can be done by a loop-wise expansion (Zinn-Justin, 1996, section 6.4). The first correction term in this perturbative expansion corresponds to finite-size fluctuations introduced by second-order cumulants, as discussed by Hildebrand, Buice and Chow for coupled Kuramoto oscillators (Hildebrand et al., 2007; Buice & Chow, 2007) and for more realistic spiking neuron models by Buice & Chow (2013). The Gaussian approximation of h_k presented here is equivalent to a saddle-point approximation, the lowest order of the loop-wise expansion, in the auxiliary field (see e.g. ref Sompolinsky & Zippelius, 1982, eq. (3.5) or ref Sompolinsky et al., 1988, eq. (3)). The close-to-Gaussian approximation considers a subset of third-order cumulants of the activity variables as the first finite-size correction term. It neglects contributions to the third-order cumulant of the summed input $h_i = \sum_j J_{ij} n_j$ that have three different unit indices. Only these cannot be expressed in terms of first and second cumulants of the involved units. This approximation scheme therefore includes all nontrivial dependencies between pairs of units (c_{ij}) and excludes all nontrivial dependencies between three or more units consistently.

A different approach treating the statistics of individual units in a single network realization analytically originates in the spin-glass literature. Initiated by the symmetrically coupled Sherrington-Kirkpatrick spin-glass model (Sherrington & Kirkpatrick, 1975), an expansion of the free energy in the coupling strength (Plefka-expansion (Plefka, 1982)) leads to the Thouless-Anderson-Palmer mean-field theory (Thouless et al., 1977; Nakanishi & Takayama, 1997; Tanaka, 1998; Nishimori, 2001). By construction this method is restricted to weak coupling and, in its original form, starting from the partition function, also to systems in thermodynamic equilibrium. An

extension to asymmetrically coupled nonequilibrium systems has been derived by invoking information theoretic arguments (Kappen & Spanjers, 2000). A related approach has been taken by Roudi & Hertz (2011) and Zeng et al. (2011). Still, the extended methods rely on the smoothness of the activation function and the smallness of the coupling constants, two restrictions that we are able to overcome here. Compared to previously mentioned works, the presented approach is more closely related to the mean-field theory by Mézard & Sakellariou (2011). Although they consider a system with a sequential Glauber update in discrete time steps, the method can be extended to the asynchronous update investigated here. Their analysis does not require weak coupling, similar to ours, but, in contrast, still relies on a smooth activation function for individual units. Moreover, their theory neglects the influence of the cross-covariance on the marginal statistics (their Eq. (4)), an important finite-size effect shown here to result in a distribution of mean activities due to correlations alone.

The mean-field methods employed in computer science, biology, artificial intelligence, social sciences, economics, and theoretical neuroscience (see e.g. Fernández-Gracia et al., 2014; Aljadeff et al., 2015; Kadmon & Sompolinsky, 2015) may be complemented by our results that go beyond population-averaged dynamics. The general formalism presented here, starting from the master equation, can be widely adapted by defining model-specific transition rates (see Tab. 1 in Gleeson, 2013) and particular types of couplings. The methods therefore also apply to equilibrium systems with Hamiltonian formulation, such as e.g. associative networks with parallel processing (Sollich et al., 2014), and may be useful to selectively probe the energy landscape for ergodicity breaking, to characterize the statistical features of individual local energy minima, and to assess how many minima exist in these systems. In more general terms, the formalism enables the study of the dynamics of systems with arbitrary symmetry in the couplings, and it complements their analysis in terms of other features such as the storage capacity (Alemi et al., 2015).

In neuroscience, with the advancement of electrophysiology, experimental data with hundreds of simultaneously recorded neurons have become readily available (Buzsaki, 2004; Brown et al., 2004). The availability of parallel data progressively changes the focus from the study of single cell responses to emergent phenomena arising through the interaction between neurons in networks (Stevenson & Kording, 2011). Pairwise correlations are moreover closely linked to fluctuations of the population activity (Tetzlaff et al., 2012), which have been shown to shape experimentally accessible signals, such as the local field potential or the electroencephalogram (EEG) (Lindén et al., 2011; Mazzoni et al., 2010). The effective description of the statistics of individual neurons, valid in the entire range of coupling strengths, forms the basis for studies of dynamics on adaptive networks (Gross et al., 2006), e.g., the interaction of neuronal dynamics with correlation-sensitive learning rules (Ocker et al., 2015). Explicit expressions, not only for zero time lag but also for the slope of covariance functions, allow the definition of plasticity rules resembling spike-timing dependent

plasticity (Morrison et al., 2008). The finding that the collective dynamics is captured by the nontrivial second-order statistics implies that theoretical descriptions of mechanisms, e.g. biologically realistic synaptic learning, which usually only rely on first- and second-order statistics, have now come into reach. Balanced networks show widely distributed correlations across pairs of neurons with small mean (Ecker et al., 2010). This robust feature is captured by the presented linear equations for the covariance matrix and is related to the properties of the underlying network structure. The presented description of the structure of fluctuations in this archetypical model of collective phenomena by a set of nonlinear equations and their linearized counterparts therefore provides a starting point for further analyses on the relation between structure and dynamics in disordered, coupled systems with large numbers of degrees of freedom.

Functional methods for disordered networks

In Chap. 2, we review the path-integral formulation of stochastic differential equations (Martin et al., 1973; Janssen, 1976; De Dominicis, 1976; De Dominicis & Peliti, 1978), and derive and extend mean-field methods for rate-based neural networks (Sompolinsky et al., 1988). The reformulation of the network dynamics in terms of generating functionals enables the transfer of methods from other domains such as statistical physics and quantum field theory. The generating functional is defined by an action which contains the neuronal dynamics and connectivity of the network. The form of the path integral as an integral of an exponential function enables approximation schemes which make use of the large dimensionality of the system. A saddle-point approximation of auxiliary fields which are determined by a large superposition of activities yields analytically tractable mean-field theories. Corrections due to finite network size can be taken into account by including mean connection strengths as well as fluctuations of auxiliary fields.

Mean-field theories rely on the limit of large network size N. When increasing N, the weights W_{ij} of connections must scale with the network size in order to keep the input ($\mathcal{O}(N \cdot W_{ij})$) and its fluctuations (variance $\mathcal{O}(N \cdot W_{ij}^2)$) within the dynamic range of neurons, that is the range in which the output of the neuron is sensitive to variability in the input (van Vreeswijk & Sompolinsky, 1996). A scaling of weights $W_{ij} \sim 1/N$ yields mean-driven activity with fluctuations as finite-size corrections (Montbrió et al., 2015). In a more realistic scaling of synaptic weights $W_{ij} \sim 1/\sqrt{N}$, fluctuations in the input persist even in the thermodynamic limit $N \to \infty$. Higher moments than the variance of the input remain finite-size corrections such that the input is approximately Gaussian. Fluctuation-driven activity can be achieved in this scaling when considering balanced networks, which provide a dynamic cancellation of excitation and inhibition to keep the mean input in the dynamic range and correlations between output signals low (Tetzlaff et al., 2012; Helias et al., 2013). The

asynchronous irregular behavior of cortical activity supports such fluctuation-driven activity based on a scaling of weights $W_{ij} \sim 1/\sqrt{N}$ and a suppression of mean inputs through a balance between excitation and inhibition (van Vreeswijk & Sompolinsky, 1996).

In physics, there are two types of mean-field theories which differ in the definition of auxiliary fields: In the common approach for ordered systems, which originates from the Curie-Weiss mean-field theory of ferromagnetism, auxiliary fields represent the magnetic field or the inputs $h_i = \sum_j W_{ij} x_j$ given by the activities x_j of presynaptic neurons (see e.g. Negele & Orland (1998)). The saddle-point approximation replaces these activities x_j by their most probable values, which, for spins, correspond to the mean magnetizations $m_j = \langle x_j \rangle$. Similarly, for nonlinear models, the input $h_i = \sum_j W_{ij} \phi(x_j)$, given by a nonlinear transform of the activities x_j, is replaced by its expectation value $\langle h_i \rangle = \sum_j W_{ij} \langle \phi(x_j) \rangle$.

A second type of mean-field theory, originating from the theory of spin glasses (Sherrington & Kirkpatrick, 1975), arises in the ensemble description of networks with different realizations of the disordered connectivity. In the disorder average, the variability of connections introduces additional terms in the action of the generating functional which can be simplified using a Hubbard-Stratonovich transformation. The saddle-point approximation of the corresponding auxiliary fields $Q = \frac{\sigma^2}{N} \sum_j \phi(x_j)\phi(x_j) \to \frac{\sigma^2}{N} \sum_j \langle \phi(x_j)\phi(x_j) \rangle$ takes into account fluctuations of the input as arising from variability σ^2 in connections and the temporal fluctuations of the input $\langle \phi(x_j)\phi(x_j) \rangle$ (Sompolinsky & Zippelius, 1982).

In both types of mean-field theory, the saddle-point approximations can be viewed as the first contribution (tree level) from a saddle-point expansion also called loop-wise expansion (Zinn-Justin, 1996), which, when considering higher-order terms, can systematically take into account arbitrary moments of fluctuations in auxiliary fields. The auxiliary-field formalism reduces the collective dynamics of the N interacting units to N pairwise independent units (Sompolinsky & Zippelius, 1981, 1982). This simplification has been employed in the seminal work by Sompolinsky et al. (1988), and, only recently, applied in a sequence of other studies which focus on chaotic dynamics in neural networks with multiple populations, bistable neurons or external noise (Stern et al., 2014; Aljadeff et al., 2015; Kadmon & Sompolinsky, 2015; Goedeke et al., 2016; Mastroguiseppe & Ostojic, 2016).

In Chap. 2, we detail the second type of mean-field theory discussed above by considering two special ensembles of networks: The first one deals with nonlinear networks of zero mean connection strength and thereby ensures fluctuation-driven activity for a weight scaling $W_{ij} \sim 1/\sqrt{N}$. The input term h_i therefore vanishes in the disorder average. We present the derivation of the effective equation of motion for a single unit in the thermodynamic limit which was the starting point of the analysis in Sompolinsky et al. (1988) and subsequent studies on chaotic dynamics in rate-based networks. The incorporation of non-zero mean connectivity has been done on the

level of heuristic mean-field equations (Kadmon & Sompolinsky, 2015; Mastrogiuseppe & Ostojic, 2016). A systematic derivation in the field-theoretic framework is left for future work.

Instead, we focus on finite-size effects in a second ensemble which is given by linear network models. These models are mathematically simple since path integrals reduce to ordinary integrals due to a factorization of the generating functional in frequency domain. The linearity of interactions further allows the inclusion of the input term h_i containing the mean connection strength into the propagator, that is, the Gaussian part of the theory. Consequently, the disorder-averaged dynamics are not decoupled and cross-covariances can be obtained in the saddle-point approximation. The auxiliary-field formulation is only used for the four-point vertex which arises from the variability of connections. The resulting Gaussian theory with homogeneous covariances across neurons and is equivalent to previous approaches taking into account mean interactions between neurons (Tetzlaff et al., 2012; Helias et al., 2014). Deviations from this Gaussian theory are obtained by taking into account fluctuations of saddle points through their source dependence. The latter causes non-trivial fourth-order cumulants which can be related to the dispersion of covariances across neurons (see Chap. 3). The calculation of fourth-order cumulants involves second derivatives of auxiliary fields, which can be obtained by solving the saddle-point equations recursively for Taylor coefficients of saddle points with respect to sources.

Mean-field theory has been successfully applied over decades in many fields of physics. However, in disordered systems it was so far applied to obtain self-averaging quantities, such as the phase diagram of the system or population-averaged fluctuations, i.e. quantities that are independent of the realization of the randomness in the system. The extension of mean-field theory for rate-based networks to non-Gaussian finite-size effects, is a stepping stone for analyses of arbitrary network dynamics beyond self-averaging. The linear network model presented here captures the dynamics of linear-response modes of nonlinear networks including binary networks, and is therefore applicable to a plethora of systems in various fields.

Distributions of covariances as a window into the operational regime of neuronal networks

In Chap. 3, we discuss the large variability of covariances across pairs of neurons in massively parallel spike recordings of cortical circuits. We derive a functional relation between the statistics of connections and the statistics of integral pairwise covariances between the linear-response modes of the network activity. Spin-glass mean-field techniques applied to a generating function representation for the joint probability distribution of the network activity as derived in Chap. 2 yield expressions that explain the divergence of mean covariances and their width at a critical

coupling where the linear network looses stability. In contrast to the average covariances, the spread of covariances is not suppressed by the inhibitory feedback, but arises predominantly from the structural variability of connections between neurons. Using the relations between structure and dynamics, distributions of covariances can provide insights into the properties of the underlying network and its operational regime. Experimental data from macaque motor cortex suggests linear-response dynamics close to criticality.

In contrast to the inference of individual network connections discussed in Chap. 1, the inverse problem discussed in Chap. 3 is on the level of the statistics of connections. The novelty of the results lies in the extraction of the neuron-to-neuron variability from a quenched average that normally leads to homogeneity across neurons. Covariances between individual pairs of neurons in finite-size networks are beyond the scope of mean-field theory for self-averaging quantities (Fischer & Hertz, 1991a). Still, the statistics of covariances is self-averaging for sufficiently large numbers of neurons. Wick's theorem applied to the linear network model relates second moments of covariances to fourth moments of the activity as derived from the disorder-averaged generating functional. Non-trivial fourth moments of the averaged system, i.e. those that do not factorize into second moments, therefore explain the dispersion of covariances across neurons. In this light, the mean-field theory derived in Sec. 2.3.3 goes beyond self-averaging individual covariances and captures the meta-statistics of activity.

The derived $\mathcal{O}(1/N)$ scaling of the variance of cross-covariances has already been observed in the seminal work by Renart et al. (2010). Their explanation for the width is, however, not based on an analytic derivation, but on the empirical finding that the presence or absence of individual synapses between the observed pair of neurons causes a difference in the covariance of order $\mathcal{O}(1/\sqrt{N})$. While this observation has been quantified analytically on the population-averaged level (Helias et al., 2013), the finding, however, is not the full explanation as this would otherwise only yield three possible values of covariances for either no connection, one excitatory connection, or one inhibitory connection respectively (mutual connections are unlikely in sparse networks). In the limit of small connection strengths this is indeed the case (Pernice et al., 2011), but for larger spectral radii, which is the regime of criticality investigated in Chap. 3, indirect connections matter for the distribution of covariances. The finding of the $\mathcal{O}(1/N)$ scaling of the variance in Renart et al. (2010) can finally be explained using the mean-field theory discussed in this thesis which is based on linear response theory and can thus also be applied to networks of binary neurons (see Chap. 1).

On the level of single network realizations, a large body of literature originating from the seminal work by Roudi & Hertz (2011) investigates the inference of covariances in networks of binary neurons with other tools such as the cavity method, or the Thouless-Anderson-Palmer (TAP) approach (Thouless et al., 1977). The limitations of these approaches prevent the solution of the problem addressed in this work: The

TAP approximation is an expansion for weak coupling strength (Plefka, 1982). In contrast, the large dispersion of experimentally observed covariances hints at a dynamic regime which requires strong coupling, where the TAP approach breaks down. Although this limitation is overcome in Chap. 1, we show there that binary models cannot reach the critical point because the spectral radius is bounded by $\simeq 0.8 < 1$. Rate models (Sompolinsky et al., 1988), in contrast, as well as spiking models (Ostojic, 2014) have been shown to reach the critical regime. Below the critical point, due to the bounded amplitude of fluctuations, the dominant and qualitative properties of these networks can be understood in small noise approximation, studying the linear response (Hawkes, 1971; Trousdale et al., 2012; Grytskyy et al., 2013b). Treating linear response modes in disordered systems therefore contains the non-trivial physics of the phenomenon and is at the same time transferable to the majority of network models.

The operational regime of cortical networks is a topic of ongoing debate in theoretical neuroscience (Ostojic, 2014; Harish & Hansel, 2015; Aljadeff et al., 2015; Kadmon & Sompolinsky, 2015; Mora et al., 2015). Here, we discuss criticality as the breakdown of linear stability and show its manifestation in the distributions of covariances across neurons. Previously, the latter have been mainly attributed to measurement noise (Ecker et al., 2010). Instead, we show that the bias and error of estimating the dispersion in experimental data is negligible for the current application and that broad distributions of covariances hint at a specific working point of the network that might have functional consequences (Toyoizumi & Abbott, 2011). In the critical regime, individual connections are strong enough to propagate information over multiple synaptic steps to allow complex processing, mixing and integration of information of various sources which could lead to the emergence of complex activity patterns.

The critical point is further shown to coincide with the transition between regular and chaotic dynamics in a nonlinear network model (Sompolinsky et al., 1988), if exposed to infinitesimal external noise, which drives the dynamics below this edge of chaos (Goedeke et al., 2016). In the slightly subcritical regime, changes in the effective connectivity by, for example, plasticity mechanisms or external inputs to the network may strongly affect the overall correlation structure and response properties of the network (Ginzburg & Sompolinsky, 1994; Dahmen et al., 2016a). Small perturbations may shift a macroscopic number of eigenvalues of the effective connectivity across the edge of stability (Wainrib & Touboul, 2013) such that chaotic dynamics emerge (Sommers et al., 1988; Sompolinsky et al., 1988). For networks receiving external noise with finite amplitude, the breakdown of linear stability acts as a lower boundary to the onset of chaotic dynamics (Goedeke et al., 2016).

The critical behavior in terms of chaotic dynamics is in contrast to strongly fluctuating population dynamics as observed in avalanches (Solovey et al., 2012; Mora et al., 2015; Tkacik et al., 2015). The latter emerge in excitation dominated networks ($\mu \approx 1/N$) and are caused by the unstable eigenvalue of the population-averaged activity (Grytskyy et al., 2013b, Fig. 3c).

The presented results for the distributions of covariances rely on a set of assumptions for the network connectivity **W** and noise covariance **D**. While the derivation can be generalized, for reasons of clarity we performed the calculations for a homogeneous random network with independent, identically distributed weights, uniform noise autocovariance and vanishing noise cross-covariance. The latter implies that the network receives uncorrelated input. In cortex, neurons are correlated in part by local connections and in part by overlapping and correlated inter-laminar or inter-area connections from outside the local network (Abeles, 1982; Stephan et al., 2001; Binzegger et al., 2004; Stepanyants et al., 2009) giving rise to non-vanishing off-diagonal elements of the noise covariance matrix. Furthermore, stationary mean activities vary strongly across neurons (Griffith & Horn, 1966; Koch & Fuster, 1989; Roxin et al., 2011) causing a dispersion of diagonal elements of the noise covariance matrix. Nevertheless, the ratio Δ between the variance of cross-covariances and the squared mean autocovariances is only mildly affected by the above sources of variability in noise covariances for sufficiently large spectral radii. The contribution to the correlation structure, which is generated by local recurrent connections, dominates over the direct influence of noise covariances. The latter still affects mean autocovariances and the variance of cross-covariances separately, but in constant proportion such that the effect approximately cancels in their ratio which determines the spectral radius.

The employed mean-field theory is based on a scaling of connections as $\mathcal{O}(1/\sqrt{N})$. This ensures that input fluctuations are independent of N and within the dynamic range of the neurons (van Vreeswijk & Sompolinsky, 1998). Cumulants of connection weights of order three in this scaling are hence $\mathcal{O}(N^{-3/2})$. Since mean cross-covariances as well as the variance of covariances are $\mathcal{O}(1/N)$ these higher-order cumulants only yield subleading terms, which can be neglected for large network size. The mean and variance of covariances can therefore by explained by the Gaussian features of the connectivity.

The calculations further rely on the assumption of uncorrelated connections. However, experimentally an overrepresentation of reciprocal connections has been observed in cortical networks (Song et al., 2005). Inclusion of such correlations between connections yields an additional term in the generating functional which introduces qualitatively new effects as it leads to a renormalization of response functions, known from spin glasses (Sompolinsky & Zippelius, 1982). Furthermore, symmetry in the network connectivity affects the spectral radius of the connectivity matrix (Sommers et al., 1988), which is a key component in our prediction of the operational regime. Using simulations, we show that the operational regime inferred from the experimental data is, however, robust against these symmetric connections. This finding justifies the derivations based on networks with uncorrelated connections.

The fact that the mean connection strength only yields subleading corrections to the ratio Δ which determines the spectral radius suggests that the inferred spectral radius is further robust against more complicated structure in the connectivity matrix as arising from multiple populations obeying Dale's law or spatial dependencies of

connections. We therefore hypothesize that the ratio between the variance of cross-covariances and the squared auto-covariances is universally applicable to uncover the operational regime of linear-response dynamics in neuronal networks. We have high hopes that generalizations of the mean-field formalism will turn out to be straight forward.

Integration of continuous-time dynamics in a spiking network simulator

Chap. 4 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 communications between simulation processes and significantly increase performance and scaling capabilities up to supercomputers (Morrison et al., 2005). In contrast, rate-based models interact in a continuous way. For delayed interactions, neuronal dynamics are still decoupled for 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 (Sec. 4.3.3) and illustrate possible generalizations to other categories of rate-based neuron models.

The current reference implementation uses an exponential Euler scheme (Adamu, 2011; Komori & Burrage, 2014) with a diagonal matrix A: The additive noise as well as the leaky dynamics of single neurons are exactly integrated while the input to the neurons is approximated as piecewise constant. The analysis in Sec. 4.3.1 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 τ, 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 4.2.3.3 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 easy customization to arbitrary gain functions. Researchers can easily create additional models, without the need for in-depth knowledge of simulator specific data structures or numerical methods. In addition, the developed 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 neural-network models. Furthermore, the generality of the model equations allows applications beyond neuronal networks, such as e.g. in computational gliascience (Amiri et al., 2012) or artificial intelligence (Haykin, 2009).

Mean-field theory has built a bridge between networks of spiking neurons and rate-based units that either represent single neurons or populations (Buice & Chow, 2007; Buice et al., 2010; Ostojic & Brunel, 2011; Bressloff, 2012; Grytskyy et al., 2013b). In the latter case, the rate-based approach comes along with a considerable reduction of the dimensionality. Due to a possibly large number of populations, the fixed-point solution of the stationary activity can generally not be determined analytically, but it can be found by evolving a pseudo-time dynamics. Within the presented framework, this approach is much faster than the spiking counter-part and thus can facilitate calibration of large-scale spiking network models (Schuecker et al., 2016).

The presented 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 and increases reproducibility in the validation of mean-field reductions of spiking networks to rate-based models. Furthermore, it is instructive to study whether and how the network dynamics changes with the neuron model (Brette, 2015). In particular, functional networks being able to perform a given task

are typically designed with rate-based neurons. Their validity can now be evaluated by going from a more abstract rate-based model to a biologically more realistic spiking neuron model. The present reference implementation does not allow for interactions between spiking and rate-based units. While this is technically trivial to implement, the proper conversion from spikes to rates and vice versa is a conceptual issue that has to be explored further by theoretical neuroscience.

The presented joined platform for spike-based and rate-based models hopefully triggers new research questions by facilitating collaboration and translation of ideas between scientists working in the two fields. This work therefore contributes to the unification of both modeling routes in multi-scale approaches combining large-scale spiking networks with functionally inspired rate-based elements to decipher the dynamics of the brain.

Hybrid scheme for modeling local field potentials from spiking point-neuron networks

In Chap. 5, we discuss the methodology underlying the hybrid scheme for modeling local field potentials from point-neuron networks. A main motivation for developing the hybrid LFP modeling scheme was to obtain the ability to compute LFPs for a key class of network models that are amenable to mathematical analysis and can provide intuitive understanding of emerging network dynamics, namely point-neuron models. Similar to networks of anatomically and biophysically detailed neuron models, point-neuron networks can generate realistic spiking activity. In addition, the hybrid modeling scheme brings a substantial computational advantage: With present-day computing and software technologies, point-neuron networks with $\sim 100,000$ neurons can be modeled with laptop computers, and networks comprising millions of point neurons can be routinely simulated on high-performance compute facilities (Helias et al., 2012; Kunkel et al., 2014). Until now, the largest simulation of LFPs based on networks of multicompartmental neuron models with reconstructed morphologies, in contrast, comprised about $12,000$ neurons and was done on a Blue Gene/P supercomputer with 4096 CPUs (Reimann et al., 2013). The linearity of electromagnetic theory allows for the implementation of the LFP hybrid modeling scheme as an "embarrassingly" parallel operation (Foster, 1995). Therefore, the results for the cortical microcircuit application with ~78,000 neurons were obtained with only 256 CPUs, and could even be acquired with much smaller computing architectures.

A full implementation of the hybrid scheme is provided by the freely available `Python` module `hybridLFPy` (http://github.com/INM-6/hybridLFPy). Our model implementation in `hybridLFPy` for calculating single-cell LFP contributions relies on the publicly available `Python` software package `LFPy` (Lindén et al., 2014) with

NEURON (Carnevale & Hines, 2006) as a simulation backend. This ensures flexibility and compatibility with a large library of existing neuron models, with or without active channels and with morphologies of arbitrary levels of detail, obtained from ModelDB (Hines et al., 2004), NeuroMorpho.org (Ascoli et al., 2007) or other resources. While we did use NEST (Eppler et al., 2015) for simulating our reference network, the hybridLFPy module can be used in combination with any other neural-network simulation software. The present hybrid LFP scheme involves several assumptions with respect to (i) the generation of realistic spiking activity, (ii) forward modeling of extracellular potentials, and (iii) the combined use of point-neuron networks and multicompartment modeling. In the following we review the main assumptions and discuss potential extensions:

(i) *Spike-train generation by point-neuron networks:* Although highly simplified, single-compartment models of individual neurons (point-neuron models) can mimic realistic spiking for a variety of cell types (Izhikevich, 2003; Kobayashi et al., 2009; Yamauchi et al., 2011) and can make accurate predictions of single-cell firing responses under in-vivo like conditions (Jolivet et al., 2008; Gerstner & Naud, 2009). Moreover, networks of point neurons can reproduce a number of activity features observed *in vivo*, such as spike-train irregularity (Softky & Koch, 1993; van Vreeswijk & Sompolinsky, 1996; Amit & Brunel, 1997; Shadlen & Newsome, 1998), membrane-potential fluctuations (Destexhe & Paré, 1999), asynchronous firing (Ecker et al., 2010; Renart et al., 2010; Ostojic, 2014), correlations in neural activity (Gentet et al., 2010; Okun & Lampl, 2008; Helias et al., 2013), self-sustained activity (Ohbayashi et al., 2003; Kriener et al., 2014) and realistic firing rates across laminar cortical populations (Potjans & Diesmann, 2014). Recently Rössert et al. (2016) demonstrated that the multicompartment-neuron network of Markram et al. (2015) could be reduced to a point-neuron network using an automated step-by-step workflow without significantly affecting the spiking statistics in the network. Nevertheless, under circumstances where simple point neurons may fail to produce plausible spike patterns for a particular input, for example in the presence of highly nonlinear shaping of post-synaptic responses or dendritic computation (reviewed e.g., by London & Häusser (2005)), more sophisticated point neurons (Breuer et al., 2014) may be required. However, the present hybrid scheme makes no assumptions on generators of spiking activity. In principle, network spikes could, as an example, be replaced by statistical models of spike generation (Lindén et al., 2011; Łęski et al., 2013), multicompartment neuron spikes or even experimentally measured spiking activity.

(ii) *Biophysical forward modeling of LFPs:* The biophysical forward model described by Eq. (5.23), implemented in LFPy (Lindén et al., 2014), underlies the presently used computational scheme for LFPs of point-neuron networks. This forward model is based on well-established volume conductor theory (Rall & Shepherd, 1968; Holt & Koch, 1999) and assumes an *infinite*, *isotropic* (same in all directions), *homogeneous* (same in all positions) and *ohmic* (frequency-independent) extracellular medium represented by a scalar conductivity σ_e. However, one could generalize the forward model in a straightforward manner to account for anisotropy (Nicholson & Freeman,

1975; Logothetis et al., 2007; Goto et al., 2010), or jumps in conductivities at tissue interfaces for example using the so-called 'method of images' (Pettersen et al., 2006; Gold et al., 2006; Hagen et al., 2015c; Ness et al., 2015). For even more complicated geometrical spatial variations of the conductivity, the forward modeling problem can always be solved by means of Finite Element Modeling (FEM) (Ness et al., 2015). Recent experiments have only found a small frequency dependence of the extracellular conductivity σ_e at LFP frequencies ($f \lesssim 500$Hz, Logothetis et al. (2007); Wagner et al. (2014)), but see Gabriel et al. (1996, 2009); Bédard & Destexhe (2009). In any case the forward model could still be applied with frequency-dependent conductivity by means of Fourier decomposition where each frequency component of the LFP signal is considered separately. For more information on possible generalizations of the biophysical forward-modeling scheme, see Pettersen et al. (2012). Finally, we assumed the so-called disc-electrode approximation and averaged the computed LFP signal across the electrode surface (Moulin et al., 2008; Lindén et al., 2014; Ness et al., 2015). Although electrode impedance will affect the measurement, it appears that confounding effects from this can easily be avoided with present-day LFP recording techniques (Nelson & Pouget, 2010), hence few compelling reasons exist to incorporate additional temporal filters.

(iii) *Combined use of point-neuron and multicompartmental models:* The key approximation in the hybrid LFP scheme comes from the combined use of point-neuron (single-compartment) and multicompartmental neuron models. The multicompartment neurons are mutually unconnected, have no outgoing (efferent) connections, and are solely used to compute the LFP. Further, due to dendritic filtering, the somatic postsynaptic potentials in the multicompartment model neurons are not identical to those of their point-neuron counterparts. This inconsistency could, at least partially, be resolved by adjusting the amplitudes and temporal shapes of the synaptic currents in either the multicompartment neurons or the point neurons (Koch & Poggio, 1985; Wybo et al., 2013, 2015).

The hybrid scheme is not limited to the example point-neuron network model and the particular multicompartment neuron models chosen here. It can be applied to networks (i) of arbitrary topology (graph structure, distance dependencies, dimensions), (ii) with any number of populations, (iii) with arbitrarily complex point-neuron (e.g. LIF, Izhikevich, MAT, Hodgkin-Huxley) and synapse dynamics (e.g., current-based, conductance-based, static, plastic), and (iv) any level of biophysical detail in multicompartment neuron models (e.g., morphologies, active channels).

While we here have focused on the computation of the LFP based on output from spiking point-neuron networks, a similar hybrid approach could be used when the network dynamics is rather modeled in terms of firing rates or even neural fields (Deco et al., 2008). For our example case of a four-layered cortical network with an excitatory and an inhibitory population in each layer, the present scheme could be adapted directly by replacing the set of spike trains for each population with the corresponding population firing rates (Schuecker et al., 2015) in the LFP-generating

step. An investigation of the feasibility and prediction accuracy of such a scheme is facilitated by our implementation of rate models in the spiking network simulator NEST (Chap. 4).

Another natural development would be to consider other measurement modalities. The present LFP scheme already incorporates the prediction of ECoG (electrocorticography) signals, i.e., the electrical signals recorded at the cortical surface, although the LFP forward-modeling scheme may have to be adjusted to account for the discontinuity in electrical conductivity at the cortical surface (Nunez & Ramesh, 2006). One straightforward method to account for discontinuities in our predictions is the so-called "method of images", as previously used with our present forward modeling scheme (Hagen et al., 2015c; Ness et al., 2015). An extension to EEG (electroencephalography) and MEG (magnetoencephalography) would in principle also be straightforward as the key variable linking single-neuron activity and the measured signal is the single-neuron current dipole moment (Hämäläinen et al., 1993; Nunez & Ramesh, 2006). This dipole moment can be computed from multicompartmental neuron models when the transmembrane or axial currents are known (Lindén et al., 2010; Ahlfors & Wreh II, 2015). Given the magnitude and orientation of the current dipole moments for all contributing neurons, the EEG and MEG signal can be computed by a linear superposition of single-cell contributions given an appropriate extracellular volume conductor model for the EEG signal. EEG dipole-source localization techniques and EEG forward models have sometimes assumed simplified spherical head models composed of N concentric shells, each with different conductivity (Nunez & Ramesh, 2006). Such models could be used also in the present hybrid scheme as well as realistic head models (Kaiboriboon et al., 2012). Note that MEG signal prediction from the current dipole moment is simplified due to its independence on conductivity, and that the variation in magnetic permeability is insignificant across soft tissue, bone and air (Hämäläinen et al., 1993). Another measurement that could be modeled is voltage-sensitive dye imaging (VSDi) where the signal largely reflects average membrane potentials of dendrites close to the cortical surface. The spatial profile of the weights in the averaging procedure of the VSDi forward-model will be determined by the spatial distribution of dye and the propagation of light in the neural tissue (Chemla & Chavane, 2010; Tian et al., 2011).

Local field potential of a cortical column

In Chap. 6, we discuss the results of the application of the hybrid scheme to simulate the LFP in a cortical column. For illustration, we used the hybrid scheme to compute LFPs along a virtual laminar multielectrode from activity in a multilayered spiking point-neuron network, modeling signal processing in a patch of primary visual cortex. The network consisted of ~78,000 neurons organized in four layers, each with an excitatory and an inhibitory population, representing a cortical patch of ~1 mm^2

(Potjans & Diesmann, 2014). Altogether 16 different cell types and 10 different morphologically reconstructed neurons were used in the LFP calculation. This cortical microcircuit model is well-suited for the illustration of the hybrid scheme due to its (i) minimum level of detail in single-neuron dynamics of both point neurons (LIF) and multicompartment neurons (passive membranes), (ii) realistic neuron density allowing investigation of the effects of correlations and scaling of network size, and (iii) its spatial organization of multiple populations across cortical layers which yields cancellation effects not captured by LFP proxies such as in Mazzoni et al. (2015).

Even though the example application was based on a generic network model biased towards cat visual cortex and not tuned to address specific experiments, its spiking activity nevertheless matched experimental findings on distributions of firing rates, asynchronous irregular activity and propagation of evoked activity (Potjans & Diesmann, 2014). We even observed the predicted LFP to be in qualitative accordance with LFP measurements in primary sensory cortices from a variety of animal species and sensory modalities in terms of (i) LFP amplitude, for both spontaneous ($\simeq$ 0.1 – 1 mV, see Fig. 4, Hagen et al. (2015c, Fig. 7a), Reyes-Puerta et al. (2016, Fig. 2A)), electrically stimulated activity ($\sim$ 1 – 3 mV, see Mitzdorf (1985, Fig. 3c), Castro-Alamancos & Connors (1996, Fig. 3)) and (ii) stimulus-evoked spatiotemporal LFP and CSD patterns (Di et al. (1990, Fig. 1), Einevoll et al. (2007, Fig. 1A), Reyes-Puerta et al. (2014, Fig. 1D-F)). The fact that our spontaneous and evoked LFP amplitudes are comparable to the observed amplitudes supports the overall biological plausibility of the hybrid LFP scheme.

For the present example the LFP was dominated by synaptic inputs, and their associated return currents, on excitatory neurons, in particular onto pyramidal cells in layers 2/3 and 6. Further, contributions from inhibitory synaptic inputs typically dominated the contributions from the excitatory inputs, particularly for LFPs stemming from spontaneous network activity. Although the main point of employing the present example was to illustrate the use of the hybrid LFP scheme and not to make predictions for specific neural systems, we note in passing that a dominance of inhibitory synaptic inputs appears to be in agreement with LFPs generated in the CA3 region of hippocampus as observed in an *in vitro* setting (Bazelot et al., 2010) and in cortex of both humans and macaque *in vivo* (Telenczuk et al., 2016). In accordance with a previous study (Lindén et al., 2011) we found that correlations in synaptic input play a major role in determining the CSD and LFP stemming from the network activity, for both spontaneous and stimulus-evoked activity. We further showed that due to inevitable correlations between synaptic input currents the main features of the LFP can only be correctly predicted by a full-scale model. As our ambition is to compute LFPs also for extended point-neuron networks with millions of neurons covering, for example, entire cortical areas, we finally demonstrated how the hybrid modeling scheme can predict LFPs from population firing rates rather than from spikes of individual neurons (Einevoll et al., 2007).

The present microcircuit model application involving simplified, passive multicompartment populations was used here to study the effect of the spatial connectivity on

the laminar pattern of spontaneous and stimulus-evoked CSD and LFP signals. The application thus represented a minimal approach incorporating spatial features in LFP predictions of multilaminar point-neuron networks. However, several of the simplifying model assumptions made in the present example application can straightforwardly be generalized. In particular, such generalizations concern (i) the synaptic connectivity between point neurons and the equivalent multicompartment neurons, (ii) the absence of active conductances, (iii) the positioning of the cells, (iv) the reconstructed morphologies, (v) the representation of external inputs, and (vi) the fact that the model only encompassed the local circuitry of a $\sim$ 1 mm^2 patch of cortex.

(i) *Synaptic connectivity:* While the population-specific connection probabilities, delay distributions, synapse time constant and mean synaptic weights were identical for the connections in the point-neuron network and those between point neurons and LFP-generating multicompartment neurons, the exact realizations of the two types of connectivities were different. In contrast to the point-neuron network, each cell of a particular type had a fixed in-degree, i.e., a fixed number of synaptic inputs, a fixed synaptic current amplitude and a layer-specific synapse positioning that depended on compartment surface area, in the LFP modeling step. The use of the hybrid LFP scheme, however, is not restricted to these or any other specific assumptions about the synaptic connectivity patterns. One could, for example, gather all point-neuron network connections and corresponding weights and delays for use with the LFP-generating multicompartment neuron populations. However, for large networks this would require additional computing and memory resources as the number of recorded connection weights and delays grow proportionally to N^2 in a network of N neurons with fixed connection probabilities. The present layer-specific synapse positioning could also be specialized further, e.g., to account for dendrite-specific connections that may be relevant for LFP predictions in particular if nonlinear synapse and dendrite models are used.

(ii) *Active conductances:* In the present study we neither included the active channels underlying spike generation, nor active dendritic conductances in the multicompartment neuron models (Remme & Rinzel, 2011). Experiments suggest that the contribution to the LFP from the former is small in stimulus-evoked recordings from sensory cortex, at least for the low frequencies of the LFP (Pettersen et al., 2008), but see Ray & Maunsell (2011). In any case the contribution of the spikes to the extracellular potentials, including the LFP, could be included in the present scheme by giving each spike produced in the point-neuron network simulations a cell-type-specific spatiotemporal signature in the computation of the extracellular potential (e.g., as calculated in Holt & Koch (1999); Hagen et al. (2015c)). Substantial effects of active dendritic conductances on the LFP were observed in a two-layered model by Reimann et al. (2013), but this should be further explored. Recent modeling results (Ness et al., 2016) suggest that for the purposes of LFP prediction, active dendritic conductances can, at least for subthreshold potentials, be effectively described by means of "quasi-active linearized" theory (Sabah & Leibovic, 1969; Koch, 1984;

Remme & Rinzel, 2011; Ness et al., 2016). This simplifies the LFP modeling substantially and makes the computation similar in complexity to the present case with purely passive membranes.

(iii) *Soma positioning:* For model conciseness, the somas of all neurons belonging to the same specific cortical network populations were set to have the same range across cortical depth, cf. Fig. 6.1E. By instead assuming a biologically more plausible distribution of soma depths, the CSD profiles are expected to be spatially smoothed compared to the profiles seen for the present distributions, e.g., in Fig. 6.1F. Also the LFP profiles will be affected, but to a lesser degree since the LFP profiles already are spatially smoothed due to volume conduction effects.

(iv) *Reconstructed morphologies:* We further chose to rely on a small number of highly detailed reconstructed dendritic morphologies from experimental preparations, and partly reuse morphologies across neural populations. Obviously, larger sets of distinct reconstructed dendritic morphologies can be used in future studies as they become available. The effect of dendritic morphologies on generated LFP can be further assessed by use of stylized (Tomsett et al., 2014; Głąbska et al., 2014) or artificially grown morphologies resembling real neurons (Cuntz et al., 2010, 2011; Torben-Nielsen & De Schutter, 2014; Mazzoni et al., 2015). While our morphologies were used as is, effects on the predicted LFP from varying morphologies could be quantified using our present measures of LFP power spectra and variance across depth. Such measures could also be extended to include effects of lateral displacement of the recording electrode (Lindén et al., 2011; Einevoll et al., 2013; Łęski et al., 2013; Tomsett et al., 2014), but also cross-channel coherences and stLFPs should be considered as well.

(v) *External input:* In our point-neuron network modified from Potjans & Diesmann (2014), we assumed the depolarizing input from surrounding cortex and remote areas to be represented by deterministic, fixed-amplitude input currents (DC inputs) rather than independent Poisson spike trains. We thus avoided the generation and storage of $\mathcal{O}(10^5)$ high-rate uncorrelated Poisson spike trains and distributing the corresponding spike events onto the multicompartment model neurons, but this can be introduced in future applications.

(vi) *Scale:* Our network model represents only an isolated cortical column under $\sim$ 1mm^2 of pial surface (Potjans & Diesmann, 2014), but work is underway to extend the network to a larger scale, e.g., by incorporating additional cortical areas (Schmidt et al., 2016a) and extending the network size in the lateral directions (Senk et al., 2015). In addition to allowing for predictions of LFPs at several lateral positions as measured by multishank electrodes, the anticipated outcome is also an altered network dynamics and consequently altered LFPs. The present columnar model lacks structured input from other parts of cortex which corresponds to about 50 % of all excitatory synapses (see Potjans & Diesmann (2014) and references therein). Multi-scale models combining microscopic spike-based descriptions for the local circuit and effective mesoscopic rate-based descriptions for remote populations could provide the

local circuit with more realistic input. An implementation of such models in NEST is enabled by the implementation of rate models discussed in Chap. 4. Accounting for different cortical areas and their interactions would, for example, allow for the detailed investigation of pathway-specific LFP contributions as recently reviewed in hippocampus by Herreras et al. (2015).

Synaptic inputs to neurons are typically correlated, and as shown in this thesis and in earlier studies (Lindén et al., 2011; Łęski et al., 2013), these correlations have a major impact on the properties of the generated LFP (and corresponding CSD). There are different contributions to these input correlations, shared presynaptic neurons and correlations in the presynaptic spiking activity, see, e.g., Renart et al. (2010), Tetzlaff et al. (2012), and correct LFP predictions require that both effects are properly taken into account. The generation of presynaptic spike trains with realistic correlation structure requires networks of realistic size (van Albada et al., 2015). The contribution from shared presynaptic input can only be accounted for by using realistic statistics of inputs to the LFP-generating multicompartment model neurons, i.e., realistic synaptic connection probabilities resulting in realistic statistics of shared inputs. If these two requirements are fulfilled, the properties of single-cell LFPs as well as the correlations between pairs of single-cell LFPs, will be correctly accounted for.

Given synaptic inputs with realistic statistics from sufficiently large point-neuron networks, do we actually need to represent the full population of LFP generating neurons in order to predict a realistic compound LFP? Or can we alternatively get a good estimate of the compound population-LFP signal from a downscaled population of neuronal LFP generators if we know the correct single-cell and pairwise LFP statistics, thereby reducing computational costs? The answer is negative: In the presence of (even tiny) synaptic-input correlations, a realistic compound LFP can only be generated by multicompartmental neuron populations with realistic size and cell densities.

The hybrid modeling scheme allows one to account for both, i.e., realistic sizes of networks to generate spike trains with correct correlation structure and realistic sizes and cell densities of the LFP-generating multicompartmental neuron populations. This can be achieved since point-neuron networks can be simulated very efficiently, and the multicompartmental neurons are independent and can be simulated serially (or in an embarrassingly parallel manner).

Outlook

The recording of extracellular potentials inside the brain is among the oldest measures of neural activity used in neuroscience. Recently the local field potential has

seen a revival, partially due to the rapid technological developments of new multielectrodes and possible applications in, for example, development of brain-machine interfaces, but also from the realization that the LFPs offer a window into activity at the level of neuronal populations. However, the interpretation of the recorded data has so far largely been qualitative (Einevoll et al., 2013). The computation of LFPs from network activity has until now been too cumbersome and computer-intensive to allow for practical exploration of the links between different types of network dynamics and the resulting LFP. Thus a validation of network models against measured LFP data has essentially been absent. With the present hybrid LFP scheme, accompanied by the simulation tool `hybridLFPy`, we believe that a significant step has been taken towards the goal of making combined modeling and measurement of the LFP signal a practical research tool for probing neural circuit activity.

Recent technological advances further allow the recording of massively parallel spike trains showing large heterogeneity at the level of individual neurons (Ecker et al., 2010). Going beyond population-averaged activity, this thesis provides theoretical tools which enable first interpretations of such data. The formulation of the dynamical system in terms of generating functionals allows systematic approximation schemes to derive consistent extensions of mean-field theories. The qualitative step from mean-field theories for self-averaging quantities to the assessment of variability beyond the self-averaging component, as derived both on the level of ensembles and single realizations of the connectivity, is universally applicable to linear-response activities of disordered systems. The developed theoretical methods therefore also pave the way to study the relation between the structure and dynamics of disordered systems in a variety of other disciplines such as physics, computer science, artificial intelligence, social sciences, and economics.

The forward model employed in the hybrid scheme is only one approach to linking brain activities at different scales. It is based on the knowledge of the network dynamics on the microscopic scale where interactions between neurons are known. The derivation of effective equations of motion for the dynamics on the mesoscopic level to date has proceeded in an ad-hoc manner by performing a spatio-temporal coarse graining of neuronal activities (Deco et al., 2008). A more systematic approach based on functional methods (Chow & Buice, 2015) follows the ideas of spin-glass mean-field techniques (Sompolinsky & Zippelius, 1981) and has recently received renewed attention. The reformulation of dynamical equations in a functional formalism follows a long and successful history in physics that allows unification of interpretations and methods across disciplines. One important example is given by renormalization techniques that are tailored to connect physical descriptions on different energy or lengths scales. Originating from ideas of Wilson (1975), the renormalization group is designed to systematically relate microscopic and macroscopic descriptions by successively incorporating finer scales into effective parameters of the next coarser levels (Gies, 2006; Metzner et al., 2012). The path-integral formulation is the natural framework for this method, which is ubiquitously employed in condensed matter physics

but so far rare in theoretical neuroscience (Steyn-Ross & Steyn-Ross, 2016). One reason for this is the lack of direct applicability of field-theoretic methods to spiking models. Further work needs to be done to generalize the methods to spiking networks in order to make the semi-systematic linear-response approximations obsolete. The renormalization group may then prove to be a powerful tool to systematically link the dynamics on the microscopic and mesoscopic scale by deriving effective equations of motion for mass signals such as the LFP. The integration of continuous-time dynamics in spiking network simulators such as NEST forms the basis for joint simulations of brain dynamics on both scales in a common framework.

This thesis follows a connectionist modeling approach which relies on the description of emergent phenomena from the architecture of the network. We focused on specific features of the correlation structure such as the low mean and large dispersion, as well as the local field potential in a cortical column. The developed methods form the basis for further investigations on the required complexity of neuron and connectivity models to explain experimentally observed features of activity on the microscopic and mesoscopic scale and the question about how important the details of individual neurons are if exposed to numerous inputs from large balanced networks.

Part VI

Appendices

Appendix A

Correlated fluctuations in strongly coupled binary networks beyond equilibrium

A.1 Derivation of moment equations up to second order

For completeness and to establish a consistent notation, here we include the derivation of equations Eq. (1.2) for the first and second moments of the activity in a binary network. We follow the notation introduced in Buice et al. (2010), which has been adapted to binary networks in Helias et al. (2014).

The stochastic system is completely characterized by the joint probability distribution $p(\mathbf{n})$ of all N binary variables $\mathbf{n}$. Knowing the joint probability distribution, arbitrary moments can be calculated, among them pairwise correlations. The occupation probability of each state follows the master equation Eq. (A.1). We denote as $\mathbf{n}_{i+} = (n_1, \ldots, n_{i-1}, 1, n_{i+1}, \ldots, n_N)$ the state, where the ith neuron is active ($n_i = 1$), and $\mathbf{n}_{i-}$ where neuron i is inactive ($n_i = 0$). Since in each infinitesimal time interval at most one neuron can change state, for each given state $\mathbf{n}$ there are N possible transitions (each corresponding to one of the N neurons changing state). The sum of the probability fluxes into the state and out of the state must equal the change of probability in the respective state (Kelly, 1979), i.e.

$$\tau \frac{\partial p(\mathbf{n})}{\partial t} = \sum_{i=1}^{N} (2n_i - 1) \left(p(\mathbf{n}_{i-}) F_i(\mathbf{n}_{i-}) - p(\mathbf{n}_{i+})(1 - F_i(\mathbf{n}_{i+})) \right) \quad \forall \mathbf{n} \in \{0,1\}^N. \tag{A.1}$$

From this equation we derive expressions for the first $\langle n_k \rangle$ and second moments $\langle n_k n_l \rangle$ by multiplying with $n_k n_l$ and summing over all possible states $\mathbf{n} \in \{0,1\}^N$, which leads to

$$\tau \frac{\partial}{\partial t} \langle n_k n_l \rangle = \sum_{\mathbf{n} \in \{0,1\}^N} \sum_{i=1}^{N} n_k n_l (2n_i - 1) \underbrace{\left(p(\mathbf{n}_{i-}) F_i(\mathbf{n}_{i-}) - p(\mathbf{n}_{i+})(1 - F_i(\mathbf{n}_{i+})) \right)}_{\equiv G_i(\mathbf{n} \setminus n_i)},$$

where we denote as $\langle f(\mathbf{n})\rangle = \sum_{\mathbf{n}\in\{0,1\}^N} p(\mathbf{n}) f(\mathbf{n})$ the average of a function $f(\mathbf{n})$ with respect to the distribution $p(\mathbf{n})$. Note that the term denoted $G_i(\mathbf{n}\backslash n_i)$ does not depend on the state of neuron i. We use the notation $\mathbf{n}\backslash n_i$ for the state of the network excluding neuron i, i.e. $\mathbf{n}\backslash n_i = (n_1, \ldots, n_{i-1}, n_{i+1}, \ldots, n_N)$. Separating the terms in the sum over i into those with $i \neq k, l$ and the two terms with $i = k$ and $i = l$, we obtain

$$\begin{aligned}
\tau \frac{\partial}{\partial t}\langle n_k n_l\rangle &= \sum_{\mathbf{n}} \sum_{i=1, i\neq k,l}^{N} n_k n_l (2n_i - 1)\, G_i(\mathbf{n}\backslash n_i) \\
&\qquad + n_k n_l (2n_k - 1)\, G_k(\mathbf{n}\backslash n_k) \\
&\qquad + n_k n_l (2n_l - 1)\, G_l(\mathbf{n}\backslash n_l) \\
&= \sum_{i=1, i\neq k,l}^{N} \sum_{\mathbf{n}\backslash n_i} n_k n_l \left(G_i(\mathbf{n}\backslash n_i) - G_i(\mathbf{n}\backslash n_i)\right) \\
&\quad + \sum_{\mathbf{n}} n_k n_l\, G_k(\mathbf{n}\backslash n_k) + \sum_{\mathbf{n}} n_k n_l\, G_l(\mathbf{n}\backslash n_l),
\end{aligned}$$

where we obtained the first term by explicitly summing over state $n_i \in \{0, 1\}$ (i.e. using $\sum_{\mathbf{n}\in\{0,1\}^N} = \sum_{\mathbf{n}\backslash n_i \in \{0,1\}^{N-1}} \sum_{n_i=0}^{1}$ and evaluating the sum $\sum_{n_1=0}^{1}$). This first sum obviously vanishes. The remaining terms are of identical form with the roles of k and l interchanged. We hence only consider the first of them and obtain the other by symmetry. The first term simplifies to

$$\begin{aligned}
\sum_{\mathbf{n}} n_k n_l\, G_k(\mathbf{n}\backslash n_k) &\overset{n_k=1}{=} \sum_{\mathbf{n}\backslash n_k} n_l\, G_k(\mathbf{n}\backslash n_k) \\
&\overset{\text{def. } G_k}{=} \begin{cases} \sum_{\mathbf{n}\backslash n_k} p(\mathbf{n}_{k-})\, F_k(\mathbf{n}_{k-}) & \\ \qquad + p(\mathbf{n}_{k+})\, F_k(\mathbf{n}_{k+}) - p(\mathbf{n}_{k+}) & \text{for } k = l \\ \sum_{\mathbf{n}\backslash n_k} p(\mathbf{n}_{k-})\, n_l\, F_k(\mathbf{n}_{k-}) & \\ \qquad + p(\mathbf{n}_{k+})\, n_l\, F_k(\mathbf{n}_{k+}) - n_l\, p(\mathbf{n}_{k+}) & \text{for } k \neq l \end{cases} \\
&= \begin{cases} \langle F_k(\mathbf{n})\rangle - \langle n_k\rangle & \text{for } k = l \\ \langle F_k(\mathbf{n})\, n_l\rangle - \langle n_k n_l\rangle & \text{for } k \neq l \end{cases}. \qquad \text{(A.2)}
\end{aligned}$$

Taken together with the mirror term $k \leftrightarrow l$, we arrive at two conditions, one for the first ($k = l$, $\langle n_k^2\rangle = \langle n_k\rangle$) and one for the second ($k \neq l$) moment

$$\tau \frac{\partial}{\partial t}\langle n_k\rangle \overset{k=l}{=} -\langle n_k\rangle + \langle F_k(\mathbf{n})\rangle, \qquad \text{(A.3)}$$

$$\tau \frac{\partial}{\partial t}\langle n_k n_l\rangle \overset{k\neq l}{=} -2\langle n_k n_l\rangle + \langle F_k(\mathbf{n})\, n_l\rangle + \langle F_l(\mathbf{n})\, n_k\rangle. \qquad \text{(A.4)}$$

The time-lagged correlation function can be derived along completely analogous lines as Eq. (A.4), as the forward time evolution equation (differential equation with respect to t) of the two-point probability distribution $p(\mathbf{n}, t, \mathbf{q}, s)$ fulfills, because of the Markov property, the same master equation Eq. (A.1) as the equal-time probabil-

ity distribution $p(\mathbf{n}, t)$. The resulting differential equation reads

$$\begin{aligned}\tau\frac{\partial}{\partial t}\langle n_k(t)n_l(s)\rangle &\equiv \tau\frac{\partial}{\partial t}\sum_{\mathbf{n},\mathbf{q}} p(\mathbf{n},t,\mathbf{q},s)\, n_k q_l \\ &= -\langle n_k(t)n_l(s)\rangle + \langle F_k(\mathbf{n}(t))\, n_l(s)\rangle. \end{aligned} \tag{A.5}$$

A.2 Cumulants of summed random variables

Let

$$y = \sum_{i=1}^{N} x_i, \tag{A.6}$$

with x_i random variables that follow an arbitrary distribution $p(\mathbf{x})$. The moment generating functions of y and $\mathbf{x}$ are then related by

$$\begin{aligned}\varphi_y(t) &= \sum_{\mathbf{x}} p(\mathbf{x})e^{ty} = \langle e^{ty}\rangle_{\mathbf{x}} \\ &= \langle e^{t\sum_{i=1}^{N} x_i}\rangle_{\mathbf{x}} \\ &= (\varphi_{\mathbf{x}} \circ \iota)(t), \end{aligned} \tag{A.7}$$

where concatenation with the function $\iota : t \mapsto (t, \ldots, t)$ replaces every t_i by a t. The cumulant generating functions therefore follow as

$$\begin{aligned}\Phi_y(t) &= \ln \varphi_y(t) \\ &= \ln(\varphi_{\mathbf{x}} \circ \iota)(t) = (\Phi_{\mathbf{x}} \circ \iota)(t), \end{aligned} \tag{A.8}$$

with $\Phi_{\mathbf{x}}(\mathbf{t}) = \ln\langle e^{\mathbf{t}\cdot\mathbf{x}}\rangle$ the cumulant generating function of the $\mathbf{x}$ and $\mathbf{t}\cdot\mathbf{x}$ the scalar product. From the expansion in cumulants $\kappa^x_{ij\ldots}$ for $\mathbf{x}$ and $\kappa_{1,2,\ldots}$ for y, we get the following relationship

$$\begin{aligned}\Phi_y(t) &\equiv \sum_{l=1}^{\infty}\frac{\kappa_l}{l!}t^l \\ &= (\Phi_{\mathbf{x}} \circ \iota)(t) \\ &= \left[\left(\sum_{i=1}^{N}\kappa_i^x t_i + \sum_{i,j=1}^{N}\frac{\kappa_{ij}^x}{2!}t_i t_j + \ldots\right)\circ\iota\right](t) \\ &= \underbrace{\sum_{i=1}^{N}\kappa_i^x}_{\kappa_1}\, t + \frac{1}{2!}\underbrace{\sum_{i,j=1}^{N}\kappa_{ij}^x}_{\kappa_2}\, t^2 + \ldots, \end{aligned} \tag{A.9}$$

so the cumulants of the summed variable κ_1, κ_2, etc are given by the sums of the cumulants of the individual variables of corresponding order.

A.3 Functions of close-to Gaussian variables

To determine the mean activity, we need to apply a nonlinear function f to a variable h_k that has a statistics close to Gaussian. For brevity, we will suppress the index k in the following. We assume that the Fourier transform $\hat{f}(\omega)$ exists. Let y be the random variable (A.6), which is the sum of a large number of individual variables x_i and is assumed to be close to Gaussian. For the expectation value $\langle f(y)\rangle_{\mathbf{x}}$ we get

$$\begin{aligned}\langle f(y)\rangle_{\mathbf{x}} &= \frac{1}{2\pi}\int d\omega\, \hat{f}(\omega)\,\langle e^{i\omega y}\rangle_{\mathbf{x}} \\ &= \frac{1}{2\pi}\int d\omega\, \hat{f}(\omega)\, e^{\Phi_y(i\omega)} \\ &= \frac{1}{2\pi}\int d\omega\, \hat{f}(\omega)\, e^{\kappa_1(i\omega)+\frac{1}{2!}\kappa_2(i\omega)^2+\frac{1}{3!}\kappa_3(i\omega)^3+\ldots},\end{aligned} \tag{A.10}$$

where Φ_y is the cumulant generating function of y (A.8) and κ_i denotes the ith cumulant of y. We are interested in an approximation that treats the dominant first and second (Gaussian) cumulants of y explicitly and separate the effect of all higher cumulants by writing

$$\begin{aligned}\langle f(y)\rangle_{\mathbf{x}} &= e^{\frac{1}{6}\kappa_3\left(\frac{\partial}{\partial\kappa_1}\right)^3+\ldots}\frac{1}{2\pi}\int d\omega\, \hat{f}(\omega)\, e^{\kappa_1(i\omega)+\frac{1}{2}\kappa_2(i\omega)^2} \\ &= e^{\frac{1}{6}\kappa_3\left(\frac{\partial}{\partial\kappa_1}\right)^3+\ldots}\langle f(y)\rangle_{y\sim\mathcal{N}(\kappa_1,\kappa_2)},\end{aligned} \tag{A.11}$$

where we identified in the last line the Fourier transform of a Gaussian with moments κ_1 and κ_2 via (A.10).

For the covariance we need to evaluate terms of the form $\langle F_k n_l\rangle$. These terms can be obtained analogously as

$$\begin{aligned}\langle f(y)x_l\rangle_{\mathbf{x}} &= \frac{1}{2\pi}\int d\omega\, \hat{f}(\omega)\,\langle x_l\, e^{i\omega y}\rangle_{\mathbf{x}} \\ &= \frac{1}{2\pi}\int d\omega\, \hat{f}(\omega)\,\left(\left(\partial_{t_l}\varphi_{\mathbf{x}}\right)\circ\iota\right)(i\omega) \\ &= \frac{1}{2\pi}\int d\omega\, \hat{f}(\omega)\,\left(\left(\left[\partial_{t_l}\Phi_{\mathbf{x}}\right]e^{\Phi_{\mathbf{x}}}\right)\circ\iota\right)(i\omega),\end{aligned}$$

where $\varphi_{\mathbf{x}}$ is the moment generating function of $\mathbf{x}$ (A.7). The derivative of the cumu-

lant generating function with (A.9) becomes

$$\begin{aligned}\partial_{t_l}\Phi_{\mathbf{x}}(\mathbf{t}) &= \kappa_l^x + \sum_{j=1}^{N}\kappa_{lj}^x t_j + \frac{1}{2}\sum_{i,j=1}^{N}\kappa_{lij}^x t_i t_j + \ldots,\\ \left(\partial_{t_l}\Phi_{\mathbf{x}}\circ\iota\right)(i\omega) &= \kappa_l^x + \underbrace{\sum_{j=1}^{N}\kappa_{lj}^x}_{=:\Delta\kappa_{1,l}}(i\omega) + \frac{1}{2}\underbrace{\sum_{i,j=1}^{N}\kappa_{lij}^x}_{=:\Delta\kappa_{2,l}}(i\omega)^2 + \ldots = \sum_{q=0}^{\infty}\Delta\kappa_{q,l}(i\omega)^q,\end{aligned}$$

where the product rule produced a factor 2 in the second term and a factor 3 in the third term and we used the symmetry of the cumulants with respect to permutations of indices. So we get

$$\begin{aligned}\langle x_l f(y)\rangle_{\mathbf{x}} &= \frac{1}{2\pi}\int d\omega\,\hat{f}(\omega)\sum_{q=0}^{\infty}\Delta\kappa_{q,l}(i\omega)^q e^{\sum_{p=0}^{\infty}\kappa_p(i\omega)^p}\\ &= \left[\sum_{q=0}^{\infty}\Delta\kappa_{q,l}\frac{\partial}{\partial\kappa_q}\right]e^{\frac{1}{3!}\kappa_3\left(\frac{\partial}{\partial\kappa_1}\right)^3+\ldots}\langle f(y)\rangle_{y\sim\mathcal{N}(\kappa_1,\kappa_2)}\\ &= \left[\sum_{q=0}^{\infty}\Delta\kappa_{q,l}\frac{1}{q!}\left(\frac{\partial}{\partial\kappa_1}\right)^q\right]\langle f(y)\rangle_{\mathbf{x}},\end{aligned}\tag{A.12}$$

where we replaced $(i\omega)^q$ by derivatives of the exponential with respect to cumulants and identified $\langle f(y)\rangle_{\mathbf{x}}$ in the last step using (A.11).

A.4 Trivial third- and fourth-order cumulants of binary variables expressed by lower-order cumulants

Third order. Let $n_l, n_i, n_j, n_r \in [0,1]$ be binary variables. The raw third moment can be written as a sum of all combinations of cumulants up to order three (1.20) $\langle n_l n_i n_j\rangle = \langle\!\langle n_l n_i n_j\rangle\!\rangle + c_{li}m_j + c_{ij}m_l + c_{jl}m_i + m_l m_i m_j$. Using $n_i^K = n_i$ for each integer $K \geq 1$, in the case of binary variables n_i we consider the two cases

$$
\begin{aligned}
l = i \neq j: \quad \langle n_l n_j \rangle &= \langle\!\langle n_l n_l n_j \rangle\!\rangle + c_{ll} m_j + c_{lj} m_l + c_{jl} m_l + m_l^2 m_j, \\
\langle\!\langle n_l n_l n_j \rangle\!\rangle &= \underbrace{\langle n_l n_j \rangle}_{c_{lj} + \langle n_l \rangle \langle n_j \rangle} - c_{ll} m_j - 2 c_{lj} m_l - m_l^2 m_j \\
&= c_{lj}(1 - 2m_l) + \underbrace{(-c_{ll} + m_l - m_l^2)}_{=0} m_j \\
&= c_{lj}(1 - 2m_l), \\
l = i = j: \quad \langle n_l \rangle &= \langle\!\langle n_l n_l n_l \rangle\!\rangle + 3 c_{ll} m_l + m_l^3, \\
\langle\!\langle n_l n_l n_l \rangle\!\rangle &= m_l - 3 m_l (1 - m_l) m_l - m_l^3 \\
&= m_l - 3 m_l^2 + 2 m_l^3 \\
&= c_{ll}(1 - 2m_l),
\end{aligned}
\tag{A.13}
$$

which together yield the expression (1.21) in Chap. 1.

The third cumulant (1.17) of h_k follows from the assumption $\langle\!\langle n_i n_j n_r \rangle\!\rangle \simeq 0$ for $i \neq j \neq r$ as

$$
\begin{aligned}
\kappa_{3,k} &= \sum_{ijr} J_{ki} J_{kj} J_{kr} \langle\!\langle n_i n_j n_r \rangle\!\rangle \\
&= \underbrace{\sum_{i \neq j \neq r}}_{N(N-1)(N-2) \text{ terms}} J_{ki} J_{kj} J_{kr} \underbrace{\langle\!\langle n_i n_j n_r \rangle\!\rangle}_{\simeq 0} \\
&\quad + 3 \underbrace{\sum_{i = j \neq r}}_{N(N-1) \text{ terms}} J_{ki}^2 J_{kr} \langle\!\langle n_i n_i n_r \rangle\!\rangle + \underbrace{\sum_{i=j=r}}_{N \text{ terms}} J_{ki}^3 \langle\!\langle n_i n_i n_i \rangle\!\rangle \\
&\simeq 3 \sum_{i=j \neq r} J_{ki}^2 J_{kr} c_{ir} (1 - 2m_i) + \sum_{i=j=r} J_{ki}^3 c_{ii} (1 - 2m_i) \\
&= 3 \sum_{i,r} J_{ki}^2 (1 - 2m_i) c_{ir} J_{kr} - 2 \sum_i J_{ki}^3 c_{ii} (1 - 2m_i),
\end{aligned}
\tag{A.14}
$$

where we included the term $i = r$ in the first sum and compensated accordingly in the latter term. With $c_{ii}(1 - 2m_i) = m_i - 3m_i^2 + 2m_i^3$ we obtain the expression (1.22) in Chap. 1. Analogously, it follows that

$$\begin{aligned}
\sum_{i,j=1}^{N} J_{ki}J_{kj}\langle\!\langle n_l n_i n_j\rangle\!\rangle &= \underbrace{\sum_{i\neq j\neq l} J_{ki}J_{kj}\langle\!\langle n_l n_i n_j\rangle\!\rangle}_{\simeq 0} + \sum_{l\neq i=j=1}^{N} J_{ki}^2\langle\!\langle n_l n_i n_i\rangle\!\rangle \\
&\quad + \sum_{i=l\neq j=1}^{N} J_{kl}J_{kj}\langle\!\langle n_l n_l n_j\rangle\!\rangle + \sum_{j=l\neq i=1}^{N} J_{ki}J_{kl}\langle\!\langle n_l n_i n_l\rangle\!\rangle \\
&\quad + J_{kl}^2\langle\!\langle n_l n_l n_l\rangle\!\rangle \\
&= \sum_{l\neq i=1}^{N} J_{ki}^2 c_{li}(1-2m_i) + 2J_{kl}\sum_{l\neq i=1}^{N} J_{ki}c_{li}(1-2m_l) \\
&\quad + J_{kl}^2 c_{ll}(1-2m_l) \\
&= \sum_{i=1}^{N} J_{ki}^2 (1-2m_i)c_{li} + 2J_{kl}\,(1-2m_l)\sum_{i=1}^{N} J_{ki}c_{li} \\
&\quad - 2J_{kl}^2\, c_{ll}(1-2m_l), \qquad \text{(A.15)}
\end{aligned}$$

which yields the expression (1.23) in Chap. 1.

Fourth order. The cumulant of fourth order is

$$\begin{aligned}
\langle n_i n_j n_r n_l\rangle &= \langle\!\langle n_i n_j n_r n_l\rangle\!\rangle + \langle\!\langle n_i n_j n_r\rangle\!\rangle m_l + \langle\!\langle n_j n_r n_l\rangle\!\rangle m_i + \langle\!\langle n_r n_l n_i\rangle\!\rangle m_j + \langle\!\langle n_l n_i n_j\rangle\!\rangle m_r \\
&\quad + c_{ij}m_r m_l + c_{ir}m_j m_l + c_{il}m_r m_j + c_{jr}m_l m_i + c_{jl}m_r m_i + c_{rl}m_i m_j \\
&\quad + c_{ij}c_{rl} + c_{ir}c_{jl} + c_{il}c_{rj} + m_i m_j m_r m_l,
\end{aligned}$$

Only those cumulants in which at most two different indices appear are fixed by first- and second-order cumulants. We need to distinguish three cases. The first case is $i = j \neq r = l$ and leads, with (A.13), to the matrix

$$\begin{aligned}
\{\langle\!\langle n_i n_i n_r n_r\rangle\!\rangle_{ir}\} &= \mathbf{C}\circledast(\mathbf{1}-2\mathbf{C}) - \mathrm{diag}(\mathbf{C})\,\mathrm{diag}(\mathbf{C})^T \\
&\quad - 2\,\mathrm{diag}(\mathbf{1}-2\mathbf{m}))\,\mathbf{C}\,\mathrm{diag}(\mathbf{m}) \\
&\quad - 2\,\mathrm{diag}(\mathbf{m})\,\mathbf{C}\,\mathrm{diag}(\mathbf{1}-2\mathbf{m}) \\
&\quad - \mathrm{diag}(\mathbf{C})\,(\mathbf{m}\circledast\mathbf{m})^T - (\mathbf{m}\circledast\mathbf{m})\,\mathrm{diag}(\mathbf{C})^T \\
&\quad - 4\,\mathrm{diag}(\mathbf{m})\,\mathbf{C}\,\mathrm{diag}(\mathbf{m}) \\
&\quad + \mathbf{m}\mathbf{m}^T\circledast(\mathbf{1}-\mathbf{m}\mathbf{m}^T). \qquad \text{(A.16)}
\end{aligned}$$

The second case $i = j = l \neq r$ yields the matrix

$$\begin{aligned}\{\langle\!\langle n_i n_i n_i n_r \rangle\!\rangle_{ir}\} &= \mathrm{diag}(\mathbf{1} - 3\,\mathrm{diag}(\mathbf{C}))\,\mathbf{C} \\ &\quad -\mathrm{diag}(\mathbf{C})\mathbf{m}^T - 3\,\mathrm{diag}(\mathbf{m})\,\mathbf{C} \\ &\quad -(\mathbf{m} \circledast \mathrm{diag}(\mathbf{C}))\;\mathbf{m}^T + 3\,\mathrm{diag}(\mathbf{m} \circledast \mathbf{m})\,\mathbf{C} \\ &\quad +\mathrm{diag}(\mathbf{1} - \mathbf{m} \circledast \mathbf{m}))\,\mathbf{m}\mathbf{m}^T \end{aligned} \tag{A.17}$$

and the third for $i = j = r = l$ yields a vector

$$\begin{aligned}\langle\!\langle n_i n_i n_i n_i \rangle\!\rangle &= \mathbf{m} \circledast (\mathbf{1} - (\mathbf{m}\circledast)^3) - 4\,\mathrm{diag}(\mathbf{C}) \circledast \mathbf{m} + 2\,\mathrm{diag}(\mathbf{C}) \circledast (\mathbf{m}\circledast)^2 \\ &\quad -3(\mathrm{diag}(\mathbf{C})\circledast)^2, \end{aligned} \tag{A.18}$$

where we use the notation$(\mathbf{x}\circledast)^n = \mathbf{x} \circledast \ldots \circledast \mathbf{x}$ for the element-wise nth power of a vector. In Eq. (1.19) we need the term

$$\begin{aligned}\Delta\kappa_{3,kl} &= \sum_{i,j,r=1}^{N} J_{ki} J_{kj} J_{kr} \langle\!\langle n_i n_j n_r n_l \rangle\!\rangle \\ &\simeq 3 \sum_i J_{ki}^2 J_{kl} \langle\!\langle n_i n_i n_l n_l \rangle\!\rangle \qquad (i{=}j,r{=}l),(i{=}r,j{=}l),(i{=}l,j{=}r) \\ &\quad + \sum_i J_{ki}^3 \langle\!\langle n_i n_i n_i n_l \rangle\!\rangle \qquad (i{=}j{=}r{\neq}l) \\ &\quad +3 \sum_i J_{ki} J_{kl}^2 \langle\!\langle n_l n_l n_l n_i \rangle\!\rangle \qquad (j{=}r{=}l{\neq}i),(r{=}l{=}i{\neq}j),(l{=}i{=}j{\neq}r) \\ &\quad + J_{kl}^3 \langle\!\langle n_l n_l n_l n_l \rangle\!\rangle \qquad (i{=}j{=}r{=}l), \end{aligned}$$

which in matrix form gives rise to the expression in Eq. (1.24) in Chap. 1.

Appendix B

Distributions of covariances as a window into the operational regime of neuronal networks

B.1 Bias and error of estimating the dispersion in experimental data

Here, we derive the dependence of experimentally estimated moments of the activities on the number of neurons and trials. Due to limited number of trials, the estimator for the dispersion of the covariance may be biased towards larger variances. We seek to find a correction for this bias here, as illustrated in Fig. B.1a. The derivation here follows the standard approach of correcting the estimation, sometimes called Bessel's correction (Barlow, 2013).

We assume a probability distribution $p(\mathbf{n}^1, ..., \mathbf{n}^{N_T})$ of activities n_i^k of neuron i in trial $k \in \{1, ..., N_T\}$. Since there are no correlations between different trials, the probability distribution factorizes and the moment generating function reads

$$\phi(\boldsymbol{l}^1, ..., \boldsymbol{l}^{N_T}) = \prod_{k=1}^{N_T} \exp\left(\mathbf{m}^{\mathrm{T}} \boldsymbol{l}^k + \frac{1}{2} \boldsymbol{l}^{k,\mathrm{T}} \mathbf{c} \boldsymbol{l}^k\right), \tag{B.1}$$

where $\mathbf{m}$ is the vector of mean activities and $\mathbf{c}$ is the covariance matrix, and higher-order correlations are neglected. We would like to obtain estimators for the mean and dispersion of the parameters $\mathbf{m}$ and $\mathbf{c}$, the meta-statistics. In line with our findings in Sec. 2.3.3, we assume the meta-statistics to be Gaussian. We denote as $\langle m_i \rangle =: \mu_i = \mu$, $\langle c_{ij} \rangle =: \sigma_{ij} = \sigma_a \delta_{ij} + \sigma_c(1 - \delta_{ij})$ the average values and as $\langle (m_i - \mu)^2 \rangle =: \delta\mu_i^2$ and $\langle (c_{ij} - \sigma_{ij})^2 \rangle =: \delta\sigma_{ij}^2 = \delta\sigma_a^2 \delta_{ij} + \delta\sigma_c^2(1 - \delta_{ij})$ the variances of the mean activities and covariances, where δ_{ij} refers to the Kronecker symbol.

Averaging Eq. (B.1) over this metastatistics introduces cross-couplings between trials due to varibilities of mean and covariances as well as fourth-order cumulants that

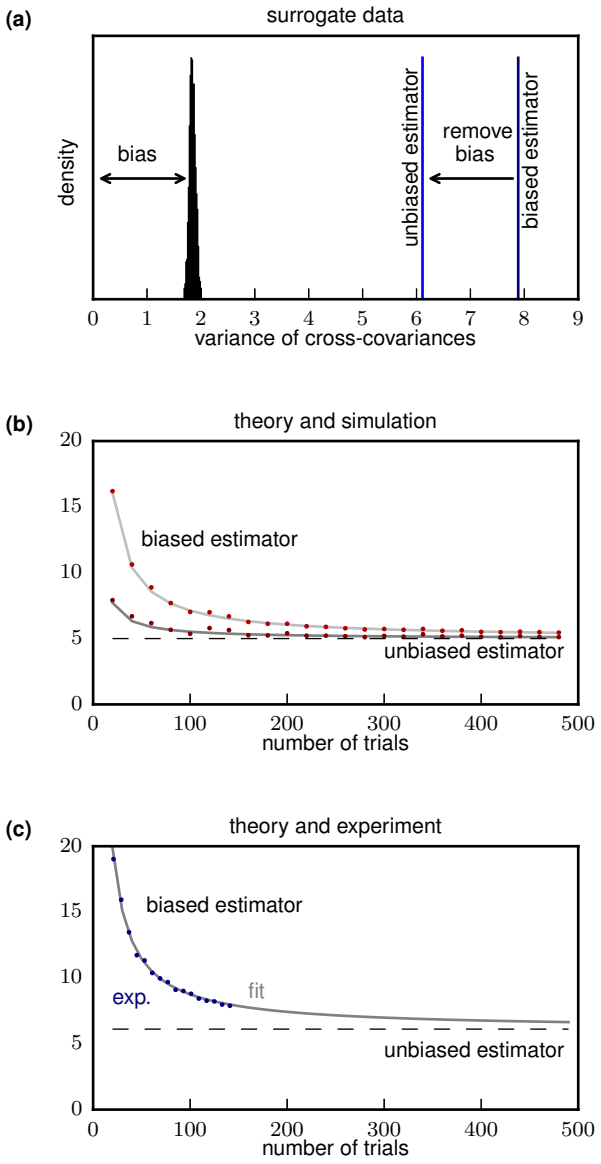


Figure B.1: Bias in estimator of variance of cross-covariances. (a) Variance of cross-covariances (black distribution) obtained from trial-shuffled data compared to the biased estimator (Eq. (B.2), dark blue) and unbiased estimator (light blue) of the true variance as obtained from experimental data. (b) Convergence of biased estimator for small (dark gray) and large (light gray) mean covariances towards true variance given by unbiased estimator (dashed line). Curves indicate the theoretical prediction Eq. (B.2), dots numerical simulations for artificial data. (c) True value 6.11 of variance of cross-covariances (dashed line) derived from fitting (gray curve) Eq. (B.2) to the biased estimator (blue dots) obtained from subsampled experimental data with different number of trials.

arise from the variability in covariances

$$\bar{\phi}(\boldsymbol{l}^1,...,\boldsymbol{l}^{N_T}) = \exp\left(\sum_{k=1}^{N_T}\left(\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{l}^k + \frac{1}{2}\boldsymbol{l}^{k,\mathrm{T}}\boldsymbol{\sigma}\boldsymbol{l}^k\right) + \sum_{k,l=1}^{N_T}\left(\frac{1}{2}\delta\mu^2\boldsymbol{l}^{k,\mathrm{T}}\boldsymbol{l}^l + \frac{1}{8}\left(\boldsymbol{l}^k \circledast \boldsymbol{l}^l\right)^{\mathrm{T}}\delta\boldsymbol{\sigma}^2\left(\boldsymbol{l}^k \circledast \boldsymbol{l}^l\right)\right)\right),$$

where $\circledast$ denotes the Hadamard product, i.e. $\left[\boldsymbol{l}^k \circledast \boldsymbol{l}^l\right]_i = l_i^k l_i^l$. It is straight-forward to show and well known that an empirical covariance $c_{ij}^{\mathrm{e}} = \frac{1}{N_T}\sum_{k=1}^{N_T}\left(n_i^k - m_i^{\mathrm{e}}\right)\left(n_j^k - m_j^{\mathrm{e}}\right)$ yields a biased estimator

$$\left\langle c_{ij}^{\mathrm{e}}\right\rangle = \frac{N_T - 1}{N_T}\sigma_{ij}$$

of the true covariance σ_{ij} (Barlow, 2013). Here, m_i denotes the empirical mean activity of neuron i across trials $m_i^{\mathrm{e}} = \frac{1}{N_T}\sum_{k=1}^{N_T} n_i^k$. The bias results from a finite sample size N_T. An unbiased definition of the empirical covariance therefore reads

$$c_{ij}^{\mathrm{e}} = \frac{1}{N_T - 1}\sum_{k=1}^{N_T}\left(n_i^k - m_i^{\mathrm{e}}\right)\left(n_j^k - m_j^{\mathrm{e}}\right).$$

Using this definition, along the same lines a lengthy, but straight-forward analogous calculation shows that the variance of cross-covariances ($i \neq j$) defined as $\overline{\delta c_{ij}^2} = \frac{1}{N(N-1)}\sum_{i\neq j=1}^{N}(c_{ij}^{\mathrm{e}} - \overline{c_{ij}})^2$ with the mean cross-covariance across neurons $\overline{c_{ij}} = \frac{1}{N(N-1)}\sum_{i\neq j=1}^{N} c_{ij}^{\mathrm{e}}$ yields the biased estimator

$$\left\langle\overline{\delta c_{ij}^2}\right\rangle = \left(1 - \frac{2}{N(N-1)}\right)\left(\delta\sigma_c^2 + \frac{\sigma_a^2 - \sigma_c^2}{N_T - 1}\right) \tag{B.2}$$

of the true variance of cross-covariances $\delta\sigma_{ij}^2$. For a finite number of trials, there is significant bias of $\left\langle\overline{\delta c_{ij}^2}\right\rangle$ caused primarily by the mean autocovariance σ_a of spike-counts across trials (see Fig. B.1a for an empirical estimate of the bias with trial-shuffled data).

The aim is hence to correct for the bias of the estimator. This can be formally done using Eq. (B.2), which we validate for synthetic data (Fig. B.1b). A fit of Eq. (B.2) to the variance of cross-covariances obtained from subsampled numbers of trials of the experimental data shows that the true variance of cross-covariances can be approximately obtained by subtracting the mean of the distribution of the surrogate data in Fig. B.1a.

In conclusion, the finite number of trials yields a significant bias to the variance of cross-covariances, but not to the mean autocovariances, if properly defined (see above). Fitting data to Eq. (B.2) and extrapolating for $N_T \to \infty$ yields the unbiased estimate used in Chap. 3. However, even neglecting this correction, the order of magnitude of the ratio between the two that determines the spectral radius (Eq. (3.8) in Chap. 3) and the operational regime is not changed by the bias.

In addition to the bias in the estimation of the variance, the estimator comes with a statistical uncertainty due to the limited numer of recorded neurons. We observe $n = N(N-1)$ cross-covariances of which we want to estimate the true variance. The relative standard error of the variance is therefore given by $\sqrt{2/(n-1)} \approx 0.009 < 1\%$ (Lehmann & Casella, 2006). Error propagation to the estimation of R^2 shows that the relative error of R^2 is even smaller for $R^2 > 1/3$, and hence in the critical regime close to $R^2 \simeq 1$ this error is negligible.

B.2 Effect of inhibitory feedback on cross-covariances

The structural coefficients

$$\begin{aligned}\gamma_{ij} &= \delta_{ij} + \gamma = \delta_{ij} + \frac{2\mu}{1-N\mu} + \left(\frac{\mu}{1-N\mu}\right)^2 N \\ \chi_{ij} &= \frac{1}{N}\gamma_{ii}\left[2\gamma_{ij}\sum_k \gamma_{ik}\gamma_{kj} + (1-\delta_{ij})\gamma_{jj}\sum_k \gamma_{ik}\gamma_{ki}\right]\end{aligned}$$

in Eq. (3.6) and Eq. (3.7) depend on the deterministic network structure, i.e. the network size N and the mean connection strength μ. Input fluctuations that are independent of network size and within the dynamic range of the neurons require $\mu \sim -1/\sqrt{N}$ in inhibitory networks (van Vreeswijk & Sompolinsky, 1998, section 10.2) which causes γ and therefore mean cross-covariances to scale as $1/N$. The inhibitory feedback, however, does not influence the dispersion of cross-covariances which can be understood by a mode decomposition of fluctuations.

In the following we give an intuitive argument, why the width of the distribution of cross-covariances is less suppressed than the mean cross-covariance. To this end consider a set of N coupled Ornstein-Uhlenbeck processes $\mathbf{x}(t)$ driven by pairwise independent white noises ξ_i with $\langle \xi_i(t)\xi_j(s)\rangle = \delta_{ij}\delta(t-s)\,D$

$$\tau\frac{dx_i(t)}{dt} = -x_i(t) + \sum_{j=1}^{N} W_{ij}x_j(t) + \xi_i(t). \tag{B.3}$$

As a first step we transform the system into a basis constructed from the population-averaged fluctuation $\bar{x}(t) = \frac{1}{N}\sum_{i=1}^{N} x_i(t) = \frac{1}{N}(1,1,..,1)\cdot\mathbf{x}(t) := \bar{\mathbf{v}}\cdot\mathbf{x}(t)$ and orthogonal fluctuations $y_\alpha(t) = \mathbf{v}_\alpha\cdot\mathbf{x}(t)$ with $\mathbf{v}_\alpha\cdot\mathbf{v}_\beta = N\delta_{\alpha\beta}$ $(\alpha,\beta \in \{0,..,N-1\})$, $\mathbf{v}_0 = \bar{\mathbf{v}}$. Furthermore, we assume the connectivity $\mathbf{W}$ to be fluctuating around the mean value μ

$$W_{ij} = \mu + \delta W_{ij}. \tag{B.4}$$

Population-averaged covariances are given by the autocovariance of the population-averaged fluctuation

$$\begin{aligned}\frac{1}{N^2}\sum_{i,j=1}^{N} c_{ij}(t,t') &= \frac{1}{N^2}\sum_{i,j=1}^{N}\langle x_i(t)x_j(t')\rangle \\ &= \left\langle\left(\frac{1}{N}\sum_{i=1}^{N} x_i(t)\right)\left(\frac{1}{N}\sum_{j=1}^{N} x_j(t')\right)\right\rangle \\ &= \langle \bar{x}(t)\bar{x}(t')\rangle.\end{aligned}$$

A decomposition into auto- and cross-covariances

$$\begin{aligned}\frac{1}{N^2}\sum_{i,j=1}^{N} c_{ij}(t,t') &= \frac{1}{N}\overline{c_{ii}}(t,t') + \frac{N-1}{N}\overline{c_{ij}}(t,t') \\ &\approx \frac{1}{N}\overline{c_{ii}}(t,t') + \overline{c_{ij}}(t,t')\end{aligned}$$

shows that population-averaged covariances $\frac{1}{N^2}\sum_{i,j=1}^{N} c_{ij}(t,t')$ can only be small if average cross-covariances $\overline{c_{ij}}(t,t')$ are small. The population fluctuation $\bar{x}(t)$ evolves according to

$$\begin{aligned}\tau\frac{d\bar{x}(t)}{dt} &= -\bar{x}(t) + \sum_{j=1}^{N}\underbrace{\left(\frac{1}{N}\sum_{i=1}^{N} W_{ij}\right)}_{\mu+O(N^{-\frac{1}{2}})} x_j(t) + \bar{\xi}(t) \\ &\approx -(1-N\mu)\bar{x}(t) + \bar{\xi}(t),\end{aligned}$$

where we defined $\bar{\xi}(t) = \frac{1}{N}\sum_{i=1}^{N}\xi_i(t)$ with $\langle\bar{\xi}(t)\bar{\xi}(s)\rangle = \frac{D}{N}\delta(t-s)$. The population fluctuations therefore have the integral covariance

$$\int d\tau\langle\bar{x}(t)\bar{x}(t+\tau)\rangle = \frac{D}{N\,(1-N\mu)^2},$$

where the term $(1-N\mu)^{-2}$ reflects the strong suppression by the inhibitory feedback in large networks ($\mu \sim -1/\sqrt{N}$) (Tetzlaff et al., 2012). The orthogonal fluctuations

($\alpha \neq 0$) follow

$$\begin{aligned}
\tau \frac{dy_\alpha(t)}{dt} &= -y_\alpha(t) + \sum_{j=1}^{N} \left(\sum_{i=1}^{N} v_\alpha^i W_{ij} \right) x_j(t) + \xi_\alpha(t) \\
&= -y_\alpha(t) + \sum_{j=1}^{N} \underbrace{\sum_{i=1}^{N} v_\alpha^i}_{\propto \mathbf{v}_\alpha \cdot \bar{\mathbf{v}} = 0} \mu\, x_j(t) + \sum_{j=1}^{N} \underbrace{\left(\sum_{i=1}^{N} v_\alpha^i \delta W_{ij} \right)}_{\delta W_{\alpha j}} x_j(t) + \xi_\alpha(t) \\
&= -y_\alpha(t) + \sum_{j=1}^{N} \delta W_{\alpha j}\, x_j(t) + \xi_\alpha(t),
\end{aligned}$$

are insensitive to the inhibitory feedback μ, and just depend on projections of fluctuations of the connectivity $\delta W_{\alpha j} = [\mathbf{v}_\alpha \cdot \delta \mathbf{W}]_{\alpha_j}$. The second moment of the population-averaged covariances depends on all projections of fluctuations $y_i(t)$

$$\begin{aligned}
\frac{1}{N^2} \sum_{i,j=1}^{N} c_{ij}^2(t,t') &= \frac{1}{N^2} \sum_{i,j=1}^{N} \langle x_i(t) x_j(t') \rangle \langle x_i(t) x_j(t') \rangle \\
&= \frac{1}{2N^2} \sum_{i,j=1}^{N} \langle x_i(t) x_j(t') x_i(t) x_j(t') \rangle - \langle x_i(t) x_i(t) \rangle \langle x_j(t') x_j(t') \rangle \\
&= \frac{1}{2} \left\langle \left(\frac{1}{N} \sum_{i=1}^{N} x_i(t) x_i(t) \right) \left(\frac{1}{N} \sum_{j=1}^{N} x_j(t') x_j(t') \right) \right\rangle \\
&\quad - \frac{1}{2} \left\langle \frac{1}{N} \sum_{i=1}^{N} x_i(t) x_i(t) \right\rangle \left\langle \frac{1}{N} \sum_{j=1}^{N} x_j(t') x_j(t') \right\rangle \\
&= \frac{1}{2} \left\langle \left(\frac{1}{N} \sum_{i=1}^{N} y_i(t) y_i(t) \right) \left(\frac{1}{N} \sum_{j=1}^{N} y_j(t') y_j(t') \right) \right\rangle \\
&\quad - \frac{1}{2} \left\langle \frac{1}{N} \sum_{i=1}^{N} y_i(t) y_i(t) \right\rangle \left\langle \frac{1}{N} \sum_{j=1}^{N} y_j(t') y_j(t') \right\rangle,
\end{aligned}$$

where we used Wick's theorem in the second step and Parceval's theorem to replace the terms $\frac{1}{N}\mathbf{x}(t) \cdot \mathbf{x}(t) = \frac{1}{N}\mathbf{y}(t) \cdot \mathbf{y}(t)$ in the last step, valid due to the choice of the unitary transform. The dispersion of covariances is therefore not suppressed by the inhibitory feedback.

Note that for excitatory networks, stable dynamics requires a scaling of connection weights with $1/N$, implying both the mean and standard deviation of cross-covariances to scale as $1/N$.

B.3 Effect of heterogeneous stationary mean activities

The calculations in Chap. 2 and Chap. 3 assume uniform stationary mean activities. These correspond to uniform autocovariances $D_{ii} = D$ of the noisy input in the model

of Ornstein-Uhlenbeck processes (Eq. (B.3)). For these networks, the spectral radius R of the effective connectivity $\mathbf{W}$ is given by the width of the distribution of spike-count cross-covariances normalized by the mean spike-count autocovariances (Eq. (3.8) in Chap. 3), which remove the influence of the noise covariance D on cross-covariances. Here, we investigate the influence of distributed stationary mean activities on the estimation of the spectral radius.

Fig. B.2f shows the experimentally obtained distribution of spike-count autocovariances. The large width of this distribution, which is on the same order of magnitude as the mean spike-count autocovariance, arises from distributed stationary mean activities (Fig. B.2b). This effect gets evident by using a geometric series expansion $(\mathbf{1}-\mathbf{W})^{-1} = \mathbf{1} + \sum_{n=1}^{\infty} \mathbf{W}^n$ for the expression of integral covariances (Eq. (3.1) in Chap. 3) (Pernice et al., 2011, 2012; Trousdale et al., 2012)

$$\begin{aligned} c_{ij} &= \left[(\mathbf{1}-\mathbf{W})^{-1}\, \mathbf{D}\, \left(\mathbf{1}-\mathbf{W}^{\mathrm{T}}\right)^{-1} \right]_{ij} \\ &= D_{ij} + \underbrace{\left[\sum_{n=1}^{\infty} \mathbf{W}^n\, \mathbf{D} \right]_{ij} + \left[\mathbf{D} \sum_{n=1}^{\infty} \mathbf{W}^{\mathrm{T}\,n} \right]_{ij} + \left[\sum_{n=1}^{\infty} \mathbf{W}^n\, \mathbf{D} \sum_{m=1}^{\infty} \mathbf{W}^{\mathrm{T}\,m} \right]_{ij}}_{=:c_{ij}^{\mathrm{loc}}} . \end{aligned} \tag{B.5}$$

Since the noise covariance enters linearly in the equation, mean covariances are unaffected by variability in noise covariances (Fig. B.2a,c). Individual autocovariances c_{ii} are primarily determined by D_{ii}. Variability of the latter therefore almost exclusively determines the variance of autocovariances for a large range of spectral radii R (Fig. B.2b). Since $D_{ij} = 0$ for $i \neq j$, this leading order effect is absent for cross-covariances. Their variance and the prediction of the spectral radius (Eq. (3.8) in Chap. 3) are therefore determined by the locally generated autocovariance c_{ij}^{loc} (see Eq. (B.5)) and only mildly affected by heterogeneity in stationary mean activities (Fig. B.2d,e).

B.4 Effect of correlated input

The theory in Chap. 2 and Chap. 3 assumes an isolated local network of N neurons. In cortex, neurons are correlated in part by local connections and in part by overlapping and correlated inter-laminar or inter-area connections from outside the local network (Abeles, 1982; Binzegger et al., 2004; Stepanyants et al., 2009). While the effect of correlated input on mean covariances was investigated in Helias et al. (2014), we here also analyze the effect on distributions of covariances. In case of external input, the dynamics of the local network is then given by

$$\frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} W_{ij} x_j + \sum_{j=1}^{N^{\mathrm{ext}}} M_{ij} y_j + \xi_i,$$

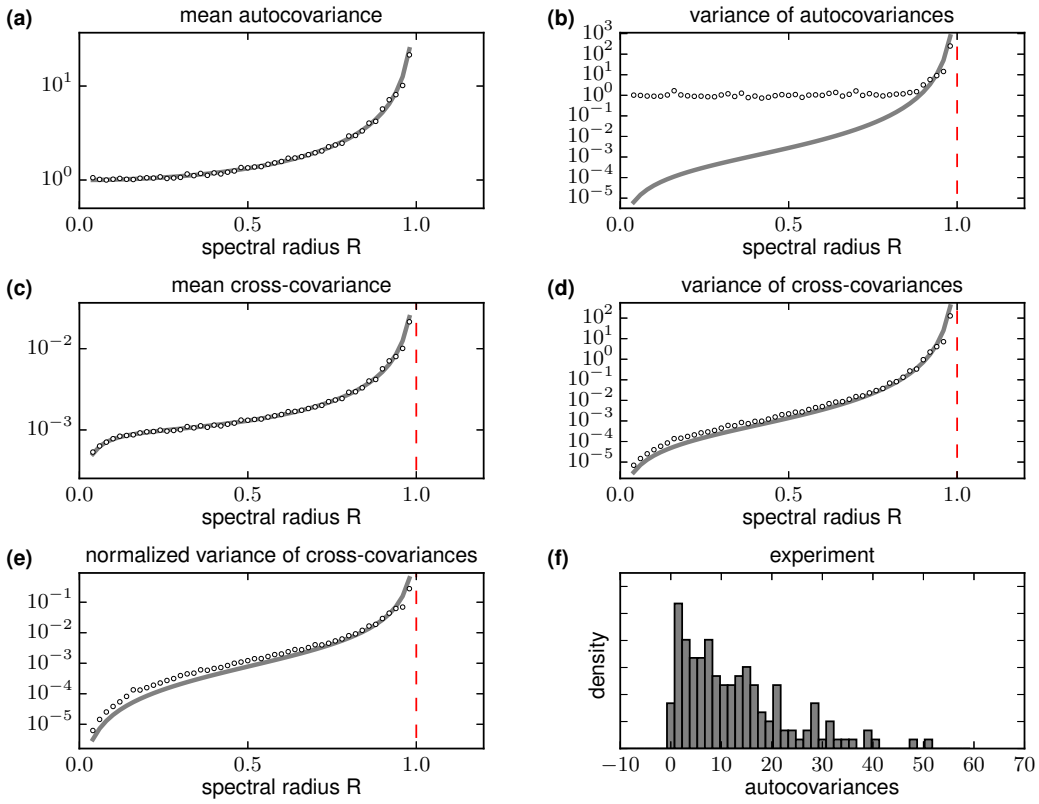


Figure B.2: Mean (absolute value) and variance of auto- (a,b) and cross-covariances (c,d), as well as normalized variance of cross-covariances Δ (e). Analytical prediction (Eq. (3.6) and Eq. (3.7)) for an inhibitory network with Gaussian connections and uniform noise autocovariance (solid line) is compared to numerical evaluation of Eq. (3.1) in Chap. 3 for inhibitory Erdős-Rényi networks with distributed noise autocovariances (circles). Parameters: $N = 1000$, $p = 0.1$, lognormal distribution of autocovariances with mean and standard deviation $\bar{D} = \delta D = 1$. Strengths of weights are varied to obtain spectral radii $R \in [0, 1]$. (f) Distribution of spike-count autocovariances $c_{ii} = \frac{1}{T}\left(\langle n_i n_i \rangle - \langle n_i \rangle \langle n_i \rangle\right)$ across trials obtained from experimental data (same dataset as in Fig. 3.1 in Chap. 3). Data courtesy of A. Riehle and T. Brochier.

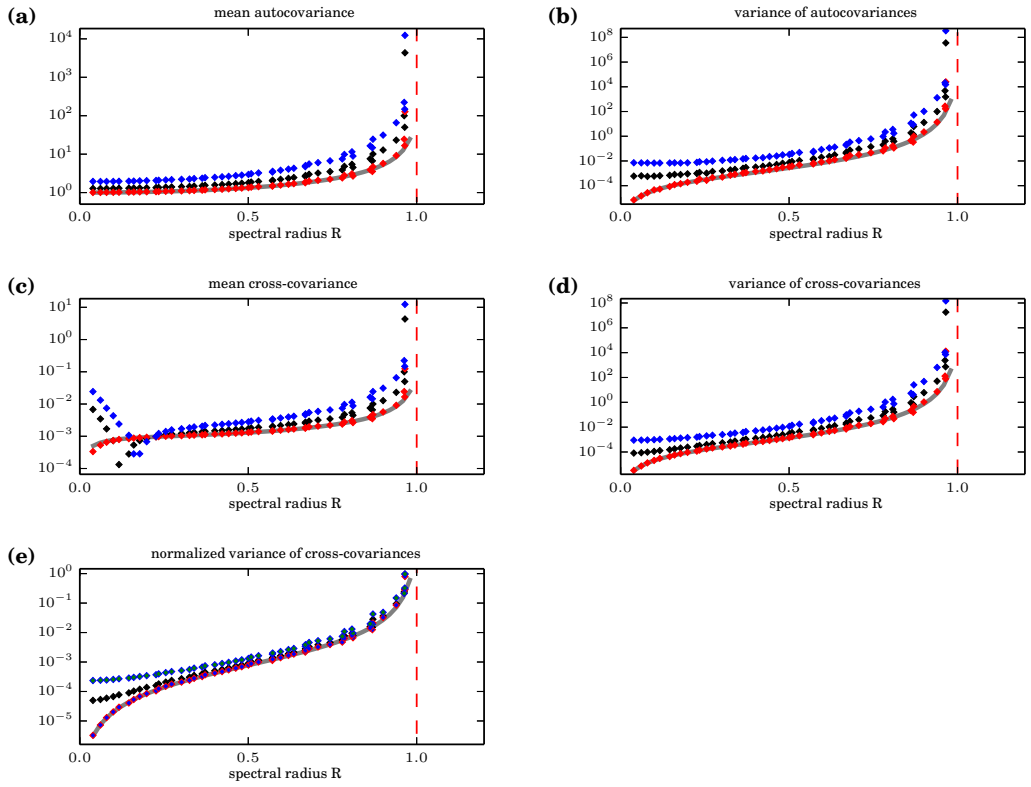


Figure B.3: Mean (absolute value) and variance of auto- (a,b) and cross-covariances (c,d), as well as normalized variance of cross-covariances Δ (e). Analytical prediction (Eq. (3.6) and Eq. (3.7)) for an inhibitory network with Gaussian connectivity (size N, spectral radius R) with uniform noise autocovariance D (solid line) is compared to numerical evaluation of Eq. (3.1) in Chap. 3 for inhibitory Erdős-Rényi networks (symbols) with uniform stationary mean activity D, receiving correlated input from an external population ($N^{\text{ext}} = N$, $R^{\text{ext}} = R$). Red, black and blue symbols show weak (spectral radius $R_M = 0.08$), intermediate ($R_M = 0.52$) and strong ($R_M = 0.96$) connectivity $\mathbf{M}$ between populations, respectively. Note that the mean cross-covariance changes sign in the low-R regime for sufficiently strong external input. Parameters: $N = 1000$, $p = 0.1$, $D = 1$. Strengths of weights are varied to obtain spectral radii $R \in [0, 1]$.

where $\mathbf{x}$ are activities of recurrently connected ($\mathbf{W}$) local neurons driven by uncorrelated noise $\boldsymbol{\xi}$ and correlated input via connections $\mathbf{M}$ to an external population of N^{ext} neurons with activities $\boldsymbol{y}$. The correlated external input adds a non-diagonal contribution to the covariance matrix $\mathbf{D}$

$$c_{ij} = \left[(\mathbf{1} - \mathbf{W})^{-1} \left(\mathbf{D} + \underbrace{\mathbf{M}\mathbf{c}^{\text{ext}}\mathbf{M}^{\mathrm{T}}}_{\tilde{\mathbf{D}}} \right) \left(\mathbf{1} - \mathbf{W}^{\mathrm{T}} \right)^{-1} \right]_{ij} .$$

Using the geometric series expansion (B.5) as in Sec. B.3 with $\mathbf{D}$ substituted by $\mathbf{D} + \tilde{\mathbf{D}} = \mathbf{D} + \mathbf{M}\mathbf{c}^{\text{ext}}\mathbf{M}^{\mathrm{T}}$ shows that variability of off-diagonal elements of $\mathbf{M}\mathbf{c}^{\text{ext}}\mathbf{M}^{\mathrm{T}}$ contributes directly to the dispersion of cross-covariances. This effect is not captured by the theory developed in Chap. 2 and Chap. 3. It arises from shared inputs and correlated activity in the external input and is, to leading order, independent of the local connectivity $\boldsymbol{W}$.

The external input to the local network typically originates from many cortical areas. Therefore, the size of the external population N^{ext} can assumed to be large and the cross-covariances c_{ij}^{ext} to be small. In the extreme case where the external population is of the same size as the local network and exhibits the same potentially strong correlation structure, i.e. $N^{\text{ext}} = N$, $\mathbf{c}^{\text{ext}} \sim \mathbf{c}$, the effect of $\tilde{\mathbf{D}}$ on the covariance structure of the local network is notable (Fig. B.3a-d), especially in the region of small spectral radius. In this region, the local network is only weakly connected and covariances are determined by shared external input. The variability of off-diagonal elements of $\tilde{D}_{ij}$ dominates the locally generated dispersion of cross-covariances c_{ij}^{loc}. The ratio Δ between the variance of cross-covariances and the mean autocovariances as obtained from the theory of an isolated local network does not predict the true ratio of the these quantities (Fig. B.3e). However, unlike the distribution of stationary mean activities (see Sec. B.3) which is potentially $\mathcal{O}(1)$, the variability of $\tilde{D}_{ij}$ is bounded to be $\mathcal{O}(1/N^{\text{ext}})$. This follows from the $\mathcal{O}(1/\sqrt{N^{\text{ext}}})$ scaling of the mean and standard deviation of the external synapses M_{ij}, which ensures that input fluctuations are independent of N^{ext} and within the dynamic range of the neurons in the local network (van Vreeswijk & Sompolinsky, 1998, section 10.2). Therefore, for larger spectral radii, locally generated covariances c_{ij}^{loc} dominate the direct influence $\tilde{D}_{ij}$ of external correlations. Even though the mean and dispersion of covariances remain affected by the external input, their effect approximately cancels in the ratio Δ. In this regime, the relation between the spectral radius of the local network and the ratio Δ can be predicted by the theory of an isolated network (Fig. B.3e). For a more realistic, larger external populations or less correlated external input, the deviation from the theory of an isolated local network becomes even smaller. Therefore, a large ratio Δ as observed in the experimental data can only be caused by a large spectral radius of the local connectivity.

B.5 Effect of correlated connections

The derivation in Chap. 2 and Chap. 3 assumes independent connections. There is, however, experimental evidence for non-random features of network connectivity (see e.g. Sporns & Kötter (2004); Song et al. (2005)). Due to the $\mathcal{O}(1/\sqrt{N})$ scaling of connection weights (van Vreeswijk & Sompolinsky (1998, section 10.2), see Sec. B.4), only connection statistics involving one or two connections are relevant as the influence of higher order motifs is suppressed by the large network size. The dominant departure from random connectivity are reciprocal connections of which a surplus was observed experimentally (Song et al., 2005). Including these in the cumulant expansion yields

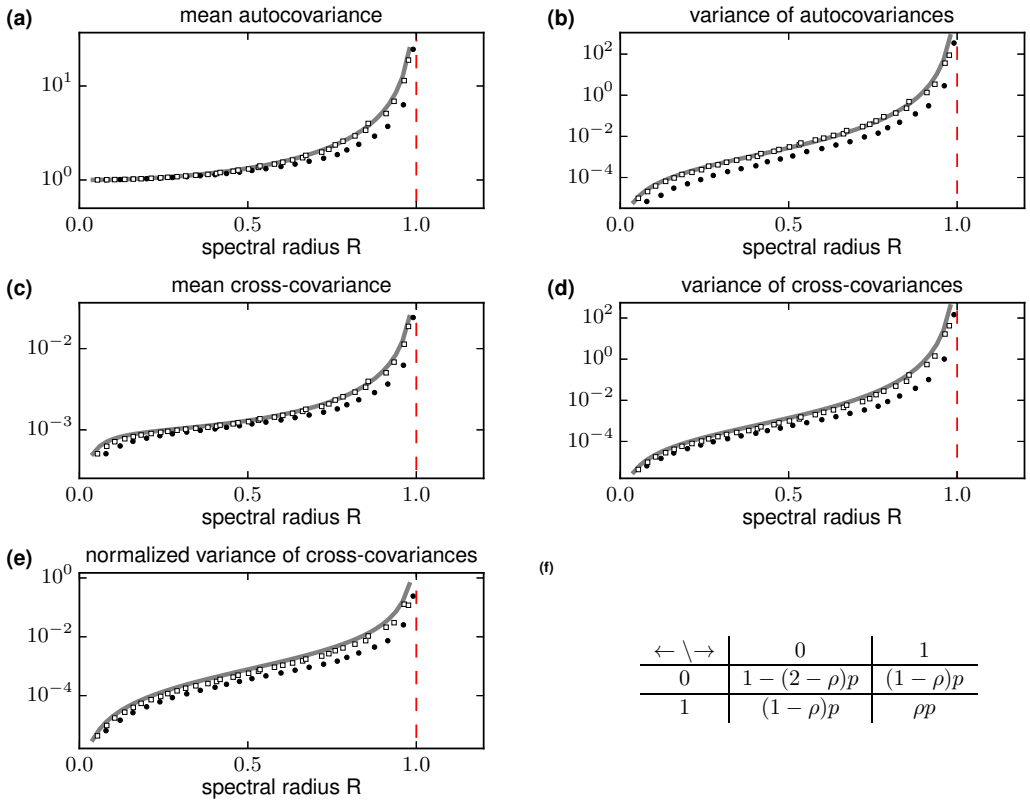
$$\left\langle e^{\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{W}\mathbf{x}} \right\rangle = \exp\left(\mu \sum_{i,j=1}^{N} \tilde{X}_i X_j + \frac{\sigma^2}{N} \sum_{i,j=1}^{N} \tilde{X}_i \tilde{X}_i X_j X_j + \frac{\lambda\sigma^2}{N} \sum_{i,j=1}^{N} \tilde{X}_i X_i \tilde{X}_j X_j \right)$$

where $\lambda\sigma^2/N$ denotes the second cumulant of reciprocal connections. This term qualitatively yields new effects as it leads to a renormalization of the response function, known from spin glasses (Sompolinsky & Zippelius, 1982). Furthermore, symmetry in the network connectivity affects the spectral radius of the connectivity matrix (Sommers et al., 1988), which is a key component in our prediction of the operational regime. In the following, we test sensitivity of our result to inclusion of reciprocal connections.

Let us assume the joint probability of two connections is given by Fig. B.4f, where $\rho \in [0,1]$ determines the degree of symmetry of connections: For $\rho = p$, we obtain the special case of independent connections, which is considered in Chap. 2 and Chap. 3. $\rho > p$ indicates a surplus of reciprocal connections with $\rho = 1$ being a fully symmetric network, $\rho = 0$ is the limit of no reciprocal connections. The parameter ρ can be estimated from the experimental data of Song et al. (2005) as the ratio between the number of observed reciprocal connections $N_{\leftrightarrow}$ and the total number of observed connections N_{tot} divided by p

$$\rho = \frac{1}{p}\frac{N_{\leftrightarrow}}{N_{\mathrm{tot}}} = \frac{1}{0.116} \cdot \frac{218}{3312 + 495 + 218} \approx 0.47 > p, \tag{B.6}$$

showing that reciprocal connections are over-represented. The variance of connections is by construction $\sigma^2/N = w^2\left(\left\langle W_{ij}W_{ij}\right\rangle - \left\langle W_{ij}\right\rangle^2\right) = p(1-p)w^2$ independent of ρ. The covariance of connections is given by $\lambda\sigma^2/N = \left(\left\langle W_{ij}W_{ji}\right\rangle - \left\langle W_{ij}\right\rangle\left\langle W_{ji}\right\rangle\right)w^2 = p\left(\rho - p\right)w^2$ resulting in $\lambda = \frac{\rho-p}{1-p} \approx 0.4$ for the ratio between the covariance of reciprocal connections and the variance of connections (Sommers et al., 1988). Numerical simulations of networks with the statistics of reciprocal connections as in Fig. B.4f with ρ given by Eq. (B.6) show that the first two cumulants of the covariance distributions are still in qualitative agreement with those in a completely random network (Fig. B.4). Even for completely symmetric networks, deviations from the theory of



$\leftarrow \backslash \rightarrow$	0	1
0	$1-(2-\rho)p$	$(1-\rho)p$
1	$(1-\rho)p$	ρp

Figure B.4: Mean (absolute value) and variance of auto- (a,b) and cross-covariances (c,d), as well as normalized variance of cross-covariances Δ (e). Analytical prediction (Eq. (3.6) and Eq. (3.7)) for an inhibitory network with Gaussian connectivity ($\rho = p$, see panel (f), corresponding to completely randomly drawn connections) with uniform noise autocovariance D (solid line) is compared to numerical evaluation of Eq. (3.1) in Chap. 3 for inhibitory Erdős-Rényi networks (symbols) with joint statistics of reciprocal connections (f) with $\rho = 0.47$ (extracted from (Song et al., 2005); squares) or $\rho = 1$ (fully symmetric; discs). Arrows in (f) denote the direction of the connection between two neurons, 0 and 1 the presence or absence of connections with probabilities given in the table. Parameters: $N = 1000$, $p = 0.1$, $D = 1$. Strengths of weights are varied to obtain spectral radii $R \in [0, 1]$.

random networks are not affecting the scale of these measures. The model prediction for the spectral radius derived from the experimental data close to 1 is therefore robust against such correlations in connections.

Appendix C

Integration of continuous-time dynamics in a spiking network simulator

C.1 Numerical evaluation of the Siegert formula

We here describe how to numerically calculate Eq. (4.22), frequently called Siegert formula in the literature (for a recent textbook see Gerstner et al., 2014), which is not straight forward due to numerical instabilities in the integral. First we introduce the abbreviations $y_\theta = (\theta - \mu)/\sigma + \gamma\sqrt{\tau_s/\tau_m}$ and $y_r = (V_r - \mu)/\sigma + \gamma\sqrt{\tau_s/\tau_m}$ and rewrite the integral as

$$\begin{aligned}
&\sqrt{\pi}\int_{y_r}^{y_\theta} e^{u^2}(1+\mathrm{erf}(u))du \\
=&2\int_{y_r}^{y_\theta} e^{u^2}\int_{-\infty}^{u} e^{-v^2}\,dv\,du \\
=&2\int_{y_r}^{y_\theta}\int_{-\infty}^{u} e^{(u+v)(u-v)}\,dv\,du\,.
\end{aligned}$$

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

$$\begin{aligned}
&\int_0^\infty e^{-w^2} \frac{e^{2y_\theta w} - e^{2y_r w}}{u}\, dw \\
=&\int_0^\infty e^{2y_\theta w - w^2} \frac{1 - e^{2(y_r - y_\theta) w}}{w}\, dw \\
=&e^{y_\theta^2} \int_0^\infty e^{-(w - y_\theta)^2} \frac{1 - e^{2(y_r - y_\theta) w}}{w}\, dw\,.
\end{aligned} \tag{C.2}$$

The integrand has a peak near y_θ. Therefore, in approximation, the lower and the upper boundary can be chosen to be left and right of y_θ, 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.

Appendix D

Hybrid scheme for modeling local field potentials from spiking point-neuron networks

Local field potential of a cortical column

This appendix contains tables specifying model parameters as well as measures and functions for data analysis.

A	**Global simulation parameters**	
Symbol	**Value**	**Description**
T	5,200 ms	simulation duration
dt	0.1 ms	temporal resolution

B	**Point-neuron network**									
Populations and external input										
Symbol	**Value**									**Description**
X	L23E	L23I	L4E	L4I	L5E	L5I	L6E	L6I	TC	Name
N_X	20,683	5,834	21,915	5,479	4,850	1,065	14,395	2,948	902	Size
k_X^{ext}	1,600	1,500	2,100	1,900	2,000	1,900	2,900	2,100		Ext. in-degree per neuron
I^{ext}	$\tau_{\text{syn}}\nu_{\text{bg}}J$, ν_{bg} = 8 Hz									DC ampl. per ext. input

Connectivity										
C_{YX}	from X									
		L23E	L23I	L4E	L4I	L5E	L5I	L6E	L6I	TC
to Y	L23E	0.101	0.169	0.044	0.082	0.032	0.0	0.008	0.0	0.0
	L23I	0.135	0.137	0.032	0.052	0.075	0.0	0.004	0.0	0.0
	L4E	0.008	0.006	0.050	0.135	0.007	0.0003	0.045	0.0	0.0983
	L4I	0.069	0.003	0.079	0.160	0.003	0.0	0.106	0.0	0.0619
	L5E	0.100	0.062	0.051	0.006	0.083	0.373	0.020	0.0	0.0
	L5I	0.055	0.027	0.026	0.002	0.060	0.316	0.009	0.0	0.0
	L6E	0.016	0.007	0.021	0.017	0.057	0.020	0.040	0.225	0.0512
	L6I	0.036	0.001	0.003	0.001	0.028	0.008	0.066	0.144	0.0196

Connection parameters		
Symbol	**Value**	**Description**
J	87.81 pA	Reference synaptic strength. All synapse weights are measured in units of J.
$\sigma_{J,\text{rel}}$		Relative width of synaptic strength distribution
	3	- for lognormal distribution
	0.1	- for Gaussian distribution
g_{YX}		Relative synaptic strength:
	1	$X \in$ {TC, L23E, L4E, L5E, L6E},
	−4	$X \in$ {L23I, L4I, L5I, L6I}, except for
	2	(X, Y) = (L4E, L23E)
	−4.5	(X, Y) = (L4I, L4E)
d_{E}	1.5 ms	Mean excitatory spike transmission delay
d_{I}	0.75 ms	Mean inhibitory spike transmission delay
$\sigma_{d,\text{rel}}$	0.5	Relative width (stdev/mean) of delay distributions
Neuron model		
Symbol	**Value**	**Description**
R_{m}	40 MΩ	Membrane resistance
C_{m}	250 pF	Membrane capacitance
τ_{m}	$R_{\text{m}}C_{\text{m}}$ (10 ms)	Membrane time constant
E_{L}	−65 mV	Resting potential
θ	−50 mV	Fixed firing threshold
$V_{\text{m}}^{\text{init}}$	[−65, −50] mV	Uniformly distributed initial membrane potential
V_{reset}	E_{L}	Reset potential
τ_{ref}	2 ms	Absolute refractory period
τ_{syn}	0.5 ms	Postsynaptic current time constant
Thalamocortical input		
Symbol	**Value**	**Description**
$\overline{\nu}_{\text{TC}}$	30 s^{-1}	Mean firing rate per thalamocortical neuron
$\Delta\nu_{\text{TC}}$	30 s^{-1}	Firing-rate modulation amplitude per TC neuron
f_{TC}	15 Hz	Frequency of sinusoidal firing-rate modulation

Table D.1: Parameters of the cortical microcircuit model.

Multicompartment model neurons		
Symbol	**Value**	**Description**
c_m	1.0 μFcm^{-2}	Membrane capacity
r_m	$\tau_\mathrm{m}/c_\mathrm{m}$	Membrane resistivity
r_a	150 Ωcm	Axial resistivity
E_L	E_L	Passive leak reversal potential
V_init	E_L	Membrane potential at t = 0 ms
λ_f	100 Hz	Frequency of AC length constant
λ_d	0.1	Factor for `d_lambda` rule (Hines & Carnevale, 2001)
σ_e	0.3 Sm^{-1}	Extracellular conductivity
r	$\sqrt{1,000^2/\pi}$ μm	Population radius
h	50 μm	Soma layer thickness
n_contact	16	Number of electrode contacts
h_elec	100 μm	Laminar-electrode intercontact distance
r_contact	7.5 μm	Electrode contact-point radius

Table D.2: Parameters of the multicompartment model neuron populations and calculations of extracellular potentials. Values for τ_m and E_L are inherited from network parameters in Tab. D.1.

Morphology files				
Cell type y	**Morphology** M_y	**File**	**Source**	**Online source**
p23	p23	oi24rpy1.hoc	(Kisvarday & Eysel, 1992)	#NMO_00851 (#NMO_10045)
b23	i23	oi38lbc1.hoc	(Stepanyants et al., 2008)	-
nb23	i23	oi38lbc1.hoc	(Stepanyants et al., 2008)	-
p4	p4	oi53rpy1.hoc	(Kisvarday & Eysel, 1992)	#NMO_00855 (#NMO_10040)
ss4(L23)	ss4	j7_L4ste.hoc	(Mainen & Sejnowski, 1996)	#MDB_2488 (#NMO_00905)
ss4(L4)	ss4	j7_L4ste.hoc	(Mainen & Sejnowski, 1996)	#MDB_2488 (#NMO_00905)
b4	i4	oi26rbc1.hoc	(Stepanyants et al., 2008)	-
nb4	i4	oi26rbc1.hoc	(Stepanyants et al., 2008)	-
p5(L23)	p5v1	oi15rpy4.hoc	(Kisvarday & Eysel, 1992)	#NMO_00850 (#NMO_10046)
p5(L56)	p5v2	j4a.hoc	(Mainen & Sejnowski, 1996)	#MDB_2488
b5	i5	oi15rbc1.hoc	(Stepanyants et al., 2008)	-
nb5	i5	oi15rbc1.hoc	(Stepanyants et al., 2008)	-
p6(L4)	p6	51-2a.CN.hoc	(Contreras et al., 1997)	#NMO_00879
p6(L56)	p5v1	oi15rpy4.hoc	(Kisvarday & Eysel, 1992)	#NMO_00850 (#NMO_10046)
b6	i5	oi15rbc1.hoc	(Stepanyants et al., 2008)	-
nb6	i5	oi15rbc1.hoc	(Stepanyants et al., 2008)	-

Table D.3: Morphology types and file names used for each cell type in the model (p - pyramidal cell, ss - spiny stellate, i - interneuron). Online source numbers #NMO_$*$ refer to NeuroMorpho.org identifiers, #MDB_$*$ refer to ModelDB identifiers.

Postsyn. pop. cell type		Layer Occurr.		Num. syn.	Presynaptic populations X and cell types x																									
					L23E	L23I		L4E		L4I		L5E			L5I		L6E			L6I		TC								
					p23	b23	nb23	ss4(L4)	ss4(L23)	p4	b4	nb4	p5(L23)	p5(L56)	b5	nb5	p6(L4)	p6(L56)	b6	nb6	TCs	TCn								
Y	y	L	F_y (%)	k_{yL}	p_{yxL} (%)																									
L23E	p23	2/3	26.7	5,800	59.9	9.1	4.4	0.6	6.9	7.7	.	0.8	7.4	.	.	.	2.3	.	.	0.8	.	.								
		1		1,306	6.3	0.1	1.1	.	.	0.1	.	.	0.1	.	.	.	.	.	.	.	.	4.1								
L23I	b23	2/3	3.2	3,854	51.6	10.6	3.4	0.5	5.8	6.6	.	0.8	6.3	.	.	.	2.1	.	.	0.7	.	0.5								
	nb23	2/3	4.3	3,307	48.6	11.4	3.3	0.5	5.5	6.2	.	0.8	5.9	.	.	.	1.8	.	.	0.6	.	0.7								
L4E	ss4(L4)	4	9.4	5,792	2.7	0.2	0.6	11.9	3.7	4.1	7.1	2	0.8	0.1	.	.	32.7	.	.	5.8	1.7	1.3								
	ss4(L23)	4	9.4	4,989	5.6	0.4	0.8	11.3	3.8	4.3	7.2	2.1	1.1	0.1	.	.	31.1	.	.	5.5	1.7	1.3								
	p4	4	9.4	5,031	4.3	0.2	0.6	11.5	3.6	4.2	7.2	2.1	1.2	0.1	.	.	31.4	0.1	.	5.9	1.7	1.3								
		2/3		866	63.1	5.1	4.1	0.6	7.2	8.1	.	0.6	7.8	.	.	.	2.5	.	.	0.8	.	.								
		1		806	6.3	0.1	1.1	.	.	0.1	.	.	0.1	.	.	.	.	.	.	.	.	4.1								
L4I	b4	4	5.5	3,230	5.8	0.5	0.8	11	3.8	4.2	8.4	2.4	1.1	.	.	.	30.3	.	.	5.4	1.6	1.2								
	nb4	4	1.5	3,688	2.7	0.2	0.6	11.7	3.6	4	8.2	2.3	0.8	0.1	.	.	32.2	.	.	5.7	1.7	1.3								
L5E	p5(L23)	5	4.8	4,316	45.9	1.8	0.3	3.3	2	7.5	.	0.9	11.7	1	0.8	1.1	2.3	2.1	.	11.5	0.1	0.4								
		4		283	2.8	0.1	0.7	12.2	3.8	4.2	5.2	1.5	0.8	0.1	.	.	33.7	.	.	5.9	1.8	1.4								
		2/3		412	63.1	5.1	4.1	0.6	7.2	8.1	.	0.6	7.8	.	.	.	2.5	.	.	0.8	.	.								
		1		185	6.3	0.1	1.1	.	.	0.1	.	.	0.1	.	.	.	.	.	.	.	.	4.1								
	p5(L56)	5	1.3	5,101	44.3	1.7	0.2	3.2	2	7.3	.	0.8	11.3	1.2	0.8	1.1	2.3	2.5	0.3	11.3	0.2	0.5								
		4		949	2.8	0.1	0.7	12.2	3.8	4.2	5.2	1.5	0.8	0.1	.	.	33.7	.	.	5.9	1.8	1.4								
		2/3		1,367	63.1	5.1	4.1	0.6	7.2	8.1	.	0.6	7.8	.	.	.	2.5	.	.	0.8	.	.								
		1		5,658	6.3	0.1	1.1	.	.	0.1	.	.	0.1	.	.	.	.	.	.	.	.	4.1								
L5I	b5	5	0.6	2,981	45.5	2.3	0.2	3.3	2	7.5	.	1.1	11.6	1	0.9	1.3	2.3	2	.	11.4	0.1	0.4								
	nb5	5	0.8	2,981	45.5	2.3	0.2	3.3	2	7.5	.	1.1	11.6	1	0.9	1.3	2.3	2	.	11.4	0.1	0.4								
L6E	p6(L4)	6	14.0	3,261	2.5	0.1	0.1	0.7	0.9	1.3	.	0.1	0.1	4.9	.	0.3	1.2	13.2	7.7	7.7	0.6	2.9								
		5		1,066	46.8	0.8	0.3	3.4	2.1	7.7	.	0.6	11.9	1	0.6	0.8	2.3	2.1	.	11.7	0.1	0.4								
		4		1,915	2.8	0.1	0.7	12.2	3.8	4.2	5.2	1.5	0.8	0.1	.	.	33.7	.	.	5.9	1.8	1.4								
		2/3		121	63.1	5.1	4.1	0.6	7.2	8.1	.	0.6	7.8	.	.	.	2.5	.	.	0.8	.	.								
	p6(L56)	6	4.6	5,573	2.5	0.1	0.1	0.7	0.9	1.3	.	0.1	0.1	4.9	.	0.3	1.2	13.2	7.8	7.8	0.6	2.9								
		5		257	46.8	0.8	0.3	3.4	2.1	7.7	.	0.6	11.9	1	0.6	0.8	2.3	2.1	.	11.7	0.1	0.4								
		4		243	2.8	0.1	0.7	12.2	3.8	4.2	5.2	1.5	0.8	0.1	.	.	33.7	.	.	5.9	1.8	1.4								
		2/3		286	63.1	5.1	4.1	0.6	7.2	8.1	.	0.6	7.8	.	.	.	2.5	.	.	0.8	.	.								
		1		62	6.3	0.1	1.1	.	.	0.1	.	.	0.1	.	.	.	.	.	.	.	.	4.1								
L6I	b6	6	2.0	3,230	2.5	0.1	0.1	0.7	0.9	1.3	.	0.1	0.1	4.9	.	0.4	1.2	13.2	7.7	7.7	0.6	2.9								
	nb6	6	2.0	3,230	2.5	0.1	0.1	0.7	0.9	1.3	.	0.1	0.1	4.9	.	0.4	1.2	13.2	7.7	7.7	0.6	2.9								

Table D.4: Spatial distribution of cell types and synapses of the cortical microcircuit model adapted from Binzegger et al. (2004) and Izhikevich & Edelman (2008). Note that occurrences are renormalized between cell types within layers 2/3, 4, 5 and 6 so that $\sum_y F_y = 100\%$.

Measurements and derived signals		
Measurements		
Symbol	**Description**	**Number of recorded units**
$I_i^{\mathrm{ex}}(t)$	Excitatory synaptic input current of neuron i	100 per population X
$I_i^{\mathrm{in}}(t)$	Inhibitory synaptic input current of neuron i	100 per population X
$V_i(t)$	Somatic voltage of neuron i	100 per population X
$s_i(t)$	Spike train of neuron i	N_X neurons
$\phi_i(\mathbf{r},t)$	Single-cell LFP generated by neuron i	N_X neurons
$\rho_i(\mathbf{r},t)$	Single-cell CSD generated by neuron i	N_X neurons
Derived signals		
Symbol	**Definition**	**Description**
$I_i(t)$	$I_i^{\mathrm{ex}}(t) + I_i^{\mathrm{in}}(t)$	Total synaptic input current of neuron i
$\overline{I}_X^{\mathrm{ex}}(t)$	$\frac{1}{n_{\mathrm{av}}} \sum_{i \in X}^{n_{\mathrm{av}}} I_i^{\mathrm{ex}}(t)$	Average excitatory synaptic input current of population X ($n_{\mathrm{av}} = 100$)
$\overline{I}_X^{\mathrm{in}}(t)$	$\frac{1}{n_{\mathrm{av}}} \sum_{i \in X}^{n_{\mathrm{av}}} I_i^{\mathrm{in}}(t)$	Average inhibitory synaptic input current of population X ($n_{\mathrm{av}} = 100$)
$\overline{I}_X(t)$	$\frac{1}{n_{\mathrm{av}}} \sum_{i \in X}^{n_{\mathrm{av}}} I_i(t)$	Average total synaptic input current of population X ($n_{\mathrm{av}} = 100$)
$\overline{V}_X(t)$	$\frac{1}{n_{\mathrm{av}}} \sum_{i \in X}^{n_{\mathrm{av}}} V_i(t)$	Average membrane voltage of population X ($n_{\mathrm{av}} = 100$)
$\nu_X(t)$	$\frac{n_X^{\mathrm{s}}(t)}{t_{\mathrm{bin}}}$	Instantaneous population (X) firing rate, with $n_X^{\mathrm{s}}(t)$ being the number of spikes in $[t, t + t_{\mathrm{bin}})$ of all cells in population X, $t_{\mathrm{bin}} = 1$ ms
$\overline{\nu}_X(t)$	$\frac{\nu_X(t)}{N_X}$	Average instantaneous firing rate of population X
$\phi_X(\mathbf{r},t)$	$\sum_{i \in X} \phi_i(\mathbf{r},t)$	Population LFP of population X
$\rho_X(\mathbf{r},t)$	$\sum_{i \in X} \rho_i(\mathbf{r},t)$	Population CSD of population X
$\phi(\mathbf{r},t)$	$\sum_X \phi_X(\mathbf{r},t)$	Compound LFP of all cells
$\rho(\mathbf{r},t)$	$\sum_X \rho_X(\mathbf{r},t)$	Compound CSD of all cells
Rescaled signals		
$\phi^{\gamma}(\mathbf{r},t)$	$\sum_X \sum_{i \in X' \subset X} \phi_i(\mathbf{r},t)$	Compound LFP signal from subset of neurons $i \in X' \subset X$ with $N_{X'} = \gamma N_X$, $\gamma \in [0,1]$
$\phi^{\gamma\xi}(\mathbf{r},t)$	$\xi \phi^{\gamma}(\mathbf{r},t)$ $(D.1)$	Compound LFP signal from subset of neurons $i \in X' \subset X$ with $N_{X'} = \gamma N_X$, $\gamma \in [0,1]$, rescaled by factor ξ

Table D.5: Measurements and derived signals obtained from the hybrid scheme for LFP simulations.

Data analysis			
Symbol	**Definition**		**Description**
ψ, ψ'	$\psi, \psi' \in \{\phi_i, \phi_X, \phi, \rho_i, \rho_X, \rho, \nu_X\}$		Signal (LFPs, CSDs, firing rates)
$\mu_\psi(\mathbf{r})$	$\frac{dt}{T}\sum_{h=1}^{T/dt}\psi(\mathbf{r}, h\,dt)$		Temporal mean of signal $\psi(\mathbf{r}, t)$
$\mathrm{cov}_{\psi\psi'}(\mathbf{r})$	$\frac{dt}{T}\sum_{h=1}^{T/dt}\psi(\mathbf{r}, h\,dt)\psi'(\mathbf{r}, h\,dt) - \mu_\psi(\mathbf{r})\mu_{\psi'}(\mathbf{r})$		Temporal covariance of signals $\psi(\mathbf{r}, t), \psi'(\mathbf{r}, t)$
$\sigma^2_\psi(\mathbf{r})$	$\mathrm{cov}_{\psi\psi}(\mathbf{r})$		Temporal variance of signal $\psi(\mathbf{r}, t)$
$cc_{\psi\psi'}(\mathbf{r})$	$\mathrm{cov}_{\psi\psi'}(\mathbf{r})/\sqrt{\sigma^2_\psi(\mathbf{r})\sigma^2_{\psi'}(\mathbf{r})}$	$(D.2)$	Zero time-lag correlation coefficient of signals $\psi(\mathbf{r}, t), \psi'(\mathbf{r}, t)$
$C_{\psi\psi'}(\mathbf{r}, f)$	$\mathcal{F}[\psi](\mathbf{r}, f)^*\mathcal{F}[\psi'](\mathbf{r}, f)$ (implemented using Welch's method)		Pairwise cross-spectral density of signals $\psi(\mathbf{r}, t), \psi'(\mathbf{r}, t)$, $\mathcal{F}[\psi]$: Fourier transform of ψ
$P_\psi(\mathbf{r}, f)$	$C_{\psi\psi}(\mathbf{r}, f)$		Power spectral density (PSD) of signal $\psi(\mathbf{r}, t)$
$\overline{P_\phi}(\mathbf{r}, f)$	$\frac{1}{N}\sum_{i=1}^{N} P_{\phi_i}(\mathbf{r}, f)$	$(D.3)$	Average single-cell LFP power spectrum
$\overline{C_\phi}(\mathbf{r}, f)$	$\frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N} C_{\phi_i\phi_j}(\mathbf{r}, f)$	$(D.4)$	Average cross-spectrum between single-cell LFPs
$\overline{\kappa_\phi}(\mathbf{r}, f)$	$\overline{C_\phi}(\mathbf{r}, f)/\overline{P_\phi}(\mathbf{r}, f)$	$(D.5)$	Average LFP coherence between cells
$P_\phi(\mathbf{r}, f)$	$N\overline{P_\phi}(\mathbf{r}, f) + N(N-1)\overline{C_\phi}(\mathbf{r}, f)$	$(D.6)$	Power spectral density (PSD) of the compound LFP
$P^0_\phi(\mathbf{r}, f)$	$N\overline{P_\phi}(\mathbf{r}, f)$	$(D.7)$	Power spectral density (PSD) of the compound LFP signal omitting pairwise cross-correlations

Table D.6: Data analysis of model output signals.

Post-processing		
Parameters		
Symbol	**Value**	**Description**
T_{trans}	200 ms	Start-up transient
dt_{ψ}	1 ms	Signal resolution
Power spectral density settings		
Method	`plt.mlab.psd`*	Welch's average periodogram
Symbol	**Value**	**Description**
T_{ψ}	5,000 ms	Signal length
NFFT	256	Number of data points used in each block for the FFT
Fs	1 kHz	Sampling rate
noverlap	128	Number of overlapping data points between blocks
window	`plt.mlab.window_hanning`*	Window filter (* `plt` denotes `matplotlib`)

Table D.7: Post-processing and data analysis parameters.

References

Abbott, L., DePasquale, B., & Memmesheimer, R.-M. (2016). Building functional networks of spiking model neurons. *Nat. Neurosci. 19*(3), 350–355. (p 81).

Abbott, L., & Kepler, T. B. (1990). Model neurons: from Hodgkin-Huxley to Hopfield. In *Statistical mechanics of neural networks*, pp. 5–18. Springer. (p 7).

Abbott, L. F., & van Vreeswijk, C. (1993). Asynchronous states in networks of pulse-coupled oscillators. *Phys. Rev. E 48*(2), 1483–1490. (p 14).

Abeles, M. (1982). *Local Cortical Circuits: An Electrophysiological Study*. Studies of Brain Function. Berlin, Heidelberg, New York: Springer-Verlag. (pp 168 and 197).

Abeles, M. (1991). *Corticonics: Neural Circuits of the Cerebral Cortex* (1st ed.). Cambridge: Cambridge University Press. (p 3 and 3).

Abramowitz, M., & Stegun, I. A. (1974). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*. New York: Dover Publications. (p 104).

Adamu, I. A. (2011). *Numerical approximation of SDEs and stochastic Swift-Hohenberg equation*. Ph. D. thesis, Heriot-Watt University. (pp 86, 87, and 169).

Aertsen, A., & Preißl, H. (1990). Dynamics of Activity and Connectivity in Physiological Neuronal Networks. In H. G. Schuster (Ed.), *Nonlinear Dynamics and Neuronal Networks*, Proceedings of the 63rd W. E. Heraeus Seminar Friedrichsdorf 1990, pp. 281–301. VCH. (p 25).

Ahlfors, S. P., & Wreh II, C. (2015). Modeling the effect of dendritic input location on MEG and EEG source dipoles. *Med Biol Eng Comput 53*(9), 879–887. (p 174).

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. (p 93).

Alemi, A., Baldassi, C., Brunel, N., & Zecchina, R. (2015). A three-threshold learning rule approaches the maximal capacity of recurrent neural networks. *PLoS Comput Biol 11*(8), 1–23. (p 162).

Aljadeff, J., Stern, M., & Sharpee, T. (2015). Transition to chaos in random networks with cell-type-specific connectivity. *Phys. Rev. Lett. 114*, 088101. (pp 13, 16, 24, 50, 70, 162, 164, and 167).

Altland, A., & B., S. (2010). *Concepts of Theoretical Solid State Physics*. Cambridge university press. (pp 50 and 51).

Amari, S.-I. (1972). Characteristics of random nets of analog neuron-like elements. *Systems, Man and Cybernetics, IEEE Transactions on* (5), 643–657. (p 13).

Amari, S.-I. (1975). Homogeneous nets of neuron-like elements. *Biol. Cybern. 17*(4), 211–220. (p 26).

Amari, S.-I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern. 27*, 77–87. (pp 13 and 80).

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. (pp 3 and 170).

Amit, D. J. (2005). *Field Theory, the Renormalization Group, and Critical Phenomena; 3rd ed.* Singapore: World Scientific. (p 65).

Amit, D. J., & Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cereb. Cortex 7*, 237–252. (pp 8, 14, 80, and 172).

Anastassiou, C., Perin, R., Markram, H., & Koch, C. (2011). Ephaptic coupling of cortical neurons. *Nat. Neurosci. 14*(2), 217. (pp 5, 113, and 129).

Anastassiou, C. A., Montgomery, S. M., Barahona, M., Buzsáki, G., & Koch, C. (2010). The effect of spatially inhomogeneous extracellular electric fields on neurons. *J. Neurosci. 30*(5), 1925–1936. (p 5).

Arbib, M. A., Érdi, P., et al. (1998). *Neural organization: Structure, function, and dynamics*. MIT press. (p 3).

Armstrong-James, M., Fox, K., & Das-Gupta, A. (1992). Flow of excitation within rat barrel cortex on striking a single vibrissa. *J Neurophysiol 68*(4), 1345–1358. (p 139).

Ascoli, G. A., Donohue, D. E., & Halavi, M. (2007). Neuromorpho.org: a central resource for neuronal morphologies. *J. Neurosci. 27*(35), 9247–9251. (pp 119, 120, 123, and 172).

Barlow, R. (2013). *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*. John Wiley & Sons. (pp 191 and 193).

Barrat, A., Barthelemy, M., & Vespignani, A. (2008). *Dynamical processes on complex networks*. Cambridge University Press. (p 24).

Bastos, A. M., Vezoli, J., & Fries, P. (2015). Communication through coherence with inter-areal delays. *Current opinion in neurobiology 31*, 173–180. (p 140).

Bazelot, M., Dinocourt, C., Cohen, I., & Miles, R. (2010). Unitary inhibitory field potentials in the CA3 region of rat hippocampus. *J Physiol 588*(12), 2077–2090. (p 175).

Bazhenov, M., Stopfer, M., Rabinovich, M., Huerta, R., Abarbanel, H. D., Sejnowski, T. J., & Laurent, G. (2001). Model of transient oscillatory synchronization in the locust antennal lobe. *Neuron 30*(2), 553–567. (p 112).

Bédard, C., & Destexhe, A. (2009). Macroscopic models of local field potentials and the apparent 1/f noise in brain activity. *Biophys J 96*(7), 2589–2603. (p 173).

Bednar, J. A. (2009). Topographica: Building and analyzing map-level simulations from python, c/c++, matlab, nest, or neuron components. *Front Neuroinform. 24*, 8. (p 80).

Ben-Yishai, R., Bar-Or, R., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proc. Natl. Acad. Sci. USA 92*, 3844. (p 80).

Bi, G., & Poo, M. (1998). Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci. 18*, 10464–10472. (p 24).

Binney, J. J., Dowrick, N. J., Fisher, A. J., & Newman, M. (1992). *The Theory of Critical Phenomena: An Introduction to the Renormalization Group*. New York, NY, USA: Oxford University Press, Inc. (p 24).

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. (pp 115, 119, 123, 124, 125, 127, 168, 197, and 211).

Blomquist, P., Devor, A., Indahl, U. G., Ulbert, I., Einevoll, G. T., & Dale, A. M. (2009). Estimation of thalamocortical and intracortical network models from joint thalamic single-electrode and cortical laminar-electrode recordings in the rat barrel system. *PLoS Comput. Biol. 5*(3), e1000328. (p 150).

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. (p 80).

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. (p 80).

Bos, H., Diesmann, M., & Helias, M. (2015). Identifying anatomical origins of coexisting oscillations in the cortical microcircuit. *arXiv preprint arXiv:1510.00642*. (pp 8, 15, 16, 50, 100, 118, 118, 140, and 142).

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. (pp 81 and 82).

Bower, J. M., & Beeman, D. (2007). GENESIS (simulation environment). *Scholarpedia 2*(3), 1383. (p 81).

Braitenberg, V., & Schüz, A. (1991). *Anatomy of the Cortex: Statistics and Geometry*. Berlin, Heidelberg, New York: Springer-Verlag. (p 5).

Braitenberg, V., & Schüz, A. (1998). *Cortex: Statistics and Geometry of Neuronal Connectivity* (second ed.). Berlin: Springer-Verlag-Verlag. (p 3).

Bressler, S. L., & Kelso, J. S. (2001). Cortical coordination dynamics and cognition. *Trends in cognitive sciences 5*(1), 26–36. (p 3).

Bressloff, P. C. (2009). Stochastic neural field theory and the system-size expansion. *SIAM J. appl. math. 70*(5), 1488–1521. (p 159).

Bressloff, P. C. (2012). Spatiotemporal dynamics of continuum neural fields. *Journal of Physics A: Mathematical and Theoretical 45*(3), 033001. (pp 13, 13, 14, 80, 81, 82, and 170).

Bressloff, P. C. (2015). Path-integral methods for analyzing the effects of fluctuations in stochastic hybrid neural networks. *Journal of Mathematical Neuroscience 5*, 4. (p 81).

Bressloff, P. C., & Cowan, J. D. (2003). A spherical model for orientation and spatial–frequency tuning in a cortical hypercolumn. *Phil. Trans. R. Soc. B 358*(1438), 1643–1667. (p 80).

Bressloff, P. C., Cowan, J. D., Golubitsky, M., Thomas, P. J., & Wiener, M. C. (2001). Geometric visual hallucinations, euclidean symmetry and the functional architecture of striate cortex. *Phil. Trans. R. Soc. B 356*(1407), 299–330. (p 80).

Brette, R. (2015). Philosophy of the spike: Rate-based vs. spike-based theories of the brain. *Frontiers in Systems Neuroscience 9*, 151. (p 170).

Brette, R., & Destexhe, A. (2012). *Handbook of neural activity measurement*. Cambridge University Press. (p 3).

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. (pp 8, 81, and 112).

Breuer, D., Timme, M., & Memmesheimer, R.-M. (2014). Statistical physics of neural systems with nonadditive dendritic coupling. *Physical Review X 4*(1), 011053. (p 172).

Brown, E. N., Kaas, R. E., & Mitra, P. P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat. Neurosci. 7*(5), 456–461. (p 162).

Brunel, N. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci. 8*(3), 183–208. (pp 8, 8, 12, 72, 80, 81, 103, 118, 118, 131, 136, and 142).

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. (pp 14, 14, 14, 15, and 81).

Brunel, N., & Sergi, S. (1998). Firing frequency of leaky intergrate-and-fire neurons with synaptic current dynamics. *J. Theor. Biol. 195*(1), 87–95. (p 14).

Buice, M. A., & Chow, C. C. (2007). Correlations, fluctuations, and stability of a finite-size network of coupled oscillators. *Phys. Rev. E 76*, 031118. (pp 161 and 170).

Buice, M. A., & Chow, C. C. (2011). Effective stochastic behavior in dynamical systems with incomplete information. *Phys. Rev. E 84*, 051120. (p 160).

Buice, M. A., & Chow, C. C. (2013). Dynamic finite size effects in spiking neural networks. *PLoS Comput Biol 9*(1), 1–21. (pp 13, 81, and 161).

Buice, M. A., & Cowan, J. D. (2007). Field-theoretic approach to fluctuation effects in neural networks. *Phys. Rev. E 75*, 051919. (pp 13, 51, 81, and 161).

Buice, M. A., Cowan, J. D., & Chow, C. C. (2010). Systematic fluctuation expansion for neural network activity equations. *Neural Comput. 22*, 377–426. (pp 13, 16, 25, 26, 27, 27, 27, 81, 161, 161, 170, and 183).

Burkitt, A. N. (2006). A review on the integrate-and-fire neuron model: I. homogenous synaptic input. *Biol. Cybern. 95*(1), 1–19. (p 7).

Buzsaki, G. (2004). Large-scale recording of neuronal ensembles. *Nat. Neurosci. 7*(5), 446–451. (pp 9 and 162).

Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents – eeg, ecog, lfp and spikes. *Nat. Rev. Neurosci. 13*(6), 407–427. (p 10 and 10).

Buzsaki, G., & Mizuseki, K. (2014). The log-dynamic brain: how skewed distributions affect network operations. *Nat. Rev. Neurosci. 15*, 264–278. (p 118).

Cabral, J., Kringelbach, M. L., & Deco, G. (2014). Exploring the network dynamics underlying brain activity during rest. *Prog. Neurobiol. 114*, 102–131. (p 3).

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. (p 80).

Camuñas Mesa, L. A., & Quiroga, R. Q. (2013). A detailed and fast model of extracellular recordings. *Neural Comput 25*(5), 1191–1212. (p 130).

Carnevale, T., & Hines, M. (2006). *The NEURON Book*. Cambridge: Cambridge University Press. (pp 81, 121, 130, 131, and 172).

Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Rev. Mod. Phys. 81*, 591–646. (p 24).

Castro-Alamancos, M. A., & Connors, B. W. (1996). Spatiotemporal properties of short-term plasticity sensorimotor thalamocortical pathways of the rat. *J Neurosci 16*(8), 2767–2779. (pp 139 and 175).

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. (pp 80 and 81).

Chemla, S., & Chavane, F. (2010). Voltage-sensitive dye imaging: Technique review and models. *J. Physiol. (Paris) 104*, 40–50. (pp 10 and 174).

Chow, C., & Buice, M. (2010). Path integral methods for stochastic differential equations. *arXiv*, 1009.5966v2. (p 51).

Chow, C., & Buice, M. (2015). Path integral methods for stochastic differential equations. *The Journal of Mathematical Neuroscience 5*. (pp 50, 51, 51, and 179).

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. (p 36).

Cocco, S., Monasson, R., & Sessak, V. (2011). High-dimensional inference with the generalized hopfield model: Principal component analysis and corrections. *Phys. Rev. E 83*, 051123. (p 45).

Cohen, M. R., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nat. Rev. Neurosci. 14*(7), 811–819. doi:10.1038/nn.2842. (pp 24 and 70).

Connors, B. W., & Gutnick, M. J. (1990). Intrinsic firing patterns of diverse neocortical neurons. *TINS 13*(3), 99–104. (pp 3 and 5).

Contreras, D., Destexhe, A., & Steriade, M. (1997). Intracellular and computational characterization of the intracortical inhibitory control of synchronized thalamic inputs in vivo. *J. Neurophysiol. 78*(1), 335–350. (pp 119, 120, 120, 123, and 210).

Coolen, A. C. C., Laughton, S. N., & Sherrington, D. (1996). Dynamical replica theory for disordered spin systems. *Phys. Rev. B 53*, 8184–8187. (pp 25 and 27).

Coombes, S. (2005). Waves, bumps, and patterns in neural field theories. *Biol. Cybern. 93*, 91–108. (p 80).

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. (pp 24, 50, and 67).

Crutchfield, J. P., & Young, K. (1989). Inferring statistical complexity. *Phys. Rev. Lett. 63*, 105–108. (pp 70 and 160).

Cugliandolo, L. F., & Kurchan, J. (1993). Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model. *Phys. Rev. Lett. 71*, 173–176. (p 24).

Cuntz, H., Forstner, F., Borst, A., & Häusser, M. (2010). One rule to grow them all: a general theory of neuronal branching and its practical application. *PLoS Comput Biol 6*(8), e1000877. (p 177).

Cuntz, H., Forstner, F., Borst, A., & Häusser, M. (2011). The TREES toolbox - probing the basis of axonal and dendritic branching. *Neuroinformatics 9*(1), 91–96. (p 177).

Dahmen, D., Bos, H., & Helias, M. (2016a). Correlated fluctuations in strongly coupled binary networks beyond equilibrium. *Phys. Rev. X 6*, 031024. (pp vii, 16, 72, 100, 100, and 167).

Dahmen, D., Bos, H., & Helias, M. (2016b). Correlated fluctuations in strongly-coupled binary networks beyond equilibrium. In *Conference Abstracts of the DPG Spring Meeting 2016*, pp. BP 53.10. (pp vii and 23).

Dahmen, D., Diesmann, M., & Helias, M. (2015). Distribution of pair-wise covariances in neuronal networks. In *Conference Abstracts of the Bernstein Conference 2015*, pp. T31. (pp vii, 49, and 69).

Dahmen, D., Diesmann, M., & Helias, M. (2016a). Distribution of pair-wise covariances in neuronal networks. In *Conference Abstracts of the DPG Spring Meeting 2016*, pp. BP 68.2. (pp vii, 49, and 69).

Dahmen, D., Diesmann, M., & Helias, M. (2016b). Distributions of covariances as a window into the operational regime of neuronal networks. *arXiv*. 1605.04153 [q-bio.NC]. (pp vii, 16, 17, and 50).

Dahmen, D., Hagen, E., Stavrinou, M. L., Lindén, H., Tetzlaff, T., van Albada, S. J., Grün, S., Diesmann, M., & Einevoll, G. T. (2013). From spiking point-neuron networks to LFPs: a hybrid approach. In *Conference Abstracts of the Bernstein Conference 2013*, pp. W56. (pp viii, 111, and 133).

Dahmen, D., Hagen, E., Stavrinou, M. L., Lindén, H., Tetzlaff, T., van Albada, S. J., Grün, S., Diesmann, M., & Einevoll, G. T. (2014). Computing local-field potentials based on a point-neuron network model of cat v1. In *2014 Neuroscience Meeting Planner. Washington, DC: Society for Neuroscience.*, pp. 532.04/GG7. (pp viii, 111, and 133).

De Dominicis, C. (1976). Techniques de renormalisation de la th'eorie des champs es dynamique des ph'enom'enes critiques. *J. Phys. Colloques 37*(C1), C1–247–C1–253. (pp 51 and 163).

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. (pp 50, 50, 51, 51, 63, 67, and 163).

de Kamps, M. (2013). A generic approach to solving jump diffusion equations with applications to neural populations. *ArXiv e-prints 1309.1654v2 [q-bio.NC]*. (p 8).

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. (p 80).

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. (p 70).

De Pittà, M., Brunel, N., & Volterra, A. (2016). Astrocytes: orchestrating synaptic plasticity? *Neuroscience 323*, 43–61. (p 3).

De Schutter, E., & Van Geit, W. (2009). Modeling complex neurons. In E. De Schutter (Ed.), *Computational Modeling Methods for Neuroscientists* (1 ed.)., Chapter 11, pp. 260–283. MIT Press, Cambridge, MA. (pp 8, 112, 120, 122, and 128).

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. (p 93).

Deco, G., & Jirsa, V. K. (2012). Ongoing cortical activity at rest: Criticality, multistability, and ghost attractors. *J. Neurosci. 32*(10), 3366–3375. (pp 8 and 80).

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. (p 80).

Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., & Friston, K. (2008). The dynamic brain: From spiking neurons to neural masses and cortical fields. *PLoS Comput. Biol. 4*(8), e1000092. (pp 6, 8, 13, 14, 81, 81, 150, 173, and 179).

DeFelipe, J. (2011). The evolution of the brain, the human nature of cortical circuits, and intellectual creativity. *Frontiers in Neuroanatomy 5*, 29. (p 4).

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*. (p 5).

Denker, M., Roux, S., Lindén, H., Diesmann, M., Riehle, A., & Grün, S. (2011). The local field potential reflects surplus spike synchrony. *Cereb. Cortex 21*, 2681–2695. (pp 150 and 153).

Dere, E., & Zlomuzica, A. (2011). The role of gap junctions in the brain in health and disease. *Neurosci. Biobehav. Rev. 36*, 206–217. (p 3).

Derrida, B. (1987). Dynamical phase transition in non-symmetric spin glasses. *J. Phys. A: Math. Gen. 20*, L721. (p 24).

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. (pp 8, 80, and 172).

Destexhe, A., Rudolph, M., & Pare, D. (2003). The high-conductance state of neocortical neurons in vivo. *Nat. Rev. Neurosci. 4*, 739–751. (p 13).

Di, S., Baumgartner, C., & Barth, D. S. (1990). Laminar analysis of extracellular field potentials in rat vibrissa/barrel cortex. *J Neurophysiol 63*(4), 832–840. (pp 138, 139, 140, 142, and 175).

Diesmann, M. (2013). The road to brain-scale simulations on K. *BioSupercomputing Newsletter 8*, 8. (pp 8 and 112).

Djurfeldt, M., Lundqvist, M., Johansson, C., Rehn, M., Ekeberg, O., & Lansner, A. (2008). Brain-scale simulation of the neocortex on the IBM Blue Gene/L supercomputer. *IBM Journal of Research and Development 52*(1/2), 31–41. (p 81).

Douglas, R. J., & Martin, K. A. C. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci. 27*, 419–451. (p 10).

Douglas, R. J., Martin, K. A. C., & Whitteridge, D. (1989). A canonical microcircuit for neocortex. *Neural Comput. 1*, 480–488. (p 10).

Eccles, J., P, F., & Koketsu, K. (1954). Cholinergic and inhibitory synapses in a pathway from motor-axon collaterals to motoneurones. *J Physiol (Lond) 126*, 524–562. (pp 5 and 24).

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. (pp 8, 17, 25, 70, 80, 163, 167, 172, and 179).

Einevoll, G. T., Kayser, C., Logothetis, N. K., & Panzeri, S. (2013). Modelling and analysis of local field potentials for studying the function of cortical circuits. *Nat. Rev. Neurosci. 14*(11), 770–785. (pp 5, 10, 10, 112, 112, 113, 118, 128, 129, 138, 142, 150, 153, 157, 177, and 179).

Einevoll, G. T., Lindén, H., Tetzlaff, T., Łęski, S., & Pettersen, K. H. (2013). Local field potentials: Biophysical origin and analysis. In Panzeri & Quiroga (Ed.), *Principles of Neural Coding*, pp. 37–59. CRC Press. (p 113).

Einevoll, G. T., Pettersen, K. H., Devor, A., Ulbert, I., Halgren, E., & Dale, A. M. (2007). Laminar population analysis: Estimating firing rates and evoked synaptic activity from multielectrode recordings in rat barrel cortex. *J. Neurophysiol. 97*, 2174–2190. (pp 10, 138, 138, 139, 139, 140, 142, 150, 150, 150, 175, and 175).

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., Albada, S., 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. (pp 26, 30, 115, 131, 131, and 172).

Erdős, P., & Rényi, A. (1959). On random graphs. *Publications Mathematicae 6*, 290–297. (p 125).

Erlhagen, W., & Schöner, G. (2002). Dynamic field theory of movement preparation. *Psychological review 109*(3), 545. (p 80).

Ermentrout, G. B., & Cowan, J. D. (1979). A mathematical theory of visual hallucination patterns. *Biol. Cybern. 34*, 137–150. (p 80).

Ermentrout, G. B., & Terman, D. H. (2010). *Mathematical foundations of neuroscience*, Volume 35. Springer Science & Business Media. (p 13).

Fan, Z. (2013). Sor waveform relaxation methods for stochastic differential equations. *Appl. Math. Comput. 219*(10), 4992–5003. (p 91).

Fee, M. S., Mitra, P. P., & Kleinfeld, D. (1996). Variability of extracellular spike waveforms of cortical neurons. *J. Neurophysiol. 76*(6), 3823–3833. (p 7).

Feldman, J. A., & Ballard, D. H. (1982). Connectionist models and their properties. *Cognitive science 6*(3), 205–254. (p 80).

Feldmeyer, D., & Lübke, J. H. (2010). *New aspects of axonal structure and function*. Springer. (p 5).

Fernández-Gracia, J., Suchecki, K., Ramasco, J. J., San Miguel, M., & Eguíluz, V. M. (2014). Is the voter model a model for voters? *Phys. Rev. Lett. 112*, 158701. (p 162).

Fischer, K., & Hertz, J. (1991a). *Spin glasses*. Cambridge University Press. (pp 12, 50, 158, and 166).

Fischer, K. H., & Hertz, J. A. (1991b). *Spin Glasses*. Cambridge University Press. (pp 63 and 73).

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn. Sci. 14*, 119–130. (p 24).

Foster, I. (1995). *Designing and building parallel programs: concepts and tools for parallel software engineering*. Reading, Mass: Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA. (pp 113 and 171).

Fourcaud, N., & Brunel, N. (2002). Dynamics of the firing probability of noisy integrate-and-fire neurons. *Neural Comput. 14*, 2057–2110. (pp 14, 104, and 104).

Fröhlich, F., & McCormick, D. A. (2010). Endogenous electric fields may guide neocortical network activity. *Neuron 67*, 129–143. (p 5).

Fukuda, T. (2007). Structural organization of the gap junction network in the cerebral cortex. *The Neuroscientist 13*(3), 199–207. (p 3).

Gabriel, C., Peyman, A., & Grant, E. H. (2009). Electrical conductivity of tissue at frequencies below 1 MHz. *Phys Med Biol 54*(16), 4863–4878. (p 173).

Gabriel, S., Lau, R. W., & Gabriel, C. (1996). The dielectric properties of biological tissues: II. measurements in the frequency range 10 Hz to 20 GHz. *Phys Med Biol 41*(11), 2251–2269. (p 173).

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. (p 206).

Gancarz, G., & Grossberg, S. (1998). A neural model of the saccade generator in the reticular formation. *IEEE Trans. Neural Netw. 11*(7), 1159–1174. (p 106).

Gardiner, C. (2009). *Stochastic Methods: A Handbook for the Natural and Social Sciences* (4th ed.). Berlin, Heidelberg: Springer. (p 51).

Gardiner, C. W. (1985). *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (2nd ed.). Number 13 in Springer Series in Synergetics. Berlin: Springer-Verlag. (pp 29, 54, 63, and 74).

Gardiner, C. W. (2004). *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (3rd ed.). Springer Series in Synergetics. Springer. (pp 83, 84, 100, and 161).

Gawne, T. J. (2010). The local and non-local components of the local field potential in awake primate visual cortex. *J Comput Neurosci 29*(3), 615–623. (p 140).

Gentet, L., Avermann, M., Matyas, F., Staiger, J. F., & Petersen, C. C. (2010). Membrane potential dynamics of GABAergic neurons in the barrel cortex of behaving mice. *Neuron 65*, 422–435. (pp 8, 80, and 172).

Gerstner, W., & Kistler, W. (2002). *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press. (pp 7, 7, 13, and 14).

Gerstner, W., Kistler, W. M., Naud, R., & Paninski, L. (2014). *Neuronal Dynamics. From single Neurons to Networks and Models of Cognition*. Cambridge University Press. (pp 3 and 205).

Gerstner, W., & Naud, R. (2009). How good are neuron models? *Science 326*(5951), 379–380. (pp 7 and 172).

Gerstner, W., & van Hemmen, J. L. (1992). Universality in neural networks: the importance of the 'mean firing rate'. *Biol. Cybern. 67*(3), 195–205. (p 14).

Gewaltig, M.-O., & Diesmann, M. (2007). NEST (NEural Simulation Tool). *Scholarpedia 2*(4), 1430. (p 82 and 82).

Gibson, J. R., Beierlein, M., & Connors, B. W. (1999). Two networks of electrically coupled inhibitory neurons in neocortex. *Nature 402*, 75–79. (p 3).

Gies, H. (2006). Introduction to the functional rg and applications to gauge theories. *arXiv*, hep–ph/0611146. (p 179).

Ginzburg, I., & Sompolinsky, H. (1994). Theory of correlations in stochastic neural networks. *Phys. Rev. E 50*(4), 3171–3191. (pp 7, 12, 16, 24, 25, 26, 27, 27, 27, 62, 72, 100, 159, 159, 161, 161, and 167).

Głąbska, H., Potworowski, J., Łęski, S., & Wójcik, D. K. (2014). Independent components of neural activity carry information on individual populations. *PloS one 9*(8), e105071. (pp 142 and 177).

Glauber, R. (1963). Time-dependent statistics of the Ising model. *J. Math. Phys. 4*(2), 294–307. (pp 7, 24, 24, 157, and 161).

Gleeson, J. P. (2011). High-accuracy approximation of binary-state dynamics on networks. *Phys. Rev. Lett. 107*(6), 068701. (pp 24 and 26).

Gleeson, J. P. (2013). Binary-state dynamics on complex networks: pair approximation and beyond. *Phys. Rev. X 3*(2), 021004. (pp 24, 26, and 162).

Goedeke, S., Schuecker, J., & Helias, M. (2016). Noise dynamically suppresses chaos in neural networks. *arXiv*. 1603.01880v1 [q-bio.NC]. (pp 13, 14, 50, 62, 83, 101, 101, 101, 164, 167, and 167).

Gold, C., Henze, D. A., & Koch, C. (2007). Using extracellular action potential recordings to constrain compartmental models. *Journal of Computational Neuroscience 23*(1), 39–58. (p 10).

Gold, C., Henze, D. A., Koch, C., & Buzsáki, G. (2006). On the origin of the extracellular action potential waveform: a modeling study. *J Neurophysiol 95*(5), 3113–3128. (pp 10, 128, and 173).

Goodman, D., & Brette, R. (2013). Brian simulator. *Scholarpedia 8*(1), 10883. (p 81).

Goto, T., Hatanaka, R., Ogawa, T., Sumiyoshi, A., Riera, J., & Kawashima, R. (2010). An evaluation of the conductivity profile in the somatosensory barrel cortex of Wistar rats. *J Neurophysiol 104*(6), 3388–412. (p 173).

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. (pp 8, 80, and 168).

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in cognitive sciences 14*(8), 357–364. (p 6).

Gross, T., D'Lima, C. J. D., & Blasius, B. (2006). Epidemic dynamics on an adaptive network. *Phys. Rev. Lett. 96*(20), 208701. (p 162).

Grossberg, S. (1973). Contour enhancement, short term memory, and constancies in reverberating neural networks. *Stud. Appl. Math. 52*(3), 213–257. (p 80).

Grytskyy, D., Tetzlaff, T., Diesmann, M., & Helias, M. (2013a). Invariance of covariances arises out of noise. *AIP Conf. Proc. 1510*, 258–262. (pp 31, 32, 39, 158, and 160).

Grytskyy, D., Tetzlaff, T., Diesmann, M., & Helias, M. (2013b). A unified view on weakly correlated recurrent networks. *Front. Comput. Neurosci. 7*, 131. (pp 15, 16, 26, 43, 50, 72, 81, 83, 100, 100, 107, 159, 160, 167, 167, and 170).

Hagen, E., Dahmen, D., Stavrinou, M. L., Lindén, H., Tetzlaff, T., van Albada, S. J., Grün, S., Diesmann, M., & Einevoll, G. T. (2015a). Hybrid scheme for modeling local field potentials from point-neuron networks. *BMC Neuroscience 16(Suppl 1)*, P67. (pp viii, 111, and 133).

Hagen, E., Dahmen, D., Stavrinou, M. L., Lindén, H., Tetzlaff, T., van Albada, S. J., Grün, S., Diesmann, M., & Einevoll, G. T. (2015b). Hybrid scheme for modeling local field potentials from point-neuron networks. In *Proceedings of the 11th Göttingen Meeting of the German Neuroscience Society*, pp. T26 7B. (pp viii, 111, and 133).

Hagen, E., Dahmen, D., Stavrinou, M. L., Lindén, H., Tetzlaff, T., van Albada, S. J., Grün, S., Diesmann, M., & Einevoll, G. T. (2016). Hybrid scheme for modeling local field potentials from point-neuron networks. *Cerebral Cortex*, bhw237. (pp viii, 17, and 18).

Hagen, E., Ness, T. V., Khosrowshahi, A., Sørensen, C., Fyhn, M., Hafting, T., Franke, F., & Einevoll, G. T. (2015c). ViSAPy: A python tool for biophysics-based generation of virtual spiking activity for evaluation of spike-sorting algorithms. *J Neurosci Methods 245*, 182–204. (pp 138, 138, 173, 174, 175, and 176).

Hagen, E., Stavrinou, M. L., Lindén, H., Dahmen, D., Tetzlaff, T., van Albada, S. J., Grün, S., Diesmann, M., & Einevoll, G. T. (2013). Hybrid scheme for modeling

LFPs from spiking cortical network models. In *Conference Abstracts of Neuroinformatics 2013*. Front. Neuroinform. 117. 10.3389/conf.fninf.2013.09.00117. (pp viii, 111, and 133).

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]. (pp vii and 17).

Hahne, J., Dahmen, D., Schuecker, J., Frommer, A., Bolten, M., Helias, M., & Diesmann, M. (2017). Integration of continuous-time dynamics in a spiking neural network simulator. *Front. Neuroinformatics 11*, 34. (pp vii and 17).

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). (pp 17, 82, 88, 89, 89, 89, 91, 91, 93, 94, and 170).

Haider, B., Duque, A., Hasenstaub, A. R., & McCormick, D. A. (2006). Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *J. Neurosci. 26*(17), 4535–4545. (p 5).

Hairer, E., & Wanner, G. (1991). *Solving Ordinary Differential Equations II*. Berlin: Springer. (p 95).

Halnes, G., Østby, I., Pettersen, K. H., Omholt, S. W., & Einevoll, G. T. (2013). Electrodiffusive model for astrocytic and neuronal ion concentration dynamics. *PLoS Comput. Biol. 9*(12), e1003386. (p 5).

Hämäläinen, M., Haari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography — theory, instrumentation, and application to noninvasive studies of the working human brain. *Rev. Mod. Phys. 65*, 413–496. (pp 10, 128, 174, and 174).

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. (p 102).

Hanuschkin, A., Kunkel, S., Helias, M., Morrison, A., & Diesmann, M. (2010). A general and efficient method for incorporating precise spike times in globally time-driven simulations. *Front. Neuroinform. 4*, 113. (p 26).

Harish, O., & Hansel, D. (2015). Asynchronous rate chaos in spiking neuronal circuits. *PLoS Comput Biol 11*(7), e1004266. (pp 160 and 167).

Hawkes, A. (1971). Point spectra of some mutually exciting point process. *J. R. Statist. Soc. Ser. B 33*(3), 438–443. (pp 50, 72, and 167).

Hay, E., Hill, S., Schuermann, F., Markram, H., & Segev, I. (2011). Models of neocortical layer 5b pyramidal cells capturing a wide range of dendritic and perisomatic active properties. *PLoS Comput. Biol. 7*(7). (p 8).

Haykin, S. S. (2009). *Neural Networks and Learning Machines* (3rd ed.). Prentice Hall. (pp 3, 80, and 170).

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: John Wiley & Sons. (pp 5 and 7).

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. (pp 8, 81, 88, 112, and 171).

Helias, M., Tetzlaff, T., & Diesmann, M. (2013). Echoes in correlated neural systems. *New J. Phys. 15*, 023002. (pp 8, 8, 15, 80, 100, 163, 166, and 172).

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. (pp 15, 17, 27, 50, 62, 72, 82, 159, 161, 165, 183, and 197).

Hermann, G., & Touboul, J. (2012). Heterogeneous connections induce oscillations in large-scale networks. *Phys. Rev. Lett. 109*(1), 018702. (pp 50 and 158).

Herreras, O., Makarova, J., & Makarov, V. (2015). New uses of LFPs: Pathway-specific threads obtained through spatial discrimination. *Neuroscience 310*, 486–503. (pp 10, 140, 144, and 178).

Hertz, J. (2010). Cross-correlations in high-conductance states of a model cortical network. *Neural Comput. 22*, 427–447. (p 146).

Hertz, J., Grinstein, G., & Solla, S. (1986). Memory networks with asymmetric bonds. *AIP Conf. Proc. 151*, 212. (p 24).

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the Theory of Neural Computation*. Perseus Books. (pp 26 and 80).

Hertz, J. A., Roudi, Y., & Sollich, P. (2016). Path integral methods for the dynamics of stochastic and disordered systems. *arXiv preprint arXiv:1604.05775*. (pp 50 and 51).

Hildebrand, E. J., Buice, M. A., & Chow, C. C. (2007). Kinetic theory of coupled oscillators. *Phys. Rev. Lett. 98*, 054101. (pp 13 and 161).

Hill, S., & Tononi, G. (2005). Modeling sleep and wakefulness in the thalamocortical system. *J. Neurophysiol. 93*(3), 1671–1698. (p 112).

Hille, B. (1992). *Ionic Channels of Excitable Membranes* (2 ed.). Sunderland (Massachusetts): Sinauer. (p 7).

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. (pp 9, 81, and 112).

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. (p 81).

Hines, M. L., & Carnevale, N. T. (2001). NEURON: a tool for neuroscientists. *Neuroscientist 7*(2), 123–135. (pp 129, 130, and 209).

Hines, M. L., Davison, A. P., & Muller, E. (2009). Neuron and python. *Front. Neuroinformatics 3*, 1–12. (pp 130, 131, and 131).

Hines, M. L., Morse, T., Migliore, M., Carnevale, N. T., & Shepherd, G. M. (2004). ModelDB: A database to support computational neuroscience. *J Comput Neurosci 17*(1), 7–11. (p 172).

Hinton, G. E., & Sejnowski, T. J. (1983). Optimal perceptual inference. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 448–453. (p 7).

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. (p 7).

Holt, A. B., & Netoff, T. I. (2013). Computational modeling of epilepsy for an experimental neurologist. *Experimental Neurology 244*, 75 – 86. Special Issue: Epilepsy. (p 3).

Holt, G. R., & Koch, C. (1999). Electrical interactions via the extracellular potential near cell bodies. *J. Comput. Neurosci. 6*(2), 169–184. (pp 10, 10, 128, 128, 129, 129, 172, and 176).

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. USA 79*, 2554–2558. (p 26).

Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik 31*(1), 253–258. (p 24).

Ito, J., Roy, S., Liu, Y., Cao, Y., Fletcher, M., Lu, L., Boughter, J., Grün, S., & Heck, D. (2014). Whisker barrel cortex delta oscillations and gamma power in the awake mouse are linked to respiration. *Nat Commun 5*(3572), 1–10. (p 140).

Iyer, R., Menon, V., Buice, M., Koch, C., & Mihalas, S. (2013). The influence of synaptic weight distribution on neuronal population dynamics. *PLoS Comput Biol 9*(10), e1003248. (p 118).

Izhikevich, E. (2003). Simple model of spiking neurons. *IEEE Trans. Neural Netw. 14*(6), 1569–1572. (p 172).

Izhikevich, E. (2004). Which model to use for cortical spiking neurons? *IEEE Trans. Neural Netw. 5*(15), 1063–1070. (p 7).

Izhikevich, E. M., & Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. *Proc. Natl. Acad. Sci. USA 105*(9), 3593–3598. (pp 115, 119, 119, 123, 125, 125, 127, and 211).

Jacobs, G., Claiborne, B., & Harris, K. (2009). Reconstruction of neuronal morphology. In E. De Schutter (Ed.), *Computational Modeling Methods for Neuroscientists* (1 ed.)., Chapter 8, pp. 187–210. MIT Press, Cambridge, MA. (p 120).

Jaeger, H. (2001). The "echo state" approach to analysing and training recurrent neural networks. Technical Report GMD Report 148, German National Research Center for Information Technology, St. Augustin, Germany. (p 160).

Janssen, H.-K. (1976). On a lagrangean for classical field dynamics and renormalization group calculations of dynamical critical properties. *Zeitschrift für Physik B Condensed Matter 23*(4), 377–380. (pp 51 and 163).

Jefferys, J. (1995). Nonsynaptic modulation of neuronal activity in the brain: electric currents and extracellular ions. *Physiological reviews 75*(4), 689–723. (pp 3 and 5).

Jiang, X., Shen, S., Cadwell, C. R., Berens, P., Sinz, F., Ecker, A. S., Patel, S., & Tolias, A. S. (2015). Principles of connectivity among morphologically defined cell types in adult neocortex. *Science 350*(6264). (p 119).

Jin, J., Wang, Y., Swadlow, H. A., & Alonso, J. M. (2011). Population receptive fields of ON and OFF thalamic inputs to an orientation column in visual cortex. *Nat Neurosci 14*(2), 232–238. (pp 138, 150, and 153).

Jirsa, V. K. (2004). Connectivity and dynamics of neural information processing. *Neuroinformatics 2*(2), 183–204. (p 3).

Jolivet, R., Schürmann, F., Berger, T. K., Naud, R., Gerstner, W., & Roth, A. (2008). The quantitative single-neuron modeling competition. *Biol. Cybern. 99*(4-5), 417–426. (p 172).

Jülich Supercomputing Centre (2015). JUQUEEN: IBM Blue Gene/Q supercomputer system at the Jülich Supercomputing Centre. *Journal of large-scale research facilities 1*. (pp 81 and 98).

Kadmon, J., & Sompolinsky, H. (2015). Transition to chaos in random neuronal networks. *Phys. Rev. X 5*, 041030. (pp 13, 16, 24, 70, 162, 164, 165, and 167).

Kaiboriboon, K., Lüders, H. O., Hamaneh, M., Turnbull, J., & Lhatoo, S. D. (2012). EEG source imaging in epilepsy–practicalities and pitfalls. *Nat Rev Neurol 8*(9), 498–507. (p 174).

Kajikawa, Y., & Schroeder, C. (2011). How local is the local field potential? *Neuron 72*(5), 847–858. (pp 10, 112, and 157).

Kandel, E. R., Markram, H., Matthews, P. M., Yuste, R., & Koch, C. (2013). Neuroscience thinks big (and collaboratively). *Nat. Rev. Neurosci. 14*(9), 659–664. (p 118).

Kappen, H. J., & Spanjers, J. J. (2000). Mean field theory for asymmetric neural networks. *Phys. Rev. E 61*(5), 5658–5663. (p 162).

Katz, P. S., & Frost, W. N. (1996). Intrinsic neuromodulation: altering neuronal circuits from within. *Trends in neurosciences 19*(2), 54–61. (p 5).

Kelley, C. T. (1995). *Iterative Methods for Linear and Nonlinear Equations*. Number 16 in Frontiers in Applied Mathematics. SIAM. (pp 85 and 96).

Kelly, F. (1979). *Reversibility and Stochastic Networks*. Cambridge University Press. (pp 24 and 183).

Kelly, R. C., Smith, M. A., Kass, R. E., & Lee, T. S. (2010). Local field potentials indicate network state and account for neuronal response variability. *J Comput Neurosci 29*(3), 567–579. (p 140).

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. (p 24).

Kilpatrick, Z. P. (2014). Wilson-Cowan model. In *Encyclopedia of Computational Neuroscience*. Springer. (p 80).

Kisvarday, Z. F., & Eysel, U. T. (1992). Cellular organization of reciprocal patchy networks in layer III of cat visual cortex (area 17). *Neuroscience 46*, 275–286. (pp 119, 119, 120, 123, 210, 210, 210, and 210).

Kloeden, P. E., & Platen, E. (1992). *Numerical Solution of Stochastic Differential Equations*. Berlin: Springer. (pp 83, 85, 95, and 95).

Knight, B. W. (1972). Dynamics of encoding in a population of neurons. *J. Gen. Physiol. 59*, 734–766. (p 13).

Kobayashi, R., Tsubo, Y., & Shinomoto, S. (2009). Made-to-order spiking neuron model equipped with a multi-timescale adaptive threshold. *Fronties in Computational Neuroscience 3*(9). doi: 10.3389/neuro.10.009.2009. (p 172).

Koch, C. (1984). Cable theory in neurons with active, linearized membranes. *Biol Cybern 50*(1), 15–33. (p 176).

Koch, C. (1999). *Biophysics of Computation. Information Processing in Single Neurons*. New York, Oxford: Oxford University Press. (p 7).

Koch, C., & Poggio, T. (1985). A simple algorithm for solving the cable equation in dendritic trees of arbitrary geometry. *J Neurosci Methods 12*(4), 303–315. (pp 123 and 173).

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. (pp 8, 80, and 168).

Komori, Y., & Burrage, K. (2014). A stochastic exponential Euler scheme for simulation of stiff biochemical reaction systems. *BIT Numerical Mathematics 54*(4), 1067–1085. (pp 86, 86, 87, and 169).

Konorski, J. (1948). *Conditioned Reflexes and Neuron Organization*. Cambridge Univ. Press, Cambridge, UK. (p 5).

Kriener, B., Enger, H., Tetzlaff, T., Plesser, H. E., Gewaltig, M.-O., & Einevoll, G. T. (2014). Dynamics of self-sustained asynchronous-irregular activity in random networks of spiking neurons with strong synapses. *Front Comput Neurosci 8*, 136. (pp 8, 80, and 172).

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. (p 81).

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. (pp 8, 8, 81, 112, 112, 170, and 171).

Laing, C. R., & Chow, C. C. (2002). A spiking neuron model for binocular rivalry. *J. Comput. Neurosci. 12*(1), 39–53. (p 80).

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. (p 80).

Lapicque, L. (1907). Recherches quantitatives sur l'excitation electrique des nerfs traitee comme une polarization. *J. Physiol. Pathol. Gen. 9*, 620–635. (p 7).

Łęski, S., Kublik, E., Świejkowski, D. a., Wróbel, A., & Wójcik, D. K. (2010). Extracting functional components of neural dynamics with Independent Component Analysis and inverse Current Source Density. *J Comput Neurosci 29*(3), 459–473. (p 142).

Łęski, S., Lindén, H., Tetzlaff, T., Pettersen, K. H., & Einevoll, G. T. (2013). Frequency dependence of signal power and spatial reach of the local field potential. *PLoS Comput. Biol. 9*, e1003137. (pp 10, 112, 138, 140, 144, 144, 147, 157, 172, 177, and 178).

Le Cun, Y., Kanter, I., & Solla, S. A. (1991). Eigenvalues of covariance matrices: Application to neural-network learning. *Phys. Rev. Lett. 66*, 2396–2399. (p 70).

Legenstein, R., & Maass, W. (2007a). Edge of chaos and prediction of computational performance for neural circuit models. *Neural Networks 20*(3), 323–334. (p 56).

Legenstein, R., & Maass, W. (2007b). *What makes a dynamical system computationally powerful?*, pp. 127–154. MIT Press. (p 160).

Lehmann, E. L., & Casella, G. (2006). *Theory of point estimation*. Springer Science & Business Media. (p 194).

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. (pp 82, 91, 169, and 169).

Li, Z., & Dayan, P. (1999). Computational differences between asymmetrical and symmetrical networks. *Network: Comput. Neural Systems 10*, 59–77. (p 24).

Lindén, H., Hagen, E., Leski, S., Norheim, E. S., Pettersen, K. H., & Einevoll, G. T. (2014). LFPy: a tool for biophysical simulation of extracellular potentials generated by detailed model neurons. *Front. Neuroinformatics 7*(41), 1–15. (pp 9, 10, 121, 128, 129, 129, 130, 130, 130, 130, 131, 171, 172, and 173).

Lindén, H., Pettersen, K. H., & Einevoll, G. T. (2010). Intrinsic dendritic filtering gives low-pass power spectra of local field potentials. *J. Comput. Neurosci. 29*(3), 423–444. (pp 8, 10, 113, 128, 140, 140, 140, 142, 144, 153, and 174).

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. (pp 10, 112, 128, 138, 144, 144, 147, 157, 162, 172, 175, 177, and 178).

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. (pp 15, 50, and 81).

Logothetis, N. K., Kayser, C., & Oeltermann, A. (2007). In vivo measurement of cortical impedance spectrum in monkeys: implications for signal propagation. *Neuron 55*(5), 809–823. (p 173 and 173).

London, M., & Häusser, M. (2005). Dendritic computation. *Annu. Rev. Neurosci. 28*, 503–532. (pp 8 and 172).

Maass, W., Natschläger, T., & Markram, H. (2002). Real-time computing without stable states: a new framwork for neural compuation based on perturbation. *Neural Comput. 14*(11), 2531–2560. (p 160).

Maier, A., Adams, G., Aura, C., & Leopold, D. (2010). Distinct superficial and deep laminar domains of activity in the visual cortex during rest and stimulation. *Front Syst Neurosci 4*(31). DOI: 10.3389/fnsys.2010.00031. (pp 11, 138, 138, and 138).

Mainen, Z. F., & Sejnowski, T. J. (1996). Influence of dendritic structure on firing pattern in model neocortical neurons. *Nature 382*, 363–366. (pp 6, 119, 120, 123, 210, 210, and 210).

Makarov, V. A., Makarova, J., & Herreras, O. (2010). Disentanglement of local field potential sources by independent component analysis. *J Comput Neurosci 29*(3), 445–457. (p 142).

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. (p 102).

Marino, A. A. (1994). Neurobiophysics. In *Encyclopedia of Applied Physics*, pp. 297–322. VCH Publishers, Inc. (p 6).

Markram, H., Lübke, J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science 275*, 213–215. (p 24).

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. (pp 8, 9, 81, 112, 119, and 172).

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., & Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nat. Rev. Neurosci. 5*(10), 793–807. (p 146).

Martin, P., Siggia, E., & Rose, H. (1973). Statistical dynamics of classical systems. *Phys. Rev. A 8*(1), 423–437. (pp 50, 51, and 163).

Mascaro, A. L. A., Silvestri, L., Sacconi, L., & Pavone, F. S. (2015). Towards a comprehensive understanding of brain machinery by correlative microscopy. *Journal of Biomedical Optics 20*(6), 061105–061105. (p 4).

Mastroguiseppe, F., & Ostojic, S. (2016). Intrinsically-generated fluctuating activity in excitatory-inhibitory networks. *arXiv*, 1605.04221. (pp 164 and 165).

Mattia, M., & Del Guidice, P. (2002). Population dynamics of interacting spiking neurons. *Phys. Rev. E 66*, 051917. (pp 14 and 81).

Mattia, M., & Del Guidice, P. (2004). Finite-size dynamics of inhibitory and excitatory interacting spiking neurons. *Phys. Rev. E 70*, 052903. (p 81).

Mazzoni, A., Brunel, N., Cavallari, S., Logothetis, N. K., & Panzeri, S. (2011). Cortical dynamics during naturalistic sensory stimulations: experiments and models. *J Physiol Paris 105*(1), 2–15. (p 112).

Mazzoni, A., Lindén, H., Cuntz, H., Lansner, A., Panzeri, S., & Einevoll, G. (2015). Computing the local field potential (LFP) from integrate-and-fire network models. *PLoS Comput. Biol. 11*, 1–38. (pp 113, 144, 175, and 177).

Mazzoni, A., Panzeri, S., Logothetis, N. K., & Brunel, N. (2008). Encoding of naturalistic stimuli by local field potential spectra in networks of excitatory and inhibitory neurons. *PLoS Comput. Biol. 4*(12), e1000239. (p 112).

Mazzoni, A., Whittingstall, K., Brunel, N., Logothetis, N. K., & Panzeri, S. (2010). Understanding the relationships between spike rate and delta/gamma frequency bands of LFPs and EEGs using a local cortical network model. *NeuroImage 52*, 956–972. (pp 112 and 162).

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in cognitive sciences 14*(8), 348–356. (p 6).

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. (p 80).

McClelland, J. L., & Rumelhart, D. E. (1989). *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*. MIT press. (p 80).

McLeod, P., Plunkett, K., & Rolls, E. T. (1998). *Introduction to Connectionist Modelling of Cognitive Processes*. Oxford: Oxford University Press. (pp 6 and 7).

Metzner, W., Salmhofer, M., Honerkamp, C., Meden, V., & Schönhammer, K. (2012). Functional renormalization group approach to correlated fermion systems. *Rev. Mod. Phys. 84*, 299–352. (p 179).

Mézard, M., Parisi, G., & Virasoro, M. (1987). *Spin Glass Theory and Beyond (World Scientific Lecture Notes in Physics, Vol 9)*. World Scientific Publishing Company. (p 24).

Mézard, M., & Sakellariou, J. (2011). Exact mean-field inference in asymmetric kinetic ising systems. *J. Stat. Mech. Theor. Exp. 2011*(07), L07001. (pp 25, 45, and 162).

Miekkala, U., & Nevanlinna, O. (1987). Convergence of dynamic iteration methods for initial value problems. *SIAM J. Sci. and Stat. Comput. 8*(4), 459–482. (p 91).

Migliore, M., Cavarretta, F., Hines, M. L., & Shepherd, G. M. (2014). Distributed organization of a brain microcircuit analyzed by three-dimensional modeling: the olfactory bulb. *Front Comput Neurosci 8*(50), 1–14. (pp 8 and 112).

Mitzdorf, U. (1985). Current source-density method and application in cat cerebral cortex: Investigation of evoked potentials and eeg phenomena. *Physiol. Rev. 65*(1), 37–100. (pp 138, 139, 139, and 175).

Mitzdorf, U., & Singer, W. (1979). Excitatory synaptic ensemble properties in the visual cortex of the macaque monkey: a current source density analysis of electrically evoked potentials. *J Comp Neurol 187*(1), 71–83. (pp 138 and 139).

Molgedey, L., Schuchhardt, J., & Schuster, H. (1992). Suppressing chaos in neural networks by noise. *Phys. Rev. Lett. 69*(26), 3717. (p 70).

Montbrió, E., Pazó, D., & Roxin, A. (2015). Macroscopic description for networks of spiking neurons. *Phys. Rev. X 5*, 021028. (pp 14, 81, and 163).

Monteforte, M., & Wolf, F. (2010). Dynamical entropy production in spiking neuron networks in the balanced state. *Phys. Rev. Lett. 105*(26), 268104. (p 56).

Mora, T., Deny, S., & Marre, O. (2015). Dynamical criticality in the collective activity of a population of retinal neurons. *Phys. Rev. Lett. 114*, 078105. (p 167 and 167).

Mora, T., Walczak, A. M., Bialek, W., & Callan, C. G. (2010). Maximum entropy models for antibody diversity. *Proc. Natl. Acad. Sci. USA 107*(12), 5405–5410. (p 45).

Moran, R. J., Stephan, K. E., Kiebel, S. J., Rombach, N., O'Connor, W. T., Murphy, K. J., Reilly, R. B., & Friston, K. J. (2008). Bayesian estimation of synaptic physiology from the spectral responses of neural masses. *Neuroimage 42*(1), 272–284. (p 150).

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., & Pouget, A. (2014). Information-limiting correlations. *Nat. Neurosci. 17*, 1410–1417. (p 70).

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. (p 90 and 90).

Morrison, A., Diesmann, M., & Gerstner, W. (2008). Phenomenological models of synaptic plasticity based on spike-timing. *Biol. Cybern. 98*, 459–478. (pp 24 and 163).

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. (pp 82, 82, 88, 89, 92, and 169).

Moshe, M., & Zinn-Justin, J. (2003). Quantum field theory in the large n limit: a review. *Physics Reports 385*, 69–228. (pp 64, 65, and 67).

Moulin, C., Glière, A., Barbier, D., Joucla, S., Yvert, B., Mailley, P., & Guillemaud, R. (2008). A new 3-D finite-element model based on thin-film approximation for microelectrode array recording of extracellular action potential. *IEEE transactions on bio-medical engineering 55*(2 Pt 1), 683–92. (p 173).

Nakanishi, K., & Takayama, H. (1997). Mean-field theory for a spin-glass model of neural networks: Tap free energy and the paramagnetic to spin-glass transition. *J. Phys. A: Math. Gen. 30*, 8085–8094. (p 161).

Nauhaus, I., Busse, L., Carandini, M., & Ringach, D. L. (2009). Stimulus contrast modulates functional connectivity in visual cortex. *Nat. Neurosci. 12*, 70–76. (pp 150 and 153).

Negele, J. W., & Orland, H. (1998). *Quantum Many-Particle Systems*. Perseus Books. (pp 12, 159, and 164).

Nelson, M. J., & Pouget, P. (2010). Do electrode properties create a problem in interpreting local field potential recordings? *J Neurophysiol 103*(5), 2315–2317. (pp 130 and 173).

Nelson, M. J., Pouget, P., Nilsen, E. A., Patten, C. D., & Schall, J. D. (2008). Review of signal distortion through metal microelectrode recording circuits and filters. *J Neurosci Methods 169*(1), 141–157. (p 130).

Ness, T. V., Potworowski, J., Łęski, S., Chintaluri, C., Głąbska, H., Wójcik, D. K., & Einevoll, G. T. (2015). Modelling and analysis of electrical potentials recorded in multielectrode arrays (MEAs). *Neuroinformatics*, 1–24. (pp 130, 130, 130, 173, 173, 173, and 174).

Ness, T. V., Remme, M. W. H., & Einevoll, G. T. (2016). Active subthreshold dendritic conductances shape the local field potential. *J Physiol Accepted article*. (pp 176 and 177).

Newman, M. E. (2003). The structure and function of complex networks. *SIAM review 45*(2), 167–256. (p 125).

Nichols, E. J., & Hutt, A. (2015). Neural field simulator: two-dimensional spatio-temporal dynamics involving finite transmission speed. *Front. Neuroinform. 9*, 25. (p 81).

Nicholson, C., & Freeman, J. (1975). Theory of current source-density analysis and determination of conductivity tensor for anuran cerebellum. *J. Neurophysiol. 2*(38), 356–368. (pp 10, 112, 114, 138, and 172).

Nicolas-Alonso, L. F., & Gomez-Gil, J. (2012). Brain computer interfaces, a review. *Sensors 12*(2), 1211–1279. (p 3).

Nishimori, H. (2001). *Statistical Physics of Spin Glasses and Information Processing An Introduction*. Clarendron Press, Oxford. (pp 24, 63, and 161).

Nordlie, E., Gewaltig, M.-O., & Plesser, H. E. (2009). Towards reproducible descriptions of neuronal network models. *PLoS Comput. Biol. 5*(8), e1000456. (p 116).

Nowak, L. G., Azouz, R., Sanchez-Vives, M. V., Gray, C. M., & McCormick, D. A. (2003). Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. *J. Neurophysiol. 89*(3), 1541–66. (p 119).

Nunez, P. L., & Ramesh, S. (2006). *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press. (pp 10, 128, 129, 174, 174, and 174).

Obien, M. E. J., Deligkaris, K., Bullmann, T., Bakkum, D. J., & Frey, U. (2015). Revealing neuronal function through microelectrode array recordings. *Frontiers in Neuroscience 8*, 423. (p 4, 4, and 4).

Ocker, G. K., Litwin-Kumar, A., & Doiron, B. (2015). Self-organization of microcircuits in networks of spiking neurons with plastic synapses. *PLoS Comput. Biol. 11*(8), e1004458. (p 162).

Ohbayashi, M., Ohki, K., & Miyashita, Y. (2003). Conversion of working memory to motor sequence in the monkey premotor cortex. *Science 301*(5630), 233–236. (pp 8, 80, and 172).

Ohira, T., & Cowan, J. (1997). *Stochastic Neurodynamics and the System Size Expansion*. Springer. (p 159).

Okun, M., & Lampl, I. (2008). Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nat. Neurosci. 11*(5), 535–537. (pp 8, 80, and 172).

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. (p 38).

Onsager, L., & Machlup, S. (1953). Fluctuations and irreversible processes. *91*(6), 1505–1512. (p 53).

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. (p 80).

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. (p 80).

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. (p 80).

Ostojic, S. (2014). Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nat. Neurosci. 17*, 594–600. (pp 8, 80, 160, 167, 167, and 172).

Ostojic, S., & Brunel, N. (2011). From spiking neuron models to linear-nonlinear models. *PLoS Comput. Biol. 7*(1), e1001056. (pp 81 and 170).

Parisi, G. (1980). The order parameter for spin glasses: A function on the interval 0-1. *Journal of Physics A: Mathematical and General 13*(3), 1101. (p 56).

Pernice, V., Staude, B., Cardanobile, S., & Rotter, S. (2011). How structure determines correlations in neuronal networks. *PLoS Comput. Biol. 7*(5), e1002059. (pp 15, 50, 72, 100, 166, and 197).

Pernice, V., Staude, B., Cardanobile, S., & Rotter, S. (2012). Recurrent interactions in spiking networks with arbitrary topology. *Phys. Rev. E 85*(3), 031916. (pp 50, 72, 72, and 197).

Pettersen, K., Hagen, E., & Einevoll, G. (2008). Estimation of population firing rates and current source densities from laminar electrode recordings. *J. Comput. Neurosci. 24*, 291–313. (pp 10, 114, 128, 129, 134, 134, 138, 138, and 176).

Pettersen, K. H., Devor, A., Ulbert, I., Dale, A. M., & Einevoll, G. T. (2006). Current-source density estimation based on inversion of electrostatic forward solution: Effects of finite extent of neuronal activity and conductivity discontinuities. *J. Neurosci. Methods 154*(1-2), 116–133. (pp 114, 138, and 173).

Pettersen, K. H., & Einevoll, G. T. (2008). Amplitude variability and extracellular low-pass filtering of neuronal spikes. *Biophys J 94*(3), 784–802. (p 10).

Pettersen, K. H., Lindén, H., Dale, A. M., & Einevoll, G. T. (2012). Extracellular spikes and CSD. In R. Brette & A. Destexhe (Eds.), *Handbook of Neural Activity Measurement*, pp. 92–135. Cambridge University Press. (pp 112, 113, 129, 153, and 173).

Plefka, T. (1982). Convergence condition of the tap equation for the infinite-ranged ising spin glass model. *J Phys A: Math Gen 15*, 1971–1978. (pp 161 and 167).

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. (p 82).

Plesser, H. E., Eppler, J. M., Morrison, A., Diesmann, M., & Gewaltig, M.-O. (2007). Efficient parallel simulation of large-scale neuronal networks on clusters of multiprocessor computers. In A.-M. Kermarrec, L. Bougé, & T. Priol (Eds.), *Euro-Par 2007: Parallel Processing*, Volume 4641 of *Lecture Notes in Computer Science*, Berlin, pp. 672–681. Springer-Verlag. (pp 8 and 112).

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. (p 90).

Poirazi, P. (2009). Information processing in single cells and small networks: Insights from compartmental models. In *Computational methods in science and engineering: advances in computational science: lectures presented at the International Conference on Computational Methods in Science and Engineering 2008 (ICCMSE 2008), Hersonissos, Crete, Greece, 25-30 September 2008*, Volume 1, pp. 158. Amer Inst of Physics. (p 8).

Poirazi, P., & Mel, B. W. (2001). Impact of active dendrites and structural plasticity on the memory capacity of neural tissue. *Neuron 29*, 779–796. (p 8).

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. (pp 5, 8, 80, 81, 83, 104, 105, 114, 115, 115, 115, 115, 115, 115, 118, 118, 118, 119, 125, 125, 127, 128, 131, 132, 134, 134, 136, 139, 140, 141, 142, 146, 172, 175, 175, 177, 177, and 177).

Preuss, T. M., & Coleman, G. Q. (2002). Human-specific organization of primary visual cortex: alternating compartments of dense cat-301 and calbindin immunoreactivity in layer 4a. *Cerebral Cortex 12*(7), 671–691. (p 4).

Rajan, K., & Abbott, L. F. (2006). Eigenvalue spectra of random matrices for neural networks. *Phys. Rev. Lett. 97*, 188104. (pp 74 and 97).

Rall, W. (1964). Theoretical significance of dendritic trees for neuronal input-output relations. In R. F. Reiss (Ed.), *Neural Theory and Modeling*. Palo Alto: Stanford University Press. (pp 8 and 122).

Rall, W. (2009). Rall model. *Scholarpedia 4*(4), 1369. revision 91692. (p 122).

Rall, W., & Shepherd, G. M. (1968). Theoretical reconstruction of field potentials and dendrodendritic synaptic interactions in olfactory bulb. *J Neurophysiol 31*(6), 884–915. (pp 10, 128, and 172).

Ray, S., & Maunsell, J. H. R. (2011). Different Origins of Gamma Rhythm and High-Gamma Activity in Macaque Visual Cortex. *PLoS Comput. Biol. 9*(4), e1000610. (p 176).

Reimann, M. W., Anastassiou, C. A., Perin, R., Hill, S. L., Markram, H., & Koch, C. (2013). A biophysically detailed model of neocortical local field potentials predicts the critical role of active membrane currents. *Neuron 79*(2), 375–390. (pp 8, 10, 112, 113, 128, 171, and 176).

Reimann, M. W., King, J. G., Muller, E. B., Ramaswamy, S., & Markram, H. (2015). An algorithm to predict the connectome of neural microcircuits. *Front Comput Neurosci 9*(120), 1–18. (p 119).

Remme, M. W. H., & Rinzel, J. (2011). Role of active dendritic conductances in subthreshold input integration. *J Comput Neurosci 31*(1), 13–30. (pp 176 and 177).

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. (pp 7, 8, 12, 16, 17, 24, 25, 26, 27, 27, 27, 28, 31, 33, 62, 70, 80, 146, 147, 158, 159, 159, 161, 166, 166, 172, and 178).

Reyes-Puerta, V., Sun, J.-J., Kim, S., Kilb, W., & Luhmann, H. J. (2014). Laminar and columnar structure of sensory-evoked multineuronal spike sequences in adult rat barrel cortex in vivo. *Cereb Cortex 25*(8), 2001–2021. (pp 139, 139, 140, and 175).

Reyes-Puerta, V., Yang, J.-W., Siwek, M. E., Kilb, W., Sun, J.-J., & Luhmann, H. J. (2016). Propagation of spontaneous slow-wave activity across columns and layers of the adult rat barrel cortex in vivo. *Brain Struct Funct*, 1–21. (p 175).

Riehle, A., Wirtssohn, S., Grün, S., & Brochier, T. (2013). Mapping the spatio-temporal structure of motor cortical lfp and spiking activities during reach-to-grasp movements. *Frontiers in Neural Circuits 7*, 48. doi: 10.3389/fncir.2013.00048. (pp 11, 11, 71, and 138).

Risken, H. (1996). *The Fokker-Planck Equation*. Springer Verlag Berlin Heidelberg. (pp 14, 14, 26, 40, 41, 43, 73, 87, 100, and 160).

Roach, J. P., Sander, L. M., & Zochowski, M. R. (2016). Memory recall and spike-frequency adaptation. *Phys. Rev. E 93*, 052307. (p 158).

Robinson, D. A. (1968). The electrical properties of metal microelectrodes. *Proceedings of the IEEE 56*(6), 1065–1071. (p 130).

Rossant, C., Kadir, S. N., Goodman, D. F., Schulman, J., Hunter, M. L., Saleem, A. B., Grosmark, A., Belluscio, M., Denfield, G. H., Ecker, A. S., Tolias, A. S., Solomon, S., Buzsáki, G., Carandini, M., & Harris, K. D. (2016). Spike sorting for large, dense electrode arrays. *Nat. Neurosci.*. (p 10).

Rössert, C., Pozzorini, C., Chindemi, G., Davison, A. P., Eroe, C., King, J., Newton, T. H., Nolte, M., Ramaswamy, S., Reimann, M. W., et al. (2016). Automated point-neuron simplification of data-driven microcircuit models. *ArXiv e-prints 1604.00087 [q-bio.NC]*. (p 172).

Rotter, S., & Diesmann, M. (1999). Exact digital simulation of time-invariant linear systems with applications to neuronal modeling. *Biol. Cybern. 81*(5/6), 381–402. (p 116).

Roudi, Y., Aurell, E., & Hertz, J. A. (2009). Statistical physics of pairwise probability models. *Front. Comput. Neurosci. 3*(22), doi: 10.3389/neuro.10.022.2009. (p 45).

Roudi, Y., & Hertz, J. (2011). Mean field theory for nonequilibrium network reconstruction. *Physical Review Letters 106*(4), 048702. (pp 162 and 166).

Rougier, N. P., & Fix, J. (2012). Dana: Distributed numerical and adaptive modelling framework. *Network: Computation in Neural Systems 23*(4), 237–253. (p 80).

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. (pp 80, 83, 102, 102, 103, 103, and 103).

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. (p 102).

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. (pp 8, 80, and 168).

Rubino, D., Robbins, K. A., & Hatsopoulos, N. G. (2006). Propagating waves mediate information transfer in the motor cortex. *Nat. Neurosci. 9*, 1549–1557. (p 3).

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. (pp 6, 7, 26, and 80).

Sabah, N. H., & Leibovic, K. N. (1969). Subthreshold oscillatory responses of the Hodgkin-Huxley cable model for the squid giant axon. *Biophys J 9*(10), 1206–1222. (p 176).

Sakata, S., & Harris, K. D. (2009). Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. *Neuron 64*, 404–418. (pp 138 and 139).

Samsonovich, A., & McNaughton, B. L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci. 17*(15), 5900–5920. (p 80).

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. (p 80).

Sarid, L., Bruno, R., Sakmann, B., Segev, I., & Feldmeyer, D. (2007). Modeling a layer 4-to-layer 2/3 module of a single column in rat neocortex: Interweaving in vitro and in vivo experimental observations. *Proc. Natl. Acad. Sci. USA 104*(41), 16353–16358. (p 118).

Schmidt, M., Bakker, R., Diesmann, M., & van Albada, S. (2016a). Full-density multi-scale account of structure and dynamics of macaque visual cortex. *arXiv preprint arXiv:1511.09364v4*. (pp 5 and 177).

Schmidt, M., Bakker, R., Diesmann, M., & van Albada, S. (2016b). Full-density multi-scale account of structure and dynamics of macaque visual cortex. *arXiv preprint arXiv:1511.09364v3*. (p 81).

Schmitz, D., Schuchmann, S., Fisahn, A., Draguhn, A., Buhl, E. H., Petrasch-Parwez, E., Dermietzel, R., Heinemann, U., & Traub, R. D. (2001). Axo-axonal coupling: a novel mechanism for ultrafast neuronal communication. *Neuron 31*(5), 831–840. (p 5).

Schneidman, E., Berry, M. J., Segev, R., & Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature 440*, 1007–1012. (p 159).

Schomburg, E. W., Anastassiou, C. A., Buzsáki, G., & Koch, C. (2012). The spiking component of oscillatory extracellular potentials in the rat hippocampus. *J Neurosci 32*(34), 11798–11811. (p 10).

Schöner, G., Spencer, J., & Group, D. (2015). *Dynamic Thinking: A Primer on Dynamic Field Theory*. Oxford Series in Developmental Cognitive Neuroscience. Oxford University Press. (p 80).

Schreiber, T., & Schmitz, A. (2000). Surrogate time series. *Phys. D 142*(3-4), 346–382. (p 136).

Schuecker, J., Diesmann, M., & Helias, M. (2015). Modulated escape from a metastable state driven by colored noise. *Phys. Rev. E 92*, 052119. (pp 8, 15, 81, and 173).

Schuecker, J., Goedeke, S., Dahmen, D., & Helias, M. (2016). Functional methods for disordered neural networks. *arXiv*. 1605.06758 [cond-mat.dis-nn]. (pp vii, 16, and 101).

Schuecker, J., Schmidt, M., van Albada, S., Diesmann, M., & Helias, M. (2016). Fundamental activity constraints lead to specific interpretations of the connectome. *arXiv preprint arXiv:1509.03162v3*. (pp 16, 83, 104, and 170).

Schurz, H., & Schneider, K. R. (2005). Waveform relaxation methods for stochastic differential equations. *International Journal of Numerical Analysis and Modeling 3*(2), 232–254. (p 91).

Segev, I., Fleshman, J. W., & Burke, R. E. (1989). Compartmental models of complex neurons. In C. Koch & I. Segev (Eds.), *Methods in Neuronal Modeling: From Synapses to Networks*, Chapter 3, pp. 63–96. Cambridge, Massachusetts: A Bradford Book, The MIT Press. (p 8).

Seifert, U. (2012). Stochastic thermodynamics, fluctuation theorems, and molecular machines. *Rep. Prog. Phys. 75*, 126001. (p 24).

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*. (pp 5 and 177).

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. (pp 8, 70, 80, and 172).

Shamir, M., & Sompolinsky, H. (2001). Correlation codes in neuronal populations. In *Advances in Neural Information Processing Systems*, pp. 277–284. MIT Press. (p 24).

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. (p 70).

Sherrington, D., & Kirkpatrick, S. (1975). Solvable model of a spin-glass. *Phys. Rev. Lett. 35*, 1792–1796. (pp 12, 24, 63, 161, and 164).

Shinomoto, S. (1987). A cognitive and associative memory. *Biol. Cybern. 57*(3), 197–206. (p 158).

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. (p 86).

Siegert, A. J. (1951). On the first passage time probability problem. *Phys. Rev. 81*(4), 617–623. (pp 14 and 81).

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. (pp 8, 80, and 172).

Sollich, P., Tantari, D., Annibale, A., & Barra, A. (2014). Extensive parallel processing on scale-free networks. *Phys. Rev. Lett. 113*, 238106. (p 162).

Solovey, G., Miller, K., Ojemann, J., Magnasco, M., & Cecchi, G. (2012). Self-regulated dynamical criticality in human ecog. *Frontiers in Integrative Neuroscience 6*, 44. (p 167).

Sommers, H., Crisanti, A., Sompolinsky, H., & Stein, Y. (1988). Spectrum of large random asymmetric matrices. *Phys. Rev. Lett. 60*(19), 1895–1898. (pp 70, 167, 168, 201, and 201).

Sompolinsky, H., Crisanti, A., & Sommers, H. J. (1988). Chaos in random neural networks. *Phys. Rev. Lett. 61*, 259–262. (pp 12, 12, 14, 16, 24, 34, 50, 56, 61, 67, 70, 70, 75, 83, 100, 101, 161, 163, 164, 164, 167, 167, and 167).

Sompolinsky, H., Yoon, H., Kang, K., & Shamir, M. (2001). Population coding in neuronal systems with correlated noise. *Phys. Rev. E 64*(5), 051904. (p 70).

Sompolinsky, H., & Zippelius, A. (1981). Dynamic theory of the spin-glass phase. *Phys. Rev. Lett. 47*, 359–362. (pp 12, 24, 50, 56, 164, and 179).

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. (pp 12, 24, 34, 50, 59, 60, 65, 161, 164, 164, 168, and 201).

Song, S., Sjöström, P., Reigl, M., Nelson, S., & Chklovskii, D. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biol. 3*(3), e68. (pp 118, 125, 168, 201, 201, 201, and 202).

Sornette, D. (2014). Physics and financial economics (1776–2014): puzzles, ising and agent-based models. *Rep. Prog. Phys. 77*(6), 062001. (p 24).

Sporns, O., & Kötter, R. (2004). Motifs in brain networks. *PLoS Biol 2*(11), e369. (p 201).

Stavrinou, M. L., Hagen, E., Dahmen, D., , Lindén, H., Tetzlaff, T., van Albada, S. J., Grün, S., Diesmann, M., & Einevoll, G. T. (2014). Computing local field potentials based on spiking cortical networks. In *Conference Abstracts of Neuroinformatics 2014*. Front. Neuroinform. 54. 10.3389/conf.fninf.2014.18.00054. (pp viii, 111, and 133).

Stein, D. L., & Newman, C. M. (2013). *Spin Glasses and Complexity*. Princeton U. Press. (p 24).

Stein, R. B. (1967). Some models of neuronal variability. *Biophys. J. 7*, 37–68. (p 7).

Stepanyants, A., Hirsch, J. A., Martinez, L. M., Kisvárday, Z. F., Ferecskó, A. S., & Chklovskii, D. B. (2008). Local potential connectivity in cat primary visual cortex. *Cereb. Cortex 18*(1), 13–28. (pp 119, 119, 119, 120, 122, 122, 123, 210, 210, 210, 210, 210, 210, 210, and 210).

Stepanyants, A., Martinez, L. M., Ferecskó, A. S., & Kisvárday, Z. F. (2009). The fractions of short- and long-range connections in the visual cortex. *Proc. Nat. Acad. Sci. USA 106*(9), 3555–3560. (pp 168 and 197).

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. (p 168).

Stern, M., Sompolinsky, H., & Abbott, L. F. (2014). Dynamics of random neural networks with bistable units. *Phys. Rev. E 90*, 062710. (pp 12, 16, 50, 105, and 164).

Stevens, J.-L. R., Law, J. S., Antolík, J., & Bednar, J. A. (2013). Mechanisms for stable, robust, and adaptive development of orientation maps in the primary visual cortex. *J. Neurosci. 33*(40), 15747–15766. (p 80).

Stevenson, I. H., & Kording, K. P. (2011). How advances in neural recording affect data analysis. *Nat. Neurosci. 14*(2), 139–142. (pp 70 and 162).

Steyn-Ross, M. L., & Steyn-Ross, D. A. (2016). From individual spiking neurons to population behavior: Systematic elimination of short-wavelength spatial modes. *Phys. Rev. E 93*, 022402. (p 180).

Sussillo, D., & Abbott, L. F. (2009). Generating coherent patterns of activity from chaotic neural networks. *Neuron 63*(4), 544–557. (p 56).

Swadlow, H. A., Gusev, A. G., & Bezdudnaya, T. (2002). Activation of a cortical column by a thalamocortical impulse. *J. Neurosci. 22*(17), 7766–7773. (pp 138, 150, and 153).

Szymanski, F. D., Garcia-Lazaro, J. A., & Schnupp, J. W. H. (2009). Current source density profiles of stimulus-specific adaptation in rat auditory cortex. *J Neurophysiol 102*(3), 1483–1490. (p 138).

Szymanski, F. D., Rabinowitz, N. C., Magri, C., Panzeri, S., & Schnupp, J. W. H. (2011). The laminar and temporal structure of stimulus information in the phase of field potentials of auditory cortex. *J Neurosci 31*(44), 15787–15801. (p 138).

Tanaka, T. (1998). Mean-field theory of boltzmann machine learning. *Phys. Rev. E 58*(2), 2302–2310. (p 161).

Telenczuk, B., Dehghani, N., Le Van Quyen, M., Cash, S. S., Halgren, E., Hatsopoulos, N. G., & Destexhe, A. (2016). Local field potentials primarily reflect inhibitory neuron activity in human and monkey cortex. *bioRxiv*. (p 175).

Teramae, J.-N., & Fukai, T. (2014). Computational implications of lognormally distributed synaptic weights. *Proceedings of the IEEE 102*(4), 500–512. (p 118).

Tetzlaff, T., Helias, M., Einevoll, G., & Diesmann, M. (2012). Decorrelation of neural-network activity by inhibitory feedback. *PLoS Comput. Biol. 8*(8), e1002596. (pp 8, 15, 15, 17, 50, 62, 70, 74, 100, 140, 146, 147, 147, 161, 162, 163, 165, 178, and 195).

Tetzlaff, T., Rotter, S., Stark, E., Abeles, M., Aertsen, A., & Diesmann, M. (2008). Dependence of neuronal correlations on filter characteristics and marginal spike-train statistics. *Neural Comput. 20*(9), 2133–2184. (p 15).

Thomson, A. M., & Bannister, P. (2003). Interlaminar connections in the neocortex. *Cereb. Cortex 13*, 5–14. (p 5).

Thouless, D. J., Anderson, P. W., & Palmer, R. G. (1977). Solution of 'solvable model of a spin glass'. *Philosophical Magazine 35*(3), 593–601. (pp 24, 45, 161, and 166).

Tian, P., Devor, A., Sakadzić, S., Dale, A. M., & Boas, D. A. (2011). Monte Carlo simulation of the spatial resolution and depth sensitivity of two-dimensional optical imaging of the brain. *J Biomed Opt 16*(1), 016006. (p 174).

Tkacik, G., Mora, T., Marre, O., Amodei, D., Palmer, S. E., Berry, M. J., & Bialek, W. (2015). Thermodynamics and signatures of criticality in a network of neurons. *Proc. Natl. Acad. Sci. USA 112*(37), 11508–11513. (p 167).

Tomsett, R. J., Ainsworth, M., Thiele, A., Sanayei, M., Chen, X., Gieselmann, M. A., Whittington, M. A., Cunningham, M. O., & Kaiser, M. (2014). Virtual Electrode Recording Tool for EXtracellular potentials (VERTEX): comparing multi-electrode recordings from simulated and biological mammalian cortical tissue. *Brain Struct Func 220*(4), 2333–2353. (pp 128, 177, and 177).

Torben-Nielsen, B., & De Schutter, E. (2014). Context-aware modeling of neuronal morphologies. *Front Neuroanat 8*(92), 1–14. (p 177).

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. (p 3).

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. (pp 56, 70, and 167).

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. (pp 15, 50, 50, 62, 72, 81, 100, 167, and 197).

Tyrcha, J., & Hertz, J. (2013). Network inference with hidden units. *arXiv* (03), 1301.7274v1. (pp 25 and 160).

Uhlenbeck, G. E., & Ornstein, L. S. (1930). On the theory of the brownian motion. *Phys. Rev. 36*, 823–841. (p 26).

Ursino, M., & Cara, G.-E. L. (2006). Travelling waves and eeg patterns during epileptic seizure: Analysis with an integrate-and-fire neural network. *Journal of Theoretical Biology 242*(1), 171–187. (p 112).

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review 108*(3), 550. (p 80).

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. (pp 39, 82, 158, 158, and 178).

van Kampen, N. G. (1992). *Stochastic Processes in Physics and Chemistry*. Amsterdam: North-Holland, Elsevier Science Publishers. (pp 36 and 159).

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. (p 140).

van Vreeswijk, C., & Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science 274*, 1724–1726. (pp 7, 7, 8, 12, 16, 25, 26, 28, 56, 63, 70, 70, 80, 160, 163, 164, and 172).

van Vreeswijk, C., & Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Comput. 10*, 1321–1371. (pp 28, 31, 40, 44, 44, 63, 158, 158, 159, 159, 159, 168, 194, 200, and 201).

Voges, N., & Perrinet, L. U. (2012). Complex dynamics in recurrent cortical networks based on spatially realistic connectivities. *Front. Comput. Neurosci. 6*, 41. (p 102).

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. (p 102).

Wagner, T., Eden, U., Rushmore, J., Russo, C. J., Dipietro, L., Fregni, F., Simon, S., Rotman, S., Pitskel, N. B., Ramos-Estebanez, C., Pascual-Leone, A., Grodzinsky, A. J., Zahn, M., & Valero-Cabré, A. (2014). Impact of brain tissue filtering on neurostimulation fields: a modeling study. *Neuroimage 85*(Pt 3), 1048–1057. (p 173).

Wainrib, G., & Touboul, J. (2013). Topological and dynamical complexity of random neural networks. *Phys. Rev. Lett. 110*(11), 118101. (pp 75 and 167).

Wang, X.-J., & Kennedy, H. (2016). Brain structure and dynamics across scales: in search of rules. *Curr. Opin. Neurobiol. 37*, 92–98. (pp 3 and 5).

Weitzenfeld, A., Arbib, M. A., & Alexander, A. (2002). *The neural simulation language: A system for brain modeling*. MIT Press. (p 80).

Welch, P. D. (1967). The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio Electroacoustics 15*(2), 70–73. (p 136).

Willshaw, D., Buneman, O., & Longuet-Higgins, H. (1969). Non holographic associative memory. *Nature 222*, 960–962. (p 26).

Wilson, H. R. (2003). Computational evidence for a rivalry hierarchy in vision. *Proc. Natl. Acad. Sci. USA 100*(24), 14499–14503. (p 80).

Wilson, H. R., & Cowan, J. D. (1972a). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal 12*(1), 1 – 24. (p 13).

Wilson, H. R., & Cowan, J. D. (1972b). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J. 12*(1), 1–24. (pp 80 and 106).

Wilson, H. R., & Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik 13*, 55–80. (pp 13 and 106).

Wilson, K. G. (1975). The renormalization group: Critical phenomena and the kondo problem. *Rev. Mod. Phys. 47*, 773–840. (p 179).

Wio, H. S., Colet, P., & San Miguel, M. (1989). Path-integral formulation for stochastic processes driven by colored noise. *Phys Rev A 40*(12), 7312–7324. (p 51).

Wong, K.-F., & Wang, X.-J. (2006). A recurrent network mechanism of time integration in perceptual decisions. *J. Neurosci. 26*(4), 1314–1328. (pp 83 and 104).

Wybo, W. A. M., Boccalini, D., Torben-Nielsen, B., & Gewaltig, M.-O. (2015). A sparse reformulation of the Green's function formalism allows efficient simulations of morphological neuron models. *Neural Comput 27*(12), 2587–2622. (pp 123 and 173).

Wybo, W. A. M., Stiefel, K. M., & Torben-Nielsen, B. (2013). The Green's function formalism as a bridge between single-and multi-compartmental modeling. *Biol Cybern 107*(6), 685–694. (pp 123 and 173).

Yamauchi, S., Kim, H., & Shinomoto, S. (2011). Elemental spiking neuron model for reproducing diverse firing patterns and predicting precise firing times. *Front Comput Neurosci 5*(42), 1–15. (p 172).

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. (p 102).

Zeng, H.-L., Aurell, E., Alava, M., & Mahmoudi, H. (2011). Network inference using asynchronously updated kinetic ising model. *Phys. Rev. E 83*, 041135. (pp 25, 45, and 162).

Zhao, L., Beverlin II, B., Netoff, T., & Nykamp, D. Q. (2011). Synchronization from second order network connectivity statistics. *Front Comput Neurosci 5*(28), 1–16. (p 125).

Zinn-Justin, J. (1996). *Quantum field theory and critical phenomena*. Clarendon Press, Oxford. (pp 51, 161, and 164).

Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature 370*, 140–143. (p 24).