Transitions in the dynamics of recurrent neural networks

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

von
Sven Ole Goedeke
aus
Göttingen

Bonn, 2025

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

Gutachter/Betreuer:	Prof. Dr. Raoul-Martin Memmesheimer
Gutachter:	Prof. Dr. Moritz Helias
Tag der Promotion:	08.09.2025
Erscheinungsjahr:	2025

Abstract

The brain processes information and performs computations through the collective activity of large populations of interconnected neurons. The dynamics of these biological neural networks can exhibit qualitatively distinct behaviors or states. Transitions between these behaviors can occur at specific points when parameters or external conditions change. Moreover, transitions between coexisting states in a system can occur due to perturbations or noise. This thesis investigates how such transitions give rise to chaos, criticality, and drifting memory representations in recurrent neural networks. In four studies, we use methods from dynamical systems theory and statistical physics to analyze biologically relevant network models.

First, we study how time-varying inputs shape the transition to chaos in large rate-based neural networks with random connectivity. Treating the input as a stochastic drive, we analyze the resulting nonautonomous dynamics using dynamic mean-field theory. Our analysis yields an exact condition for the transition and the phase diagram. The input dynamically suppresses chaos: the transition shifts to significantly higher coupling strengths than predicted by local stability analysis. This leads to a distinctive regime of locally expansive yet stable, nonchaotic dynamics with optimal memory for past inputs.

Second, turning to spiking neural networks, we study how the intrinsic dynamics of individual neurons influence the collective irregular spiking dynamics in networks where recurrent inhibition balances constant excitation. We combine mean-field approaches and stability analysis to characterize both the activity statistics and the detailed dynamical stability of mixed networks with two types of inhibitory neurons. Remarkably, even a single neuron whose intrinsic dynamics accelerates towards spike generation induces a transition from stable to chaotic dynamics.

Third, we investigate how critical dynamics, neuronal avalanches, can emerge during development. With a simple model of activity-dependent network formation, we show that networks of stochastically spiking neurons can robustly self-organize into a critical state near the onset of unstable recurrent excitation. The resulting activity generates spike avalanches with power-law statistics. We map the activity dynamics to a self-exciting stochastic point process and analytically derive the distributions of avalanche sizes and durations, which exhibit power-law exponents consistent with experimental observations.

Fourth, we examine how assemblies of strongly interconnected neurons can drift due to synaptic plasticity or turnover while still stably representing a memory. We demonstrate such drifting assemblies in several spiking network models, where the same plasticity mechanisms, driven by ongoing activity in the absence of structured inputs, lead to both drift and self-organized compensation. We describe the gradual exchange of neurons in a drifting assembly as noise-activated transitions between metastable states and construct reduced models that explain a transition from drifting to static assemblies.

Contents

CHAPTER 1

Introduction

A central goal of theoretical neuroscience is to uncover principles that shape the emergent dynamics of neural circuits in the brain. The behavior of recurrent neural networks is governed by the interplay between neuronal dynamics, network connectivity, external inputs, and plasticity mechanisms. Recurrent networks can exhibit a remarkable diversity of collective dynamics, including stable fixed points, oscillations, chaos, and criticality. Such dynamical behaviors have direct implications for neural network function: chaos can support computational flexibility but must be controlled to ensure reliability, criticality has been linked to optimal information processing, and stable states enable memory storage and representation. Indeed, experimental studies suggest that neural circuits operate near the edge of chaos, show signatures of criticality, and exhibit representational drift over time—phenomena that are only partially understood from a theoretical perspective.

This thesis investigates qualitative changes or transitions in the dynamical behavior of recurrent neural networks. We study how external inputs dynamically suppress chaos, how intrinsic neuronal dynamics induce chaos in spiking neural networks, how criticality emerges through self-organization, and how memory representations can drift while maintaining stable function. The research lies at the interface of neuroscience and physics, where biological constraints meet the theoretical tools needed to analyze complex collective dynamics. Using tools from dynamical systems theory, stochastic processes, and statistical physics, we analyze biologically relevant models of both rate-based and spiking neural networks. Our results reveal how different aspects of neural circuits, from time-varying inputs to single-neuron dynamics to synaptic plasticity, can fundamentally reshape network dynamics and function.

To provide a solid foundation for the investigations presented in this thesis, this chapter introduces models of biological neural networks (Section 1.1) and relevant mathematical concepts (Section 1.2). The final Section 1.3 gives an overview of the structure of the thesis.

1.1 Neural network models

1.1.1 Biological neural networks

The behavior of humans and other animals arises from the activity in their nervous systems, especially in their brains. The brain contains an enormous number of nerve cells, called neurons, that are interconnected to form neural networks [1]. Typical neurons have a cell body, the soma, and two types of cell extensions: dendrites, which receive signals, and an axon, which sends signals. Most neurons communicate with

each other through short electrical pulses called action potentials or spikes. At rest, neurons maintain a negative electrical membrane potential (voltage) with respect to the extracellular medium. The membrane potential changes in response to incoming spikes from other neurons. Such input can make the membrane potential either less negative (depolarization) or more negative (hyperpolarization), and the input is called excitatory or inhibitory, respectively. When the depolarization exceeds a certain threshold, voltage-gated sodium ion channels open, initiating the generation of an action potential: a short, stereotypically shaped voltage excursion that propagates along the axon to other neurons. During this excursion, voltage-gated potassium ion channels open, restoring the membrane potential to a value near or below its resting value. Action potentials are often triggered in the initial segment of the axon, close to the soma. The generation of an action potential is commonly referred to as spiking or firing. Because of their stereotypical shape, only the number and timing of the spikes of a neuron are thought to convey information.

Neurons form networks through synaptic connections. Chemical synapses are directed contacts between the axon of a presynaptic neuron and, most often, a dendrite or the soma of a postsynaptic neuron. At chemical synapses, the arrival of an action potential generated by the presynaptic neuron is signaled to the postsynaptic neuron by the release of neurotransmitter molecules. Binding of these molecules to receptors induces the opening of ion channels in the postsynaptic cell membrane, allowing electrical charge to flow across the membrane. The membrane's reversal potential of a specific ion species, due to its concentration difference, determines whether the resulting postsynaptic current is positive (excitatory) or negative (inhibitory). This input current, which for ligand-gated ion channels rises rapidly and then decays on a characteristic time scale, drives changes in the postsynaptic membrane potential. At excitatory synapses, the postsynaptic potential is positive, whereas at inhibitory synapses, it is negative. The postsynaptic potentials resulting from presynaptic spiking activity at the numerous synapses of a neuron are spatiotemporally integrated along the dendrites and at the soma of the cell. Multiple excitatory synaptic inputs, occurring more or less simultaneously, are usually required to elicit an output spike. Inhibitory synaptic inputs can prevent output spikes.

The efficacy or strength of a synapse can be measured by the size of the corresponding postsynaptic potential. Synaptic strengths effectively serve as "weights" that determine how patterns of presynaptic spiking activity are combined in the postsynaptic neuron. Synapses in biological neural networks can change over time: they are plastic. Long-term changes in synaptic strength, as well as the creation or deletion of synapses, are thought to underlie learning and memory formation; see also Section 1.1.6 below. The remarkable flexibility (and molecular complexity) of chemical synapses is believed to be crucial for the adaptive function of neural networks.

Neural circuits in the brain consist of networks of interacting neurons that can be divided into populations of different neuron types. The most fundamental classification is the distinction between excitatory and inhibitory neurons. According to Dale's principle [2, 3], neurons release the same neurotransmitters at all their outgoing synapses, making them either exclusively excitatory or inhibitory. There are many more excitatory than inhibitory neurons in the brain. The mammalian cortex, and more generally the vertebrate forebrain, is important for higher brain functions, and neural circuits in the cortex exhibit characteristic connectivity properties. Cortical neurons form a large number of synaptic connections, with each neuron receiving input from up to thousands of other neurons and projecting to a similar number. However, local cortical networks are surprisingly sparse: the connection probability between nearby neurons typically ranges from 5% to 20%, meaning that a cell is connected to only a fraction of the other cells in its vicinity. Like many other biological neural networks, cortical networks are recurrently connected and contain loops through which neuronal activity can influence itself; see Section 1.1.5. The large number of relatively weak synaptic connections suggests that information is

represented and processed by the coordinated activity of neuronal populations rather than by individual neurons operating independently.

1.1.2 Integrate-and-fire neuron models

As basic mathematical models for the electrical activity dynamics of single neurons, we consider integrate-and-fire neuron models [4, 5]. These widely employed models idealize the generation of action potentials, i.e. spikes, by a threshold-crossing mechanism, acting on top of the neuron's membrane potential dynamics. The resulting spiking activity thus consists of discrete events in time. This level of abstraction captures essential aspects of how neurons interact with each other in neural circuits and yields phenomenological yet mechanistic spiking neuron models with only small numbers of biophysically interpretable variables and parameters. Here, we introduce such integrate-and-fire models together with a simple model for the synaptic input current driving the neuron.

The state of the neuron is described by its membrane potential or voltage at the soma, assuming that the voltage there determines spike initiation. The standard leaky integrate-and-fire (LIF) neuron model takes the passive leak current of the membrane into account. The neuron's voltage $V(t)$ integrates input currents in a leaky manner and obeys the differential equation

$$\tau_m \frac{dV}{dt} = -V + V_{\text{rest}} + RI(t)\,, \tag{1.1}$$

where $I(t)$ is the (total) input current at the soma. The membrane time constant $\tau_m = RC$ is an intensive[1] combination of the input resistance R and capacitance C of the membrane and specifies the time scale of the voltage dynamics. In particular, without input, the membrane potential $V(t)$ decays exponentially with a time constant τ_m to its resting value V_{rest}. In the following, we will always measure the input current in terms of its effect on the somatic membrane potential and, thus, we set $R = 1$. Inputs drive changes of $V(t)$ according to Eq. (1.1) as long as $V(t)$ is below the spike threshold V_{th}. When $V(t)$ reaches or crosses V_{th}, the neuron emits a spike, which is transmitted to other neurons. The membrane potential is then reset to the value $V_{\text{re}} < V_{\text{th}}$, where it is kept fixed for an absolute refractory period $\tau_{\text{ref}} \geq 0$. Thereafter, $V(t)$ evolves again according to Eq. (1.1) with initial value V_{re}. The temporal sequence of threshold crossings of this integrate-and-fire mechanism determines the neuron's output spiking activity or spike train. Mathematically, we represent a spike train as a sum of Dirac delta functions

$$S(t) = \sum_k \delta(t - t_k)\,, \tag{1.2}$$

where $t_k\,,\ k = 1, 2, \ldots$, are the firing times.

As a simple example, we consider a constant input current $I(t) = I$. Neglecting the spike threshold, the voltage $V(t)$ starting from $V(0) = V_{\text{re}}$ is then given by

$$V(t) = \mathrm{e}^{-t/\tau_m}\,(V_{\text{re}} - V_\infty) + V_\infty$$

with $V_\infty = V_{\text{rest}} + I$. If the asymptotic value V_∞ is above the threshold V_{th}, the neuron spikes at time t_1 determined by $V(t_1) = V_{\text{th}}$. After the refractory period, $V(t)$ starts again at the reset value V_{re}. Hence,

[1] Both the resistance R and the capacitance C depend on the area of the considered membrane, whereas their product $\tau_m = RC$ does not.

for $V_\infty > V_{\text{th}}$, the voltage dynamics is periodic with interspike interval (ISI)

$$t_{\text{ISI}} = \tau_{\text{ref}} + t_1 = \tau_{\text{ref}} + \tau_m \ln \frac{V_\infty - V_{\text{re}}}{V_\infty - V_{\text{th}}} \,. \tag{1.3}$$

The inverse value of the ISI gives the spike frequency or firing rate defined as the number of spikes in a long time interval divided by its length. Thus, $\rho_{\text{LIF}} = 1/t_{\text{ISI}}$ with t_{ISI} given by Eq. (1.3) yields the firing rate of the standard (deterministic) LIF neuron dynamics (1.1) with a constant input current I; the rate vanishes for $V_\infty = V_{\text{rest}} + I \leq V_{\text{th}}$.

It is sometimes useful to add a stochastic noise term to the neuron dynamics, for example, to include inevitable noise in biological cells or to mimic fluctuating input. For this, we add $\sqrt{2\tau_m}\sigma_V\,\xi(t)$ to the right-hand side of Eq. (1.1), where $\xi(t)$ is a standard Gaussian white noise [6] and the parameter $\sigma_V > 0$ (with units of voltage) determines the noise strength. The resulting fluctuations of the membrane potential can then lead to threshold crossings even for input currents below the threshold, i.e. for $V_{\text{rest}} + I < V_{\text{th}}$. In this case, the neuron spikes at irregular (random) points in time. In Section 1.2.3 below, we derive the firing rate of the stochastic LIF neuron model.

A possible way to generalize the LIF neuron model is to write Eq. (1.1) as

$$\begin{aligned} \dot{V} = \frac{dV}{dt} &= -\frac{1}{\tau_m}(V - V_{\text{rest}}) + \frac{1}{\tau_m} I(t) \\ &= f(V) + \tilde{I}(t) \end{aligned} \tag{1.4}$$

and use a more general function $f(V)$ that describes the membrane potential's intrinsic rate of change. The intrinsic dynamics in Eq. (1.4) is driven by the rescaled input current $\tilde{I}(t)$ with units of voltage over time. The exponential integrate-and-fire neuron model with $\tau_m f(V) = -V + V_{\text{rest}} + \Delta \exp\left((V - V_{\text{rheo}})/\Delta\right)$, $\Delta > 0$, is an important example [5]; it implements the dynamic spike generation mechanism by introducing a positive feedback. For a constant input current $\tilde{I}$ with $V_{\text{rest}} + \tau_m \tilde{I} > V_{\text{rheo}} - \Delta$, the neuron spikes periodically, and the membrane potential accelerates towards spiking when $V(t) > V_{\text{rheo}}$.

In neural networks, the input current $I(t)$ driving changes in the neuron's membrane potential $V(t)$ originates from the integration of postsynaptic currents at the soma. We consider only a simple linear model of synaptic transmission, also known as a current-based model. In this model, a spike of a presynaptic neuron at time t_0 generates the effective postsynaptic somatic current $w\, g(t - t_0)$, whose strength and shape are given by the weight w and the function $g(t)$, respectively, and postsynaptic currents evoked by different presynaptic spikes superpose linearly. A jump-like increase followed by an exponential decay, $g(t) = \exp(-t/\tau_{\text{syn}})\Theta(t)$ with the Heaviside step function $\Theta(t) = \int_{-1}^{t} \delta(s)\, ds$, captures in a basic manner the time scale of synaptic currents via the synaptic time constant $\tau_{\text{syn}} > 0$. Integrating Eq. (1.1) for the synaptic input current $I(t) = w\, g(t)$ in a resting neuron, i.e. with $V(0) = V_{\text{rest}}$, yields the resulting postsynaptic potential (PSP)

$$V_{\text{PSP}}(t) = w \frac{\tau_{\text{syn}}}{\tau_m - \tau_{\text{syn}}} \left(\mathrm{e}^{-t/\tau_m} - \mathrm{e}^{-t/\tau_{\text{syn}}} \right) \Theta(t) \tag{1.5}$$

as a transient excursion from the resting potential V_{rest}. Importantly, the PSP (1.5) has a nonzero rise time and resembles typical PSPs observed in biological neurons. A positive synaptic weight w gives rise to an excitatory postsynaptic potential (EPSP) driving the voltage towards the threshold while a negative one leads to an inhibitory postsynaptic potential (IPSP).

The weight w in Eq. (1.5) has by construction units of voltage. Specifically, $w\,\tau_{\mathrm{syn}} = \int_0^\infty V_{\mathrm{PSP}}(t)\,dt$. Nevertheless, it may be preferable to express synaptic weights in terms of the peak amplitude $\hat{w}$ of the PSP:

$$\hat{w} = \mathrm{sgn}(w) \max_t |V_{\mathrm{PSP}}(t)| = w \exp\left(\frac{\log(\tau_{\mathrm{syn}}/\tau_m)}{1-\tau_{\mathrm{syn}}/\tau_m}\right), \tag{1.6}$$

where the translation factor depends only on the ratio $\tau_{\mathrm{syn}}/\tau_m$. Hence, to obtain a given peak PSP $\hat{w}$, one has to choose w according to Eq. (1.6). For $\tau_{\mathrm{syn}} \ll \tau_m$, we have $\hat{w} \approx w\,\tau_{\mathrm{syn}}/\tau_m$ asymptotically. If we therefore write the postsynaptic current as $\hat{w}\tau_m\, g(t)/\tau_{\mathrm{syn}}$ and take the limit $\tau_{\mathrm{syn}} \to 0$ (for fixed τ_m), the current can be approximated by a weighted delta function, $I(t) = \hat{w}\tau_m\delta(t)$, because $g(t)/\tau_{\mathrm{syn}} \to \delta(t)$ in this limit. This delta-function shaped postsynaptic current results in a jump-like change of the membrane potential $V(t)$ given by the weight $\hat{w}$ both in the standard LIF model (1.1) as well as in its generalization (1.4). For the former, the resulting PSP reads

$$V_{\mathrm{PSP}}(t) = \hat{w}\,\mathrm{e}^{-t/\tau_m}\Theta(t)\,, \tag{1.7}$$

consistent with the definition of $\hat{w}$ as the PSP peak amplitude. We briefly note that $x(t) = \exp(-t/\tau)\Theta(t)$ is the causal impulse response or Green's function solving the inhomogeneous linear differential equation

$$\tau\frac{d}{dt}x(t) = -x(t) + \tau\delta(t) \tag{1.8}$$

with time constant $\tau > 0$.

In our linear model of synaptic integration, the total input current at the soma equals the sum of postsynaptic currents generated by the presynaptic spiking activity at different synapses. Thus, the input current $I_i(t)$ for a neuron labeled with index i is given by

$$I_i(t) = \sum_j \sum_l w_{ij} g(t - t_{j,l})\,, \tag{1.9}$$

where the sum over the index j runs over the neurons that are presynaptic to neuron i, and for each presynaptic neuron j, the sum over $l = 1, 2, \ldots$ runs over this neuron's spike times $t_{j,l}$. The synaptic weight w_{ij} determines the strength of the postsynaptic current in neuron i generated by a spike of presynaptic neuron j as described above; this convention will be used throughout the thesis. In Eq. (1.9) we assume that all synaptic connections give rise to the same postsynaptic current shape $g(t) = \exp(-t/\tau_{\mathrm{syn}})\Theta(t)$; in particular, they have the same synaptic time constant τ_{syn}. Typically, the time scale depends on the type of synaptic connection [5]. For example, excitatory and inhibitory postsynaptic currents can have different synaptic time constants. In the simple case of Eq. (1.9), the total input current $I_i(t)$ obeys the differential equation

$$\tau_{\mathrm{syn}}\frac{dI_i}{dt} = -I_i + \tau_{\mathrm{syn}} \sum_j w_{ij} S_j(t) \tag{1.10}$$

driven by a weighted sum of the presynaptic spike trains $S_j(t) = \sum_l \delta(t - t_{j,l})$; this results from Eq. (1.8) and linearity. In the LIF model, the subthreshold evolution of the membrane potential $V_i(t)$ of neuron i then follows Eq. (1.1) with input current $I(t) = I_i(t)$, which itself evolves according to Eq. (1.10).

Finally, we state the dynamics of $V_i(t)$ in the generalized model (1.4) for delta-function shaped postsynaptic currents:

$$\dot{V}_i = f(V_i) + \sum_j \hat{w}_{ij} S_j(t)\,, \tag{1.11}$$

where a spike of presynaptic neuron j leads to a jump-like change of size $\hat{w}_{ij}$ in neuron i's membrane potential.

1.1.3 Stochastically spiking neuron models

The fundamental stochastic model for the irregular spiking activity of cortical and other neurons is the Poisson point process [4, 7, 8]. In general, a (temporal) stochastic point process models a sequence of discrete events, i.e. points, that occur randomly in time [9]. Such a point process is specified by the probabilities of sequences of event times. In the case of a neuronal spike train $S(t) = \sum_k \delta(t - t_k)$, the events are the spikes at times t_k. In the Poisson process model of a spike train, the number of spikes in a finite time interval has a Poisson probability distribution. Furthermore, the number of spikes in disjoint time intervals are independent random variables. More intuitively, spikes occur independently of each other at an instantaneous rate $r(t) \geq 0$, where $r(t)\,dt$ gives the probability of a spike in a sufficiently short time interval of length dt. If the instantaneous rate is constant in time, $r(t) = r > 0$, the Poisson process is called homogeneous and the ISIs, i.e. the intervals between consecutive spikes, have an exponential probability density with mean $1/r$. The standard deviation of the exponential distribution equals the mean. Hence, the ISIs are quite variable with a coefficient of variation, defined as the standard deviation divided by the mean, equal to one. Since the ISI distribution of the activity of cortical neurons under stationary conditions often resembles an exponential distribution, the homogeneous Poisson process, which is fully specified by its rate, is used as a standard model for such irregular spiking activity.

When the threshold crossings of an integrate-and-fire neuron model are driven by fluctuations, as in the above-mentioned case of the LIF neuron with added stochastic noise, the neuron's output spike train appears rather irregular and we may approximate it to some degree by a Poisson point process. In the modeling and theory of spiking neural networks, irregular spiking activity is often replaced by Poisson processes. For instance, we do this to model the input spikes to a neuron, which in turn leads to fluctuations in the neuron dynamics.

The Poisson process model of a neuron's spike train can be regarded as a stochastically spiking neuron model with a given instantaneous, possibly time-dependent, rate. To include interactions with other neurons, we assume that this instantaneous rate is modulated by inputs to the modeled neuron. In the linear Poisson or Hawkes process neuron model [10–13], a spike of presynaptic neuron j arriving at time $\hat{t}_j$ changes the instantaneous firing rate $f_i(t)$ of the postsynaptic neuron i by adding the causal function $G_{ij}(t - \hat{t}_j)$ to it. This function describes the average response of a stochastically spiking neuron to a single input spike; in particular, $\int_0^\infty G_{ij}(t)\,dt$ equals the expected number of induced response spikes and thus yields the effective synaptic coupling strength in the model. The changes of $f_i(t)$ due to different presynaptic spikes superpose linearly. For simplicity, we use the same function shape for all synaptic connections, $G_{ij}(t) = w_{ij} g(t)$ with $g(t) = \exp(-t/\tau)\Theta(t)$, which models the response of the instantaneous rate by a jump-like increase followed by an exponential decay. The decay time constant τ is similar to the neuron's membrane time constant and determines the time scale of the dynamics. Taken

together, the instantaneous rate of neuron i driven by presynaptic spiking activity is given by

$$f_i(t) = f_0 + \sum_j \sum_{\hat{t}_j} w_{ij} g\left(t - \hat{t}_j\right) , \tag{1.12}$$

where $f_0 \geq 0$ is a constant baseline rate, the sum over j runs over the presynaptic neurons, whereas the sum over $\hat{t}_j$ runs over the spike times of neuron j, i.e. $\hat{t}_j = t_{j,l}$, $l = 1, 2, \ldots$, with the notation of Eq. (1.9). The nonnegative synaptic weights w_{ij} in this model have units of the firing rate; a single spike of the presynaptic neuron j induces on average τw_{ij} additional spikes in the postsynaptic neuron i. For given (deterministic) presynaptic spike times, the neuron model spikes stochastically as an inhomogeneous Poisson point process with the time-dependent rate (1.12). If we also model the input spiking activity as stochastic, the neuron's instantaneous rate $f_i(t)$ becomes itself a random function, and the output spike train forms an example of a doubly stochastic Poisson process [9].

Analogously to the total postsynaptic current in Eq. (1.10), we can rewrite Eq. (1.12) into a differential equation for the instantaneous rate,

$$\tau \frac{d}{dt} f_i(t) = -f_i(t) + f_0 + \tau \sum_j w_{ij} S_j(t) , \tag{1.13}$$

which is excited by the input spike trains $S_j(t) = \sum_{\hat{t}_j} \delta\left(t - \hat{t}_j\right)$. In the absence of input, $f_i(t)$ decays exponentially to f_0 with time constant τ.

We note that the LIF neuron with added stochastic noise can also be viewed as a stochastically spiking neuron model, since the threshold crossings occur at random points in time due to the noise. The corresponding instantaneous rate of firing, however, depends on the synaptic input currents and the neuron's last spike time in a rather complicated way [14]. In a fluctuation-driven regime, where the average input current is below threshold and the neuron spikes at a low rate, the response of the instantaneous rate to a single presynaptic spike with small synaptic weight is approximately linear and, hence, determined by a causal linear response function. This function is also rather complicated, even for the standard LIF neuron model [15–17]. Therefore, we opt for the more explicit and analytically tractable linear Poisson neuron with instantaneous rate following Eq. (1.13) to directly model the stochastic spiking dynamics in neural networks.

1.1.4 Firing-rate neuron models

In rate-based neural network models, we describe the activity of neurons or neuron-like units by an analog firing rate or rate-like variable. Generally, firing rates quantify the number of spikes per unit of time [4, 5]. We usually think of a possibly time-dependent firing rate as the density of spikes as they occur in time; its definition for neuronal spiking activity requires some form of average. The activity of large populations of spiking neurons is often successfully characterized by the population-averaged firing rate [5], ignoring the individual spike trains of the neurons. This more macroscopic perspective allows one to theoretically study the activity dynamics of neural networks in terms of firing rates. For instance, we can determine the self-consistent firing rate of a recurrently coupled population of spiking neurons in a mean-field approximation [5, 18]. Further reduced, low-dimensional rate-based models with dynamics similar to the one given below in Eq. (1.14) are commonly used in theoretical neuroscience to

generically describe interacting neuron populations in neural circuits, yielding both mechanistic and functional insights.

In more top-down approaches, networks of abstract neuron-like units that interact through analog variables are also successfully employed to understand brain function. These networks are similar to the artificial neural networks used in machine learning; however, one usually considers models operating in continuous time instead of discrete steps. They can serve as generic dynamical systems, which still have some commonalities with biological neural networks. Constructing or optimizing such rate-based networks to perform tasks or model experimentally observed activity patterns can uncover possible computational mechanisms in neural circuits [19–21]. Therefore, it is also important to theoretically analyze these more abstract models.

Instead of a spike train, a firing-rate neuron directly outputs its momentary activity in terms of an analog rate $r(t)$, which is nonnegative for a physical rate but can also be negative in more abstract models. In the latter case, it could represent the deviation from a baseline. The output rate $r_i(t) = \phi(x_i(t))$ of neuron i is obtained through a generally nonlinear transfer or activation function $\phi(x)$ from a local variable $x_i(t)$ that corresponds to the neuron's membrane potential or total input current [4, 22]. Specifically, we consider rate-based models where $x_i(t)$ defines the state of a neuron-like unit and this state evolves by integrating inputs similarly to the total postsynaptic current in Eq. (1.10):

$$\tau \frac{dx_i}{dt} = -x_i + \sum_j J_{ij} r_j(t)\,, \tag{1.14}$$

where the sum over j runs, as before, over the presynaptic input units whose activity is now given directly by their rates $r_j(t)$. The time constant τ sets the time scale of the leaky integration in Eq. (1.14). Here, the synaptic coupling weights J_{ij} specify how strongly and with which sign the rate of neuron j affects the state variable of neuron i. Formally, the dynamics (1.14) can be obtained from Eq. (1.10) or Eq. (1.13) by replacing the input spike trains $S_j(t)$ with firing rates $r_j(t)$, which may vary in time. The evolution of $x_i(t)$ determines by definition the neuron's time-dependent output rate $r_i(t) = \phi(x_i(t))$. It is common to assume that both $x_i(t)$ as well as $r_i(t)$ are dimensionless. In this case, typical examples for the dimensionless transfer function are the rectified-linear function $\phi(x) = \max(x, 0)$, the threshold power-law $\phi(x) = \max(x, x_{\text{th}})^\nu$ with threshold x_{th} and exponent $\nu > 0$ or the sigmoidal, more classical, hyperbolic tangent $\phi(x) = \tanh(x)$. These nonlinear, monotonically increasing functions capture different aspects of the neuronal input-output relation such as a threshold for output or saturation at strong inputs; they may lead to nonlinear dynamics in neural networks. We note that for a dimensionless right-hand side in Eq. (1.14), the units of the input rates $r_j(t)$ can be absorbed into the synaptic weights J_{ij}.

To derive a firing-rate model from a spiking model, some kind of averaging is required [4, 22, 23]. For instance, the (weighted) sum over the presynaptic spike trains in the total input current (1.10) provides such an average. If the number of inputs is large (as in the cortex) and the presynaptic spike trains are uncorrelated stochastic point processes[2], the sum is close to its expectation value with respect to the probabilities of the spike trains. As in the inhomogeneous Poisson process model, we define the instantaneous firing rate $r_j(t)$ of a spike train $S_j(t)$ via the probability $r_j(t)\,dt$ of a spike in a short time interval of length dt. The instantaneous firing rate also equals the expected density of spikes, $r_j(t) = \langle S_j(t) \rangle$, where the angular brackets denote the expectation value. We therefore approximate the

[2] Two point processes are uncorrelated if the probability of a point in one process does not depend on the occurrence of a point in the other. This may be a reasonable assumption when the presynaptic spiking activity is asynchronous.

total input current in its integral form (1.9) as

$$\begin{aligned} I_i(t) &= \sum_j w_{ij} \int g(t-t') S_j(t')\, dt' \\ &\approx \sum_j w_{ij} \int g(t-t') r_j(t')\, dt' \end{aligned} \tag{1.15}$$

with the postsynaptic current function $g(t) = \exp(-t/\tau_{\text{syn}})\Theta(t)$. A slow synaptic time scale, i.e. a large time constant τ_{syn}, improves the approximation due to temporal averaging [24]. To complete the rate-based description, we need to determine the time-dependent output rate $r_i(t)$ of a spiking neuron for a total input current $I_i(t)$ given by Eq. (1.15). If $I_i(t)$ varies only slowly in time, we can approximate $r_i(t)$ by the neuron's firing rate for a constant input current, also called f-I (frequency-current) curve, which is the above-mentioned neuronal input-output relation in this context. In the case of the LIF neuron model with membrane potential dynamics (1.1), this rate $\rho_{\text{LIF}}(I)$ for a constant input current I is given by the inverse value of Eq. (1.3) when $I > V_{\text{th}} - V_{\text{rest}}$ and is equal to zero otherwise. Hence, for slow postsynaptic currents, i.e. for $\tau_{\text{syn}} \gg \tau_m$, we can use $r_i(t) = \rho_{\text{LIF}}\left(I_i(t)\right)$ together with

$$\tau_{\text{syn}} \frac{dI_i}{dt} = -I_i + \tau_{\text{syn}} \sum_j w_{ij} r_j(t)$$

as a rate-based model analogous to Eq. (1.14) (with $J_{ij} = \tau_{\text{syn}} w_{ij}$) to describe slow firing-rate dynamics in networks of LIF spiking neurons [24]. Additional noise on the fast time scale of the membrane potential would only modify the LIF neuron's transfer function $\rho_{\text{LIF}}(I)$, see Section 1.2.3.

1.1.5 Recurrent neural networks

Neural networks in the brain typically exhibit recurrent connections, meaning that the activity of the neurons can affect itself through feedback loops in the network. The resulting self-referential dynamics can give rise to intricate activity patterns in local neural circuits. Recurrent neural network models often consist of one or more populations of neurons with similar properties. For simplicity, we first focus on a single population of $N \in \mathbb{N}$ neurons with identical intrinsic parameters. We label the individual neurons in the recurrent network by an index, usually denoted by i or j, running from 1 to N. So far, the origin of the input driving the spiking and firing-rate neuron models introduced above has been unspecified. Here, the input originates from the other neurons in the network. Further input may originate from external neurons whose activity is not influenced by the activity in the recurrent network. The external neurons are sometimes not explicitly modeled.

Concretely, for a recurrent network of N LIF model neurons with postsynaptic currents following Eq. (1.9), the voltage of neuron $i = 1, \ldots, N$ obeys

$$\tau_m \frac{dV_i}{dt} = -V_i + V_{\text{rest}} + I_i + I_i^{\text{ext}}(t)\,, \tag{1.16}$$

together with its total recurrent postsynaptic current

$$\tau_{\text{syn}} \frac{dI_i}{dt} = -I_i + \tau_{\text{syn}} \sum_{j=1}^{N} w_{ij} S_j(t) \,, \tag{1.17}$$

where $S_j(t) = \sum_{\hat{t}_j} \delta\left(t - \hat{t}_j\right)$ denotes the spike train of neuron j while $I_i^{\text{ext}}(t)$ represents an additional input current originating from external neurons or other sources. As explained after Eq. (1.1), the voltage dynamics (1.16) is augmented by a threshold-crossing and reset mechanism: Neuron i emits a spike when $V_i(t)$ reaches the threshold V_{th}; subsequently, the voltage is reset to $V_{\text{re}} < V_{\text{th}}$ and possibly kept fixed there for a refractory period τ_{ref}. Taken together, the integrate-and-fire dynamics of the neurons in the recurrent network determine their output spike trains, which in turn drive the neurons' input currents in Eq. (1.17). The synaptic weights w_{ij} , $i, j = 1, \ldots, N$, describe the strength of the postsynaptic current in neuron i evoked by a spike of neuron j [see also Eq. (1.9)] and are often collected into a matrix of (directional) coupling weights. In the absence of a connection from neuron j to neuron i, we set $w_{ij} = 0$, and usually there are no self-couplings, $w_{ii} = 0$. For given external inputs $I_i^{\text{ext}}(t)$ in Eq. (1.16), the synaptic weights determine how the activity in the recurrent spiking neural network evolves.

We note briefly that we have neglected synaptic transmission delays, the time it takes for a spike generated at the soma of the presynaptic cell to cause a noticeable input at the soma of the postsynaptic cell. In local circuits, axonal and dendritic conduction delays are typically short and can be partially accounted for by the shape of the effective postsynaptic current [25].

Most neurons in the brain have either an excitatory or an inhibitory impact on all the postsynaptic neurons to which they project (Dale's principle).[3] In our models, this implies that the nonzero synaptic output weights w_{ij} of a neuron j have the same sign, i.e. $w_{ij} \geq 0$ for all i or $w_{ij} \leq 0$ for all i. Accordingly, we distinguish between excitatory and inhibitory neurons and often consider them as separate populations in neural networks. A standard recurrent neural network model of activity dynamics in local cortical circuits consists of an excitatory and an inhibitory neuron population interacting with each other; they further receive feed-forward input from an external excitatory population. In this model, total excitatory and inhibitory postsynaptic currents balance each other on average, giving rise to a fluctuation-driven regime with asynchronous irregular spiking activity [26, 27].

How do we choose the structure and the weights of the synaptic connections in our models? The detailed connectivities of biological neural circuits are mostly unknown. Furthermore, we assume that the relevant dynamics and function of larger networks are at least partly independent of the precise instance of their synaptic connections. Put differently, we are typically interested in such robust properties of neural networks. A common approach in theoretical neuroscience is to study network models with randomly generated connectivity using statistical rules, which can incorporate existing experimental knowledge. For example, one assumes that a synaptic connection from one neuron to another is present with a certain probability that may depend on the pre- and postsynaptic neuron population. Distributions of synaptic weights in neural circuits have also been characterized experimentally [28]. We can include this heterogeneity in random neural network models by selecting the coupling weights of synaptic connections from a given probability distribution. In general, the random neural network approach allows us to theoretically study effects of the seemingly disordered and complex connectivity observed in the brain, independently of the precise instance. Additionally, random connectivity often forms the starting point for plastic network models, in which synapses change, e.g., due to learning.

[3] Dale's principle states that a neuron releases the same neurotransmitters at all of its outgoing synapses [2, 3].

In a rate-based recurrent neural network, the states $x_i(t)$, $i = 1, \ldots, N$, of the neuron-like units evolve according to the system of coupled differential equations

$$\tau \frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} J_{ij} \phi\left(x_j\right) + \xi_i(t) \,, \tag{1.18}$$

see Eq. (1.14). Here, $\xi_i(t)$ represents a possible external input, which may vary between neurons and in time; our notation anticipates the modeling of this input as a stochastic drive. Such rate-based networks are useful for studying the dynamics of firing rates in neural circuits, see Section 1.1.4. In the influential random neural network model introduced by Sompolinsky, Crisanti, and Sommers [29], the individual coupling weights J_{ij} in Eq. (1.18) are Gaussian-distributed with mean zero and variance g^2/N. The authors study the resulting dynamics (1.18) without input for the sigmoidal transfer function $\phi(x) = \tanh(x)$ in the limit of large network size N. This model exhibits an important transition from a stable fixed-point state to chaotic dynamics when the coupling strength parameter g is increased above a critical value, which will be discussed in Section 1.2.2. As mentioned above, rate-based recurrent neural networks are employed as generic dynamical systems in more top-down approaches. There, the dynamics are adapted by changing the weights, e.g., to optimize task performance. Some approaches also harness the potentially rich dynamics of large random neural networks to construct functional models [30].

1.1.6 Plasticity of synaptic connections

In biological neural networks, many forms of plasticity can lead to changes in the synaptic connections between the neurons. We focus on forms of long-term and structural plasticity; they induce changes that in principle persist for tens of minutes or longer. Such plasticity mechanisms are assumed to underlie adaptations in neural circuits during development, learning, and memory formation [31, 32]. Structural plasticity refers to changes in the connectivity structure of the network, i.e. the presence or absence of different synaptic connections. During neural network development, for example, neurons extend their processes and create new synapses. Long-term synaptic plasticity, on the other hand, often refers to modifications in the strength of existing synapses, typically measured by the amplitude of the postsynaptic response. Important forms of both long-term and structural plasticity depend on or are regulated by neuronal activity, which in turn is affected by changes in the synaptic connections. There are, however, also synaptic changes that occur spontaneously, even in the absence of activity [33–35]. In spike-timing-dependent plasticity (STDP), modifications in synaptic strength depend on the relative timing of the pre- and postsynaptic spiking activity [36, 37]. Below, we describe a simple phenomenological model of STDP. Thereafter, homeostatic synaptic plasticity mechanisms will be briefly addressed.

To model plasticity in neural networks, we need to specify how the synaptic connections change over time. For this, one often abstracts a mathematical model of synaptic changes from experimental findings [4, 5]; such a model is also called plasticity or learning rule, especially if the changes depend on neuronal activity. In more top-down approaches, plasticity rules are postulated or derived from certain objectives. Activity-dependent plasticity rules give rise to a dynamic interplay between a network's collective activity and its synaptic connectivity, which can be studied in plastic neural network models.

In a simple phenomenological STDP model [5, 36], changes of synaptic weights are driven by pairs of pre- and postsynaptic spikes that occur in relatively close temporal proximity, typically on a time scale of at most tens of ms. Specifically, the weight w_{ij} of an existing synaptic connection from neuron j to neuron i changes by an amount $\Delta w_{ij}\left(\hat{t}_i - \hat{t}_j\right) = \eta\, h\left(\hat{t}_i - \hat{t}_j\right)$ for each pair $\hat{t}_j$ and $\hat{t}_i$ of pre- and postsynaptic spike times, where contributions from different pairs are summed and by causality, the update is applied at the respective later time. The function $h(\Delta t)$ determines the dependence on the relative timing between the post- and the presynaptic spike, i.e. their time difference $\Delta t = \hat{t}_i - \hat{t}_j$, whereas the parameter $\eta > 0$ sets the overall amplitude of the weight changes. Together, $\eta\, h(\Delta t)$ is called STDP or learning window of the plasticity rule. We choose the STDP window shape $h(\Delta t)$ to be dimensionless; hence, the amplitude parameter η has units of the synaptic weights but nevertheless acts as a "learning rate". The parameter η is often assumed to be small giving rise to slow plasticity through accumulating rapid but small weight changes.

To express the described STDP model as a dynamic equation, we represent as before the spike trains of the presynaptic neuron j and the postsynaptic neuron i by $S_j(t) = \sum_{\hat{t}_j} \delta\left(t - \hat{t}_j\right)$ and $S_i(t) = \sum_{\hat{t}_i} \delta\left(t - \hat{t}_i\right)$, respectively. The changes of the synaptic weight w_{ij} generated by the pre- and postsynaptic spiking activity then follow

$$\frac{d}{dt} w_{ij}(t) = S_i(t) \int_{-\infty}^{t} \eta\, h(t-s) S_j(s)\, ds + S_j(t) \int_{-\infty}^{t} \eta\, h(s-t) S_i(s)\, ds\,; \tag{1.19}$$

the delta functions in front of the integrals lead to stepwise changes at the spike times while the integrals themselves evaluate to sums over the corresponding previous spike times, e.g.,

$$\int_{-\infty}^{\hat{t}_i} h\left(\hat{t}_i - s\right) S_j(s)\, ds = \sum_{\hat{t}_j < \hat{t}_i} h\left(\hat{t}_i - \hat{t}_j\right)\,.$$

The STDP window function $h(\Delta t)$ usually vanishes for time differences Δt with large magnitude. Thus, the integrals in Eq. (1.19) effectively extend only over the recent past of the current spike time. Such an STDP model can be introduced for synaptic weights in networks of spiking neurons, independent of the specific neuron model.

We consider STDP at excitatory synaptic connections. In this case, a positive value of $h(\Delta t)$ implies that the excitatory synaptic weight $w_{ij} \geq 0$ is strengthened leading to long-term potentiation (LTP), whereas a negative value implies that it is weakened leading to long-term depression (LTD). The first experiments on STDP induced by temporally close spike pairs found asymmetric window functions with LTP for time differences $\Delta t = \hat{t}_i - \hat{t}_j > 0$ and LTD for $\Delta t < 0$ [36]. Accordingly, the excitatory synaptic weight is increased when a presynaptic spike precedes a postsynaptic one, i.e. when $\hat{t}_i > \hat{t}_j$, while it is decreased when the timing is reversed. A simple STDP window implementing this asymmetric dependence on the relative timing is given by

$$\eta\, h(\Delta t) = \begin{cases} \eta_{\mathrm{LTP}}\, \mathrm{e}^{-\Delta t/\tau_{\mathrm{LTP}}} & \text{for } \Delta t > 0 \\ -\eta_{\mathrm{LTD}}\, \mathrm{e}^{\Delta t/\tau_{\mathrm{LTD}}} & \text{for } \Delta t \leq 0 \end{cases}$$

with possibly different amplitudes η_{LTP} and η_{LTD} as well as decay time constants τ_{LTP} and τ_{LTD} for LTP and LTD. The time constants are typically on the order of 10 ms. More recently, STDP with a symmetric window exhibiting maximal LTP around $\Delta t = 0$ was found in the hippocampus region CA3 [38]. A

simple choice for such a symmetric STDP window is

$$\eta\, h(\Delta t) = \frac{\eta}{1-\beta}\left[\exp\left(-\left|\Delta t\right|/\tau_{\mathrm{LTP}}\right) - \beta\, \exp\left(-\left|\Delta t\right|/\tau_{\mathrm{LTD}}\right)\right] , \tag{1.20}$$

with $0 \leq \beta < 1$ and $\tau_{\mathrm{LTP}} < \tau_{\mathrm{LTD}}$; centrally peaked LTP is accompanied by LTD for larger $|\Delta t|$ (if $\beta > 0$). Hence, coactive pre- and postsynaptic spiking in close temporal proximity induces weight strengthening, irrespective of the temporal order.

Synaptic plasticity or, more generally, learning that is driven by correlations between pre- and postsynaptic neuronal activity is often called Hebbian, since Donald Hebb [39] postulated that neural circuits in the brain are modified in this way. Hebbian plasticity can organize and stabilize activity patterns during development and associative learning. We may regard STDP as a form of Hebbian plasticity driven by temporal correlations between pre- and postsynaptic spiking activity.

The phenomenological STDP model, Eq. (1.19), can result in an average weight change even for uncorrelated pre- and postsynaptic spike trains, as pairs of temporally close spikes also occur by chance. The resulting weight change depends on the pre- and postsynaptic firing rates. To illustrate this, we assume that the two spike trains follow independent Poisson processes with (instantaneous) rates $r_j(t) = \left\langle S_j(t)\right\rangle$ and $r_i(t) = \langle S_i(t)\rangle$; as before, the angular brackets denote the expectation value with respect to the stochastic spike trains. By independence, we have $\left\langle S_i(t)S_j(s)\right\rangle = \langle S_i(t)\rangle \left\langle S_j(s)\right\rangle = r_i(t)r_j(s)$. Taking the expectation value of Eq. (1.19) then yields the expected weight change per time,

$$\begin{aligned}
\frac{d}{dt}\left\langle w_{ij}(t)\right\rangle &= \int_{-\infty}^{t} \eta\, h(t-s)\left\langle S_i(t)\, S_j(s)\right\rangle \, ds + \int_{-\infty}^{t} \eta\, h(s-t)\left\langle S_j(t)\, S_i(s)\right\rangle \, ds \\
&= \int_{-\infty}^{t} \eta\, h(t-s)\, r_i(t)\, r_j(s)\, ds + \int_{-\infty}^{t} \eta\, h(s-t)\, r_j(t)\, r_i(s)\, ds \\
&\approx \eta \left(\int_{-\infty}^{t} h(t-s)\, ds + \int_{-\infty}^{t} h(s-t)\, ds\right) r_i(t)\, r_j(t) \\
&= \eta \left(\int_{-\infty}^{\infty} h(u)\, du\right) r_i(t)\, r_j(t) ,
\end{aligned} \tag{1.21}$$

where we assume in the third line that the rates $r_j(t)$ and $r_i(t)$ vary slowly on the time scale during which $h(\Delta t)$ is effectively nonzero. Therefore, the integral of the STDP window controls the resulting average weight change in Eq. (1.21). Importantly, this rate-based contribution is also present for irregular but correlated spiking activity [13] and it may dominate the weight changes accumulated over time. To study how the interplay of the temporal correlations with the shape of the STDP window drives weight changes, one often considers windows with a negative or vanishing integral [13, 40, 41].

The weight dynamics described by Eq. (1.19) is usually augmented by boundaries restricting the range of possible weights. A natural lower bound for excitatory synaptic weights is $w_{ij} \geq 0$, whereas an upper bound like $w_{ij} \leq w_{\max}$ prevents unlimited growth, which can occur in such STDP models, e.g., due to random diffusion or positive feedback.

Hebbian plasticity or learning tends to be unstable because the weight changes typically amplify the activity patterns that caused them, driving further similar changes [4]. This positive feedback loop can lead to instability and runaway activity in neural networks. In addition to bounds on the individual weights, plasticity models often include other, compensatory mechanisms [4, 42–44]. A simple example is the normalization of a neuron's synaptic weights; it acts as a heterosynaptic mechanism in which

changes at different synapses influence each other [45]. For instance, assuming that the sum of the (excitatory) input weights to neuron i is constrained to a given value, i.e. $\sum_j w_{ij} = w_{\text{sum}}$, we can normalize these weights as follows:

$$w_{ij} = w_{\text{sum}} \frac{w'_{ij}}{\sum_k w'_{ik}} ,$$

where w'_{ij} are the corresponding unnormalized weights. Such multiplicative normalization preserves the relative weight strengths and gives rise to competition between the synapses [46–48], since the increase of a single weight leads to the decrease of the other, unchanged ones. Synaptic competition is usually required for neurons to form and maintain selectivity in their response to different input patterns. We note that STDP alone can also induce competition [13, 49, 50].

Homeostatic synaptic plasticity refers to mechanisms that act to maintain activity at a certain level [51].[4] This is often necessary for neural circuits to function properly; for example, spiking neurons have a rather limited operating range in which they can respond well to variable inputs. Homeostatic plasticity may counteract the inherently unstable nature of Hebbian plasticity [43, 52]. A prominent form of homeostatic plasticity is synaptic scaling [53]: Neurons slowly increase or decrease the strength of their excitatory synaptic connections in response to deviations from a target firing rate level, giving rise to firing rate homeostasis through slow negative feedback. The scaling occurs multiplicatively, similar to the above-mentioned multiplicative normalization. Models of plasticity in neural networks found that synaptic scaling can function as an important compensatory mechanism [44, 54]. Although normalization of synaptic weights typically acts rapidly, it is often considered as a homeostatic plasticity mechanism.

Structural plasticity may also depend on neuronal activity in a homeostatic manner [55, 56]; it appears that neurons create or delete synaptic connections to attain a certain activity level. In particular, models of neural network formation during development are based on homeostatic structural plasticity [57, 58]. Such plastic models describe how homeostatic regulation can direct the growth of initially unconnected neurons into networks in which activity is maintained at a particular target level.

1.2 Deterministic and stochastic dynamics

1.2.1 Deterministic dynamical systems

The mathematical models of neuronal activity and synaptic plasticity that we presented above in Section 1.1 are formulated in terms of how variables change over time: the models are dynamical. The mathematics of dynamical systems provides a general framework to describe and analyze systems evolving in time, such as recurrent neural networks. We start here with deterministic dynamical systems and explain some basic concepts [59]. Furthermore, we introduce linear stability analysis and Lyapunov exponents, partially following [60].

A deterministic dynamical system consists of a space of possible system states, also called phase space, and a dynamical rule that uniquely determines the future states given the present state and thereby the time evolution of the system.[5] The evolution can occur in discrete or in continuous time. In the

[4] Homeostasis in biology refers to the active maintenance of stable physiological conditions, e.g., body temperature, and typically involves regulation via negative feedback.

[5] Modeling a real system as a deterministic dynamical system often requires the choice of appropriate state variables ("degrees of freedom"); they should determine their time evolution in a closed way that is sufficiently isolated from fluctuations and perturbations.

following, we consider continuous-time dynamics, which is more natural for modeling physical systems and biological neural networks. Therefore, we assume that the dynamics of the state $\mathbf{x}(t) \in \mathbb{R}^N$, with $N \in \mathbb{N}$, is generated by the (deterministic) ordinary differential equation

$$\frac{d}{dt}\mathbf{x}(t) = \mathbf{f}(\mathbf{x}(t); t)\,, \tag{1.22}$$

where the sufficiently regular function $\mathbf{f}(\mathbf{x}; t) \in \mathbb{R}^N$ uniquely determines the change, i.e. the velocity, of the state $\mathbf{x}$ at time t; this function is often called a vector field. Since the phase space has N dimensions, we say that the dynamical system (1.22) is N-dimensional. For example, the dynamics (1.18) of a rate-based recurrent neural network has the form of Eq. (1.22) and its dimension equals the number of neurons, which can be large. The solution $\mathbf{x}(t)$ of Eq. (1.22) for $t \geq t_0$ with initial condition $\mathbf{x}(t_0)$ at time t_0 yields the time evolution of this initial state as a curve or trajectory in the system's N-dimensional phase space. For nonlinear dynamics, closed-form analytical solutions are typically not available. Instead, we aim to understand the behavior of the dynamics by investigating qualitative and quantitative properties of trajectories without explicit formulas for them.

The dynamical system (1.22) is called autonomous if the vector field on the right-hand side, $\mathbf{f}(\mathbf{x}; t)$, does not explicitly depend on time and is thus expressed as $\mathbf{f}(\mathbf{x})$; otherwise, it is called nonautonomous. In the autonomous case, the dynamics can be regarded as intrinsic to the system because it generates itself in a time-invariant manner; the theory of deterministic dynamical systems mainly focusses on this situation. In particular, closed physical systems evolve in this way. Furthermore, systems with temporally constant external fields or inputs can be described by autonomous dynamics. In contrast, time-dependent external drive or input gives rise to nonautonomous dynamics.

The following types of trajectories are important because they govern typical long-term behaviors of autonomous dynamical systems and thereby may provide structure in the system's phase space: A fixed point is a solution that is constant in time, $\mathbf{x}(t) = \bar{\mathbf{x}}$; fixed points are given by states $\bar{\mathbf{x}}$ with $\mathbf{f}(\bar{\mathbf{x}}) = \mathbf{0}$. This fundamental behavior describes steady states in dynamical systems, which are often considered as a baseline from which more complicated dynamics could emerge. Periodic trajectories (or orbits) are time-dependent solutions that repeat themselves after some period in time and they can describe oscillations like that of a swinging pendulum. Moreover, there are quasiperiodic, never-repeating trajectories consisting of two or more periodic components with incommensurate frequencies. Finally, trajectories corresponding to chaotic dynamics are bounded aperiodic solutions that are statistically stationary and exhibit sensitive dependence on the initial condition [59, 61].

To understand the behavior of a dynamical system, we also need to know the stability properties of particular trajectories. For example, a perfectly inverted pendulum is at a fixed point state but an arbitrarily small perturbation of this state will lead to a different dynamical behavior. In general, stability characterizes how the dynamics changes in response to a small perturbation of the initial state and whether nearby trajectories will remain close during the time evolution of the system [62]. There are different notions of stability. We focus on the linear stability of trajectories, which yields insight into both the dynamics near a fixed point and the stability properties of time-dependent solutions [60].

To perform linear stability analysis, we linearize the dynamics (1.22) around a reference trajectory $\mathbf{x}(\mathbf{t})$ with initial condition $\mathbf{x}(t_0)$. For this, we consider a small perturbation of the initial condition given by $\mathbf{x}(t_0) + \epsilon\,\mathbf{y}(t_0)$, where $\mathbf{y}(t_0) \in \mathbb{R}^N$ determines the perturbation's phase space direction and $\epsilon > 0$ controls its size. For $\epsilon \to 0$, the trajectory with the perturbed initial condition can be expanded to linear order in ϵ as $\mathbf{x}(t) + \epsilon\,\mathbf{y}(t)$, again assuming sufficient regularity. The vector $\mathbf{y}(t) \in \mathbb{R}^N$ evolves from its

initial value $\mathbf{y}(t_0)$ at time t_0 according to the linearized dynamics

$$\frac{d}{dt}\mathbf{y}(t) = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}(\mathbf{x}(t); t)\, \mathbf{y}(t)\,, \tag{1.23}$$

with the Jacobian matrix $\partial \mathbf{f}/\partial \mathbf{x} = \left(\partial f_i/\partial x_j\right) \in \mathbb{R}^{N\times N}$ of the vector field $\mathbf{f}(\mathbf{x}; t)$ in Eq. (1.22), evaluated at the reference trajectory $\mathbf{x}(t)$ and time t. This linear differential equation describes the time evolution of infinitesimal perturbations of the initial condition; it is sometimes called variational equation. The asymptotic behavior of $\mathbf{y}(t)$ as $t \to \infty$, which depends on $\mathbf{y}(t_0)$, determines the linear stability properties of the reference trajectory.

In the important case of a fixed point $\mathbf{x}(t) = \bar{\mathbf{x}}$ of an autonomous dynamical system with vector field $\mathbf{f}(\mathbf{x})$, the Jacobian is a constant matrix $\mathbf{A} = (\partial \mathbf{f}/\partial \mathbf{x})(\bar{\mathbf{x}})$, and the linearized dynamics (1.23) takes the form

$$\frac{d}{dt}\mathbf{y} = \mathbf{A}\,\mathbf{y}\,, \tag{1.24}$$

a linear differential equation with constant coefficients. Its general solution can be written as $\mathbf{y}(t) = \exp(\mathbf{A}t)\,\mathbf{y}(0)$ using the matrix exponential function [63]. Here and in the following, we set $t_0 = 0$ for simplicity. The eigenvalue spectrum of the matrix $\mathbf{A}$ controls the asymptotic behavior of $\mathbf{y}(t)$ as $t \to \infty$. If all eigenvalues have negative real parts, every solution $\mathbf{y}(t)$ asymptotically decays to zero and we have linear asymptotic stability [63].[6] This implies that small perturbations of the fixed point $\bar{\mathbf{x}}$ decay; it is asymptotically stable and attracts nearby trajectories. On the other hand, if $\mathbf{A}$ has an eigenvalue with a positive real part, there exists a direction of initial values $\mathbf{y}(0)$ for which $\mathbf{y}(t)$ grows unboundedly as $t \to \infty$. In this situation, the fixed point $\bar{\mathbf{x}}$ is linearly unstable because generic arbitrarily small perturbations will grow until the dynamics becomes non-negligibly nonlinear. To summarize, the largest real part of the eigenvalues of the linearized dynamics around a fixed point determines its stability. Furthermore, the largest real part gives the asymptotic rate of exponential decay or growth of generic infinitesimal perturbations.

Characterizing linear stability properties in terms of asymptotic growth or shrinkage rates of infinitesimal perturbations can be extended to time-dependent trajectories with the help of Lyapunov exponents [60, 61]. In this case, the Jacobian matrix generating the linearized dynamics (1.23) usually varies in time, even for autonomous systems. The respective matrices at different time points often do not commute. Then, a simultaneous eigendecomposition is unavailable and in contrast to the case of a constant matrix, stability cannot be determined from the eigenvalue spectra of the (local) Jacobian matrix. Nevertheless, we can define the largest or maximum Lyapunov exponent as the asymptotic average rate of exponential growth or shrinkage of a generic infinitesimal perturbation:

$$\lambda_{\max} = \lim_{t\to\infty} \frac{1}{t} \log\left(\frac{||\mathbf{y}(t)||}{||\mathbf{y}(0)||}\right)\,, \tag{1.25}$$

where $\mathbf{y}(t)$ follows Eq. (1.23) with a generic initial perturbation vector $\mathbf{y}(0)$ for a given reference trajectory $\mathbf{x}(t)$ with initial condition $\mathbf{x}(0)$, i.e., we consider the perturbed initial condition $\mathbf{x}(0) + \epsilon\,\mathbf{y}(0)$ in the limit of small ϵ. Here, it is assumed that $\mathbf{y}(t)$ will align with the (time-dependent) fastest-growing perturbation direction; non-generic, specific $\mathbf{y}(0)$ may yield further, smaller Lyapunov exponents [60, 61]. Because

[6] The corresponding stability criterion for discrete-time dynamics is that the absolute values of all eigenvalues are smaller than 1.

the linearized dynamics (1.23) acts multiplicatively on $\mathbf{y}(t)$ and the logarithm transforms products into sums, the maximum Lyapunov exponent (1.25) effectively measures the time-averaged rate of the exponential divergence or convergence of nearby trajectories. Hence, it quantifies how sensitively the dynamics depends on the initial conditions. At a fixed point of an autonomous system, the maximum Lyapunov exponent equals the largest real part of the eigenvalues of the constant Jacobian matrix.

In the case of (quasi-)periodic or statistically stationary aperiodic trajectories, Lyapunov exponents, being time averages, characterize the stability properties of the trajectory as a whole. If the maximum nontrivial[7] Lyapunov exponent is negative, small perturbations decay and the trajectory is asymptotically stable. An important example is a stable limit cycle, a periodic trajectory that attracts neighboring trajectories. It describes stable oscillatory dynamics, which are self-sustained in autonomous systems. If the maximum Lyapunov exponent is positive, generic arbitrarily small perturbations of the trajectory will grow and the dynamics is linearly unstable.

A bounded aperiodic trajectory with a positive maximum Lyapunov exponent realizes chaotic dynamics if most trajectories with nearby initial conditions first separate exponentially from it but show the same statistical behavior in the long term. Assuming ergodicity, these trajectories are then asymptotically representative of the chaotic dynamics as well; in particular, they have the same Lyapunov exponents [60, 61]. The positivity of the maximum Lyapunov exponent indicates that the dynamics exhibits sensitive dependence on initial conditions, a defining property of deterministic chaos. It leads to long-term unpredictability and seemingly random behavior. Chaotic dynamics in dissipative systems, in which phase space volume contracts on average, takes place on a chaotic strange attractor, a lower-dimensional, complicated set in phase space. By attracting neighboring trajectories, the chaotic attractor governs the long-term behavior of the system, whereas the dynamics on it is unstable to arbitrarily small perturbations. Nonchaotic, stable dynamics, indicated by a non-positive maximum Lyapunov exponent, are sometimes called regular, especially when the corresponding trajectories depend on time.

Analogously to the eigenvalue spectrum of a constant matrix, one can introduce a full spectrum of Lyapunov exponents, which describe the asymptotic average growth or shrinkage rates of infinitesimal perturbations in different phase space directions. To define the Lyapunov exponents, we consider N linearly independent initial perturbation vectors $\mathbf{y}_k(0)$, $k = 1, \ldots, N$, and their simultaneous time evolution $\mathbf{y}_k(t)$ according to the linearized dynamics (1.23). By linearity, it suffices to use the canonical basis vectors $\mathbf{y}_k(0) = (\delta_{ik}) \in \mathbb{R}^N$, with the Kronecker delta δ_{ik}, as initial perturbation directions. Collecting the corresponding solutions into a matrix $\mathbf{Y}(t) = (\mathbf{y}_k(t)) \in \mathbb{R}^{N\times N}$ yields the fundamental matrix or propagator of Eq. (1.23); the solution with a general $\mathbf{y}(0)$ then reads $\mathbf{y}(t) = \mathbf{Y}(t)\,\mathbf{y}(0)$. Hence, it contains the complete information on the evolution of infinitesimal perturbations around a reference trajectory of the dynamical system (1.22). In this context, the matrix $\mathbf{Y}(t)$ is also known as the long-term Jacobian, since it equals the $N \times N$ Jacobian matrix of the reference trajectory $\mathbf{x}(t)$ considered as a (differentiable) function of the initial condition $\mathbf{x}(0)$, i.e.,

$$\mathbf{Y}(t) = \frac{\partial\,\mathbf{x}(t)}{\partial\,\mathbf{x}(0)} = \left(\frac{\partial x_i(t)}{\partial x_j(0)}\right) .$$

Generalizing the maximum Lyapunov exponent (1.25), the spectrum of Lyapunov exponents $\lambda_1 = \lambda_{\max} \geq \lambda_2 \geq \ldots \geq \lambda_N$ can be defined as the logarithms of the eigenvalues of the positive-semidefinite symmetric

[7] Time-dependent bounded trajectories of continuous-time autonomous systems always have a zero Lyapunov exponent due to the dynamics' time-translation invariance [60]; its perturbation direction corresponds to the direction of the vector field.

matrix

$$\mathbf{\Lambda} = \lim_{t\to\infty} \left(\mathbf{Y}^\top(t)\,\mathbf{Y}(t)\right)^{\frac{1}{2t}} , \tag{1.26}$$

provided that the limit exists. Oseledets' multiplicative ergodic theorem establishes the existence of the limit and the Lyapunov exponents for statistically stationary, ergodic trajectories [60, 61]. In the case of a fixed point with a constant Jacobian matrix $\mathbf{A}$ as in Eq. (1.24), the Lyapunov exponents are given by the real parts of the eigenvalues of the matrix $\mathbf{A}$.

The Lyapunov exponents are also the long-term rates of exponential growth or shrinkage of infinitesimal perturbations with different directions: for every Lyapunov exponent λ_k, there exists an initial perturbation vector $\mathbf{y}(0)$ such that

$$\lambda_k = \lim_{t\to\infty} \frac{1}{t} \log\left(\frac{||\mathbf{y}(t)||}{||\mathbf{y}(0)||}\right) = \lim_{t\to\infty} \frac{1}{t} \log\left(\frac{||\mathbf{Y}(t)\,\mathbf{y}(0)||}{||\mathbf{y}(0)||}\right) .$$

The specific (local) perturbation directions leading to the different Lyapunov exponents are known as covariant Lyapunov vectors [60]. They indicate the stable and unstable phase space directions of the dynamics.

For continuous-time dynamics generated by a differential equation like (1.22) with a smooth vector field, the trajectories are continuously differentiable curves. In contrast, the state trajectories in spiking neural networks with integrate-and-fire neuron models exhibit jumps: at spike events, the continuous evolution is interrupted and the network state is mapped to another state in a discrete manner. For instance, when a neuron's membrane potential reaches the spike threshold, it is mapped to the reset potential while the emitted spike may change the state variables of other, postsynaptic neurons. Recurrent networks of integrate-and-fire neurons with dynamics such as those given by Eqs. (1.16) and (1.17) can be described as dynamical systems with discontinuities [60]; they combine continuous evolution with aspects of discrete-time dynamics. With temporally constant external input and without noise, such spiking neural networks have autonomous deterministic dynamics with discontinuities at spike events. We can study the dynamical behavior in a similar way as introduced above. In particular, assuming sufficiently regular mappings between states at spike events, linear stability analysis can be applied to trajectories generated by the recurrent network dynamics. This allows us to characterize their stability properties using Lyapunov exponents. We note that the timings of (internal) spike events are implicitly determined by the network state, which has to be taken into account in the evolution of small perturbations [60, 64].

1.2.2 Chaos in recurrent neural networks

The collective dynamics of randomly connected recurrent neural networks can be chaotic [26, 29, 65], i.e. the trajectories of (microscopic) network states are aperiodic, statistically stationary, and exhibit sensitive dependence on the initial conditions. In these networks, the single units have rather simple dynamics and chaos arises through heterogenous and relatively strong recurrent connectivity, which is modeled as random. Chaos is a potential mechanism by which neural circuits can generate irregular, seemingly random patterns of activity, even when fully deterministic network dynamics are assumed. Therefore, chaos may contribute to the observed irregularity and variability of neuronal activity in the brain. Furthermore, the underlying stability properties of the network dynamics may be important for information processing.

Here, we illustrate the transition to self-generated, chaotic activity fluctuation in rate-based recurrent neural networks with random connectivity, using the "standard" model by [29]. Thus, we consider the following autonomous version of the network dynamics (1.18) (setting $\tau = 1$):

$$\frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} J_{ij}\phi\left(x_j\right) , \tag{1.27}$$

for $i = 1, \ldots, N$. The model employs as nonlinearity the sigmoidal transfer function $\phi(x) = \tanh(x)$. The synaptic coupling weights J_{ij} are chosen randomly: they are independent and identically distributed Gaussian variables with mean 0 and variance g^2/N. The weights remain fixed ("quenched") during the time evolution and each (independent) random choice of them gives rise to a different realization of the network, i.e. to a different dynamical system. Nonetheless, large networks show typical behavior, especially a transition to chaotic dynamics when the intensive coupling strength parameter $g \geq 0$ is increased. By writing $J_{ij} = g\tilde{J}_{ij}$, where the $\tilde{J}_{ij}$ are now Gaussian-distributed with mean 0 and variance $1/N$, we can view g as an explicit parameter of Eq. (1.27).

For $g = 0$, the neurons are unconnected and the linear (one-dimensional) dynamics of each unit has a stable fixed point at $x_i = 0$ with eigenvalue -1. Since $\phi(0) = 0$, the state $x_i = 0$ for all $i = 1, \ldots, N$ of the autonomous network dynamics (1.27) is a fixed point for any value of g. To determine the stability of this fixed point, we linearize the dynamics around it, see Eqs. (1.23) and (1.24). With $\phi'(0) = 1$, this yields the linear dynamics governing small perturbations,

$$\frac{dy_i}{dt} = -y_i + \sum_{j=1}^{N} J_{ij}y_j ,$$

for $i = 1, \ldots, N$. Hence, the Jacobian matrix of the nonlinear dynamics (1.27) at the zero fixed point $x_i = 0$ for all i equals $\mathbf{A} = -\mathbf{I} + \mathbf{J}$, where $\mathbf{J} = (J_{ij}) \in \mathbb{R}^{N\times N}$ is the matrix of the recurrent coupling weights and $\mathbf{I}$ denotes the identity matrix. The fixed point is stable if the real parts of all eigenvalues of the matrix $\mathbf{A}$ are negative, and it becomes linearly unstable if there is an eigenvalue with a positive real part. Because the negative identity matrix in $\mathbf{A}$ only shifts the eigenvalues by -1, it suffices to consider the eigenvalues of the coupling matrix $\mathbf{J}$.

In the random network model, $\mathbf{J}$ is a matrix with random entries. Therefore, its eigenvalues are randomly distributed points in the complex plane. The statistics of the eigenvalues of random matrices are studied by random matrix theory [66]. In the limit of large N, the eigenvalues of an $N \times N$ matrix whose entries are independent Gaussian random variables with mean 0 and variance $1/N$ are uniformly distributed over the unit disk in the complex plane [67–70]. This is known as the circular law and also holds for more general distributions of the entries. Multiplying a matrix by $g > 0$ scales its eigenvalues by this factor. Hence, for large network size N, the eigenvalues of the coupling matrix $\mathbf{J}$ are approximately uniformly distributed over a centered disk with radius g. The spectral radius of a matrix is the largest absolute value of its eigenvalues. Thus, the spectral radius of $\mathbf{J}$ is approximately given by the coupling strength g.

From the radius of the eigenvalue disk of $\mathbf{J}$, we obtain that for large N the largest real part of the eigenvalues (or spectral abscissa) of the Jacobian matrix $\mathbf{A}$ approximately equals $-1 + g$. Therefore, in the limit of large N, the zero fixed point is expected to be stable for $g < 1$ and linearly unstable for $g > 1$, which defines the critical coupling strength $g_{\text{crit}} = 1$. In simulations of large but finite networks, the fixed

point typically appears to be globally attractive for values of g below 1. When g is increased above 1, it becomes unstable giving rise to other dynamical behaviors, such as other fixed points or limit cycles [29].[8] However, increasing g further typically leads to dynamics with aperiodic trajectories characterized by asynchronous temporal fluctuations of the neuronal activity. By measuring the maximum Lyapunov exponent (1.25) in simulations, this dynamics can be shown to be chaotic. In the limit of $N \to \infty$, the transition from the stable fixed point to chaos in the autonomous random network model with dynamics (1.27) becomes sharp and occurs at $g_{\text{crit}} = 1$ [29]. Dynamic mean-field theory analytically describes the transition and the maximum Lyapunov exponent. The transition to chaos occurs precisely when the fixed point becomes linearly unstable according to random matrix theory, demonstrating an important connection to this field. We note that the application of random matrix theory to the linear stability analysis of large complex systems was pioneered by Robert M. May [71] in the context of theoretical ecology [72].

The transition to chaos is quite generic and happens also in random network models with biologically more realistic features, such as non-negative firing rates [24, 73]. If one considers rate-based recurrent networks as models of biological neural circuits, then the emerging chaotic activity fluctuations correspond to time-varying (instantaneous) firing rates of the neurons. A dynamical regime with chaotically fluctuating firing rates, sometimes called asynchronous rate chaos [24], is, however, difficult to define for networks of spiking neurons [24, 74, 75].

The statistically stationary activity in a randomly connected recurrent network of spiking neurons with deterministic autonomous dynamics may be well described by a fixed point of the neurons' firing rates, in the sense that each neuron spikes at a certain constant rate. Nevertheless, the collective dynamics can be chaotic, exhibiting sensitive dependence on the initial conditions [65, 76–78]. This means that the trajectories are unstable to small perturbations of the microscopic network state, such as the membrane potentials of the neurons. Such a small perturbation then grows and leads to a different sequence of spike timings in the network activity, while the firing rates remain stable. Hence, chaos in spiking neural networks gives rise to spike timings that are unpredictable in the long term. Since asynchronous irregular spiking activity as observed in the cortex appears unpredictable, the question was asked if chaotic neural network dynamics underlies this activity state, in which the neurons presumably operate in a fluctuation-driven regime. The balanced random network model introduced by van Vreeswijk and Sompolinsky [26, 79] showed that an asynchronous irregular activity state can robustly emerge through the dynamical balance of strong excitation and inhibition, even with deterministic dynamics. This work uses binary threshold units as simplified neuron models and finds a very fast growth of small perturbations, i.e. the change of a single unit's binary state.

Balanced spiking neural networks with deterministic dynamics can exhibit a fluctuation-driven regime with asynchronous irregular activity as well [80, 81]. While the balanced network mechanism is quite insensitive to the single neuron dynamics, the microscopic stability properties of the asynchronous irregular activity state depend substantially on them [65, 76–78, 82, 83]. In particular, statistically similar spiking activity can be stable or chaotic depending on the neurons' intrinsic dynamics.

1.2.3 Stochastic dynamical systems

Complex biological systems often have multiple scales that interact with each other, and on some of these scales, inevitable fluctuations or perturbations may be present. Hence, using deterministic dynamics

[8] To observe these behaviors, we need to start the dynamics from a nonzero initial condition, since $x_i = 0$ for all i is always a solution.

to model such systems is often only a coarse approximation or even impossible. Like in statistical physics, stochastic descriptions can provide effective mathematical models in this situation [6, 84]. In Section 1.1.3, we introduced stochastic point processes as probabilistic models of neuronal spiking activity. It can be useful to model also other aspects of dynamics in neural networks stochastically; for instance, the membrane potential of a neuron driven by the superposed inputs from many irregularly spiking neurons fluctuates in a random fashion.[9] Here, we briefly introduce dynamical systems whose time evolution includes a stochastic component.

Specifically, we consider the stochastic dynamics given by a differential equation that is perturbed or driven by noise:

$$\frac{dx}{dt} = f(x) + \sigma(x)\xi(t)\,, \tag{1.28}$$

where $\xi(t)$ represents a stochastic noise process. For simplicity, we focus on the case of a one-dimensional state variable $x(t) \in \mathbb{R}$. The vector field $f(x) \in \mathbb{R}$ specifies the deterministic part of the dynamics (see Section 1.2.1), assumed to be autonomous. The coefficient $\sigma(x)$ determines the possibly state-dependent noise strength, translating $\xi(t)$ into a velocity of $x(t)$. In general, different kinds of noise processes can be used to define a stochastic differential equation such as (1.28). For example, using a Poisson point process in the form of a spike train yields dynamics driven by so-called shot noise, which leads to jump-like changes and is important for stochastic models of neuronal dynamics [85, 86]. As another fundamental model of stochastic dynamics evolving in continuous time [6, 84], we consider here the stochastic differential equation (1.28) with Gaussian white noise $\xi(t)$. Additionally, we present a stochastic integrate-and-fire neuron model driven by this kind of noise.

With $\xi(t)$ being a Gaussian stochastic process, its probability distribution is fully specified by the vanishing mean $\langle \xi(t) \rangle = 0$ and the delta-function shaped autocorrelation $\langle \xi(t)\xi(s) \rangle = \delta(t - s)$ for all $t, s \in \mathbb{R}$, where angular brackets denote the expectation value. For this definition of standard Gaussian white noise, $\xi(t)$ has units of $t^{-1/2}$; hence, the noise strength $\sigma(x)$ in Eq. (1.28) has units of $x\, t^{-1/2}$. If $\sigma(x) = \sigma$ is independent of the state $x(t)$, the noise in Eq. (1.28) is called additive; otherwise, the noise is called multiplicative [6]. Continuous-time Gaussian white noise is highly irregular since it formally has infinite variance and its values at different times are statistically independent. Heuristically, Gaussian white noise corresponds to a limit of infinitely many and infinitely small random perturbations occurring during a short time interval. Nevertheless, one can construct solutions $x(t)$ of Eq. (1.28) with trajectories that are continuous functions of time.

The solution $W(t)$ for $t \geq 0$ with initial condition $W(0) = 0$ of the arguably simplest stochastic differential equation

$$\frac{dW}{dt} = \xi(t)$$

is called Wiener process [6], also referred to as (mathematical) Brownian motion; it can be formally obtained as the integral of $\xi(t)$,

$$W(t) = \int_0^t \xi(t')\, dt'\,. \tag{1.29}$$

The Wiener process $W(t)$ has continuous but nowhere differentiable trajectories. According to properties

[9] As mentioned in the previous Section 1.2.2, the deterministic dynamics of spiking neural networks can generate seemingly random fluctuations. On the level of a single neuron in the network, however, a stochastic description is appropriate. This is similar to the Brownian motion of a small pollen grain suspended in water, one of the historical origins leading to the study of stochastic dynamics [6].

of Gaussian random variables, $W(t)$ at time $t > 0$ is Gaussian-distributed with mean zero and variance t. Furthermore, $W(t)$ is a Markov process, i.e., given the present state, the future (stochastic) time evolution of the process is independent of the past [6].

From a physical perspective, the Wiener process models the one-dimensional diffusive[10] random motion of a particle with negligible inertia, i.e. in the overdamped limit; this motion could be caused by thermal fluctuations. The stochastic differential equation (1.28) with Gaussian white noise generalizes this picture: The deterministic part of the dynamics, specified by the vector field $f(x)$ and also called drift in this context, leads to directed motion of the state $x(t)$, while the noise simultaneously leads to diffusive motion, at least for short times. For this reason, solutions of such equations are known as diffusion processes [6], continuous-time Markov processes with continuous trajectories.

Mathematically, the stochastic differential equation (1.28) with Gaussian white noise is defined by requiring that its solution $x(t)$ for $t \geq 0$ satisfies the stochastic integral equation

$$x(t) = x(0) + \int_0^t f\big(x(t')\big)\, dt' + \int_0^t \sigma\big(x(t')\big)\, dW(t') \,. \tag{1.30}$$

The second term on the right-hand side is a stochastic integral with respect to a Wiener process $W(t)$, i.e. formally $dW(t) = \xi(t)\, dt$, see Eq. (1.29). There are different ways to define this stochastic integral as the limit of approximations by finite sums, leading in general to different results. The two most common definitions are the Itô integral and the Stratonovich integral [6]; they give rise to the corresponding Itô and Stratonovich interpretations of the stochastic differential equation (1.28). The choice of an interpretation is part of the mathematical modeling. Writing the stochastic integral as in Eq. (1.30) usually refers to the Itô integral, and in this case, the stochastic differential equation with Gaussian white noise is interpreted as an Itô equation. We note that for additive noise, $\sigma(x) = \sigma$, the two conventions yield identical results [6].

Due to the noise, the solution of a stochastic differential equation is a random quantity even for a deterministic (non-random) initial condition, calling for a probabilistic description of the dynamics. The Fokker-Planck equation governs how a probability distribution of initial conditions evolves in time. The time-dependent probability density $p(x,t)$ of the one-dimensional diffusion process $x(t)$ generated by the stochastic dynamics (1.28), interpreted as an Itô equation, evolves according to the following Fokker-Planck equation [6]:

$$\begin{aligned} \frac{\partial}{\partial t}\, p(x,t) &= -\frac{\partial}{\partial x}\left[f(x)p(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[\sigma^2(x)p(x,t)\right] && (1.31)\\ &= -\frac{\partial}{\partial x}\left[f(x)p(x,t) - \frac{1}{2}\frac{\partial}{\partial x}\,\sigma^2(x)p(x,t)\right] \\ &= -\frac{\partial}{\partial x}\, j(x,t)\,, && (1.32) \end{aligned}$$

where the third equality defines the probability current or flux $j(x,t)$. This definition puts the Fokker-Planck equation (1.31) into the form of a continuity equation (1.32), which describes the conservation of probability. For diffusion processes in multiple dimensions, the associated Fokker-Planck equation generalizes Eqs. (1.31) and (1.32), see [6].

[10] By "diffusive" we mean that the typical distance traveled from the origin scales as $t^{1/2}$ with time t, as is the case for the Wiener process.

As an example, we consider the stochastic LIF neuron model (Section 1.1.2) with Gaussian white noise and derive the probability density of the fluctuating membrane potential and the firing rate in a statistically stationary state. Driven by a constant input current $\bar{I}$ and an additive Gaussian white noise $\sqrt{2\tau_m}\sigma_V\,\xi(t)$, the neuron's subthreshold membrane potential $V(t)$ follows the stochastic differential equation

$$\begin{aligned}\frac{dV}{dt} &= -\frac{1}{\tau_m}(V - V_{\text{rest}}) + \frac{1}{\tau_m}\bar{I} + \sqrt{\frac{2}{\tau_m}}\sigma_V\,\xi(t) \\ &= f(V) + \sigma\,\xi(t)\,, \end{aligned} \tag{1.33}$$

with $f(V) = -(V - \mu_V)/\tau_m$, $\mu_V = V_\infty = V_{\text{rest}} + \bar{I}$ and $\sigma^2 = 2\sigma_V^2/\tau_m$, where we included $\bar{I}$ into the intrinsic voltage dynamics given by $f(V)$. This stochastic model arises through the commonly employed diffusion approximation [18, 85], in which the synaptically weighted sum of asynchronous and irregular input spike trains to a neuron is replaced by a mean input and fluctuations in the form of Gaussian white noise. The strengths of mean and noise can be obtained from the mean and variance of the weighted sum of spike trains integrated over a short time interval. The diffusion approximation becomes adequate for a large number of inputs with small synaptic weights. The one-dimensional stochastic dynamics (1.33) corresponds to the case of the LIF neuron model with delta-function shaped postsynaptic currents, see Eq. (1.11).[11] In the absence of a spike threshold, the solution $V(t)$ of Eq. (1.33) constitutes an Ornstein-Uhlenbeck process [6], a well-known diffusion process; its stationary probability distribution is Gaussian with mean μ_V and variance σ_V^2, as we will also show below. Thus, the chosen parameters in Eq. (1.33) can be interpreted intuitively using this Ornstein-Uhlenbeck process.

Like in a deterministic integrate-and-fire neuron model, the stochastic model considered here emits a spike when the membrane potential $V(t)$ reaches the threshold V_{th}, followed by a reset to the value $V_{\text{re}} < V_{\text{th}}$. In between threshold crossings, $V(t)$ evolves, after an optional refractory period $\tau_{\text{ref}} \geq 0$, according to the stochastic differential equation (1.33). The noise driving the dynamics has two important consequences. First, the neuron's firing times are random variables; their sequence, i.e. the output spike train, forms a stochastic point process (Section 1.1.3).[12] Second, even for constant input currents $\bar{I}$ that give rise to a subthreshold asymptotic voltage $\mu_V = V_{\text{rest}} + \bar{I} < V_{\text{th}}$, the membrane potential will reach the threshold due to fluctuations, and the neuron spikes in a rather irregular manner with a nonzero firing rate.

The threshold-crossing and reset mechanism of the stochastic integrate-and-fire model can be incorporated into the Fokker-Planck equation governing the probability density $p(V, t)$ of the membrane potential $V(t)$ [18, 25]. Intuitively, upon reaching the threshold V_{th}, a trajectory is removed from the probability density, and then it is (instantaneously or after the optional refractory period) reinserted at the reset potential V_{re}. The resulting flux of probability through the threshold gives the temporal density of output spikes, i.e. the neuron's instantaneous firing rate.

Considering the Fokker-Planck equation as a continuity equation, see (1.32), the threshold and reset

[11] For postsynaptic currents with dynamics as Eq. (1.9), the diffusion approximation results in a multi-dimensional stochastic differential equation.

[12] The output spike train is not a Poisson process, as the spikes occur not completely independently of each other due to the reset. In the present model, it can be described as a renewal process, in which the probability of a spike depends only on the passed time since the last spike (reset) [5, 85].

act as delta-function sink and source terms [87, 88]:

$$\frac{\partial}{\partial t}\, p(V,t) + \frac{\partial}{\partial V}\, j(V,t) = -\rho_{\mathrm{th}}(t)\, \delta(V - V_{\mathrm{th}}) + \rho_{\mathrm{re}}(t)\, \delta(V - V_{\mathrm{re}})\,, \tag{1.34}$$

where

$$j(V,t) = f(V)p(V,t) - \frac{\sigma^2}{2}\frac{\partial}{\partial V}\, p(V,t) \tag{1.35}$$

is the voltage probability current due to drift and diffusion given by the stochastic dynamics (1.33). The rate-like quantities $\rho_{\mathrm{th}}(t)$ and $\rho_{\mathrm{re}}(t)$ multiplying the delta functions in Eq. (1.34) determine the jump-like changes of the probability current $j(V,t)$ (as a function of V) at V_{th} and V_{re}, respectively. Equation (1.35) implies that the probability density $p(V,t)$ itself is continuous in V; otherwise, the left-hand side of Eq. (1.34) would contain an unbalanced derivative of a delta function. Since the trajectories are "absorbed" at the threshold, we have $p(V,t) = 0$ for $V \geq V_{\mathrm{th}}$ and thus $j(V,t) = 0$ for $V > V_{\mathrm{th}}$. Taken together, the jump-like changes of $j(V,t)$ at V_{th} and V_{re} read

$$j(V_{\mathrm{th}}^-, t) = -\frac{\sigma^2}{2}\frac{\partial}{\partial V} p(V_{\mathrm{th}}^-, t) = \rho_{\mathrm{th}}(t)$$

and

$$j(V_{\mathrm{re}}^+, t) - j(V_{\mathrm{re}}^-, t) = -\frac{\sigma^2}{2}\left[\frac{\partial}{\partial V} p(V_{\mathrm{re}}^+, t) - \frac{\partial}{\partial V} p(V_{\mathrm{re}}^-, t)\right] = \rho_{\mathrm{re}}(t)\,,$$

where V^- and V^+ indicate the limit from below and above, respectively. In particular, $\rho_{\mathrm{th}}(t)$ is equal to the instantaneous firing rate $\rho(t) = j(V_{\mathrm{th}}^-, t)$, and the reset mechanism sets $\rho_{\mathrm{re}}(t) = \rho(t - \tau_{\mathrm{ref}})$. We note that instead of adding the sink and source terms in the continuity equation (1.34), these jump conditions together with the required continuity of $p(V,t)$ have been directly applied as boundary conditions at V_{th} and V_{re} on the solution of the Fokker-Planck equation [18, 25, 87].

In a statistically stationary state, the voltage probability density is time-invariant, $p(V,t) = p(V)$, and the firing rate is constant in time, $\rho(t) = \rho$. With these assumptions, Eqs. (1.34) and (1.35) reduce to the ordinary differential equation

$$\frac{d}{dV} j(V) = \frac{d}{dV}\left[f(V)p(V) - \frac{\sigma^2}{2}\frac{d}{dV} p(V)\right] = -\rho\, \delta(V - V_{\mathrm{th}}) + \rho\, \delta(V - V_{\mathrm{re}})\,. \tag{1.36}$$

Hence, the time-invariant probability current $j(V)$ is piecewise constant, jumping by $-\rho$ at V_{th} and by $+\rho$ at V_{re}. Since the probability current through the threshold equals the firing rate, $j(V_{\mathrm{th}}^-) = \rho$, it follows that $j(V) = \rho$ if $V_{\mathrm{re}} < V < V_{\mathrm{th}}$ and $j(V) = 0$ if $V < V_{\mathrm{re}}$. Integrating Eq. (1.36), we obtain the remaining first-order linear differential equation

$$\frac{d}{dV} p(V) - \frac{2}{\sigma^2} f(V) p(V) = -\frac{2}{\sigma^2} j(V)\,, \tag{1.37}$$

whose (normalized) solution $p(V)$ for $V \leq V_{\mathrm{th}}$ with $p(V_{\mathrm{th}}) = 0$ yields the stationary probability density.

To solve Eq. (1.37) by variation of constants, we first consider the solution $\psi(V)$ of the homogeneous equation, i.e. with $j(V) = 0$. This solution can be written as

$$\psi(V) = A \, \exp\left(\frac{2}{\sigma^2}\int_{V_0}^{V} f(v)\, dv\right) = A \, \exp\left(-\frac{2}{\sigma^2}U(V)\right), \tag{1.38}$$

with a constant A and a potential function $U(V)$ such that $f(V) = -dU(V)/dV$. For the LIF neuron dynamics with $f(V) = -(V - \mu_V)/\tau_m$, the potential

$$U(V) = \frac{1}{2\tau_m}\,(V - \mu_V)^2 \tag{1.39}$$

together with $\sigma^2 = 2\sigma_V^2/\tau_m$ lead to the Gaussian-shaped function

$$\psi(V) = A \, \exp\left(-\frac{1}{2\sigma_V^2}\,(V - \mu_V)^2\right), \tag{1.40}$$

with mean μ_V and standard deviation σ_V of the corresponding Gaussian probability density, given by the normalization $A = 1/\left(\sqrt{2\pi}\sigma_V\right)$. This Gaussian density also yields the stationary probability distribution of the Ornstein-Uhlenbeck process generated by Eq. (1.33) in the absence of threshold and reset, because it is the normalized time-invariant solution of the associated Fokker-Planck equation [6].

The solution of the inhomogeneous equation (1.37) for $V \le V_{\text{th}}$ with $p(V_{\text{th}}) = 0$ is obtained through variation of constants as

$$\begin{aligned} p(V) &= \frac{2}{\sigma^2}\,\psi(V)\int_{V}^{V_{\text{th}}} \frac{j(V')}{\psi(V')}\, dV' \\ &= \frac{\rho\,\tau_m}{\sigma_V^2}\,\exp\left(-\frac{(V-\mu_V)^2}{2\sigma_V^2}\right)\int_{V}^{V_{\text{th}}} \Theta(V' - V_{\text{re}})\,\exp\left(\frac{(V'-\mu_V)^2}{2\sigma_V^2}\right) dV' \\ &= \frac{\sqrt{2}\,\rho\,\tau_m}{\sigma_V}\,\exp\left(-\frac{(V-\mu_V)^2}{2\sigma_V^2}\right)\int_{\frac{V-\mu_V}{\sqrt{2}\sigma_V}}^{\frac{V_{\text{th}}-\mu_V}{\sqrt{2}\sigma_V}} \Theta\left(y - \frac{V_{\text{re}}-\mu_V}{\sqrt{2}\sigma_V}\right) \mathrm{e}^{y^2}\, dy\,, \end{aligned} \tag{1.41}$$

which depends on the voltages V, V_{th} and V_{re} only in the form $(V - \mu_V)/(\sqrt{2}\sigma_V)$; we recall that $\Theta(x)$ denotes the Heaviside step function. By construction, the solution is continuous at V_{re}.

The stationary probability density $p(V)$ fulfills the normalization condition

$$\int_{-\infty}^{V_{\text{th}}} p(V)\, dV = 1 - \rho\,\tau_{\text{ref}}\,,$$

where the subtracted term on the right-hand side accounts for the probability of finding the neuron in the

refractory period. This normalization condition determines the stationary firing rate ρ as follows:

$$\begin{aligned}\frac{1}{\rho} &= \tau_{\text{ref}} + 2\tau_m \int_{-\infty}^{\frac{V_{\text{th}}-\mu_V}{\sqrt{2}\sigma_V}} \mathrm{e}^{-x^2} \int_{-\infty}^{\frac{V_{\text{th}}-\mu_V}{\sqrt{2}\sigma_V}} \Theta\left(y-x\right) \Theta\left(y - \frac{V_{\text{re}}-\mu_V}{\sqrt{2}\sigma_V}\right) \mathrm{e}^{y^2}\, dy\, dx \\ &= \tau_{\text{ref}} + 2\tau_m \int_{\frac{V_{\text{re}}-\mu_V}{\sqrt{2}\sigma_V}}^{\frac{V_{\text{th}}-\mu_V}{\sqrt{2}\sigma_V}} \mathrm{e}^{y^2} \int_{-\infty}^{y} \mathrm{e}^{-x^2}\, dx\, dy\,, \end{aligned} \tag{1.42}$$

where we used the substitution $x = (V - \mu_V)/(\sqrt{2}\sigma_V)$. The right-hand side of Eq. (1.42) is the mean interval between consecutive threshold crossings. In particular, the second term equals the expected time interval until the voltage $V(t)$ starting at the reset V_{re} reaches the threshold V_{th} for the first time, which is called mean first-passage time [6, 85]. With $\mu_V = V_\infty = V_{\text{rest}} + \bar{I}$, the firing rate ρ as a function of the constant input current $\bar{I}$ for fixed σ_V yields the transfer function of the stochastic LIF neuron model with dynamics (1.33).

1.2.4 Random dynamical systems

In addition to a fully probabilistic description, we can consider a stochastic dynamical system such as the stochastic differential equation (1.28) for a fixed or "frozen" realization of the noise. This provides a way to extend concepts from deterministic dynamics to stochastic systems, treating them as (random) nonautonomous dynamical systems, see Section 1.2.1. This approach is founded on the theory of random dynamical systems [89], see [90, 91] for a brief overview. In this theory, the dynamical rule that uniquely determines the future states given the present state is assumed to vary randomly in time. In particular, the dynamical stability of time-dependent trajectories can be characterized using Lyapunov exponents, as discussed in Section 1.2.1. For this, one considers the exponential divergence or convergence of nearby trajectories that evolve under the same noise realization. Since Lyapunov exponents are long-term averages, they usually do not depend on the specific realization of the noise and are thus non-random quantities that characterize the system dynamics.

The maximum Lyapunov exponent $\lambda_{\max}$, see Eq. (1.25), of a stochastic differential equation treated as a random dynamical system exposes the following dichotomy [90, 91]. If $\lambda_{\max}$ is negative, the dynamics is asymptotically stable and small perturbations of the initial condition decay. Furthermore, under certain conditions, trajectories forget their initial condition as they are globally attracted by a single time-dependent trajectory determined by the noise realization. Put differently, the time-dependent probability density in phase space describing an ensemble of trajectories given by a probability density of initial conditions will collapse to a delta function centered on this trajectory, which is also called a random sink [90, 91]. In contrast, a positive maximum Lyapunov exponent indicates sensitive dependence on the initial condition and therefore chaotic dynamics. Indeed, for dissipative systems, the dynamics will settle on a random strange attractor [90, 91], the time-dependent analogue of a strange attractor governing chaotic dynamics in autonomous systems. Despite being driven by the same noise, trajectories remain (partially) unpredictable due to the sensitive dependence on initial conditions. The maximum Lyapunov exponent is hence an important diagnostic measure to distinguish between stable and chaotic dynamics in stochastically driven systems.

Neural networks are often driven by time-dependent inputs or signals, for example when they process time-varying information, rendering their dynamics nonautonomous. In this context, dynamical stability is closely related to the reliability of the network response [90, 91]: How reproducible is the response

to the same input in different trials? Such trial-to-trial variability of neurobiological systems could be caused by noise or chaotic dynamics [5]. If we model the time-dependent input as a stochastic drive, as anticipated in Eq. (1.18), the random dynamical systems approach can be applied to study the response reliability of driven neural networks [90, 91]. Assuming that the network's initial state differs between trials, a negative maximum Lyapunov exponent is associated with reliability, whereas a positive one indicates (partial) unreliability.

1.2.5 Transition phenomena

Transition phenomena constitute an important aspect of the behavior of dynamical systems. Here, we will consider two different kinds of transitions. The first kind refers to qualitative changes in the dynamics that occur at particular values of system parameters or external conditions when these are varied. In Section 1.2.2, we already discussed the transition from a stable fixed point state to chaotic dynamics in rate-based random neural networks when the coupling strength is increased. Mathematically, such transitions in dynamical systems are known as bifurcations [59, 92, 93]. In statistical physics, similar qualitative changes in the collective and often macroscopic behavior of systems with many interacting degrees of freedom are called phase transitions. From a perspective of dynamics, phase transitions can be related to bifurcations, in particular in non-equilibrium and complex systems [93–98].

The second kind of transitions that will be considered are transitions between different stable dynamical states or attractors that coexist in a single system with given parameters. Such coexistence is termed multistability [99, 100], with the simplest case being bistability. For autonomous deterministic dynamics, the initial condition fully determines which attractor is selected, and only an external perturbation to a different initial condition can lead to a switch between attractors. In stochastic systems, noise can give rise to transitions between coexisting metastable dynamical states [6, 101, 102]. This is sometimes called noise-activated escape or switching. Finally, varying properties of the noise can also induce transitions in the dynamical behavior that are analogous to bifurcations and phase transitions [103–105].

To introduce bifurcations, we consider autonomous deterministic dynamical systems (Section 1.2.1) depending on parameters:

$$\frac{d}{dt}\mathbf{x} = \mathbf{f}(\mathbf{x};\boldsymbol{\alpha})\,,$$

where the autonomous vector field $\mathbf{f}(\mathbf{x};\boldsymbol{\alpha}) \in \mathbb{R}^N$ is a smooth function of both the state variable $\mathbf{x} \in \mathbb{R}^N$ and the parameter vector $\boldsymbol{\alpha} \in \mathbb{R}^P$; the parameters are sometimes regarded as control parameters. In other words, we are interested in a parameterized family of dynamical systems. A bifurcation occurs at a critical parameter value $\boldsymbol{\alpha}_0$, called a bifurcation point, if an arbitrarily small change of the parameters can lead to a qualitative change in the dynamical behavior, meaning that the dynamics are not topologically equivalent [59, 92, 106].[13] As pointed out in Section 1.2.1, fixed points and periodic trajectories provide important structure in the phase space of autonomous dynamical systems. If the number or the stability properties of such solutions change as parameters are varied, the phase space structure changes topologically and a bifurcation occurs.

Local bifurcations of fixed points are fundamental [59, 92, 93, 106]; they describe qualitative changes or transitions in the steady states of a dynamical system. For a given fixed point $\bar{\mathbf{x}}$ with $\mathbf{f}(\bar{\mathbf{x}};\boldsymbol{\alpha}) = \mathbf{0}$, such a bifurcation can be detected from the local linearization around $\bar{\mathbf{x}}$, see Eq. (1.24), specifically from the

[13] Given the assumed smoothness of the vector field, a slight change of the parameters at a generic parameter value leads to topologically equivalent dynamics.

eigenvalue spectrum of the Jacobian matrix $(\partial \mathbf{f}/\partial \mathbf{x})(\bar{\mathbf{x}}; \boldsymbol{\alpha})$. A bifurcation of the fixed point can only occur if the Jacobian matrix has a zero eigenvalue or a purely imaginary pair of complex conjugate eigenvalues [92, 106], and the fixed point is then called non-hyperbolic. Following the eigenvalues of the local linearization as parameters are varied, we expect a bifurcation of the fixed point when a real eigenvalue or a pair of complex conjugate eigenvalues reaches the imaginary axis. Either of these conditions can be met by tuning the parameters along a one-dimensional curve to a bifurcation point; hence, such a bifurcation can arise in a family of dynamical systems with a single (control) parameter. If all other eigenvalues, i.e. those with a nonzero real part, have negative real parts, the bifurcating fixed point is situated at the boundary between stability and instability.

A system that operates at, or very close to, a transition between qualitatively different dynamics is said to be in a critical state or regime [93, 97]. In statistical physics, critical phenomena refer to the peculiar behavior of systems with many interacting units near a continuous phase transition [95, 107, 108], such as the onset of ferromagnetism in iron below a critical temperature. Phase transitions typically involve a qualitative change of a system's collective order or organization when a control parameter, such as the temperature, is varied. At a continuous phase transition, a qualitatively different collective state (phase) emerges from another state in a gradual manner. An order parameter quantifying this emergence changes continuously but not smoothly as a function of the control parameter at the transition point [95, 107, 108]; it often equals zero on one side of the transition. Important aspects of critical behavior near a continuous phase transition are fluctuations on many scales, scale invariance, power-law dependencies, and universality, meaning that the behavior on larger scales is strikingly independent of microscopic details of the system [108]. Some aspects of critical behavior and their consequences are shared by the behavior of dynamical systems, even low-dimensional ones, near a bifurcation point, i.e. in a critical state [93, 97]. The analogue of continuous phase transitions are bifurcations where the qualitatively different dynamics emerges gradually. The continuous onset allows a system to exist right at the edge of an instability, which is sometimes taken as tantamount to criticality.

As an example, we present the transcritical bifurcation, which can model the onset of an epidemic or, more generally, the transition to a state with self-sustained activity [59, 93]. Specifically, a simple model exhibiting a transcritical bifurcation is given, in dimensionless form, by the one-dimensional dynamics

$$\begin{aligned} \frac{dx}{dt} &= -x + \alpha x(1-x) \\ &= (\alpha - 1)x - \alpha x^2 = f(x;\alpha)\,, \end{aligned} \tag{1.43}$$

with a scalar parameter $\alpha \geq 0$. This differential equation corresponds to the deterministic mean-field equation of the contact process [95, 97], a simple stochastic model for the spreading of activity, such as an infection, in a system of interacting units with binary states. In this context, x equals the fraction of active units in the system, i.e. $0 \leq x \leq 1$ and $1 - x$ equals the fraction of inactive units; active units become inactive at a rate set to one, while inactive units become active at a rate αx, which is proportional to the fraction of active units. Hence, activity generates more activity through positive feedback, as in the spread of an infectious disease or in an excitatory coupled neural network.

The dynamics (1.43) has two fixed points, $\bar{x} = 0$ and $\bar{x} = 1 - 1/\alpha$ (assuming $\alpha > 0$). In particular, consistent with the above interpretation of the dynamics, the quiescent steady state $\bar{x} = 0$ exists for

all parameter values.[14] The local linearization around the fixed points is given by the Jacobian factor $f'(\bar{x};\alpha) = \alpha - 1 - 2\alpha\bar{x}$. What happens when the parameter α is increased? For $\bar{x} = 0$, $f'(\bar{x};\alpha) = \alpha - 1$ changes sign from negative to positive at the critical parameter value $\alpha_0 = 1$. Hence, the quiescent steady state $\bar{x} = 0$ is stable for $\alpha < 1$ and linearly unstable for $\alpha > 1$. For the other steady state $\bar{x} = 1 - 1/\alpha$, we have $f'(\bar{x};\alpha) = 1 - \alpha$, and the stability changes in the opposite way at $\alpha_0 = 1$. This exchange of stability between two fixed points that coincide at the bifurcation point is characteristic of a transcritical bifurcation [59, 93]. For $\alpha > 1$, $\bar{x} = 1 - 1/\alpha$ becomes positive and thus describes a stable steady state with sustained activity.

Let us consider the time evolution of a small positive perturbation $\delta x(0)$ at time $t = 0$ of the quiescent steady state $\bar{x} = 0$. For $\alpha < 1$, the perturbation decays approximately exponentially with (decay) rate $1 - \alpha$, i.e. $\delta x(t) \approx \mathrm{e}^{-(1-\alpha)t}\,\delta x(0)$. For $\alpha > 1$, assuming that $\delta x(0)$ is much smaller than the active steady state $\bar{x} = 1 - 1/\alpha$, the perturbation initially grows approximately exponentially with (growth) rate $\alpha - 1$; indeed, it will grow as a logistic function towards the active steady state. The two dynamical regimes, separated by the bifurcation at the critical parameter value $\alpha_0 = 1$, are also referred to as subcritical and supercritical, respectively [97, 109]. As α approaches $\alpha_0 = 1$ from below, the exponential decay becomes increasingly slow, and at the bifurcation point, the exponential decay rate vanishes. This behavior is known as critical slowing down [93, 97]. Right at $\alpha_0 = 1$, solving the differential equation $\dot{x} = -x^2$ yields that the perturbation will decay as $\delta x(t) = (1/\delta x(0) + t)^{-1}$, in particular $\delta x(t) \propto t^{-1}$ asymptotically as $t \to \infty$. This power-law decay is a simple instance of critical behavior in dynamical systems operating at the boundary between stability and instability [93, 97].

Next, we consider the second kind of transitions mentioned above, which can occur in systems with given parameter values. We focus on the simple case of one-dimensional dynamics with two coexisting stable fixed points. Such bistability is exhibited by the (dimensionless) differential equation $\dot{x} = \alpha x - x^3 = f(x)$ [59, 110].[15] For $\alpha > 0$, the dynamics has two stable fixed points at $\bar{x} = \pm\sqrt{\alpha}$, whose domains of attraction are separated by an unstable fixed point at $\bar{x} = 0$. Depending on the initial condition, the dynamics will settle in one of the two stable steady states. Switching between them requires a sufficiently strong external perturbation. For example, an additional constant input term $+\mu$ in the differential equation would shift $f(x)$ vertically up or down. For sufficiently positive or negative values of μ, one of the stable fixed points and the unstable one coalesce and disappear together, while the other stable fixed point becomes globally attractive. Consequently, a transiently applied input that is strong enough can lead to a switch to the other stable steady state.

It is useful to picture the dynamics as the overdamped motion in a potential, i.e. $\dot{x} = -U'(x) = -dU(x)/dx$ with a potential function $U(x)$. This is always possible in the considered case of one-dimensional dynamics given by a continuous vector field on the real line. Specifically, the quartic potential $U(x) = -\alpha x^2/2 + x^4/4$ yields $f(x) = -U'(x) = \alpha x - x^3$; for $\alpha > 0$, it is a double-well potential. A local minimum (maximum) of $U(x)$ corresponds to a stable (unstable) fixed point of the dynamics. Assuming $\alpha > 0$ as before, the two local minima of $U(x)$ correspond to the stable steady states at $\bar{x} = \pm\sqrt{\alpha}$. They are separated by a local maximum at $\bar{x} = 0$. This potential barrier must be overcome for a transition between the two stable states located in the potential wells.

[14] In stochastic models for the spreading of activity, the completely inactive state is often a so-called absorbing state, which means that once the dynamics reaches this state, it cannot be left. Transitions involving absorbing states underlie important forms of non-equilibrium phase transitions [95, 97].

[15] This equation with parameter $\alpha \in \mathbb{R}$ is the normal form of the supercritical pitchfork bifurcation, in which a stable fixed point becomes unstable while two stable ones emerge from it in a symmetrical manner.

To briefly introduce metastability and noise-activated escape over a potential barrier [6, 101, 102], we consider the dynamics in the presence of additive Gaussian white noise (Section 1.2.3):

$$\begin{aligned}\frac{dx}{dt} &= \alpha x - x^3 + \sigma\,\xi(t) \\ &= -U'(x) + \sigma\,\xi(t)\,,\end{aligned} \tag{1.44}$$

with standard Gaussian white noise $\xi(t)$, i.e. $\langle\xi(t)\rangle = 0$ and $\langle\xi(t)\xi(s)\rangle = \delta(t-s)$ for all $t, s \in \mathbb{R}$, and noise strength $\sigma > 0$. In contrast to the bistability of the deterministic dynamics discussed above, the solution $x(t)$ of the stochastic differential equation (1.44) for $t \to \infty$ exhibits a unique stationary probability density $p_{\mathrm{st}}(x)$, independent of the initial condition [6, 110]. This probability density can be obtained as the normalized time-invariant solution of the Fokker-Planck equation (1.31) associated to Eq. (1.44). In the present one-dimensional case without any sinks or sources (cf. Eqs. (1.36) and (1.37)), the time-invariant probability current must vanish and the stationary probability density can be written as

$$p_{\mathrm{st}}(x) = A\,\exp\left(-\frac{2}{\sigma^2}U(x)\right) \tag{1.45}$$

with the quartic potential $U(x) = -\alpha x^2/2 + x^4/4$ and a normalization constant $A > 0$ such that $\int_{\mathbb{R}} p_{\mathrm{st}}(x)\,dx = 1$; see also [6, 110]. We note that the form of Eq. (1.45) agrees with the homogenous solution (1.38) used to determine the stationary voltage probability density of the stochastic LIF neuron model in Section 1.2.3.

Since the exponential function is strictly increasing, a local minimum (maximum) of $U(x)$ gives rise to a local maximum (minimum) of the probability density $p_{\mathrm{st}}(x)$ in Eq. (1.45); in the limit of small noise strength σ, the local maxima become increasingly sharp. For $\alpha > 0$, the symmetric double-well potential $U(x)$ yields a bimodal $p_{\mathrm{st}}(x)$ with two equal local maxima at $\bar{x} = \pm\sqrt{\alpha}$, i.e. at the two stable fixed points of the deterministic dynamics. In the stationary state of the stochastic dynamics (1.44), $x(t)$ can thus be found with equal probability near either of the two local minima of $U(x)$. This implies that there are noise-activated transitions between the potential wells.

If the potential barrier separating the two wells is high enough and the noise is not too strong, transitions are expected to be rare compared to the dynamics within the wells. In this case, the peaks of the bimodal stationary probability density $p_{\mathrm{st}}(x)$ correspond to metastable states of the stochastic dynamics [6, 102]: If the process $x(t)$ is initially inside one of the wells, it typically stays there for a long time, settling to a quasi-stationary state, before a transition occurs. In such a quasi-stationary state, the time to escape is exponentially distributed. Since transitions are rare events, the quasi-stationary probability density of $x(t)$ in the metastable state should have a shape similar to the peak of $p_{\mathrm{st}}(x)$ that corresponds to the state. The expected time until a transition occurs becomes exponentially large in the limit of weak noise, as we show below.

For the one-dimensional diffusion process $x(t)$ given by Eq. (1.44), the time to escape over a potential barrier can be obtained as a first-passage or exit time from an interval [6, 110]. For this, we consider the interval $(-\infty, x_1)$ and assume that $x(t)$ starts in $x(0) = x_0 < x_1$ at time $t = 0$. The mean of the random time at which $x(t)$ reaches x_1 for the first time and thereby exits the interval is then given by

$$t_1(x_0; x_1) = \frac{2}{\sigma^2}\int_{x_0}^{x_1} \exp\left(\frac{2}{\sigma^2}U(y)\right)\int_{-\infty}^{y}\exp\left(-\frac{2}{\sigma^2}U(x)\right)\,dx\,dy\,. \tag{1.46}$$

Notice that this expression for the mean first-passage time from x_0 to x_1 generalizes the second term in Eq. (1.42) for the inverse stationary firing rate of the stochastic LIF neuron. We set the initial point x_0 close or equal to $U(x)$'s left local minimum point $\bar{x} = -\sqrt{\alpha}$; for a metastable state, the precise position should have negligible influence. If we further set the exit point x_1 to the local maximum point $\bar{x} = 0$, Eq. (1.46) yields the mean exit time from the left potential well. Due to symmetry, the expected escape or transition time over the potential barrier is twice this time.[16]

In the limit of weak noise, i.e. as $\sigma \to 0$, Eq. (1.46) can be approximated asymptotically using Laplace's method [111], which is a particular instance of what is known as a saddle-point approximation in theoretical physics. The basic idea is that integrals with an integrand of the form $\exp(n\,\phi(x))$ are dominated as $n \to \infty$ by the contribution from a small neighborhood around the point where $\phi(x)$ has a maximum. In case of the double-well potential $U(x)$ in Eq. (1.46), the relevant points are the left local minimum point $\bar{x} = -\sqrt{\alpha}$ for the inner integral and the local maximum point $\bar{x} = 0$ for the outer integral. Since $U'(\bar{x}) = 0$, the quadratic Taylor approximation around these points reads $U(x) \approx U(\bar{x}) + 1/2\,U''(\bar{x})(x - \bar{x})^2$, where the second derivative $U''(\bar{x})$ of the potential is positive (negative) at the local minimum (maximum) point. This approximation will be used in the exponents of the integrands. As discussed above, we set x_0 equal to $\bar{x} = -\sqrt{\alpha}$ and x_1 equal to $\bar{x} = 0$. Taken together, we obtain the asymptotic approximation as $\sigma \to 0$ of the mean exit time (1.46):

$$\begin{aligned}
t_1(x_0; x_1) &\approx \frac{2}{\sigma^2}\, \mathrm{e}^{2U(x_1)/\sigma^2} \int_{x_0}^{x_1} \mathrm{e}^{U''(x_1)(y-x_1)^2/\sigma^2} \int_{-\infty}^{x_1} \mathrm{e}^{-2U(x)/\sigma^2}\, dx\, dy \\
&\approx \frac{2}{\sigma^2}\, \mathrm{e}^{2(U(x_1)-U(x_0))/\sigma^2} \int_{-\infty}^{0} \mathrm{e}^{-|U''(x_1)|\, y^2/\sigma^2}\, dy \int_{-\infty}^{x_1} \mathrm{e}^{-U''(x_0)(x-x_0)^2/\sigma^2}\, dx \\
&\approx \frac{1}{\sigma} \sqrt{\frac{\pi}{|U''(x_1)|}}\, \mathrm{e}^{2(U(x_1)-U(x_0))/\sigma^2} \int_{-\infty}^{\infty} \mathrm{e}^{-U''(x_0)(x-x_0)^2/\sigma^2}\, dx \\
&= \frac{\pi}{\sqrt{U''(x_0)\,|U''(x_1)|}}\, \mathrm{e}^{2(U(x_1)-U(x_0))/\sigma^2}\,,
\end{aligned} \tag{1.47}$$

where the integration limits are extended after applying the quadratic approximation in the integrands; see also [6, 110]. Setting the exit point x_1 sufficiently above $\bar{x} = 0$ would result in a factor of two. Equation (1.47) demonstrates that the expected time to escape from a metastable state over a potential barrier increases as an exponential function of the barrier height $U(x_1) - U(x_0)$ relative to $\sigma^2/2$; for the quartic double-well potential $U(x)$, the height is $\alpha^2/4$. The inverse value of the escape time gives the rate of noise-activated transitions. The exponential dependence of this rate has the form of the Arrhenius equation for chemical reaction rates, where the barrier height acts as the required activation energy and the squared noise strength is proportional to the temperature [6, 112].

The two coexisting metastable states in the previous example originated from the two stable fixed points of the underlying deterministic dynamics. Bistability in the form of a bimodal stationary probability density can also be induced by noise [103, 105]. We illustrate this effect with a simple example [110]:

[16] Alternatively, one could consider the mean exit time (1.46) as a function of $x_1 > x_0$ for fixed $x_0 = -\sqrt{\alpha}$. In the case of metastability, this function has a plateau for x_1 near the right minimum point of $U(x)$ and the function value of this plateau directly gives the transition time [6].

Let $x(t) \in [0, 1]$ be the solution of the one-dimensional stochastic differential equation

$$\begin{aligned} \frac{dx}{dt} &= -x + \frac{1}{2} + \beta\, x(1-x)\, \xi(t) \\ &= f(x) + \sigma(x)\, \xi(t)\,, \end{aligned} \tag{1.48}$$

with multiplicative (state-dependent) Gaussian white noise (Section 1.2.3). The parameter $\beta > 0$ scales the state-dependent noise strength $\sigma(x) = \beta x(1-x)$, which vanishes at the boundaries of the interval $[0, 1]$. If $x(t)$ is initially in the interval, it stays there for all future times [110]. The deterministic part of the dynamics (1.48), given by the vector field $f(x) = -x + 1/2$, has a single stable, globally attractive fixed point at $\bar{x} = 1/2$. For weak noise, we therefore expect that $x(t)$ will be attracted to this fixed point, exhibiting only small fluctuations around it in the stationary state. In this situation, the stationary probability density $p_{\mathrm{st}}(x)$ of $x(t)$ will be unimodal with a peak at $\bar{x} = 1/2$. This is indeed the case if the parameter β, which controls the noise strength, is small enough. However, when β is increased above a critical value, the stationary probability density $p_{\mathrm{st}}(x)$ undergoes a qualitative change from unimodal to bimodal, indicating a noise-induced transition to bistability [103, 105, 110].

This qualitative change can be seen using an explicit expression for $p_{\mathrm{st}}(x)$ [110]. The normalized time-invariant solution of the Fokker-Planck equation (1.31) associated to Eq. (1.48), interpreted as an Itô equation, yields the stationary probability density

$$\begin{aligned} p_{\mathrm{st}}(x) &= \frac{\tilde{A}}{\sigma^2(x)} \exp\left(\int_{x_0}^{x} \frac{2f(x')}{\sigma^2(x')}\, dx' \right) \\ &= \frac{\tilde{A}}{\beta^2 \left(x(1-x)\right)^2} \exp\left(\frac{1}{\beta^2} \int_{x_0}^{x} \frac{1-2x'}{\left(x'(1-x')\right)^2}\, dx' \right) \\ &= A\, \exp\left(-\frac{1}{\beta^2\, x(1-x)} - 2\log\left(x(1-x)\right) \right) \end{aligned} \tag{1.49}$$

for $x \in [0, 1]$; the normalization constant $A > 0$ absorbs the integration constant in the exponent. We note that $p_{\mathrm{st}}(0) = p_{\mathrm{st}}(1) = 0$. The expression (1.49) is similar to Eq. (1.45), but the function in the exponent includes the effect of the state-dependent noise strength $\sigma(x)$ on the stochastic dynamics (1.48). As mentioned above, local extrema of the function in the exponent give rise to local extrema of the probability density $p_{\mathrm{st}}(x)$. The derivative of this function provides the following necessary condition for a local extremum at $x \in (0, 1)$:

$$\frac{1-2x}{x(1-x)} \left(\frac{-1}{\beta^2\, x(1-x)} + 2 \right) = 0\,.$$

The fixed point $\bar{x} = 1/2$ of the deterministic dynamics yields a local extremum for all values of the parameter β, which can also be expected due to the symmetry $p_{\mathrm{st}}(1-x) = p_{\mathrm{st}}(x)$. If $0 < \beta < \sqrt{2}$, $\bar{x} = 1/2$ is the only extremum point in the open interval $(0, 1)$, and $p_{\mathrm{st}}(x)$ is thus unimodal with a maximum at that point. However, if β exceeds the critical value $\beta_0 = \sqrt{2}$, two additional extrema of $p_{\mathrm{st}}(x)$ appear at $\bar{x} = (1 \pm \sqrt{1 - 2/\beta^2})/2$. They are local maxima of the same height, separated by a local minimum at $\bar{x} = 1/2$. Therefore, $p_{\mathrm{st}}(x)$ becomes bimodal, with peaks located away from the deterministically stable fixed point. Interpreting the two emerging peaks as coexisting metastable states,

we can say that these states are generated by the noise.[17] As β is increased further, the peaks move closer to the interval boundaries and become more pronounced. This is in contrast to the effect of additive noise on bistable deterministic dynamics, see Eq. (1.44), where stronger noise leads to broader, less pronounced peaks. Taken together, increasing the parameter β, which scales the state-dependent noise in the stochastic dynamics (1.48), above the critical value $\beta_0 = \sqrt{2}$ yields a change in the number of local maxima of the stationary probability density (1.49).[18]

The underlying mechanism in this example is that the inhomogeneous noise strength $\sigma(x) = \beta x(1-x)$ is largest at $\bar{x} = 1/2$ and decreases towards the interval boundaries. Stronger noise leads to faster undirected random motion away from the current state (see also Section 1.2.3). Consequently, in the absence of directed motion, trajectories spend longer time in states with weaker noise, giving rise to a higher stationary probability density there. If the noise is inhomogeneous enough, this "repulsive" effect can even outweigh the attraction to $\bar{x} = 1/2$ due to the deterministic part $f(x) = -x + 1/2$ of the dynamics. This is achieved by increasing the parameter β above a critical value.

Such qualitative changes in the stationary probability density caused by stronger noise are called noise-induced transitions [103, 110]. They should not be confused with noise-activated transitions (switching) between metastable states, which are sometimes also termed "noise-induced", see e.g. [101]. A noise-induced transition can be considered as a kind of bifurcation in a stochastic dynamical system [89, 103, 110]. More generally, a qualitative change in the stationary probability density that occurs when a parameter is varied is called a phenomenological stochastic bifurcation, or P-bifurcation [89]. Since it relies solely on the stationary density, this concept of a stochastic bifurcation is rather static. An alternative concept is a dynamical stochastic bifurcation, or D-bifurcation [89]; it is based on random dynamical systems theory (see Section 1.2.4) and focuses on qualitative changes of the dynamical stability.

1.3 Structure of the thesis

In the present Chapter 1, we have introduced the mathematical models and fundamental concepts that will be used throughout the thesis to study the collective dynamics and self-organization of recurrent neural networks. The main body of the thesis, starting with Chapter 2, consists of four chapters. Each chapter begins with a summary, in which we outline the scientific questions to be addressed and highlight the key results obtained. In Chapter 2, we investigate how a time-varying input, modeled as a generic stochastic drive, shapes the transition to chaos and the memory of past inputs in large rate-based neural networks with random connectivity. In particular, we develop the dynamic mean-field theory for stochastically driven networks, including derivations of the maximum Lyapunov exponent and the memory capacity. Turning to spiking neuron models, in Chapter 3 we study the collective dynamics in inhibitory balanced networks that contain two types of integrate-and-fire neurons with different intrinsic membrane potential dynamics. The spectrum of Lyapunov exponents reveals that inserting a single neuron that accelerates towards the spike threshold into the network causes a transition from stable to chaotic dynamics. In Chapter 4, we show that developing networks of stochastically spiking neurons can self-organize through

[17] To assess the metastability of the two peaks, one should additionally examine the rate of transitions between them, as we did before in the case of additive noise using Eq. (1.46). However, since the peaks emerge only above a critical noise strength, the weak noise limit is not relevant here.

[18] If we interpret the stochastic differential equation (1.48) as a Stratonovich equation, we observe the same behavior with the critical parameter value $\beta_0 = 2$ [110].

homeostatic structural plasticity into a state close to the transition from stable to unstable recurrent excitation. We analytically characterize the critical avalanches of spikes that arise in this state. Finally, Chapter 5 presents neural network models in which synaptic plasticity during ongoing spiking activity leads to drifting assemblies of strongly interconnected neurons; at the same time, input and output neurons compensate for the drift. We describe the gradual exchange of the assembly neurons underlying the drift as noise-activated transitions between metastable states. Appendices A, B, and C, which supplement Chapters 2, 3, and 5, respectively, provide further details and additional derivations. In the final Chapter 6, we give an overall summary and conclusion of the thesis.

CHAPTER 2

Transition to chaos and optimal memory in driven random neural networks

This chapter is based on the following published article [113]:

J. Schuecker*, **S. Goedeke***, and M. Helias
(* equal contribution)
Optimal sequence memory in driven random networks
Phys. Rev. X **8** (2018) 041029 DOI: 10.1103/PhysRevX.8.041029

We begin our investigations in this thesis by studying a classical transition in rate-based recurrent neural network models. The autonomous collective dynamics in large randomly coupled networks of neuron-like units exhibits a transition from a stable fixed point state to chaotic fluctuations at a critical coupling strength. This transition is particularly relevant because close to the transition, near the "edge of chaos", the rich intrinsic dynamics may serve as a reservoir for information processing and for modeling activity patterns in the brain. Seminal previous work, employing dynamic mean-field theory, analytically described the activity statistics and the transition to chaos in autonomous networks. However, neural networks often do not operate in isolation but are driven by external inputs, which can also vary in time. Because of its interplay with the nonlinear recurrent network dynamics, the input will affect the dynamical state and thereby the network function. In this chapter, we investigate the effect of a time-varying input modeled as a generic stochastic drive on the onset of chaos and the consequences for information processing. For the latter, we evaluate the memory of past inputs in the present state of the driven network.

The randomly distributed synaptic coupling weights between the neurons in the network can lead to variability of the total recurrent inputs to different neurons that does not average out in the limit of large network size. Dynamic mean-field theory characterizes the resulting activity fluctuations in terms of their self-consistent temporal autocorrelations. We derive and solve the dynamic mean-field equations for stochastically driven random neural networks. The derivation is based on a path (or functional) integral formulation of the stochastic network dynamics. The solution shows how the activity statistics are shaped by the stochastic drive. The time-dependent input in the form of irregular noise gives rise to a

nonzero and decaying activity autocorrelation function for any coupling strength. Hence, in contrast to the autonomous case, the emergence of activity fluctuations with decaying autocorrelations cannot be used as an indicator for the transition to chaos.

A defining property of chaos in deterministic systems is that the dynamics exhibits sensitive dependence on the initial conditions, leading to long-term unpredictability. In the autonomous random neural network model, chaos gives rise to non-periodic, potentially rich, activity patterns. To identify the transition to chaos in the driven, nonautonomous network, we quantify the stability or instability of the dynamics by the maximum Lyapunov exponent, which measures the asymptotic average rate of exponential convergence or divergence of trajectories with nearby initial conditions. For this, we consider two identical copies of the network that are driven by the same realization of the stochastic input, sometimes referred to as "frozen" noise. A negative Lyapunov exponent then implies that at least locally the copies forget their initial conditions and asymptotically converge (synchronize) to a time-dependent activity trajectory determined by the input; such response reliability may be important for network function. In contrast, a positive Lyapunov exponent indicates that tiny differences in the initial conditions of the two copies grow exponentially; this instability of the microscopic dynamics ultimately leads to differing activities in the network copies, even though they are driven by the same input.

To obtain the Lyapunov exponent in the limit of large network size, we develop the self-consistent dynamic mean-field theory for the two copies, which are known as replicas in the theory of disordered systems. Linearizing the resulting coupled dynamic mean-field equations about the fully synchronized solution yields the evolution of the infinitesimal distance between the activity trajectories of the two copies. The asymptotic growth rate of this distance determines the Lyapunov exponent. We find an exact condition for the transition from stable to chaotic dynamics in networks driven by Gaussian white-noise input. The time-varying input decreases the Lyapunov exponent and thus suppresses chaos by shifting the transition to a larger coupling strength that depends on the input amplitude. Beyond the previously known static effect, whereby the input can reduce the effective slope of the nonlinear transfer function in the network, we uncover a dynamic mechanism for suppressing chaos that is related to how fast the input decorrelates the network state: The resulting decorrelation of the dynamics' unstable directions can prevent the growth of infinitesimal perturbations. This dynamic mechanism gives rise to a regime in which the local dynamics is expansive at every point in time, while the Lyapunov exponent is still negative.

How does the interplay of the input with the recurrent network dynamics relate to functional properties? In autonomous networks, the emerging chaotic fluctuations become arbitrarily slow at the transition, which is considered functionally important. Here, we demonstrate that for white-noise input this divergence of the intrinsic network timescale is rather strongly suppressed. As a concrete information-processing capability, we study the above-mentioned dynamic memory of a past time-varying input signal. Employing again dynamic mean-field theory for the driven random network, we analytically determine the corresponding memory capacity. The contribution of the recurrent network dynamics to this capacity is optimal in the locally expansive but asymptotically stable, i.e. nonchaotic, regime. Our results therefore indicate that this regime can be functionally beneficial.

The following sections of this chapter contain the published manuscript with only minor clarifications and changes to adapt the contents to the thesis. Details of the derivations and calculations are presented in the Appendix A belonging to this chapter. I contributed with equal share to all parts of the research and the writing of the article. Of particular importance was the synthesis of concepts from dynamical systems theory with statistical physics approaches such as taking the disorder average in the path integral representation of the network dynamics.

Starting from a dynamical systems perspective, I derived the linear partial differential equation that describes the evolution of the infinitesimal distance between the trajectories of two network copies; it holds both for autonomous and for driven random networks. In a stationary state, the equation can be reduced to a Schrödinger eigenvalue equation, whose lowest eigenvalue determines the maximum Lyapunov exponent. For autonomous dynamics, this Schrödinger equation was stated in previous work, but its derivation remained opaque. The dynamic mean-field theoretical approach to the Lyapunov exponent developed here provides a transparent derivation and yields further insight, making it an important result of the article. By analyzing the Schrödinger equation, we concertedly discovered the exact condition for the transition to chaos in the presence of white-noise input. Furthermore, I led the implementation of a standard algorithm for measuring the Lyapunov exponent in simulations of large but finite networks, which allowed us to confirm the prediction from dynamic mean-field theory.

The obtained analytical expression for the memory capacity is based on a linear partial differential equation as well. I derived this equation and worked out the calculations leading to the memory capacity. The article's analytical results are of particular importance because they allowed us to determine the phase diagram of the different dynamical regimes and relate the memory capacity to it.

2.1 Introduction

Large random networks of neuron-like units can exhibit collective chaotic dynamics [26, 29, 65, 91]. Their information processing capabilities have been a focus in neuroscience [114] and in machine learning [115], and they show optimal performance close to the transition to chaos [30, 116, 117]. Because of its rich chaotic dynamics, the seminal random neural network model by Sompolinsky, Crisanti, and Sommers [29] underlies models for various activity patterns observed in working memory tasks [118–121], motor control [122], and perceptual decision making [123]. The interplay between a time-dependent input and the dynamical state of the network, however, is poorly understood, notwithstanding consequences for information processing.

In the absence of time-varying inputs, the network dynamics is autonomous. Networks of randomly coupled model neurons which interact through analog rate-like variables display a transition from a fixed point to chaotic fluctuations at a critical coupling strength [29], illustrated in Fig. 2.1(a). The transition is well understood by dynamic mean-field theory, originally developed for spin glasses [124, 125]. The onset of chaos is equivalent to the emergence of a nonzero, decaying autocorrelation function, whose decay time diverges at the transition. This equivalence has been used in several subsequent studies [24, 126, 127]. Furthermore, for autonomous systems, a tight relationship exists to random matrix theory [69, 128]: The transition from a fixed point to the chaotic state happens precisely when the fixed point becomes linearly unstable [29], which identifies the spectral radius of the random coupling matrix as the parameter controlling the transition in such systems.

In stochastically driven systems, a direct relation between a decaying autocorrelation function, chaotic dynamics, and linear stability does not exist: Stochastic drive *per se* decorrelates the network activity even if the dynamics is stable [Fig. 2.1(b)], so a decaying autocorrelation function does not necessarily indicate chaos. The stochastic drive, furthermore, causes perpetual fluctuations in a stable regime as well. Therefore, a transition to chaos, if it exists at all, must be of a qualitatively different kind than the transition from the silent fixed point in the autonomous case. Time-dependent driving has indeed been found to stabilize neural network dynamics [126, 131]. However, the mechanism is only understood for low-dimensional systems in the context of chaos synchronization by noise [132], in networks driven

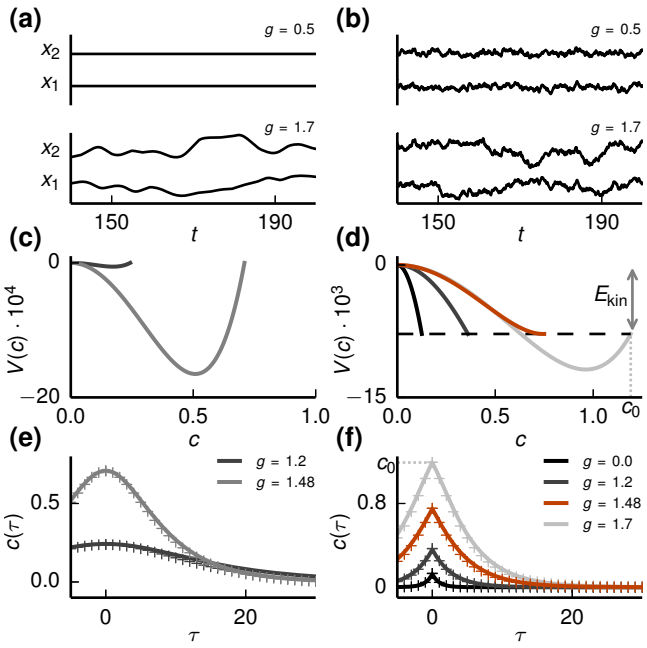

Figure 2.1: Activity statistics of autonomous and driven networks. The autonomous case is for $\sigma = 0$ (left column), and the stochastically driven case is for $\sigma = 0.35$ (right column). (a) and (b): Simulated trajectories of two example neurons for subcritical $g = 0.5$ (upper part of vertical axis) and supercritical $g = 1.7$ (lower part of vertical axis). (c) and (d): Classical potential (2.10) with self-consistently determined variance c_0 following from energy conservation (2.11) for different coupling strengths g [corresponding legends in panels (e) and (f)]. The dashed horizontal line indicates negative initial kinetic energy $E_{\text{kin}} = \sigma^4/2$. In the driven case, the critical coupling $g_c = 1.48$ from Eq. (2.20) is shown in red. (e) and (f): Self-consistent autocorrelation function (solid curve) compared to simulations (crosses). The variance (peak height) c_0 corresponds to the largest value of c at which the potential [panels (c) and (d)] is defined, indicated for $g = 1.7$ with gray dotted lines in panels (d) and (f). The network size in the simulations is $N = 10\,000$. Code for the dynamic mean-field theory is available as part of [129, 130].

by periodic signals [126], and in networks with discrete-time dynamics [131]. In the latter model, the effect of the stochastic input on the transition to chaos is completely captured by its influence on the spectral radius of the Jacobian matrix. Its single neuron dynamics, moreover, does not possess nontrivial temporal correlations. But these temporal correlations are indeed essential for the transition to chaos and for information processing in continuous-time systems, as we show here.

Realistic continuous-time random network models can generate complex but controlled responses to input [30] that resemble activity patterns observed in the motor cortex. In particular, the dynamical state of the network plays a crucial role during the involved learning process. However, the effect of the input on the dynamical state has remained obscure.

To investigate the generic influence of external input on the collective network dynamics, we include additive white noise in the seminal model by Sompolinsky, Crisanti, and Sommers [29] and develop the dynamic mean-field theory for the resulting stochastic continuous-time dynamics. In contrast to the original work, here we reformulate the problem in terms of the generating functional formalism for stochastic differential equations [133–137]. The application of the auxiliary field formulation known from large N field theory [138] then allows us to derive the mean-field equations by a saddle-point approximation (Section 2.2). In Section 2.3 we show that the autocorrelation function is formally

identical to the motion of a classical particle in a potential as in the original model [29]. In our model, the noise amounts to an initial kinetic energy. In Section 2.4 we then determine the maximum Lyapunov exponent [61] by considering two copies of the system with different initial conditions [139] in a replica calculation. Our main result is an exact condition for the transition from stable to chaotic dynamics. We find that the input suppresses chaos significantly more strongly than expected from time-local linear stability, the criterion valid in discrete-time networks [131]. This observation is explained by a dynamic effect: The decrease of the maximum Lyapunov exponent is related to the sharpening of the autocorrelation function by the stochastic drive. The regime in the phase diagram between local instability, as indicated by the spectral radius of the Jacobian, and transition to chaos, corresponding to a positive maximum Lyapunov exponent, constitutes an as yet unreported dynamical regime that combines locally expansive dynamics with asymptotic stability. Moreover, in contrast to the autonomous case, the decay time of the autocorrelation function does not diverge at the transition. Its peak is strongly reduced by the input and occurs slightly above the critical coupling strength.

We investigate the contribution of the dynamic effect to chaos suppression in Section 2.5 by considering static and colored noise inputs. Faster varying inputs suppress chaos more strongly if we keep the input's static effect on the spectral radius of the Jacobian fixed. In Section 2.6 we extend our findings to biologically more realistic random networks with nonvanishing mean coupling and non-negative transfer function, in particular, the firing-rate function of a spiking neuron model.

To study information-processing capabilities, in Section 2.7 we evaluate the capacity to reconstruct a past input signal by a linear readout of the present state, the so-called memory curve [140]. Dynamic mean-field theory and a replica calculation lead to a closed-form expression for the memory curve. We find that the memory capacity peaks within the expansive, nonchaotic regime, indicating that locally expansive, while asymptotically stable dynamics is beneficial to store input sequences in the dynamics of the neural network.

2.2 Dynamic mean-field equation

We study the continuous-time dynamics of a random network of N neurons, whose activity states $x_i(t) \in \mathbb{R},\ i = 1, \ldots, N$, evolve according to the system of stochastic differential equations

$$\frac{dx_i}{dt} = -x_i + \sum_{j=1}^{N} J_{ij}\phi(x_j) + \xi_i(t)\,. \tag{2.1}$$

The J_{ij} are independent and identically Gaussian-distributed random coupling weights with zero mean and variance g^2/N, where the intensive gain parameter g controls the recurrent coupling strength or, equivalently, the weight heterogeneity of the network. We further exclude self-coupling, setting $J_{ii} = 0$. The time-varying inputs $\xi_i(t)$ are independent Gaussian white-noise processes with correlation functions $\langle \xi_i(t)\xi_j(s)\rangle = 2\sigma^2\delta_{ij}\delta(t-s)$. We choose the sigmoidal transfer function $\phi(x) = \tanh(x)$ so that without input, i.e. for $\sigma = 0$, the model agrees with the autonomous one studied in Ref. [29].

The dynamical system (2.1) contains two sources of randomness: the quenched disorder due to the random coupling weights and the temporally fluctuating drive. A particular realization of the random couplings J_{ij} defines a fixed network configuration, and its dynamical properties usually vary between different realizations. For large network size N, however, certain quantities are self-averaging, meaning that their values for a typical realization can be obtained by an average over network configurations,

i.e. the quenched disorder [141]. An important example is the population-averaged autocorrelation function.

Here, we derive a dynamic mean-field theory that describes the statistical properties of the system under the joint distribution of disorder, noise, and possibly random initial conditions in the limit of large network size $N \to \infty$. The theory can be derived via a heuristic "local chaos" assumption [142] or by using a generating functional formulation [125, 143]. We follow the latter approach because it casts the problem into the established language of statistical field theory for which a wealth of approximation techniques is available [144]. A mathematically rigorous proof uses large deviation techniques [145]. The general idea is that for large network size N, the local recurrent input $\sum_{j=1}^{N} J_{ij}\phi(x_j)$ in Eq. (2.1) approaches a Gaussian process with self-consistently determined statistics.

We interpret the stochastic differential equations in the Itô convention [6] and formulate the problem (2.1) in terms of a moment-generating functional Z. Using the Martin–Siggia–Rose–de Dominicis–Janssen path integral formalism [133, 136, 146, 147], we obtain

$$Z[\mathbf{l}](\mathbf{J}) = \int \mathcal{D}\mathbf{x} \int \mathcal{D}\tilde{\mathbf{x}} \, \exp\left(S_0[\mathbf{x}, \tilde{\mathbf{x}}] - \tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{J}\phi\left(\mathbf{x}\right) + \mathbf{l}^{\mathrm{T}}\mathbf{x}\right) \tag{2.2}$$

with

$$S_0[\mathbf{x}, \tilde{\mathbf{x}}] = \tilde{\mathbf{x}}^{\mathrm{T}}\left(\partial_t + 1\right)\mathbf{x} + \sigma^2\tilde{\mathbf{x}}^{\mathrm{T}}\tilde{\mathbf{x}}\,, \tag{2.3}$$

where the field $\mathbf{x} = \left(x_i(t), t \in \mathbb{R}\right)_{i=1}^{N}$ is the vector of trajectories (paths), $\mathbf{x}^{\mathrm{T}}\mathbf{y} = \sum_i \int dt\, x_i(t) y_i(t)$ denotes the scalar product in time and in neuron space, and $\mathbf{l} = \left(l_i(t),\, t \in \mathbb{R}\right)_{i=1}^{N}$ represents a source field. The response field $\tilde{\mathbf{x}} = \left(\tilde{x}_i(t),\, t \in \mathbb{R}\right)_{i=1}^{N}$ appears as a result of a Hubbard-Stratonovich transformation, representing, at each time point, a Dirac delta as $\delta(x) = (2\pi\mathrm{i})^{-1}\int_{a-\mathrm{i}\infty}^{a+\mathrm{i}\infty} d\tilde{x}\, \mathrm{e}^{\tilde{x}x}$ [137, 147]. The measures are defined as $\int \mathcal{D}\mathbf{x} = \lim_{M\to\infty}\Pi_{k=1}^{N}\Pi_{l=1}^{M}\int_{-\infty}^{\infty} dx_k^l$ and $\int \mathcal{D}\tilde{\mathbf{x}} = \lim_{M\to\infty}\Pi_{k=1}^{N}\Pi_{l=1}^{M}\int_{-\mathrm{i}\infty}^{\mathrm{i}\infty}(2\pi\mathrm{i})^{-1} d\tilde{x}_k^l$, with the subscript k denoting the kth unit and the superscript l denoting the lth time slice. The action S_0 in Eq. (2.3) contains all single-unit properties whereas the remaining coupling term $-\tilde{\mathbf{x}}^{\mathrm{T}}\mathbf{J}\phi\left(\mathbf{x}\right)$ is written explicitly in Eq. (2.2).

Assuming that the dynamics is self-averaging, we average over the quenched disorder in the coupling matrix $\mathbf{J} = (J_{ij})_{i,j=1}^{N}$ and perform a saddle-point approximation (Appendix A.1). The resulting functional factorizes into N terms

$$\bar{Z}^* \propto \int \mathcal{D}x \int \mathcal{D}\tilde{x} \, \exp\left(S_0[x, \tilde{x}] + \frac{g^2}{2}\tilde{x}^{\mathrm{T}} C_{\phi(x)\phi(x)}\tilde{x}\right) \tag{2.4}$$

with $C_{\phi(x)\phi(x)}(t, s) = \langle\phi(x(t))\phi(x(s))\rangle$ denoting the average autocorrelation function of the nonlinearly transformed activity of the units, see Eq. (A.10), and

$$\tilde{x}^{\mathrm{T}} C_{\phi(x)\phi(x)}\tilde{x} = \iint dt\, ds\, \tilde{x}(t)\, C_{\phi(x)\phi(x)}(t, s)\, \tilde{x}(s)\,.$$

The factorization reduces the network to N noninteracting units, each on a background of an independent Gaussian noise with identical, self-consistently determined statistics. At this level of approximation, the problem is hence equivalent to a single-unit system. The effective equation of motion corresponding to

this system follows from Eq. (2.4) (Appendix A.1):

$$\frac{dx}{dt} = -x + \eta(t) + \xi(t) \,. \tag{2.5}$$

Without the additional input $\xi(t)$, Eq. (2.5) is the starting point of the analysis in Ref. [29]. Here, $\xi(t)$ is a Gaussian white-noise process as in Eq. (2.1), independent of $\eta(t)$. The centered Gaussian process $\eta(t)$ is fully specified by its autocorrelation function

$$\langle \eta(t)\eta(s) \rangle = g^2 C_{\phi(x)\phi(x)}(t,s) \,. \tag{2.6}$$

2.3 Effective equation of motion of the autocorrelation function

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

$$\ddot{c} = \frac{d^2c}{d\tau^2} = c - g^2 f_\phi(c, c_0) - 2\sigma^2 \delta(\tau) \tag{2.7}$$

with $c_0 = c(0)$. The Dirac delta inhomogeneity originates from the white-noise autocorrelation function of the time-varying input. Without the delta inhomogeneity, Eq. (2.7) has been derived in Ref. [29], while the first derivation including the inhomogeneity, to our knowledge, has been presented by Cabana and Touboul [145], using the mathematically rigorous framework of large deviations. The same inhomogeneity arises from Poisson spiking noise with $2\sigma^2 = g^2 r$ [73], where r is the population-averaged firing rate. One may obtain the same result with a direct dynamic mean-field assumption [148]. In Eq. (2.7) we write $f_\phi(c(\tau), c_0) = C_{\phi(x)\phi(x)}(t+\tau, t)$, introducing the notation

$$f_u(c, c_0) = \iint u\left(\sqrt{c_0 - \tfrac{c^2}{c_0}}\, z_1 + \tfrac{c}{\sqrt{c_0}}\, z_2 \right) u\left(\sqrt{c_0}\, z_2 \right) \, Dz_1 Dz_2 \tag{2.8}$$

for an arbitrary function $u(x)$, where $Dz_i = \exp(-z_i^2/2)/\sqrt{2\pi}\, dz_i$ denotes the standard Gaussian integration measure. This representation holds since $x(t)$ is itself a Gaussian process. We note that Eq. (2.8) reduces to one-dimensional integrals for $f_u(c_0, c_0) = \langle u(\sqrt{c_0} z_1)^2 \rangle$ and $f_u(0, c_0) = \langle u(\sqrt{c_0} z_1) \rangle^2$.

Following the approach in Ref. [29], we formulate Eq. (2.7) as the one-dimensional motion of a classical particle in a potential:

$$\ddot{c} = -V'(c) - 2\sigma^2 \delta(\tau) \,, \tag{2.9}$$

where we define

$$V(c) = V(c; c_0) = -\frac{1}{2}c^2 + g^2 f_\Phi(c, c_0) - g^2 f_\Phi(0, c_0) \tag{2.10}$$

with $\Phi(x) = \int_0^x \phi(y)\, dy$ and $\partial/\partial c\, f_\Phi(c, c_0) = f_\phi(c, c_0)$ following from Price's theorem [149, 150]. The autocorrelation $c(\tau)$ here plays the role of the position of the particle and the time lag τ the role of time. The potential (2.10) depends on the initial value c_0, which has to be determined self-consistently.

We obtain c_0 from classical energy conservation $\dot{c}^2/2 + V(c) = \text{constant}$. Considering $\tau \geq 0$ and the symmetry of $c(\tau)$, the delta inhomogeneity in Eq. (2.9) amounts to an initial velocity $\dot{c}(0+) = -\sigma^2$ and thus to the kinetic energy $\dot{c}^2(0+)/2 = \sigma^4/2$. Since $|c(\tau)| \leq c_0$ is required for all time lags τ, the solution $c(\tau)$ and its first derivative must approach zero as $\tau \to \infty$. Thus, we obtain the self-consistency condition for c_0 as

$$\frac{1}{2}\sigma^4 + V(c_0; c_0) = V(0; c_0) = 0\,. \tag{2.11}$$

For the autonomous case, Figs. 2.1(c) and 2.1(e) show the resulting potential and the corresponding self-consistent autocorrelation function $c(\tau)$ in the chaotic regime. Approaching the transition from above, $g \to g_c = 1$, the amplitude c_0 vanishes and the asymptotic decay time of $c(\tau)$ given by $\tau_\infty = 1/\sqrt{1 - g^2\langle\phi'(x)\rangle^2}$ diverges [29]. This picture breaks down in the stochastically driven case [Figs. 2.1(d) and 2.1(f)], where c_0 is always nonzero and $c(\tau)$ decays with finite timescale and has a kink at zero. The mean-field prediction is in excellent agreement with the population-averaged autocorrelation function obtained from numerical simulations of one network instance showing that the self-averaging property is fulfilled. In the next section, we derive a condition for the transition from stable to chaotic dynamics in the presence of the time-varying input.

2.4 Effect of input on the transition to chaos

The maximum Lyapunov exponent quantifies how sensitively the dynamics depends on the initial conditions [61]. It measures the asymptotic growth rate of infinitesimal perturbations. For stochastic dynamics, the stability of the solution for a fixed realization of the noise or, equivalently, the stochastic input is also characterized by the maximum Lyapunov exponent: If it is negative, trajectories with nearby initial conditions converge to the same time-dependent solution determined by the input; the dynamics is stable. If it is positive, the distance between two initially arbitrary close trajectories grows exponentially in time; the dynamics exhibits sensitive dependence on initial conditions and is hence chaotic. The theory of random dynamical systems makes this more precise; a brief overview is given in Ref. [91].

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

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

We now follow an idea by Derrida and Pomeau [139] and exploit the self-averaging property of population-averaged correlation functions, i.e, $(1/N)\sum_{i=1}^N x_i^\alpha(t)x_i^\beta(s) \approx c^{\alpha\beta}(t,s)$, where $c^{\alpha\beta}(t,s)$ denotes the mean-field correlation functions. We express the mean-squared Euclidean distance as

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

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

first consider

$$d(t,s) = c^{11}(t,s) + c^{22}(t,s) - c^{12}(t,s) - c^{21}(t,s) \tag{2.12}$$

with the obvious property $d(t) = d(t,t)$. We then determine the temporal evolution of $d(t,s)$ for infinitesimally perturbed initial conditions $||\mathbf{x}^1(0) - \mathbf{x}^2(0)|| = \epsilon$. To this end, it is again convenient to use a generating functional that captures the joint statistics of the two systems and, in addition, allows averaging over the quenched disorder [see also 144, Appendix 23, last remark]. The generating functional describing the two copies is defined analogously to the single system functional (2.2) as

$$\begin{aligned} Z[\mathbf{l}^1, \mathbf{l}^2](\mathbf{J}) = \left\{ \Pi_{\alpha=1}^2 \int \mathcal{D}\mathbf{x}^\alpha \int \mathcal{D}\tilde{\mathbf{x}}^\alpha \, \exp\left(S_0[\mathbf{x}^\alpha, \tilde{\mathbf{x}}^\alpha] - \tilde{\mathbf{x}}^{\alpha\mathrm{T}} \mathbf{J} \phi\left(\mathbf{x}^\alpha\right) + \mathbf{l}^{\alpha\mathrm{T}} \mathbf{x}^\alpha \right) \right\} \\ \times \exp\left(2\sigma^2 \tilde{\mathbf{x}}^{1\mathrm{T}} \tilde{\mathbf{x}}^2 \right) \end{aligned} \tag{2.13}$$

with the single-system "free action" $S_0[\mathbf{x}, \tilde{\mathbf{x}}]$ defined in Eq. (2.3). The last factor results from the identical external input in the two copies and effectively couples the two systems. We also note that the coupling matrix $\mathbf{J}$ is the same in both copies.

Averaging Eq. (2.13) over the quenched disorder of the random coupling matrix $\mathbf{J}$ and performing a saddle-point approximation, we obtain a pair of effective equations of motion (Appendix A.3),

$$\left(\frac{d}{dt} + 1 \right) x^\alpha(t) = \xi(t) + \eta^\alpha(t) \,, \quad \alpha \in \{1, 2\} \,, \tag{2.14}$$

together with a set of self-consistency equations for the correlations of the Gaussian noises $\eta^\alpha(t)$,

$$\langle \eta^\alpha(t) \, \eta^\beta(s) \rangle = g^2 \, \langle \phi(x^\alpha(t)) \phi(x^\beta(s)) \rangle \,. \tag{2.15}$$

Now, there are two terms that introduce correlations between the two copies. First, the common fluctuating drive $\xi(t)$ is injected into both systems. Second, the effective noises $\eta^\alpha(t)$ and $\eta^\beta(t)$ can be correlated between replicas, see Eq. (2.15), because the two systems have the same coupling matrix $\mathbf{J}$ in each realization. The origin of the latter coupling is hence of static nature.

The squared distance (2.12) between the two copies is given by the autocorrelations of the single systems and the cross-correlations between them. We consider the case where both copies are prepared with identical initial conditions and thus are fully synchronized: The cross-correlation $c^{12}(t,s)$ initially equals the autocorrelations $c^{11}(t,s)$, $c^{22}(t,s)$. The latter are identical to the single-system autocorrelation function because the marginal statistics of each subsystem is not affected by the mere presence of the other system. An increase of the squared distance $d(t)$, by Eq. (2.12), amounts to a decline of $c^{12}(t,s)$ away from its initial value $c(t-s)$. Here, c is the stationary autocorrelation, as we are interested in the Lyapunov exponent averaged over initial conditions drawn from the stationary distribution. To determine the growth rate for infinitesimal distances between the two copies, we consider $d(0) \propto \epsilon$ with $\epsilon \to 0$ and therefore expand the cross-correlation around its stationary solution $c^{12}(t,s) = c(t-s) + \epsilon \, k^{(1)}(t,s)$, which, to linear order, leads to an equation of motion for the deflection (Appendix A.3.1)

$$(\partial_t + 1)\,(\partial_s + 1)\, k^{(1)}(t,s) = g^2 f_{\phi'}\big(c(t-s), c_0\big) k^{(1)}(t,s) \tag{2.16}$$

with $d(t) = -2\epsilon \, k^{(1)}(t,t)$.

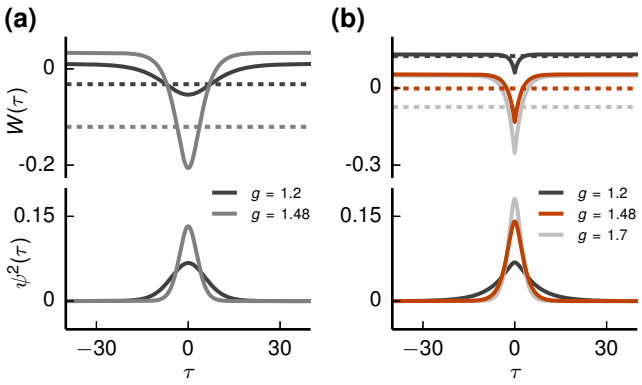

Figure 2.2: Ground state of the Schrödinger equation, which determines the maximum Lyapunov exponent. Upper part of vertical axis: Quantum potential $W(\tau)$ (solid curve) and ground-state energy E_0 (dashed line) for the autonomous case (a) and the driven case (b) for $\sigma = 0.35$. Lower part of vertical axis: Corresponding squared ground-state wave function. The parameters are the same as in Fig. 2.1 (with the driven case for $g = 0$ left out).

A separation ansatz in the coordinates $\tau = t - s$ and $T = t + s$ then yields an eigenvalue problem in the form of a time-independent Schrödinger equation [29, 73] (Appendix A.3.2),

$$\left[-\partial_\tau^2 + W(\tau)\right]\, \psi(\tau) = E\, \psi(\tau)\,, \tag{2.17}$$

where τ now plays the role of a spatial coordinate. Sompolinsky, Crisanti, and Sommers [29] state this Schrödinger equation for the Lyapunov exponent, but no details are provided. As in their case, the quantum potential $W(\tau) = -V''(c(\tau)) = 1 - g^2 f_{\phi'}(c(\tau), c_0)$ is the negative second derivative of the classical potential $V(c)$ evaluated along the self-consistent autocorrelation function $c(\tau)$. The ground-state energy E_0 of Eq. (2.17) determines the asymptotic growth rate of $k^{(1)}(t,t)$ as $t \to \infty$ and, hence, the maximum Lyapunov exponent via $\lambda_{\max} = -1 + \sqrt{1 - E_0}$. Therefore, the dynamics is predicted to become chaotic if $E_0 < 0$. The quantum potential, together with the solution for the ground-state energy and wave function, is shown in Fig. 2.2. The latter are obtained as solutions of a finite-difference discretization of Eq. (2.17).

In the autonomous case, a decaying autocorrelation function corresponds to a positive maximum Lyapunov exponent [29]. This follows from the observation that for $g > 1$ the derivative of the self-consistent autocorrelation function $\dot{c}(\tau)$ solves the Schrödinger equation with $E = 0$. But as $\dot{c}(\tau)$ is an eigenfunction with a single node, it cannot be the ground state, which must have zero nodes. The ground-state energy, which is necessarily lower, must therefore be negative, $E_0 < 0$. So the dynamics is chaotic, and $\lambda_{\max}$ crosses zero at $g = 1$ [Fig. 2.3(a)].

In the presence of a fluctuating drive, the maximum Lyapunov exponent becomes positive at a critical coupling strength $g_c > 1$; with increasing input amplitude, the transition shifts to larger values [Fig. 2.3(a)]. The mean-field prediction $\lambda_{\max} = -1 + \sqrt{1 - E_0}$ shows excellent agreement with the Lyapunov exponent obtained in simulations using a standard algorithm [61]. Since the ground-state energy E_0 must be larger than the minimum $W(0) = 1 - g^2\langle[\phi'(x)]^2\rangle$ of the quantum potential, an upper bound for $\lambda_{\max}$ is provided by $-1 + g\sqrt{\langle[\phi'(x)]^2\rangle}$, leading to a necessary condition

$$\rho = g\sqrt{\langle[\phi'(x)]^2\rangle} \geq 1 \tag{2.18}$$

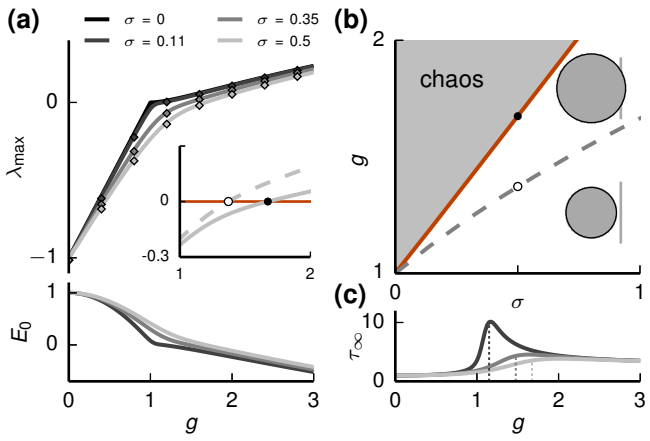

Figure 2.3: Transition to chaos. (a) Upper part of vertical axis: Maximum Lyapunov exponent $\lambda_{\max}$ as a function of the coupling strength g for different input amplitude levels. Mean-field prediction (solid curve) and simulation results (diamonds). We show a comparison to the upper bound $-1 + g\sqrt{\langle[\phi'(x)]^2\rangle}$ (dashed curve) for $\sigma = 0.5$ in the inset. Zero crossings are marked with dots. Lower part of vertical axis: Ground-state energy E_0 as a function of g. (b) Phase diagram with transition curve (solid red curve) obtained from Eq. (2.20) and necessary condition [Eq. (2.18) with equal sign] (gray dashed curve). Dots correspond to zero crossings in the inset in panel (a). We show the disk of eigenvalues of the Jacobian matrix in Eq. (2.19) for $\sigma = 0.8$ and $g = 1.25$ (lower) and $g = 2.0$ (upper) centered at -1 in the complex plane (gray dots). The radius $\rho = g\sqrt{\langle[\phi'(x)]^2\rangle}$ is obtained from random matrix theory (black circle). The vertical line is at zero. (c) Asymptotic decay time τ_∞ of the autocorrelation function. Vertical dashed lines mark the transition to chaos. The color code is the same as in panel (a). The network size in the simulations is $N = 5000$.

for chaotic dynamics. However, close to the transition, $\lambda_{\max}$ is clearly smaller than the upper bound, which is a good approximation only for small g [Fig. 2.3(a), inset]: The actual transition occurs at substantially larger coupling strengths. In contrast, for memoryless discrete-time dynamics, the necessary condition found here is also sufficient for the transition to chaos [131, Eq. 13].

The local linear stability of the dynamical system (2.1) is analyzed via the variational equation

$$\frac{d}{dt}y_i(t) = -y_i(t) + \sum_{j=1}^{N} J_{ij}\phi'(x_j(t))\, y_j(t)\,, \tag{2.19}$$

$i = 1, \ldots, N$, describing the temporal evolution of an infinitesimal deviation $y_i(t)$ about a reference trajectory $x_i(t)$. Interestingly, $\rho = g\sqrt{\langle[\phi'(x)]^2\rangle}$ in Eq. (2.18) equals also the radius of the disk formed by the eigenvalues of the Jacobian matrix in the variational equation (2.19) estimated by random matrix theory [69, 128]. In the following, we refer to ρ as the eigenvalue radius. Therefore, the dynamics is expected to become locally unstable if this radius exceeds unity, as shown in Fig. 2.3(b) displaying ρ and the eigenvalues at an arbitrary point in time. But even for the case with $\rho > 1$, the system is not necessarily chaotic. Hence, contrary to the autonomous case [29], the transition to chaos is not predicted by random matrix theory.

To derive an exact condition for the transition, we determine a ground state with vanishing energy $E_0 = 0$. As in the autonomous case, $\dot{c}(\tau)$ solves Eq. (2.17) for $E = 0$, except at $\tau = 0$ where it exhibits a jump, because $c(\tau)$ has a kink due to the white-noise input, see Eq. (2.7). By linearity, $|\dot{c}(\tau)|$ is a

continuous and symmetric solution with zero nodes. Therefore, if its derivative is continuous as well, requiring $\ddot{c}(0+) = 0$, it constitutes the ground state that we searched for. This is in contrast to the autonomous case, where $\dot{c}(\tau)$ corresponds to the first excited state. Consequently, with Eq. (2.7), we find the condition for the transition

$$g_c^2 \, f_\phi(c_0, c_0) - c_0 = 0\,, \tag{2.20}$$

in which c_0 is determined by the self-consistency condition (2.11) resulting in the transition curve (g_c, σ_c) in parameter space [Fig. 2.3(b)]. This reveals the relationship between the onset of chaos, the statistics of the random coupling matrix, and the input amplitude.

From Eq. (2.20), it follows that the system becomes chaotic precisely when the variance c_0 of a typical single unit equals the variance of its recurrent input from the network $g^2\langle\phi^2(x)\rangle$. At the transition, the classical self-consistent potential $V(c; c_0)$ has a horizontal tangent at c_0, while in the chaotic regime, a minimum emerges [Fig. 2.1(d)]. With Eq. (2.7) it follows that the curvature $\ddot{c}(0+)$ of the autocorrelation function at zero changes sign from positive to negative [Fig. 2.1(f)]. Close to the transition, a standard perturbative approach shows that $\lambda_{\max}$ is proportional to $g^2\langle\phi^2(x)\rangle - c_0$, indicating a self-stabilizing effect: Since both terms grow with g, the growth of their difference is attenuated, explaining why $\lambda_{\max}(g)$ bends down as the transition is approached [Fig. 2.3(a)].

In the autonomous case, the timescale of fluctuations diverges at the transition to chaos [29]. Here, we consider the effect of the input on the asymptotic decay time τ_∞ of the autocorrelation function [Fig. 2.3(c)]. For weak input amplitude, the decay time peaks at the transition, corresponding to the diverging timescale in the autonomous case. For larger input amplitudes, the peak is strongly reduced, and the maximum decay time is attained above the transition.

2.5 Static and dynamic suppression of chaos

The condition (2.20) predicts the transition at significantly larger coupling strengths than the necessary condition $\rho = g\sqrt{\langle[\phi'(x)]^2\rangle} \geq 1$ [Fig. 2.3(b)]. In this section, we show that chaos suppression can be decomposed into a static and a dynamic mechanism. We first consider the static effect, which can be fully attributed to the increase of the variance c_0 caused by the additional input. Because $\phi'(x)$ is maximal at the origin, increasing the variance c_0 reduces the averaged squared slope in $g^2\langle[\phi'(x)]^2\rangle$ with $x \sim \mathcal{N}(0, c_0)$, thereby stabilizing the dynamics. The same effect can be obtained if we replace the white-noise input in Eq. (2.1) by a static (constant in time) heterogeneous input $\xi_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma_{\text{q}}^2\right)$, called quenched noise in the following. In this case, the calculation of the saddle-point solution, analogous to Section 2.3, leads to the differential equation (Appendix A.4)

$$\ddot{c} = c - g^2 f_\phi(c, c_0) - \sigma_{\text{q}}^2 \tag{2.21}$$

for the stationary autocorrelation function $c(\tau)$.

A possible self-consistent solution of Eq. (2.21) is given by a constant autocorrelation function $c(\tau) = c_0$, which describes the variance c_0 of a heterogeneous fixed point of the network dynamics in the presence of heterogeneous static input. Such a fixed point loses stability when the eigenvalue radius of the Jacobian matrix $\rho = g\sqrt{\langle[\phi'(x)]^2\rangle}$ evaluated at the fixed point exceeds unity. This determines a critical coupling strength, which coincides with our necessary condition. Hence, we denote its value by g_{nec}.

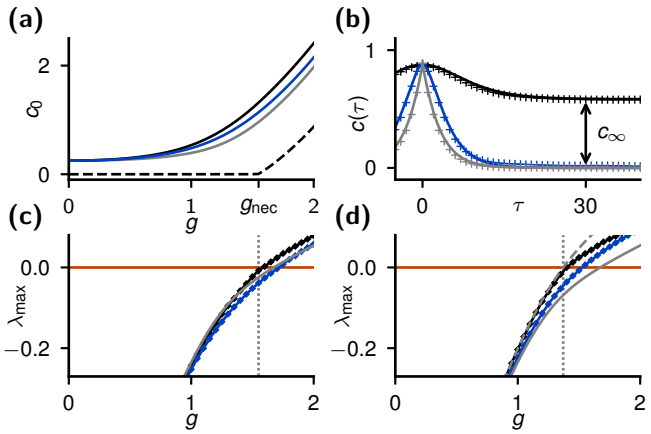

Figure 2.4: Effect of correlation time of the input on the transition to chaos. The cases of white-noise, colored noise with $\tau_n = 2$, and quenched noise input correspond to the curves shown in gray, blue, and black, respectively. (a) Self-consistent variance c_0 depending on the coupling strength g. At vanishing coupling, the inputs are chosen such that $c_0 = \sigma^2 = 0.25$ is identical in all three cases of white, colored, and quenched noise. Resulting dynamic contribution $c_0 - c_\infty$ (dashed black curve) of the variance in the case of quenched noise. (b) Self-consistent autocorrelation function (solid curves) compared to simulations (crosses) for $g = 1.5$. The amplitude of colored and quenched noise is adjusted to obtain the same variance (peak height) as for white noise. (c) Maximum Lyapunov exponent as a function of g for colored and quenched noise measured in simulations (diamonds connected by lines) compared to white noise case [identical to Fig. 2.3(a) with $\sigma = 0.5$]. (d) Same as in panel (c) but for identical variance c_0 for each value of g. Gray curves correspond to the Lyapunov exponent (solid) and the upper bound (dashed) from the white noise case. The necessary condition (2.18) is marked as a gray dotted vertical line in panels (c) and (d).

For $g > g_{\mathrm{nec}}$, a decaying autocorrelation function $c(\tau)$ emerges, which decays to a nonzero asymptotic value c_∞ (Appendix A.4). The function can be constructed by again writing Eq. (2.21) in the form of a classical motion in a potential: $\ddot{c} = -V_q'(c)$. The modified potential $V_q(c; c_0)$ is given by Eq. (2.10) plus the additional linear term $\sigma_q^2 c$ originating from the quenched noise. The asymptotic value requires $V_q'(c_\infty; c_0) = 0$, which we solve together with classical energy conservation, $V_q(c_\infty; c_0) = V_q(c_0; c_0)$, to self-consistently determine the initial value c_0 [Fig. 2.4(a)]. By integrating Eq. (2.21), we obtain the corresponding self-consistent autocorrelation function $c(\tau)$, which agrees with direct simulations [Fig. 2.4(b)]. The Schrödinger equation for the Lyapunov exponent (2.17) also holds in the case of quenched-noise input, since the explicit form of the noise term vanishes in the linearized equation (2.16) (Appendix A.4). As in the case without input [29], $\dot{c}(\tau)$ is an eigenfunction with a single node, and therefore, the dynamics corresponding to a decaying autocorrelation function must be chaotic; we hence conclude that the transition to chaos for quenched input occurs at g_{nec}, which is confirmed by the numerically estimated maximum Lyapunov exponent [Fig. 2.4(c)].

In the following, we consider how the variance c_0 depends on g for the different cases of white- and quenched-noise input. This is important because many of the relevant quantities like the eigenvalue radius ρ depend on c_0. For $g = 0$, we have $c_0 = \sigma_q^2$; so, choosing $\sigma_q^2 = \sigma^2$ leads to the same variance c_0 as for the white-noise input in Eq. (2.1). With increasing coupling strength g, however, the variance c_0 grows significantly faster for quenched-noise input [Fig. 2.4(a)]: The resulting internally generated fluctuations are stronger in this case, which will have important consequences for information-processing

capabilities as we investigate below in Section 2.7.

To isolate the dynamic effect on the transition to chaos, we need to keep the static effect identical for quenched- and white-noise input. To maintain the same variance c_0 also for nonzero couplings, we decrease the input variance σ_q^2 with increasing g. Because the variance c_0 is now identical in both cases, the eigenvalue radius is identical, too, as is the minimum of the quantum mechanical potential in Eq. (2.17).

Comparing the Lyapunov exponents in the two cases shows that the system with quenched input becomes chaotic at a smaller coupling than the system with white-noise input [Fig. 2.4(d)]. As expected, the point of the transition coincides with the critical value g_{nec}. Therefore, a dynamical regime between the loss of local stability and the transition to chaos does not exist in the case of quenched input.

The comparison between quenched- and white-noise input demonstrates two points: First, the variability of the input, be it static or dynamic, increases the variance c_0. It therefore reduces the eigenvalue radius ρ and increases the minimum of the quantum mechanical potential and thus suppresses chaos by a purely static effect that is identical in both cases. This mechanism corresponds to the one observed for discrete-time dynamics [117, 131]. Second, the remaining difference between the Lyapunov exponents in Fig. 2.4(d) for quenched input and white-noise input is a dynamic effect caused by the temporal variations of the input. It occurs for the continuous-time dynamics considered here and is also expected to appear in discrete-time networks with leaky single-unit dynamics.

To gain some insight into this dynamic mechanism, we return to the variational equation (2.19) describing the evolution of infinitesimal perturbations. Its fundamental solution can be written as a product of short-time propagator matrices:

$$\begin{aligned}\mathbf{y}(t) &= \exp\left(\int_0^t \left[-1 + \mathbf{J}\phi'\left(\mathbf{x}(s)\right)\right] ds\right) \mathbf{y}(0) \\ &= \lim_{h\to 0} \prod_{l=0}^{t/h-1} \left[1 - h + h\,\mathbf{J}\phi'(\mathbf{x}(lh))\right] \mathbf{y}(0),\end{aligned} \tag{2.22}$$

where the matrix $\mathbf{J}\phi'(\mathbf{x})$ has the entries $\left(\mathbf{J}\phi'(\mathbf{x})\right)_{ij} = J_{ij}\phi'(x_j)$ and the matrix product is time-ordered with earlier times on the right. Here, each factor has the same stability properties since the fraction of its eigenvalues with positive real parts stays approximately constant. However, the corresponding unstable directions may vary in time: They are determined by the matrix $\mathbf{J}\phi'(\mathbf{x}(lh))$ at the respective discretized time point lh. The rate of this variation depends on the timescale on which $\phi'(x(t))$ varies, which for fixed c_0 is monotonically related to the decay of the autocorrelation function $c(\tau)$. We conjecture that a faster variation of the unstable directions as indicated by a sharper autocorrelation function decreases the growth of generic perturbations. If this temporal variation is induced by time-varying input, it is even possible that perturbations shrink despite the presence of unstable directions at each point in time.

Is the dynamic mechanism also present for nonwhite time-varying input? How does the timescale of the autocorrelation function affect the transition? To answer these questions, we investigate the case of colored Gaussian noise inputs in the form of Ornstein-Uhlenbeck processes with finite correlation time (Appendix A.4). This setting constitutes an intermediate case between the two extremes discussed above. Again, we keep the variance c_0 at the same level as for white-noise input; hence, the static effect is identical in all three cases. The Lyapunov exponent for colored noise input lies between the two limiting cases [Fig. 2.4(d)], and the dynamic effect is still present as the transition to chaos occurs at a larger coupling strength compared to the case of quenched input. Indeed, Fig. 2.4(d) shows that for fixed c_0,

the Lyapunov exponent decreases if the input varies faster, which is in line with a sharper autocorrelation function [Fig. 2.4(b)]. Note that for colored noise input, the autocorrelation function does not exhibit a kink at the origin, so this is not a prerequisite for the dynamic effect.

We can also understand the dynamic effect on the basis of the Schrödinger equation (2.17). As noted above, the form of the input only implicitly appears by shaping the quantum potential $W(\tau) = -V''(c(\tau)) = 1 - g^2 f_{\phi'}(c(\tau), c_0)$, where $c(\tau)$ is the self-consistent autocorrelation function in the presence of the quenched or colored noise input. Increasing the correlation time of the input while keeping the variance c_0 constant leads to a slower decay of the autocorrelation function [Fig. 2.4(b)] and thereby to a wider quantum potential with identical minimum $W(0)$ because the latter only depends on c_0. Therefore, the ground-state energy decreases, and the Lyapunov exponent increases. The sign of the minimum also determines the local stability. So, for both cases of time-varying noise, white noise and colored noise, there is an offset between local stability and the transition to chaos [Fig. 2.4(d)]; an intermediate regime exists, and its extent decreases with increasing correlation time of the input. The term shaping the quantum potential is $f_{\phi'}(c(\tau), c_0) = \langle \phi'(x(t+\tau))\phi'(x(t))\rangle$, the autocorrelation function of $\phi'(x(t))$. This directly links the temporal variation of $\phi'(x(t))$ to the Lyapunov exponent and hence to the growth of generic perturbations as conjectured above.

In low-dimensional systems, the suppression of chaos by external fluctuations is understood: Noise forces the system to visit regions of the phase space with locally contracting dynamics more frequently [132], so contraction dominates expansion, in total yielding stable asymptotic behavior. This mechanism is similar to the static stabilization effect described above, where fluctuating drive causes the system to sample regions of the phase space with smaller eigenvalues of the Jacobian. The self-averaging high-dimensional system, however, has a constant eigenvalue radius over time, and hence, the dynamics is either locally contracting or locally expanding for all times. While the previous effects are explained by local stability, the dynamic suppression of chaos found here is a genuinely time-dependent mechanism, explained by the time evolution of the Jacobian in Eq. (2.22).

2.6 Nonvanishing mean coupling and non-negative transfer functions

The random coupling weights J_{ij} with vanishing mean and the odd transfer function $\phi(x) = \tanh(x)$ in Eq. (2.1) are certainly not biologically realistic. We considered this case before to explain the mechanism of chaos suppression by fluctuating drive in the simplest setting; it allowed us to directly compare our findings to the classical results by Sompolinsky, Crisanti, and Sommers [29]. In this section, we extend our analysis to the biologically more realistic case of random networks with nonvanishing mean coupling and with non-negative transfer functions.

2.6.1 Rectified-linear transfer function

We first examine the rectified-linear transfer function $\phi(x) = \max(x, 0)$, for which the transition to chaos has been investigated recently [24, 73, 148]. However, the effect of time-varying input in these networks has not been studied yet. We consider Gaussian-distributed coupling weights $J_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\bar{g}/N, g^2/N)$ or the sparse connectivity of a directed Erdős-Rényi random network with connection probability p and nonzero weight $J_0/\sqrt{N}$. Both cases lead to the same mean-field equations by identifying the parameters for the mean as $\bar{g} = \sqrt{N}J_0$ and the variance as $g^2 = J_0^2 p(1-p)$ [shown in Appendix A.5; see also

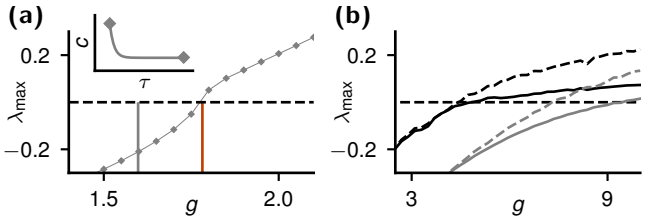

Figure 2.5: Transition to chaos for nonvanishing mean coupling and more realistic transfer functions. (a) Rectified-linear transfer function $\phi(x) = \max(x, 0)$: Maximum Lyapunov exponent as a function of g measured in simulations (diamonds connected by lines), transition criterion (2.20) (red vertical line) and necessary condition (2.18) (gray vertical line) for $\bar{g} = -1$ and $\sigma = 0.5$. The inset shows the autocorrelation function in simulations (solid curve) and c_0, c_∞ (diamonds) from the self-consistent solution for $g = 1.5$. (b) Leaky integrate-and-fire neuron transfer function: Maximum Lyapunov exponent (solid curve) and eigenvalue radius ρ minus one (dashed curve) as a function of g for the autonomous (black) and driven cases (gray). The network size is $N = 1000$ and $\bar{g} \in [-0.05, -30]$ varies linearly with g. Other parameters are $J_{\text{ext}} = 0.1$, $\nu_{\text{ext}} = 600$ Hz and $\sigma = 7.1$. Parameters of the transfer function [for definitions, see, e.g., [152, Eq. 27]] are as follows: $\tau_{\text{m}} = 20$ ms, synaptic time constant $\tau_{\text{s}} = 1$ ms, threshold $\theta = 1$, reset $V_{\text{r}} = 0$, and refractory period $\tau_{\text{r}} = 2$ ms. Simulations were carried out with NEST [153].

Ref. [73]]. The resulting equation for the stationary autocorrelation function $c(\tau)$ of $\delta x(t) = x(t) - \langle x \rangle$ is identical to Eq. (2.7), but with $f_{\phi(\circ+\langle x \rangle)}(c_0, c(\tau))$: We need to displace the transfer function by the nonzero mean $\langle x \rangle$ of x as $\phi(\circ + \langle x \rangle)$. The additional equation for the mean reads

$$\langle x \rangle = \bar{g} \, \langle \phi(x) \rangle = \bar{g} \int Dz \, \phi\left(\langle x \rangle + \sqrt{c_0} \, z \right) , \tag{2.23}$$

where Dz is again the standard Gaussian measure.

In structure, these equations resemble those that arise in the Sherrington-Kirkpatrick spin-glass model [151, Eqs. (2.17)-(2.18)], which is not surprising given that the physical problems are related and a similar mean-field approximation of the auxiliary fields has been employed. Importantly, $C_{\phi(x)\phi(x)}(\tau) = f_{\phi(\circ+\langle x \rangle)}(c_0, c(\tau))$ approaches a nonzero asymptotic value for $\tau \to \infty$ (Appendix A.5). It reflects the variability of the average input across neurons in the network that arises from the variability of the incoming connections and the nonzero mean output $\langle \phi(x) \rangle$ of the neurons. Qualitatively, it is the same as the addition of a quenched noise with variance $\sigma_q^2 = \lim_{\tau \to \infty} g^2 C_{\phi(x)\phi(x)}(\tau)$, which we have studied above (Section 2.5). Correspondingly, the autocorrelation function $c(\tau)$ also approaches a nonzero asymptotic value c_∞; we determine it simultaneously with the variance c_0 and the mean $\langle x \rangle$ by self-consistently solving Eq. (2.23) together with the condition for the asymptotic value $V'(c_\infty; c_0) = 0$ and energy conservation (2.11). These statistics are in excellent agreement with direct simulations [inset in Fig. 2.5(a)].

We now consider the effect of the nonvanishing mean coupling and the non-negative transfer function on the transition to chaos. The variability in the average input across neurons, as reflected by $c_\infty > 0$, leads to an additional static contribution to the suppression of chaos as in the quenched-noise case in Section 2.5. Does the dynamic suppression of chaos, as well as the intermediate regime between local instability and the transition to chaos, persist in this setting? The replica calculation proceeds completely analogously to the case of Gaussian-distributed couplings with vanishing mean (Appendix A.6). The

result is a pair of effective equations of the form

$$\left(\frac{d}{dt}+1\right)x^{\alpha}(t)=\bar{g}\,\langle\phi\,(x^{\alpha}(t))\rangle+\xi(t)+\eta^{\alpha}(t)\,, \tag{2.24}$$

$\alpha\in\{1,2\}$, together with a set of self-consistency equations for the correlations of the noises $\eta^{\alpha}(t)$,

$$\langle\eta^{\alpha}(t)\,\eta^{\beta}(s)\rangle=g^{2}\,\langle\phi(x^{\alpha}(t))\phi(x^{\beta}(s))\rangle. \tag{2.25}$$

Compared to Eq. (2.14), the only difference is the term $\bar{g}\,\langle\phi\,(x^{\alpha})\rangle$ in Eq. (2.24), which represents the nonzero mean input from the network. Importantly, the additional term does not introduce a coupling between the two replicas and therefore does not directly affect their correlation. It is determined independently for each replicon by a separate self-consistency equation of the form (2.23). Therefore, the mean coupling enters the calculation of the maximum Lyapunov exponent only indirectly by shaping the autocorrelation function: the shift of the transfer function $\phi(\circ+\langle x\rangle)$ in Eq. (2.7) and the nonzero asymptotic value c_{∞}. Furthermore, the same transition criterion (2.20) still applies; this prediction is confirmed by direct simulations in Fig. 2.5(a). As a consequence, the dynamic mechanism of chaos suppression is identical to the case of coupling with vanishing mean and the locally unstable but globally stable regime emerges [Fig. 2.5(a)].

2.6.2 Leaky integrate-and-fire neuron transfer function

Here, we demonstrate that the mechanism of dynamic chaos suppression is also present in random networks of biologically more realistic neuron models. To this end, we consider a nonlinear rate model [17, 148]:

$$\frac{d\nu_i}{dt}=-\nu_i+\phi(\mu_i,\sigma_i) \tag{2.26}$$

with

$$\mu_i=\tau_{\mathrm{m}}\sum_j J_{ij}\nu_j+\tau_{\mathrm{m}}J_{\mathrm{ext}}\nu_{\mathrm{ext}}+\xi_i(t)\,,$$
$$\sigma_i^2=\tau_{\mathrm{m}}\sum_j J_{ij}^2\nu_j+\tau_{\mathrm{m}}J_{\mathrm{ext}}^2\nu_{\mathrm{ext}}\,.$$

The transfer function ϕ in Eq. (2.26) is the firing rate function of the leaky integrate-and-fire (LIF) neuron model with synaptic filtering [154]. Each ν_i can be interpreted as the instantaneous firing rate of a LIF neuron (or an uncoupled population thereof), and Eq. (2.26) represents the rate dynamics of a random network of such neurons. More precisely, the firing rate response of these neurons is approximated by a first-order low-pass characteristic, qualitatively capturing the dominant behavior [15, 16].

As above, the coupling weights are Gaussian distributed, $J_{ij}\overset{\text{i.i.d.}}{\sim}\mathcal{N}(\bar{g}/N,g^2/N)$. We also include a time-homogeneous external input to the neurons with firing rate ν_{ext} and weight J_{ext}, since this is usually present in biological networks. Here, τ_{m} is the membrane time constant of the corresponding LIF neurons, and it sets the scale of the firing rates.

Our goal is to study the effect of the external time-varying drive $\xi_i(t)$ on the transition to chaos; it models an additional fluctuating input driving the membrane potential of the corresponding LIF neurons and therefore contributes to μ_i in Eq. (2.26). The Gaussian white-noise input in Eq. (2.1) is low-pass filtered before it enters the nonlinear transfer function. Consequently, for $\xi_i(t)$ we use independent Gaussian colored noise with correlation time $\tau_n = 1$ and standard deviation σ.

Using simulations, we show that both the transition to chaos and the dynamic suppression of chaos exist in these networks [Fig. 2.5(b)]. To assess the local linear stability, we simulate the network dynamics (2.26) and determine the eigenvalue radius ρ by numerically evaluating the eigenvalues of the Jacobian matrix $\left(\partial\phi_i/\partial v_j\right)$ [155, Eqs. A.2 and A.3]. The eigenvalue radius ρ minus one and the maximum Lyapunov exponent $\lambda_{\max}$ both become positive when the variability of the couplings g is increased [Fig. 2.5(b)]. To keep the transfer function from complete saturation, and thus the rates of the neurons in their dynamical range, we balance the increase in g by a more negative mean coupling strength $\bar{g}$. For an autonomous network, i.e. for $\xi_i \equiv 0$, the transition to chaos precisely coincides with the local linear instability. Below the transition, the Lyapunov exponent $\lambda_{\max}$ equals $\rho - 1$ as expected for a dynamical system at a stable fixed point. For the driven, nonautonomous case, however, a dynamical regime emerges where the eigenvalue radius ρ exceeds unity, while the Lyapunov exponent $\lambda_{\max}$ is still negative. Hence, the mechanism of dynamic chaos suppression also persists in this more realistic setup.

2.7 Information-processing capabilities

We expect the expansive, nonchaotic regime to be beneficial for information processing: The local instability of the network ensures sufficient initial amplification of an impinging external signal. The asymptotic stability is required for the driving signal to be uncorrupted by the amplification of small variations of the input; it is hence necessary to ensure generalization. In the following, we investigate these ideas quantitatively by considering the sequential memory capacity of the network.

We focus on the component $z(t) = 1/N \sum_{i=1}^{N} \xi_i(t)$ of the input that is received by all neurons with equal strength. In other words, the total input to each neuron is decomposed into the signal $z(t)$ and the remaining inputs $\xi_i(t) - z(t)$, which act as noise. In Appendix A.7 we treat the more general case of signals given by an arbitrary linear combination of the inputs. We consider the dynamic short-term or sequential memory defined as the capacity to reconstruct the signal $z(t)$ from the network state at a later time $t + \tau$ using a linear readout, $\sum_{i=1}^{K} w_i x_i(t + \tau)$, where $K \leq N$ is the number of readout neurons. The reconstruction capacity as a function of the time lag τ yields the memory curve $m(\tau) = 1 - \epsilon(\tau)$ [140], where $\epsilon(\tau)$ is the minimal relative mean-squared error between readout and signal. Alternatively, this measure quantifies the fidelity by which a sequence of past inputs can be reconstructed from the current network activity.

For optimal readout weights w_i that minimize $\epsilon(\tau)$, the memory curve is given by [140, 156]

$$m(\tau) = \frac{\langle \mathbf{x}(t+\tau) z(t) \rangle^{\mathrm{T}} \langle \mathbf{x}(t) \mathbf{x}(t)^{\mathrm{T}} \rangle^{-1} \langle \mathbf{x}(t+\tau) z(t) \rangle}{\langle z(t)^2 \rangle} . \tag{2.27}$$

Here, the vector $\mathbf{x}(t)$ contains only the states of the K readout neurons. We follow the approach by Toyoizumi and Abbott [117] and neglect the off-diagonal entries in $\langle \mathbf{x}\mathbf{x}^{\mathrm{T}} \rangle$, which is justifiable for a sparse readout with $K \ll N$. Additionally, for large N, the diagonal entries $\langle x_i^2 \rangle$ are given by their mean-field value c_0, which is identical for all units. Determining the memory curve (2.27) then amounts

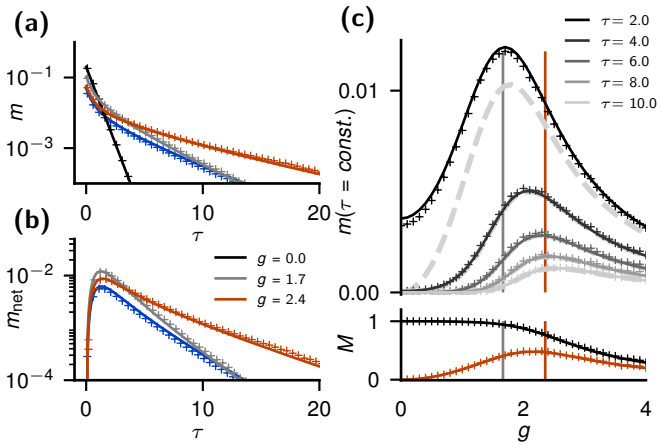

Figure 2.6: Sequential memory. Mean-field prediction (solid curves) and simulation results (crosses). (a) Memory (2.28) as a function of time lag τ between signal and readout for different coupling strengths encoded in color [legend in (b)] for $\sigma = 1$ and additionally for colored background noise with $\tau_n = 2$ for $g = 2.4$ and $\sigma = 1$ (blue). (b) Corresponding network contribution (2.29) to memory. (c) Upper part of vertical axis: Memory at different time lags τ as a function of coupling strength g. Network contribution to memory shown as dashed thick light-gray curves, which coincide with total memory curves for $\tau \geq 4$. The vertical gray line marks local instability (2.18), and the vertical red line marks the transition to chaos [cf. Fig. 2.3(b)]. Lower part of vertical axis: Memory capacity M (black), Eq. (2.30), and network contribution to memory capacity, M_{net} (red).

to computing the sum of squared correlation functions $\sum_i \langle x_i(t+\tau) z(t) \rangle^2$ between the signal and the network activity, which we obtain by a replica calculation (Appendix A.7). The technique to perform a replica calculation, as it is known from disordered systems [141], is closely tied to the field-theoretical formulation chosen here. The key idea is to express the correlation functions $\langle x_i z \rangle$ as a sum of response functions $\langle x_i \tilde{x}_j \rangle$; this is possible because z is Gaussian distributed. The calculation is similar to the derivation of the Schrödinger equation (Appendix A.3) with the difference, however, that the two replicas receive independent realizations of the inputs. The memory curve follows from a differential equation for the correlation between the two systems and is measured in units of the readout ratio K/N, Eq. (A.63):

$$m(\tau) = \frac{2\sigma^2}{c_0} \, \mathrm{e}^{-2\tau} \, I_0 \left[2g \langle \phi'(x) \rangle \, \tau \right] \, \Theta(\tau) \, d\tau \tag{2.28}$$

with the modified Bessel function of the first kind I_0 and the Heaviside step function $\Theta(\tau)$. The memory curve (2.28) has two contributions [see Eq. (A.63)]: memory due to the collective network dynamics and local memory due to the leaky integration of the single units. The latter effect is trivial and is reflected in the initial steep falloff as $e^{-2\tau}$ of the memory curves with time lag τ, independent of the coupling strength [Fig. 2.6(a)]. Its decay time is half the time constant of the neurons, which is set to unity in Eq. (2.1). With increasing coupling strength, the variance c_0 increases, so the memory curve (2.28) decreases at zero time lag. For time lags that are large compared to the single-unit time constant, the network contribution to the memory dominates. A nonvanishing memory capacity for longer time lags therefore results solely from the reverberation of the input through the recurrent network. The analytical results are in excellent agreement with direct simulations [Figs. 2.6(a) and 2.6(b)].

We isolate the interesting network memory by subtracting the single-unit contribution:

$$m_{\text{net}}(\tau) = m(\tau) - \frac{2\sigma^2}{c_0}\, \mathrm{e}^{-2\tau}\, \Theta(\tau)\, d\tau\,. \tag{2.29}$$

This quantity is particularly important in situations where the readout does not have access to the neurons receiving the signal. The network memory curve vanishes for the uncoupled case. We compare the performance at two different coupling strengths: g_{nec}, defined by $\rho = g_{\text{nec}}\sqrt{\langle[\phi'(x)]^2\rangle} = 1$ and corresponding to the onset of the local instability, and g_c, marking the onset of chaos [see Fig. 2.3(b)]. For short time lags, here for $\tau < 4$, the network memory curve is larger for g_{nec}, while for longer time lags, it is larger for g_c because of a slower decay of the memory curve. This behavior is confirmed by considering the memory capacity as a function of g, shown for different time lags [Fig. 2.6(c), upper panel]. For $\tau \geq 4$, the memory capacity m is entirely given by m_{net}, which is in line with the fast decay of the single-unit contribution. Moreover, while for small time lags the memory is maximal around g_{nec}, for larger time lags it peaks near g_{c}, indicating that the intermediate, expansive, nonchaotic regime supports storage of the input.

The total memory capacity is defined as the integral over the memory curve:

$$M = \int_0^\infty m(\tau) = \frac{\sigma^2}{c_0}\frac{1}{\sqrt{1 - g^2\langle\phi'(x)\rangle^2}} = \frac{\sigma^2}{c_0}\,\tau_\infty\,, \tag{2.30}$$

where the result follows directly from the Laplace transform of the Bessel function. Here, τ_∞ is the asymptotic decay time constant of the autocorrelation function, as defined above in Section 2.3 and shown in Fig. 2.3(c). Correspondingly, the total network memory capacity is defined as $M_{\text{net}} = \int m_{\text{net}} = M - \sigma^2/c_0$. Usually, the memory capacity is bounded by the number of neurons N [140, 156]. The signal z in our situation, however, can be seen as one out of N independent inputs, and $m(\tau)$ is its corresponding memory curve. The expressions are therefore independent of N [157], and the memory capacity satisfies $M \leq 1$. While the memory capacity decreases in the chaotic regime, the network memory peaks within the expansive, nonchaotic regime [Fig. 2.6(b), lower panel].

So far, we have considered the memory capacity at a fixed amplitude σ of the input. In the following, we investigate the memory capacity over the whole phase diagram. The total memory capacity $M = \tau_\infty\,\sigma^2/c_0$, Eq. (2.30), shows a steep falloff directly above the onset of chaos [Fig. 2.7(a)]. This is expected because the variance c_0 increases as the network dynamics becomes more chaotic. The contour lines of the memory capacity are nearly parallel to the transition criterion, the curve with a vanishing Lyapunov exponent. This observation closely links the transition to chaos to the memory capacity: The direction in the phase diagram in which the system most quickly enters the chaotic regime is accompanied by the steepest decline of memory. The contour lines of the network memory capacity $M_{\text{net}} = (\tau_\infty - 1)\,\sigma^2/c_0$ show a ridge running through the expansive, nonchaotic regime [Fig. 2.7(b)] confirming the results found above: The memory is optimal in the dynamical regime of local instability and asymptotic stability. Moreover, the network capacity has a substantial contribution to the total memory capacity of about 50%.

The optimal network memory in the hitherto unreported regime between g_{nec} and g_{c} can be understood in an intuitive manner. Two conditions must be met for good memory. First, individual units must be susceptible to the signal; the susceptibility equals the fluctuation-averaged slope $\langle\phi'(x)\rangle$ of the transfer function. Second, the signal must propagate effectively through the network, requiring a sufficiently

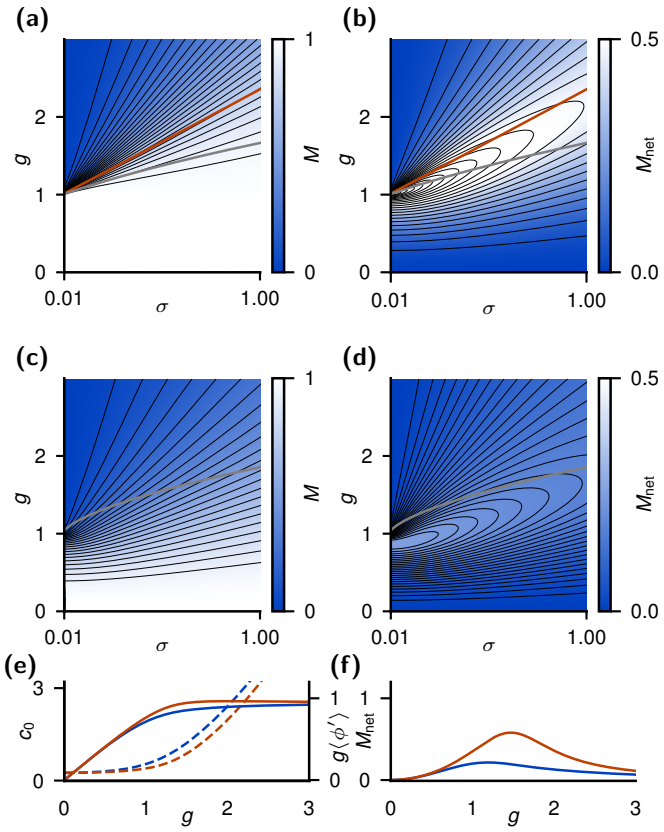

Figure 2.7: Memory capacity in different phases of the network dynamics. (a) Total memory capacity (2.30) encoded in color. We show the transition curve (2.20) between regular and chaotic dynamics (red) and the necessary condition (2.18) of local instability (gray) as in Fig. 2.3. Contour lines of memory are shown in black. (b) Same as panel (a) for the network contribution to memory capacity. Panels (c) and (d) are the same as panels (a) and (b) but for quenched background noise. (e) Self-consistent c_0 (dashed curve) and $g\langle\phi'\rangle$ (solid curve) for white-noise input (red) and quenched noise input (blue) for $\sigma_q = \sigma = 0.5$. (f) Network contribution to memory capacity M_{net} for white-noise input (red) and quenched noise input (blue).

strong coupling g. These two requirements are reflected in the monotonic increase of the memory curve (2.28) with the effective gain $g\,\langle\phi'\rangle$, independent of the time lag τ. Increasing the coupling strength g, however, elevates internally generated fluctuations as well. The corresponding increase of the variance c_0 [Fig. 2.7(e)] has a twofold effect on the memory capacity. First, it decreases $\langle\phi'\rangle$, so $g\,\langle\phi'\rangle$ assumes a maximum [Fig. 2.7(e)], which is also reflected by the maximum of the decay time τ_∞ of the autocorrelation function [Fig. 2.3(c)]. Second, it reduces the signal-to-noise ratio of the readout, the prefactor σ^2/c_0 in Eqs. (2.28) and (2.30). The interplay of these two effects determines the location of the optimal memory within the phase diagram.

To further relate optimal memory to the dynamical behavior and to demonstrate the robustness of our results, we close this section by considering colored background noise with correlation time τ_n as in Section 2.5. So far, the memory task was to reconstruct a white-noise signal $z(t)$ on the background of the remaining white-noise inputs $\xi_i(t) - z(t)$. Since a signal with temporal correlations contains memory on its own, we keep the signal $z(t)$ as a white noise with amplitude $2\sigma^2/N$, while the independent colored Gaussian background inputs $\xi_i(t)$ with autocorrelation function (A.26) act as noise. This setup allows a direct comparison of the memory capacity in the driven network to the case of white background noise. In the uncoupled network, $g = 0$, the colored background inputs lead to a variance $c_0 = \sigma^2$ that is independent of the correlation time τ_n; the variance then increases with g (see Section 2.5). Hence, the input signal-to-noise ratio is $1/N$, and the variance c_0 is predominantly determined by the background.

We find that the analytically derived memory curves (2.28) and (2.29) also hold for colored background noise because the autocorrelation function of the background does not enter the derivation of Eq. (A.53). This finding is confirmed by simulations [Fig. 2.6(a)]. Indeed, the temporal correlations of the background noise affect the memory only by changing the variance c_0. The qualitative behavior can be understood by comparing the peaks of the network memory within the phase diagrams for the two limiting cases of quenched and white-noise background input (Fig. 2.7).

For quenched background input, the network memory capacity peaks below the transition to local instability, which in this case also marks the transition to chaos [Fig. 2.7(c)]. To understand the emergence of the peak, we plot the decisive quantities c_0 and $g\langle\phi'\rangle$ as a function of g in Fig. 2.7(e); the corresponding network memory is shown in Fig. 2.7(f). The peak emerges at the point where $g\langle\phi'\rangle$ begins to saturate, while the variance c_0 starts to increase significantly. Now, compare this to the case of white-noise input: A corresponding increase of c_0 takes place only at larger values of g, and thus $g\langle\phi'\rangle$ assumes higher values. The latter effect combined with the appearance of c_0 in the denominator of the memory leads to a substantial increase in the peak height as well as a shift of the maximum to higher coupling values [Fig. 2.7(f)]. Furthermore, the decrease in c_0, by Eq. (2.18), shifts the line of local instability to lower values of g [compare Figs. 2.7(b) and (d)]. In combination, these effects interchange the relative positions of the local stability line and the peak of the memory when going from quenched to white noise. Finally, due to the dynamic chaos suppression, the transition to chaos is located at even higher values of g, so the optimal memory is located in the expansive, nonchaotic regime.

2.8 Discussion

In this chapter, we present a completely solvable random network model that allows us to investigate the effect of time-varying input on the transition to chaos and information-processing capabilities. Adding time-varying stochastic drive to the seminal model introduced by Sompolinsky, Crisanti, and Sommers [29] yields a stochastic continuous-time dynamical system. Contrary to the original model [29], here we

reformulate the stochastic differential equations as a field theory [133–136]. This formal step allows us to develop the dynamic mean-field theory by standard tools: a saddle-point approximation of the auxiliary-field generating functional [125, 138, 143]. As in the original model, this procedure reduces the interacting system to the dynamics of a single unit. The self-consistent solution of the effective equation yields a standard physics problem: The autocorrelation function of a typical unit is given by the motion of a classical particle in a potential. We find that the amplitude of the input corresponds to the initial kinetic energy of the particle.

Using the field-theoretical formulation, we then perform a replica calculation to determine the maximum Lyapunov exponent; the problem formally reduces to finding the ground-state energy of a single-particle quantum mechanical problem. The transition to chaos appears at the point where the ground state energy changes sign, which allows us to obtain an exact condition relating the coupling strength and the input amplitude at the transition. We find a simple hallmark of the transition in the single-unit activity: At the transition point, the variance of the recurrent input to a single unit equals the variance of its own activity. Correspondingly, the autocorrelation function at zero time lag changes its curvature from convex to concave at the transition point. These features can readily be measured in most physical systems. The assessment of chaos by these passive observations, in particular, does not require a perturbation of the system.

The transition criterion allows us to map out the phase diagram spanned by the coupling strength and the input amplitude. It shows that stochastic drive shifts the transition to chaos to significantly larger coupling strengths than predicted by time-local linear stability analysis. The transition in the stochastic system is thus qualitatively different from the transition in the autonomous system, where loss of local stability and transition to chaos are equivalent [29, 69]. The discrepancy of these two measures in the driven system is explained by a dynamic effect: The decrease of the maximum Lyapunov exponent is related to the sharpening of the autocorrelation function by the stochastic input. The displacement between local instability and transition to chaos leads to an intermediate regime that is absent in discrete-time networks of memoryless units [131]. This hitherto unreported dynamical regime combines locally expansive dynamics with asymptotic stability.

The seminal works [29, 69] have established a tight link between the fields of random matrix theory and autonomous neural networks with random couplings: Deterministic chaos emerges if the spectral radius of the coupling matrix exceeds unity. In contrast, we find in stochastically driven networks that the spectral radius only yields a necessary condition for a positive Lyapunov exponent; it determines the minimum of the quantum mechanical potential whose ground-state energy relates to the Lyapunov exponent. The presented exact relation between input strength, the statistics of the random matrix, and the onset of chaos, Eq. (2.20), generalizes the well known link to nonautonomous, stochastic dynamics.

It is controversially discussed whether the instability of deterministic rate dynamics leads to a transition to chaos in networks of spiking neurons [74, 75, 158]. It was argued that such a transition is absent in spiking models because the correlation time does not peak at the point where the corresponding deterministic rate dynamics becomes unstable [75, 148, 158]. For the analysis of oscillations [25] and correlations [155, 159] in these networks, the irregular spiking activity of the neurons can be approximated by effective stochastic rate equations, whereby the realization of the spikes is represented by an explicit source of noise. In this setting, the white-noise input in Eq. (2.1) can be interpreted as such spiking noise, explicitly investigated by Kadmon and Sompolinsky [73]. For weak noise, one may neglect its impact on the location of the transition. In this limit, noise suppresses the divergence of the correlation time [73]. Moreover, the authors obtained scaling laws for the decay time and amplitude of the autocorrelation function, indicative of a second-order phase transition. Our work is not limited to

small noise amplitudes, and it suggests that a diverging timescale in spiking networks does not occur at the instability for two reasons. First, we have shown that in a stochastic system the transition to chaos is not predicted by local instability. If a diverging timescale at the transition existed, it would not occur at the instability but at a larger coupling strength. But the presented analysis shows that the decay time of the autocorrelation function does not even peak at the transition to chaos but rather in the chaotic regime. While these results strictly only hold for the rate dynamics considered here, they still strongly suggest the absence of a diverging timescale in networks of spiking neurons. Indeed, the absence of a diverging timescale has been observed in simulation of spiking neurons [75] as well as in an iterative approach solving for the self-consistent autocorrelation function [160, 161].

A complementary view on the transition to chaos in autonomous systems has been given by Wainrib and Touboul [162], showing that the number of fixed points of the dynamics, the topological complexity of the system, diverges at the transition in tight relation to the Lyapunov exponent, which measures the dynamical complexity. It is intriguing to ask how this view extends to the stochastic systems considered here. Investigating topological features of the fixed points of the effective action instead of the fixed points of the actual action may allow for addressing this question, drawing on the analogy between stochastic systems and their effective deterministic counterparts, which is exposed by the field-theoretical formulation presented here.

To assess whether the found interplay of the input with the nonlinear recurrent network dynamics has functional consequences, we investigate dynamic short-term or sequential memory [140]. The obtained closed-form expression for the network memory capacity exhibits a peak within the expansive, nonchaotic regime. We identify two mechanisms whose partly antagonistic interplay causes optimal memory: local amplification of the signal and intrinsically generated noise. Local instability of the network ensures sensitivity to the external input, so that on short timescales the incoming signal is amplified and can therefore be read out more reliably. But stronger coupling also increases network-intrinsic fluctuations, which, in turn, reduce the susceptibility as well as the signal-to-noise ratio of the readout. Therefore, it is plausible that the optimal memory appears at a point where local amplification of the external input is large enough, but intrinsic chaoticity is still limited.

Sequential memory has been studied in a discrete-time neural network model [117], which receives a single, weak, external input. In contrast, here we investigate the memory of a single signal in the presence of multiple simultaneous inputs with arbitrary amplitude σ. Without additional observation noise, Toyoizumi and Abbott [117] find that sequential memory does not possess a maximum; it is constant and optimal for subcritical coupling values $0 < g < 1$ and falls off in the chaotic regime $g > 1$ due to intrinsically generated fluctuations. Perfect reconstruction in the nonchaotic regime is possible because the one-step delayed activity is a direct linear function of the input. In our setting, the single neuron memory due to leaky integration has a similar effect [Fig. 2.7(a)]. In the discrete-time system without additional input noise, optimal memory close to the transition arises only in the presence of observation noise [117]. In the chaotic regime, memory falls off more slowly than in the direction of regular dynamics, so fine-tuning is not needed if the network dynamics is slightly chaotic. The authors mention that the situation changes qualitatively in the presence of a small additional noise in the input: The signal-to-noise ratio then stays finite and peaks slightly below the transition to chaos.

In the continuous-time model studied here, we do not consider observation noise, but the remaining inputs act in a similar manner as the input noise mentioned in Ref. [117]. The network component of the memory or, equivalently, the memory for longer delay times τ, also shows nonmonotonic behavior. For small signal amplitude σ, our analytical results show that memory is optimal right below the transition to chaos and steeply falls off above [Fig. 2.7(b)]; this is in line with the observation in the discrete-time

setting with input noise [117]. For larger σ, the falloff is weaker in the chaotic regime, qualitatively more similar to the situation with observation noise.

A negative maximum Lyapunov exponent for nonautonomous dynamics indicates the echo state property, the reliability of the network response to input [163]. We could indeed show, for the analytically tractable model here, that memory capacity quickly declines in the chaotic regime due to intrinsically generated chaotic fluctuations. Echo state networks show long temporal memory near the edge of chaos [164–167]. Typically, these networks operate in discrete time, and thus, the onset of chaos is directly linked to the spectral radius of the Jacobian. This relation is used in the design of these systems, exploiting the fact that a spectral radius close to local instability ensures long memory times. Here, we show, for the driven and continuous-time system, that the edge of chaos and local instability are two different concepts and that memory capacity is a third, distinct measure: Memory is not optimal at either of the two other criteria but rather in between. In particular, our analytical results for the memory capacity can be used to determine the optimal coupling strength for a given input amplitude.

The finding that the edge of chaos does not necessarily coincide with the optimality of measures of computation, as reported here, qualitatively alters the edge-of-chaos hypothesis leading to implications for methods that tune networks towards the edge of criticality with the goal to optimize performance [168]: The fundamental assumption of an equivalence of these *a priori* different measures is true for some discrete-time network dynamics [117]. But it certainly fails for nonautonomous networks operating in continuous time, as we show here. Moreover, the divergence of the timescale of the autocorrelation function in the autonomous setting [Fig. 2.3(c)] is indicative of critical slowing down [169] and could therefore point to a continuous phase transition. This phenomenon, however, vanishes in the nonautonomous case; the nature of the transition may hence be different. The presence of a continuous phase transition is an often-made second assumption to identify optimal network parameters [168, 170].

Sussillo and Abbott [30] proposed an algorithm to train a random network as given by Eq. (2.1) to produce a wide range of activity patterns. They find that learning shows the best performance if the random network without input is initially in the chaotic state. It has been argued that such networks have a rich dynamic repertoire and are able to produce a wide variety of outputs. In the training phase, the input to the network needs to suppress chaos so that learning converges. The procedure therefore requires the choice of an initial coupling that is large enough to ensure chaos but not too large so that the input can suppress chaos. Our quantitative criterion for the transition easily enables a proper choice of parameters and facilitates the design, control, and understanding of functional networks.

In this work, we have considered sequential memory of the input signal. An important task of the brain, however, is not only to maintain the input but also to perform nonlinear transformations on it. We expect the locally unstable but globally stable dynamics to be beneficial for such a task: The expansive behavior can project the input into a high-dimensional space, which is crucial for nonlinear computations or discrimination tasks [165]. Thus, this dynamical regime not only provides memory but might serve as a basis for more complex computations.

To show the generic effect of input, we added a time-varying drive to the seminal model by Sompolinsky, Crisanti, and Sommers [29]. Even though the original model makes some simplifying assumptions, such as the all-to-all Gaussian connectivity and a sigmoidal odd transfer function, the transition to chaos occurs in a similar manner in networks with biologically more realistic features [24, 73, 148]. To focus on the new physics arising in nonautonomous systems, we start from the simplest but nontrivial and yet application-relevant model. Subsequently, we show that the results also hold for more realistic connectivity with nonvanishing mean coupling and for non-negative transfer functions, as they arise,

for example, in rate model abstractions of spiking neurons. Preliminary results indicate that the same mechanism also persists in networks of populations of spiking neurons. Here, however, a suitable definition of rate chaos still needs to be found for a more quantitative advancement.

Our work predicts that time-varying input to the network modulates the distance to the transition to chaos and it elucidates the corresponding mechanisms that control the transition. Consequently, we expect that the ongoing dynamics moves further away from the transition to chaos in periods in which the fluctuating drive to a network is elevated, thus potentially altering the processing capabilities along with the dimensionality of the intrinsic dynamics [171].

The reformulated derivation of the dynamic mean-field theory using established methods from field theory [125, 138, 143, 147, 172] here allowed us to obtain the sequential memory by a replica calculation. In general, the presented formulation opens the study of recurrent neural networks to the rich and powerful set of field-theoretical methods developed in other branches of physics. This language leads to straightforward extension of our results in various directions. Among these directions are more biologically realistic settings, such as sparse connectivity respecting Dale's law [2] or multiple populations. The over-representation of bidirectional connections [173] can also be treated in this framework [174]. The extension to multiple populations would enable the study of the interesting case in which the population receiving the signal is separated from the readout population. Such a situation would most likely emerge in the cortex, where the input population of local microcircuits typically differs from the output population.

More generally, the stability of complex dynamical systems plays an important role in various other fields of physics, biology, and technology. Examples include oscillator networks [175], disordered soft-spin models [124], power grids [176], food webs [72], and gene-regulatory networks [177]. Presenting exact results for a prototypical and solvable model, the work in this chapter contributes to the understanding of chaos and signal propagation in such high-dimensional systems.

CHAPTER 3

Dynamics of balanced networks containing neurons that accelerate towards spiking

This chapter is based on the following published article [178]:

P. Manz, **S. Goedeke**, and R.-M. Memmesheimer
Dynamics and computation in mixed networks containing neurons that accelerate towards spiking
Phys. Rev. E **100**.4 (2019) 042404 DOI: 10.1103/PhysRevE.100.042404

In this chapter, we move to recurrent networks of spiking model neurons. Like neurons in the brain, they interact with short activity pulses called action potentials or spikes, in contrast to the more abstract continuous rate dynamics we investigated in the previous chapter. The state of each neuron is specified by its membrane potential (voltage), which changes in response to input spikes from other neurons and evolves according to its intrinsic dynamics. When the membrane potential of a neuron reaches a threshold, the neuron emits a spike, which is transmitted to other neurons. Thereafter, the neuron's membrane potential is reset to a value below the threshold. Here, we study the collective dynamics in networks of such integrate-and-fire model neurons. In particular, we investigate a transition from stable to chaotic dynamics that arises when a single neuron with a different type of intrinsic dynamics is inserted into the network.

The ongoing activity in cortical neural networks is often characterized by asynchronous irregular spiking occurring at low rates. We consider a network model in which the deterministic dynamics of recurrently connected spiking neurons can give rise to such asynchronous irregular activity. In this model, strong excitatory (positive) and inhibitory (negative) input currents balance each other, such that the average membrane potential is close to the threshold and spikes are generated by fluctuations of the membrane potential. This balanced state can be realized in recurrent networks with purely inhibitory synaptic couplings: strong inhibitory couplings result in a negative feedback that dynamically balances a strong excitatory external input. Because of the neuronal integrate-and-fire dynamics, the external input current alone would lead to regular spiking at a high rate. The emerging recurrent input counteracts the external input, which reduces the neurons' rates of spiking and makes them susceptible to fluctuations in their input currents. As the neurons in the network are randomly connected, different

neurons receive different inhibitory recurrent inputs. Therefore, temporal input fluctuations that are only weakly correlated between different neurons can self-consistently generate asynchronous irregular spiking activity in these networks.

The resulting asynchronous irregular spiking appears unpredictable even for fully deterministic network dynamics. This has raised the question whether the underlying microscopic dynamics are chaotic and exhibit sensitive dependence on the initial conditions. The mechanism of the balanced network model described above is quite general and insensitive to the specific intrinsic dynamics of the neurons. It has been shown that inhibitory balanced networks of pulse-coupled leaky integrate-and-fire (LIF) neurons have irregular dynamics that are stable against small perturbations and thus non-chaotic. The LIF neuron is a canonical spiking neuron model, which takes a passive leak current of the neuronal membrane into account. The leak current acts as a restoring force in the neuron's intrinsic membrane potential dynamics. In biological neurons, however, action potentials are generated through a positive feedback loop that is activated above a certain membrane potential value. This causes the membrane potential to accelerate towards spiking. A simple way to implement this intrinsic positive feedback is a "leak" current with reversed sign, which, for a constant excitatory input current, leads to a membrane potential that rises as a convex function of time. We thus introduce such "anti-leaky" (with convex rise function) integrate-and-fire (XIF) neurons in inhibitory balanced networks of LIF neurons and study the collective dynamics of mixed networks containing both types of spiking neuron models.

We first demonstrate that such mixed networks can generate asynchronous irregular spiking activity for any number of XIF neurons in the network. The XIF neurons are endowed with a voltage-dependent modulation of their recurrent synaptic inputs, which prevents unphysical intrinsic dynamics towards negative infinity in these neurons. We obtain neuron parameters that yield identical self-consistent firing rates for both types of neurons by employing a mean-field approximation. To do so, we analytically derive the steady-state membrane potential probability distributions and the output firing rates of LIF and XIF neurons in a shot-noise approach. This approach assumes that the superposed spike inputs form a Poisson process and includes the finite-amplitude postsynaptic impact of individual spikes. It is therefore applicable to networks of intermediate size, as we mostly consider here.

The obtained setup allows us to study the effect of the two different neuron types on the stability properties of the deterministic microscopic dynamics. We characterize the stability properties in detail using the spectrum of Lyapunov exponents. These are the asymptotic average growth or shrinkage rates of infinitesimal perturbations in different directions of the dynamics' phase space (the space spanned by the membrane potentials of all neurons in the network). The largest Lyapunov exponent describes the stability for generic infinitesimal perturbations; if it is positive, the dynamics depend sensitively on the initial conditions. We find that irrespective of the network size, a single XIF neuron in the network already gives rise to a positive largest Lyapunov exponent, indicating chaotic dynamics. A simple mean-field approach to the full spectrum of Lyapunov exponents explains this finding. In particular, the approach reveals that in our inhibitory balanced networks, the microscopic linear stability properties are largely determined by the individual neurons' intrinsic membrane potential dynamics and only to a lesser extent by the spreading of perturbations in the network. In the article, we show as a concrete application of our results that the different stability properties in mixed networks can be useful for computations based on spike timings [178].

My most important contribution to the article is the analytical derivation of the output firing rates and membrane potential distributions of both types of neuron models (LIF and XIF) when they receive purely inhibitory shot noise as recurrent input. This enabled us to determine self-consistent firing rates in inhibitory balanced networks containing both LIF and XIF neurons and, thus, to find a self-consistent state

with asynchronous irregular activity in such mixed networks of intermediate size. Specifically, the overall characteristics of the asynchronous irregular activity remain unchanged for any ratio of the numbers of LIF and XIF neurons. The derivation also demonstrates the necessity of the voltage-dependent input modulation for XIF neurons. The shot-noise approach ultimately allowed us to give a complete mean-field approximation of the spectrum of Lyapunov exponents. Furthermore, I contributed to the derivations of the single-spike Jacobian factors and matrices on which the computations of the Lyapunov exponents are based. These Jacobians incorporate the jumps in the dynamics at spike events. We derived them for both LIF and XIF neurons by considering infinitesimal perturbations of their membrane potentials. The resulting mean-field approximation of the Lyapunov exponents in mixed networks lead us to the above-mentioned insight regarding the microscopic stability properties. As an appendix to this chapter (Appendix B), I present modified derivations of the single-spike Jacobians, which further clarify how the evolution of infinitesimal perturbations is governed by two separate factors corresponding to the neurons' intrinsic dynamics and the discrete spike events, respectively. Finally, I wrote to a large extent the part of the article that presents the shot-noise approach and I contributed to the writing of the other parts. The following sections of this chapter contain the parts of the published manuscript that include significant contributions from me. I added clarifications and changes to adapt the contents to the scope and the format of the thesis. In particular, I refer to Appendix B for the derivations of the single spike Jacobians and the mean-field Lyapunov exponents.

3.1 Introduction

In this chapter, we study purely inhibitory networks of oscillating integrate-and-fire type spiking neurons with postsynaptic currents modeled as pulses of infinitesimally short duration. It has been shown numerically [76, 179] and analytically [77, 82] that if such networks contain only leaky integrate-and-fire (LIF) neurons, the networks' irregular balanced state dynamics are stable against infinitesimal and small finite size perturbations and are thus not chaotic but a realization of stable chaos [180, 181]. The dynamics ultimately converge to a periodic orbit; the durations of the preceding irregular transients, however, grow exponentially with system size. The stability of the network dynamics is robust against introducing excitatory connections and considering postsynaptic currents of finite temporal extent and there is a continuous transition to chaos upon increasing the number of excitatory connections and the duration of synaptic currents [76, 77]. The computational capabilities of the stable precise spiking dynamics have not yet been explored, even though the specific structure of the phase space, which is composed of "flux tubes", may be beneficial and exploitable [83].

LIF neurons incorporate a leak current as found in biological neurons [4]. It increases linearly with increasing membrane potential and leads to dissipation (contraction of phase space volume) in the subthreshold dynamics. When driven by a constant excitatory input current, the membrane potential therefore has a negative second derivative; the neuron has a purely concave so-called rise function. In the considered class of network models, this implies the stability of the microscopic dynamics if only LIF neurons are present [77, 82]. In biological neurons as well as in neuron models that explicitly model spike generation, such as the quadratic and the exponential integrate-and-fire neuron [5], the membrane potential accelerates towards a spike at higher voltages. The rise function thus has a convex part. Monteforte and Wolf [65] showed that networks of quadratic integrate-and-fire neurons that are otherwise similar to those considered in Refs. [76, 77, 82] exhibit extensive chaos. Specifically, Monteforte and Wolf [65] computed the spectrum of Lyapunov exponents (LEs) and quantities that are

derivable from them, as well as the statistics of the first covariant Lyapunov vector, which points into the directions to which a generic perturbation vector aligns in the long term.

Motivated by these results and by the fact that there are many different types of cortical inhibitory neurons [182], here we investigate the impact of inserting a different type of neuron, with non-concave rise function, into inhibitory networks of LIF neurons. To be specific, we insert "anti-leaky" integrate-and-fire (XIF) neurons with purely convex rise function. We choose the letter "X" in the abbreviation to highlight the convexity and the expansion of phase space volume by the flow of the subthreshold dynamics. XIF neurons may be interpreted as a simplified model for biological neurons whose membrane potential lingers in a region where it accelerates towards spiking. Simultaneously, these neurons maintain similar analytical tractability as their leaky counterparts because of their mostly linear subthreshold dynamics. Our neuron and network models are described in detail in the following Section 3.2. Thereafter, we analytically derive self-consistent firing rates and membrane potential probability distributions for both types of neurons, assuming Poisson input spike trains with finite size impacts (Sec. 3.3). We then consider the dynamical stability properties and local phase space structures of the network dynamics in Sec. 3.4. In a mean-field approach, we obtain analytical approximations of the entire spectrum of LEs for mixed networks containing both LIF and XIF neurons. We also compute the spectra of LEs in simulations of the full network dynamics and compare them to the mean-field ones.

3.2 Mixed networks of integrate-and-fire neurons with concave and convex rise function

We consider recurrent networks of N integrate-and-fire type neurons. The kth spike of neuron j, which is emitted at time t_{jk}, generates a postsynaptic current pulse $h_i(V_i^-)C_{ij}\delta\left(t - t_{jk}\right)$ in neuron i. Here, $C_{ij} \leq 0$ is the coupling strength of the inhibitory synaptic connection and $h_i(V_i^-)$ is a possible voltage-dependent modulation, which depends on the membrane potential of neuron i just before input arrival, given by the left-hand side limit $V_i^- = V_i(t^-) = \lim_{\varepsilon \searrow 0} V_i(t - \varepsilon)$. We assume that all excitatory inputs to neuron i are gathered into a constant excitatory external input current $I_i^{\text{ext}} > 0$ and that the remaining explicitly modeled recurrent inhibition is fast [82, 83, 183]. We further assume that there is a "leak" term with prefactor $\gamma_i \neq 0$, which has units of t^{-1}. Taken together, we model the subthreshold membrane potential dynamics of neuron i by

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

When V_i reaches the spike threshold at time t, $V_i^- = V_{\text{th}} > 0$, it is reset, $V(t) = V_{\text{re}} = 0$, and a spike is emitted. This, in turn, generates in a postsynaptic neuron l a current pulse as introduced above, which causes V_l to decrease in a jump-like manner from V_l^- to $V_l^- + h_l(V_l^-)C_{li}$. The rise function, i.e., the membrane potential dynamics with $V_i(0) = 0$ in absence of recurrent inhibitory input [184, 185], reads

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

It is concave for $\gamma_i > 0$ and convex for $\gamma_i < 0$.

There are two types of neurons in our networks: LIF neurons with subthreshold dissipation and

concave rise function, which obey Eq. (3.1) with $\gamma_i > 0$, and anti-leaky XIF neurons with convex rise function, which obey Eq. (3.1) with $\gamma_i < 0$ (see Fig. 3.1). In the absence of the threshold for spike generation and the inhibitory input, the membrane potential dynamics of an LIF neuron has a globally attracting fixed point at $V_{\infty,i} = I_i^{\text{ext}}/\gamma_i$. We assume $V_{\infty,i} > V_{\text{th}}$, so without inhibitory input, LIF neurons periodically spike and reset. For our study, it is sufficient to endow the LIF neurons with a current-based synaptic input model, setting $h_i(V_i^-) = 1$. A coarse approximation of the membrane potential dynamics without threshold and neglecting input fluctuations yields $\dot{\bar{V}}_i = -\gamma_i \bar{V}_i + I_i^{\text{ext}} + I_i^{\text{inh}}$, where I_i^{inh} is the average inhibitory input current. In the balanced state, its attractor at $\bar{V}_{\infty,i} = \left(I_i^{\text{ext}} + I_i^{\text{inh}}\right)/\gamma_i$ is below or close to the spike threshold, such that spikes are always or typically generated by input fluctuations, more specifically by periods of less than average inhibition.

In the absence of inhibitory input, XIF neurons have an unstable, repelling fixed point at $V_{-\infty,i} = I_i^{\text{ext}}/\gamma_i < 0$. If the membrane potential starts above this separatrix, it increases exponentially towards the threshold [Fig. 3.1(d)]. When the threshold is reached, the neuron spikes and the membrane potential resets to zero, from where it increases again towards the threshold: Without inhibitory input, XIF neurons oscillate and spike periodically for any $I_i^{\text{ext}} > 0$. If the membrane potential starts below the separatrix, it decreases exponentially to $-\infty$. In the presence of recurrent inhibitory inputs, an XIF neuron is unrecoverably switched off once its membrane potential falls below $V_{-\infty,i}$, since inhibitory inputs will only decrease the membrane potential further. Averaging over the inhibitory inputs as for the LIF neuron dynamics above yields an effective separatrix at $\bar{V}_{-\infty,i} = \left(I_i^{\text{ext}} + I_i^{\text{inh}}\right)/\gamma_i$. Membrane potentials falling below it have a tendency to further decrease, causing the neuron to effectively switch off. This can also be seen from the phase response curve [186–188] of XIF neurons, which gets steeper for negative phases, in contrast to that of LIF neurons which becomes flatter [Fig. 3.1(e)]. In other words, for XIF neurons inhibitory input arriving at a low potential still above the separatrix (and thus at a low phase) has a larger effect in the sense that it delays the next spike time more than the same input arriving at a higher potential. As a consequence, we observe in balanced networks containing XIF neurons with purely current-based synaptic input, i.e. with $h(V_i^-) = 1$, that many of these neurons are first effectively and then unrecoverably switched off, if the network dynamics are irregular and the inhibitory inputs are therefore strongly fluctuating. In order to prevent this biologically implausible phenomenon, we introduce a simple voltage-dependent modulation

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

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

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

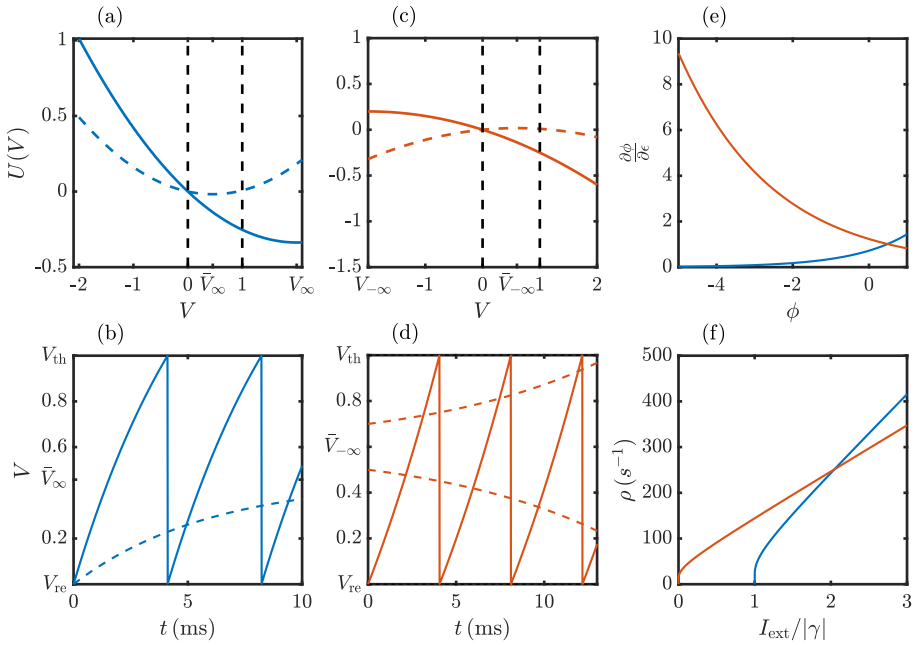

Figure 3.1: LIF and XIF neuron dynamics with constant input current. Blue and red indicate LIF and XIF neurons, respectively. Solid curves in panels (a)–(d) indicate constant excitatory input I^{ext} only, dashed lines inclusion of an average inhibitory input I^{inh} (cutoff for inputs to XIF neurons is neglected). Panels (a) and (c) show a potential function $U(V)$ of the membrane potential (voltage) V; V follows the negative gradient of U, $\dot{V}(t) = -U'(V(t))$ up to the threshold at $V_{\text{th}} = 1$. (a) $U(V)$ for an LIF neuron is an upward parabola; V tends to the stable fixed point at U's minimum (at V_∞ or $\bar{V}_\infty$). (c) $U(V)$ for an XIF neuron is a downward parabola; V tends to $-\infty$ if started to the left or to $+\infty$ if started to the right of U's maximum (at $V_{-\infty}$ or $\bar{V}_{-\infty}$). A monotonically decreasing $U(V)$ between reset and threshold (left and right vertical dashed lines) indicates mean-driven periodic spiking [solid curves in panels (a)–(d)]. In the balanced state, the spiking is fluctuation-driven with (a) U's minimum below threshold for LIF neurons and (c) U's maximum above reset for XIF neurons (dashed curves). Panels (b) and (d) show example trajectories for LIF and XIF dynamics including threshold and reset. Without inhibition, V is periodically driven to the threshold and reset ("free dynamics"). When averaged inhibition is included, the LIF voltage (b) converges to the subthreshold fixed point at $\bar{V}_\infty$, while the XIF voltage (d) is repelled from $\bar{V}_{-\infty}$. (e) Infinitesimal phase response curves. Pulse inputs to LIF (XIF) neurons have a smaller (larger) spike delaying effect at lower phases, corresponding to lower V. (f) Firing rates of free LIF and XIF neurons at different strengths of the scaled external input current. Neuron parameters and (if applicable) values for constant excitatory input and average inhibitory input are as in our network simulations.

since the spike frequency of their oscillating "free dynamics" [Fig. 3.1(b) and (d)] reads

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

whereas in LIF neurons, it is

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

As a consequence, we observe that in a mixed network, the XIF neurons suppress the LIF neurons, which become quiescent. Using the analytical results of the next section, we therefore rescale γ_{LIF} such that the rates in both populations are identical, see also Fig. 3.1(f). Further, we fix the neurons' indegree to the same number K, implying that $\sum_j C_{ij} = KC$ is identical for each neuron i. This reduces quenched variability [27, 79] and avoids strong differences of individual firing rates and quiescent neurons.

With the described network model setup, we observe asynchronous irregular spiking activity indicating a dynamically balanced state for any ratio of the numbers of LIF and XIF neurons, which is illustrated in Fig. 3.2.

3.3 Self-consistent firing rates and membrane potential distributions

In this section, we analytically determine the steady-state firing rates and membrane potential distributions for LIF and XIF neurons in mixed networks with asynchronous irregular spiking activity (Fig. 3.2). We use the results to obtain neuron parameters that lead to the same firing rates for both neuron types and thus to homogeneous firing rates in the entire network. In addition, we employ the firing rates to analytically approximate the Lyapunov spectrum of the network dynamics using a mean-field approach in Sec. 3.4.1.

We approximate the superposed input spike trains to a neuron by a Poisson spike train with a given rate, i.e., we assume that all input spikes occur independently of each other. The assumption that the input neurons spike independently of each other is the mean-field approximation in this context. A common approach is to additionally consider the limit of a large number of small-sized inputs. The neuron dynamics can then be approximated by a diffusion process, which allows one to compute firing rates and membrane potential distributions [85, 86]. This diffusion approximation assumes that the inputs have (infinitesimally) small amplitude and arrive at (infinitely) high rate. Here, we use a shot-noise approach, which accounts for the finite input rate and size of individual inputs [85, 86], in the recent formulation of Olmi et al. [183] and Richardson and Swarbrick [189]. This allows us to more accurately obtain the firing rates and membrane potential distributions. In particular, the fact that in our networks the voltage probability density does not go to zero at the threshold is reflected. We shortly review the approach for LIF neurons [183, 189, 190] and then extend it to XIF neurons with the voltage-dependent coupling of Eq. (3.3).

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

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

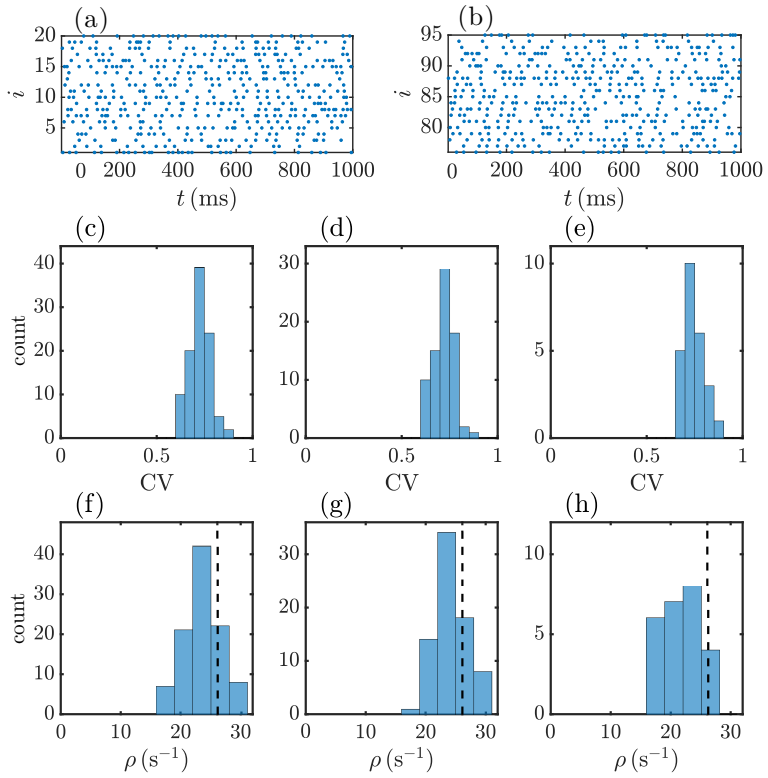

Figure 3.2: Mixed networks of LIF neurons with concave rise function and XIF neurons with convex rise function can generate a balanced state with asynchronous irregular spiking activity in both types of neurons. Panels (a) and (b) show the spiking activity for a subset of the LIF (a) and for the XIF (b) neurons in a network with 75 LIF and 25 XIF neurons ($N = 100$). (c)–(e) Histograms of coefficients of variation (CV) of measured interspike intervals for all neurons (c) and for LIF (d) and XIF (e) neurons separately. (f)–(h) Distributions of the firing rates of all neurons (f) and of LIF (g) and XIF (h) neurons separately. The analytically derived rate $\rho \approx 26.1\,\mathrm{s}^{-1}$ [Eqs. (3.18) and (3.19)] is indicated by the black dashed vertical line. We use $\gamma_{\mathrm{XIF}} = -0.1\,\mathrm{ms}^{-1}$, $\gamma_{\mathrm{LIF}} = 0.169\,\mathrm{ms}^{-1}$, $V_{\infty,\mathrm{LIF}} = -V_{-\infty,\mathrm{XIF}} = 2$ and a randomly connected network with fixed indegree $K = 50$ and nonzero synaptic strength $C = -0.2$.

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

For the LIF neuron with the current-based synaptic input model, inhibitory input spikes arriving when the considered neuron is at a voltage V give rise to a sink at V, whereas spikes arriving when the neuron is at a voltage $V - C > V$ give rise to a source at V. We therefore have a first source term

$$\sigma_{\mathrm{inh}}(V,t) = r(t)\,[p(V-C,t) - p(V,t)] \tag{3.7}$$

with the rate $r(t)$ of input spikes. We note that in Refs. [183, 189, 190] this term is included in the probability current. The second source term is due to the spike and reset mechanism of the neuron model. Its threshold and reset act as Dirac delta sink and source at the corresponding discrete voltages,

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

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

We investigate stationary network dynamics in a steady state, which are described by constant r and ρ and time-independent $p(V)$. In this case, Eq. (3.6) reduces to the linear delay differential equation (or differential-difference equation)

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

Dividing Eq. (3.9) by $\rho > 0$ yields an equation for the rescaled density $q(V) = p(V)/\rho$, which is independent of the unknown steady-state firing rate ρ. This equation can be integrated, for example, with the method of steps [191]. The integration starts with the "initial conditions" $q(V) = 0$ for $V > V_{\text{th}}$ and thus $q(V_{\text{th}}) = 1/\left(-\gamma V_{\text{th}} + I^{\text{ext}}\right)$ slightly below V_{th}. The normalization of $p(V)$ allows us to compute ρ via

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

To obtain a more explicit expression for ρ, we apply a bilateral Laplace transform $\tilde{f}(s) = \int_{-\infty}^{\infty} f(V) \mathrm{e}^{sV}\, dV$ [183, 189]. The value $\tilde{q}(0)$ of $\tilde{q}(s) = \int_{-\infty}^{\infty} q(V)\mathrm{e}^{sV}\, dV$ yields ρ^{-1}; it suffices to consider $s \geq 0$. The Laplace transform of the rescaled Eq. (3.9) results in a linear first-order ordinary differential equation for $\tilde{q}(s)$,

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

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

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

with an arbitrary constant A and

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

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

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

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

side of Eq. (3.14) needs to vanish for large s. We thus have

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

For an XIF neuron without the voltage-dependent input modulation, there is no stationary membrane potential probability density $p(V)$. This is because for any time $t > 0$ there is a finite probability that the membrane potential of a neuron jumps below I^ext/γ and thereafter moves unrecoverably to negative infinity. In contrast, for the XIF neuron with the voltage dependence Eq. (3.3), $p(V)$ exists and we may use the same approach as for the LIF neuron to determine it together with the firing rate. Since membrane potentials do not drop below $V_\mathrm{cutoff} + C$, we focus on the interval $[V_\mathrm{cutoff} + C, V_\mathrm{th}]$, where $p(V)$ can be nonzero. The couplings' voltage dependence enters the source term σ_inh in Eq. (3.6): If V is below V_cutoff, incoming spikes have no effect and the sink term due to them vanishes. Equation (3.7) therefore changes to

$$\sigma_\mathrm{inh}(V,t) = r(t)\,[p(V-C,t) - h(V)p(V,t)]\,, \tag{3.16}$$

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

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

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

We apply the above results to find mixed networks in which LIF and XIF neurons have similar firing rates. The expression for ρ, Eq. (3.15), provides a map G_LIF from the input to the output rate for LIF neurons, $G_\mathrm{LIF}(r) = \rho$, while Eq. (3.17) together with Eq. (3.10) implicitly defines such a map G_XIF for XIF neurons. The firing rate ρ of the LIF and XIF neurons in the desired mixed network needs to solve both self-consistency equations:

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

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

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

Figure 3.3 compares the stationary probability densities $p(V)$ and rates ρ obtained from the shot-noise approach with those of an LIF and an XIF neuron that receive input spike trains as they are generated in the recurrent network of Fig. 3.2. In case of an LIF neuron, there is a pronounced discrepancy between the densities and rates for the considered indegree $K = 50$ and network size $N = 100$, indicating that both the individual [see Fig. 3.2(c)–(e)] and the superposed input spike trains in these dense networks are more regular than Poisson spike trains. Such input spike trains are less variable and thereby generate a $p(V)$ that is more concentrated around the value $\left(I_i^\mathrm{ext} + I_i^\mathrm{inh}\right)/\gamma_i$, where I_i^inh is the average inhibitory input current as discussed in Sec. 3.2. Removing correlations between neurons, for example, by increasing N reduces the discrepancy [Fig. 3.3(b)]. Such "spatial" correlations appear to be the main cause of the

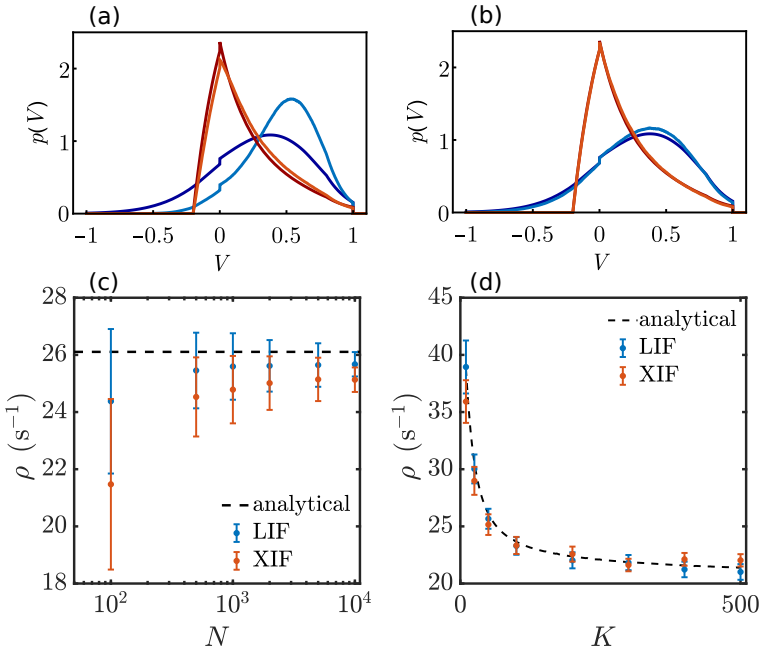

Figure 3.3: Analytically and numerically estimated membrane potential distributions and firing rates of LIF and XIF neurons. (a), (b) Stationary voltage probability densities $p(V)$ for networks of (a) $N = 100$ and (b) $N = 10\,000$ neurons. The dark blue and dark red curves show the analytical results (3.9) and (3.17) for LIF and XIF neurons, where the rate ρ is obtained self-consistently from Eqs. (3.18) and (3.19). The light blue and light red curves show representative numerically sampled voltage densities of an LIF and an XIF neuron, where the input spike trains are superpositions of simultaneous output spike trains of $K = 50$ neurons in the recurrent network. (c), (d) Firing rates of neurons in networks of (c) different size N and fixed indegree $K = 50$ and (d) size $N = 10\,000$ and different indegree K. In panel (d) the synaptic couplings are scaled with $1/K$ such that their sum is independent of K. Numerically measured average rates of LIF and XIF neurons in the different networks are shown by blue and red dots. Error bars display the standard deviations of the rate distributions. Analytical results obtained from Eqs. (3.18) and (3.19) are displayed by dashed black lines. Remaining parameters are as in Fig. 3.2.

discrepancy observed for smaller network size, as we further discuss in the article [178]. In case of an XIF neuron, the input spike train statistics have less impact on $p(V)$. Presumably, this is because negative voltage excursions due to input fluctuations are suppressed by the voltage dependence of the input (for values near V_{cutoff}) and by the increased drive towards threshold (for higher voltages). We note that the assumption of Poisson input spike trains is the only approximation in the chosen approach. For both types of neurons, the numerically measured firing rates in the networks become closer to the self-consistently obtained ones as we increase the network size N for fixed indegree K [Fig. 3.3(c)]. This illustrates how the underlying mean-field approximation improves for larger and sparser networks.

3.4 Dynamical growth or shrinkage of small perturbations

After obtaining the firing rates and membrane potential distributions using a statistical mean-field approximation, we investigate the collective dynamics of mixed networks from a deterministic dynamical systems perspective. Irregular dynamics are often chaotic and thus sensitive to perturbations: initially

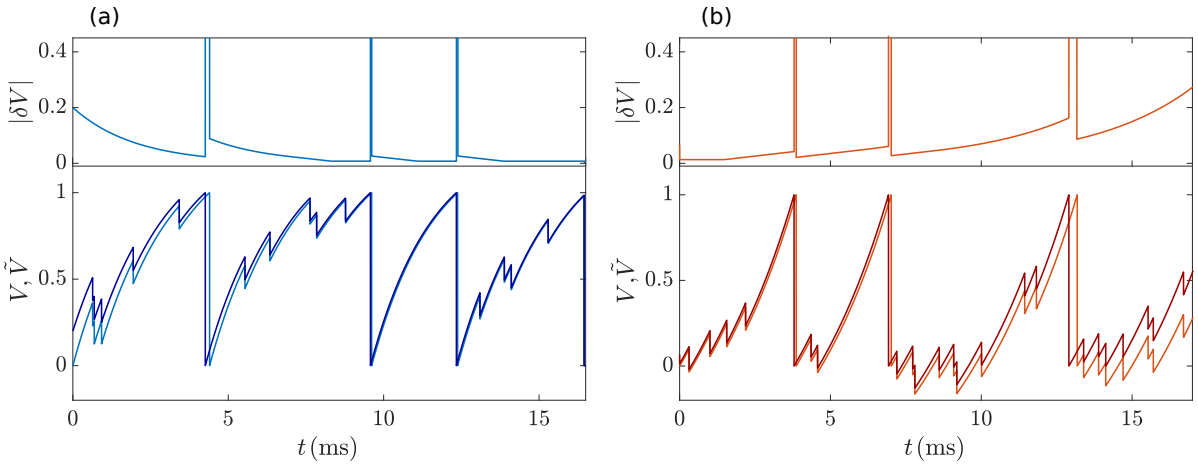

Figure 3.4: Evolution of small perturbations during subthreshold dynamics and spiking in (a) an LIF and (b) an XIF neuron. During the subthreshold evolution, the distance (perturbation size) $|\delta V(t)|$ between two neighboring trajectories shrinks for LIF and grows for XIF neurons, while spike generation partially resets it. Since the neurons receive only inhibitory input spikes, the intervals between the neuron's spike times are generally longer than the period of the oscillating free dynamics. The impact of the subthreshold dynamics therefore dominates and on average the perturbation in the LIF neuron shrinks while that in the XIF neuron grows. We use $V_{\infty,\mathrm{LIF}} = 1.33$, $V_{-\infty,\mathrm{XIF}} = -1$, $\gamma_{\mathrm{LIF}} = 0.5\,\mathrm{ms}^{-1}$, and $\gamma_{\mathrm{XIF}} = -0.3\,\mathrm{ms}^{-1}$ for better illustration of the mechanism.

very small ones can grow with time resulting ultimately in large differences between trajectories. An important tool to quantify this sensitivity and the local phase space structure are the LEs and associated with them, the covariant Lyapunov vectors [60, 192, 193]. The largest LE equals the asymptotic average rate of exponential growth or shrinkage of generic infinitesimal perturbations. A positive largest LE implies sensitive dependence on initial conditions and is usually taken to indicate chaotic dynamics. The full spectrum of LEs describes the asymptotic growth or shrinkage rates of infinitesimal perturbations in non-generic phase space directions, which are given by their corresponding covariant Lyapunov vectors. The covariant Lyapunov vectors therefore indicate unstable or stable directions of the dynamics at a given point in phase space.

3.4.1 Mean-field approach

We first develop a simple mean-field approach to the spectrum of LEs. It focuses on the evolution of small perturbations to the membrane potential of a single neuron and treats the spike input from other neurons as external. Concretely, we disregard perturbations of the rest of the network including those originating from the considered neuron's changed spike times. Inputs thus arrive at the same times in the perturbed and in the unperturbed system and do not change the neuron's perturbation. Figure 3.4 illustrates this and compares the resulting evolution of a small perturbation for an XIF neuron and for an LIF neuron: The perturbation of the XIF neuron's voltage gradually increases as long as it is not spiking, while that of the LIF neuron's voltage decreases. Conversely, spiking and resetting reduces perturbations in the XIF neuron, while it increases them in the LIF neuron; compare the values of the distance $|\delta V(t)| = |\tilde{V}(t) - V(t)|$ between two neighboring trajectories $\tilde{V}(t)$ and $V(t)$ in Fig. 3.4 before and after a spike and reset event has taken place in both the perturbed and the unperturbed dynamics. To assess the impact of these two processes, we first note that in an autonomous, freely oscillating neuron

they need to cancel each other such that perturbations neither grow or shrink on average and the LE is zero. We then note that the additive inhibitory input spikes do not affect perturbations but prolong the subthreshold evolution between spikes. The impact of the subtheshold evolution therefore dominates, and perturbations in XIF neurons grow over time, while they shrink in LIF neurons. This does not depend on the specifics of the LIF and XIF dynamics but is a consequence of the curvature of the rise function and the additive inhibitory spike inputs.

The gained intuitive understanding is made precise in App. B.1 by considering the growth or shrinkage of infinitesimal perturbations and the resulting LE. For this, we describe the neuron's dynamics by a sequence of discrete maps from the state (membrane potential) shortly after generation of a spike at time t_k to the state shortly after generation of the next spike at time t_{k+1}. The associated linear evolution of infinitesimal perturbations is then given by "single spike Jacobians" [65, 83]. For the effective single neuron dynamics we consider here, they reduce to scalar factors

$$J_{\mathrm{mf}}(t_{k+1}) = \lim_{\varepsilon \searrow 0} \frac{\partial V(t_{k+1}+\epsilon)}{\partial V(t_k+\epsilon)} = \exp\left(\frac{\gamma}{\rho_{\mathrm{free}}} - \gamma(t_{k+1}-t_k)\right) \tag{3.20}$$

with the neuron's leak or anti-leak rate γ and its free firing rate ρ_{free} [Eq. (3.4) or (3.5)]. We note that in Eq. (3.20) the first term in the exponent originates from the neuron's spike and reset (at time t_{k+1}) while the second term results from the subthreshold evolution of the membrane potential (see App. B.1).

The asymptotic growth rate of infinitesimal perturbations and thus the single neuron mean-field LE is given by the long-term average of the logarithms of Eq. (3.20),

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

with the long-term average firing rate ρ of the neuron [see also Eq. (B.6)]. This result confirms the intuitive understanding that for unperturbed spike inputs the LE of a single neuron depends (i) on the growth rate during subthreshold dynamics and (ii) on the prolongation ($\rho < \rho_{\mathrm{free}}$) or shortening ($\rho > \rho_{\mathrm{free}}$) of the time between spiking relative to the free neuron dynamics. In particular, we obtain $\lambda_{\mathrm{mf}} = 0$ in the free case, while $\lambda_{\mathrm{mf}} = -\gamma$ if the neuron is not spiking at all. In our inhibitory networks, we have $\rho < \rho_{\mathrm{free}}$ and hence $\lambda_{\mathrm{mf}} > 0$ for XIF and $\lambda_{\mathrm{mf}} < 0$ for LIF neurons. For networks in the dynamically balanced state, the actual firing rate is much smaller than the firing rate of a neuron if only external excitation is present. Since in our networks, the latter equals the firing rate of the freely oscillating neuron, we have $\rho/\rho_{\mathrm{free}} \ll 1$. Thus, the mean-field approach shows that the growth rate of small perturbations is mainly governed by the subthreshold dynamics. If we use the mean-field approach to approximate the full spectrum of LEs, it further indicates that a single XIF neuron renders the entire network dynamics unstable and that the number of unstable directions equals the number of XIF neurons in the network, while the number of stable directions equals the number of LIF neurons. This, however, does not give rise to a zero LE, which is present for the autonomous dynamics of the full network due to time-translation symmetry [60]. The mean-field spectrum and the stated rule for the number of stable and unstable directions can thus only be an approximation to the spectrum of the full recurrent network dynamics.

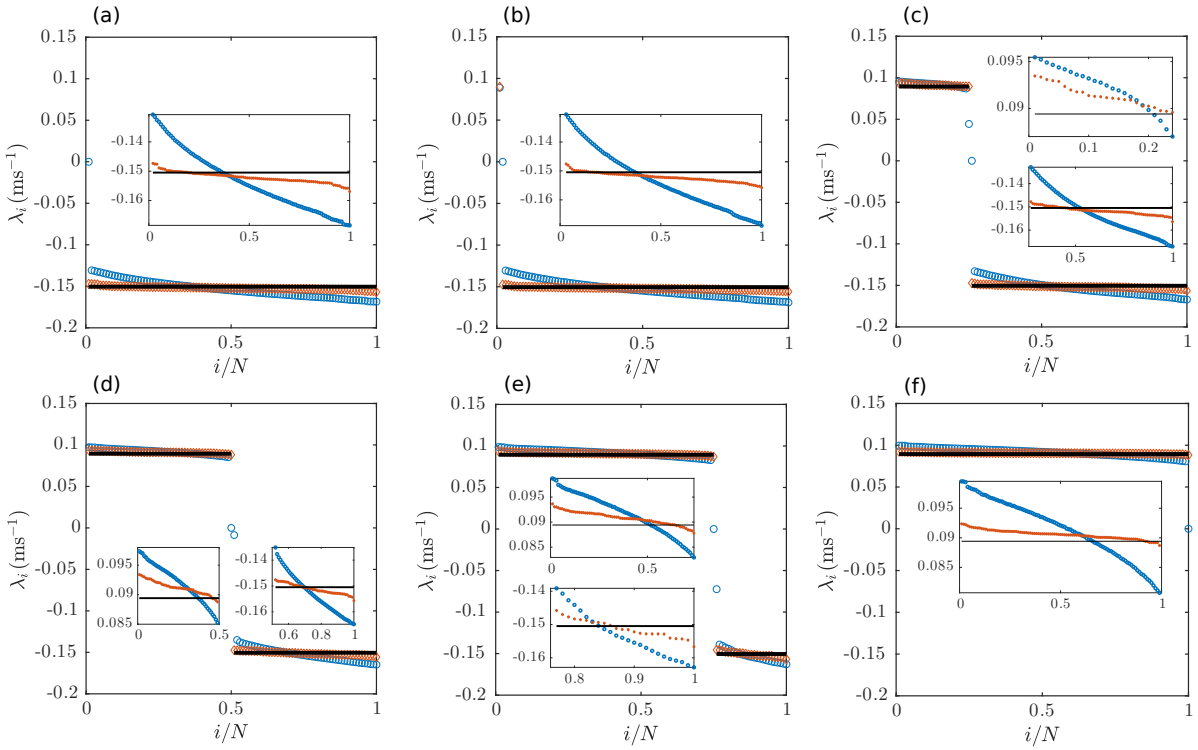

Figure 3.5: Lyapunov spectra of mixed networks of integrate-and-fire neurons. The numbers of LIF and XIF neurons in the networks are (a) $N_{\mathrm{LIF}} = 100$ and $N_{\mathrm{XIF}} = 0$, (b) $N_{\mathrm{LIF}} = 99$ and $N_{\mathrm{XIF}} = 1$, (c) $N_{\mathrm{LIF}} = 75$ and $N_{\mathrm{XIF}} = 25$,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

Employing Eq. (3.21) for each neuron together with the self-consistent firing rate ρ obtained from Eqs. (3.18) and (3.19) gives a fully analytical approximation of the network's spectrum of LEs. Since all LIF or XIF neurons have the same leak or anti-leak rate, γ_{LIF} or γ_{XIF}, respectively, the spectrum consists of N_{LIF} identical negative and N_{XIF} identical positive exponents (Fig. 3.5). In our finite size recurrent networks, the individual neurons' firing rates differ form the prescribed self-consistent rate and display a distribution of values. The latter is presumably due to remaining quenched variability originating from the random connectivity. To account for this distribution, we use for each neuron its numerically measured firing rate in place of ρ in the formula for the mean-field LE, Eq. (3.21). Sorting the resulting LEs yields a further approximation of the Lyapunov spectrum, which we also show in Fig. 3.5.

3.4.2 Spectrum of Lyapunov exponents of the recurrent network dynamics

To numerically compute the spectrum of Lyapunov exponents of the full recurrent network dynamics, we need to take into account the spreading of perturbations in the network. For this, we first derive the single spike Jacobian matrix $J_{ij}(t_{k+1})$, $i, j \in \{1, \ldots, N\}$ [65, 83], which describes the evolution of infinitesimal perturbations of the full network's state between two subsequent spike events, where we

now consider the spikes of all neurons in the network. This Jacobian matrix is defined as the linear map of infinitesimal voltage perturbations (tangent vectors) $\delta V_i(t)$, $i \in \{1, \ldots, N\}$, from a time $t_k^+ = t_k + \varepsilon$ shortly after a spike event to a time $t_{k+1}^+ = t_{k+1} + \varepsilon$ shortly after the next spike event (with arbitrarily small $\varepsilon > 0$), where t_k, $k \in \mathbb{N}$, denote now all spike times of the network dynamics. In App. B.2, we show that the single spike Jacobian matrices for our network dynamics (3.1) can be expressed as

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

where $l \in \{1, \ldots, N\}$ is the index of the neuron emitting the spike at time t_{k+1}. It holds for both LIF and XIF neurons in the network; the neuron type enters through the parameters γ_i and I_i^{ext}. The step-function voltage dependence, Eq. (3.3), for XIF neurons has the effect that, if the neuron's membrane potential $V_i(t_{k+1}^-)$ is below V_{cutoff} immediately before receiving a spike, the spike will not affect the neuron's potential and thus it will also not transmit an infinitesimal perturbation.

Owing to the simple form of the single spike Jacobians (3.22), we can find an analytical expression for the sum of the LEs of the full network dynamics. This sum yields the average rate of expansion or contraction of infinitesimal phase space volume [60]. The result in terms of the individual neurons' firing rates ρ_i in the network is exact. It allows one to analytically compute the Lyapunov exponents for a two-neuron network and offers a test for the accuracy of their numerical estimates in larger networks.

The sum of LEs equals the long-term average of the logarithms of the determinants of the Jacobians [60]. The single spike Jacobian matrices (3.22) have the specific form of a sum of a diagonal matrix and a matrix with a single nonzero column, which allows us to compute their determinants $\det\left(J_{ij}(t_k)\right) = \exp\left(\gamma_l/\rho_{\mathrm{free},l} - \sum_{i=1}^N \gamma_i (t_k - t_{k-1})\right)$ in App. B.2 with the help of the matrix determinant lemma. We recall that $l \in \{1, \ldots, N\}$ is the index of the neuron that spikes and resets at time t_k while $\rho_{\mathrm{free},l}$ is the free firing rate of this neuron [Eq. (3.4) or (3.5)]. We thus have for the long-term average of the logarithms:

$$\begin{aligned} \sum_{i=1}^N \lambda_i &= \lim_{L\to\infty} \frac{1}{t_L} \sum_{k=1}^L \ln \left|\det\left(J_{ij}(t_k)\right)\right| \\ &= \lim_{L\to\infty} \frac{1}{t_L} \sum_{k=1}^L \left(\frac{\gamma_l}{\rho_{\mathrm{free},l}} - \sum_{i=1}^N \gamma_i (t_k - t_{k-1}) \right) \\ &= \lim_{L\to\infty} \left(\sum_{i=1}^N \frac{\gamma_i}{\rho_{\mathrm{free},i}} \frac{L_i}{t_L} - \sum_{i=1}^N \gamma_i \frac{t_L - t_0}{t_L} \right) \\ &= \sum_{i=1}^N \frac{\gamma_i}{\rho_{\mathrm{free},i}} \left(\lim_{L\to\infty} \frac{L_i}{t_L} \right) - \sum_{i=1}^N \gamma_i \end{aligned} \tag{3.23}$$

with $t_0 = 0$. Here, $L_i \in \mathbb{N}_0$ denotes the number of spikes of neuron i up to time t_L, the time of the Lth

spike event in the network. The limit in Eq. (3.23) equals the long-term average firing rate of the neuron,

$$\lim_{L\to\infty} \frac{L_i}{t_L} = \rho_i \,.$$

Hence, we can express the sum of LEs of the full network dynamics and therefore the average rate of infinitesimal volume expansion or contraction exactly as

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

Remarkably, this exact result agrees with the sum of the effective single neuron mean-field LEs, $\sum_{i=1}^{N} \lambda_{\mathrm{mf},i}$, where $\lambda_{\mathrm{mf},i}$ is given by Eq. (3.21) with the neuron's firing rate ρ_i. Our mean-field approach to the Lyapunov spectrum therefore leads to the exact rate of phase space volume expansion or contraction. We also note that the mean-field approach corresponds to ignoring the off-diagonal entries of the single spike Jacobian matrices (3.22) while keeping the spike times of the asynchronous irregular activity in the recurrent balanced network.

The single spike Jacobians (3.22) allow us to numerically compute the largest LE and the full Lyapunov spectrum [60, 65, 78] in event-based simulations of the network dynamics. In short, for the largest LE, one iterates an initial random perturbation vector by multiplication with the sequence of single spike Jacobian matrices along a trajectory, stores its growth or shrinkage every few steps and thereafter renormalizes it to its initial magnitude. The long-term average rate of exponential growth or shrinkage equals the largest Lyapunov exponent λ_1. For the full spectrum, a system of N orthonormal perturbation vectors is iterated simultaneously. Every few steps, one reorthogonalizes the vectors, always in the same order, and records the growth or shrinkage of the corresponding orthogonal vectors. Thereafter, the vectors are renormalized. The long-term average growth rate of the first vector then equals λ_1, that of the second equals λ_2, etc. Engelken [78] suggested an efficient method to compute the Lyapunov spectrum and applied it to large networks; we use some of the ideas in our implementation.

For networks consisting only of LIF neurons, we find in agreement with previous work [76, 77, 82, 179] and our mean-field approach that the largest nonzero LE is negative [Fig. 3.5(a)]. However, the presence of already a single XIF neuron in the network renders the largest LE positive [Fig. 3.5(b)] indicating chaotic dynamics. The numerical computations of the Lyapunov spectrum demonstrate that the destabilization of a network by a single XIF neuron is a special case of a general rule, namely that each XIF neuron introduces about one positive LE (Fig. 3.5). The trivial zero exponent following from time-translation symmetry is an exception to the rule. The numerical results suggest that it replaces a negative exponent if there are more LIF than XIF neurons in the network and a positive exponent otherwise. The computed spectra also show good quantitative agreement with our mean-field approximation of the Lyapunov spectrum, in particular the nontrivial LEs (except a single one) are close to either $-\gamma_{\mathrm{LIF}} < 0$ or $-\gamma_{\mathrm{XIF}} > 0$. But even if we use the numerically measured firing rates in Eq. (3.21) and sort the resulting mean-field LEs, some deviations remain, see Fig. 3.5 insets. These deviations show that the spreading of perturbations in the recurrent network through the transmission of infinitesimal perturbations between neurons has an effect on their growth and hence on the shape of the Lyapunov spectrum.

3.5 Discussion

In this chapter, we investigate the self-consistent firing rates and membrane potential distributions as well as the dynamical stability properties of balanced inhibitory networks containing both conventional LIF and novel XIF model neurons, where XIF neurons accelerate towards spiking. We develop two analytical mean-field approaches, one for the statistics and one for the dynamical stability properties; simulations of the network dynamics yield additional features and a better understanding of the mean-field approximations.

We investigate spiking networks in a dynamically balanced state. To establish it in our mixed networks, we introduce a voltage dependence in the synaptic input of the XIF neurons: below a certain membrane potential, the inhibitory input spikes have no impact. This simple model of a conductance-based synaptic coupling prevents XIF neurons from switching off and provides a good-natured nonlinearity, which leaves the dynamics analytically tractable.

The balanced state is typically investigated using network models with an excitatory and an inhibitory neuron population or with a single population of hybrid excitatory-inhibitory or inhibitory neurons [26, 27, 65, 83, 194, 195]. While detailed models of small circuits with specific abilities such as central pattern generators commonly consider multiple types of neurons [196], studies on the impact of mixed populations containing multiple neuron types on the collective dynamics of larger networks are rare. Savin, Ignat, and Muresan [197] simulated networks with excitatory and inhibitory populations containing resonator and integrator type neurons. These mixed networks both persistently generated activity and quickly changed their overall firing rate in response to inputs, thereby combining abilities of their pure counterparts. Other works considered models for working memory and visual processing with different types of inhibitory neurons that are grouped into distinct populations with different connectivities [198, 199].

We first characterize the collective dynamics of inhibitory mixed LIF and XIF networks from a statistical perspective, adopting a shot-noise approach that accounts for the finite rate and amplitude of the spike inputs from the other neurons [85, 86, 183, 189]. We extend this approach to XIF neurons and derive their steady-state firing rate and membrane potential probability density. In contrast to the case of LIF neurons, the resulting stationary continuity equation needs to be integrated numerically, due to the voltage dependence in the XIF inputs. We apply the results to obtain neuron parameters that lead to equal self-consistent firing rates for the LIF and XIF neurons in our mixed networks. The networks exhibit asynchronous irregular spiking activity, indicating a dynamically balanced state for any ratio of the numbers of LIF and XIF neurons.

We then study the dynamical stability properties of the mixed networks. While the irregular dynamics of inhibitory balanced networks of LIF neurons are stable against infinitesimal and small perturbations [76, 77, 82, 83, 179], we find that introducing already a single XIF neuron gives rise to a positive largest LE indicating chaotic dynamics, in contrast to the robustness against introducing excitatory connections [76, 77]. We give an analytical argument for this finding and expand it to a mean-field approximation of the entire Lyapunov spectrum. Heuristically, the destabilizing effect of excitatory spike inputs in LIF neurons can be compensated by inhibitory ones, if the latter dominate and, thus, the neurons spend on average longer times between spiking compared to their "free" oscillation period in the absence of recurrent inputs. In contrast, inhibitory spike inputs always destabilizes an XIF neuron leading to the exponential growth of infinitesimal perturbations, which then can spread in the recurrent network. In the considered randomly connected networks exhibiting irregular dynamics, we do not expect that infinitesimal perturbations transmitted via input spikes from other neurons will cancel the growth of a

XIF neuron's perturbation. We note that in the phase representation of LIF neurons used in Ref. [77], in contrast to our voltage representation, an excitatory input spike explicitly increases a perturbation, while an inhibitory one decreases it, unless the excitatory input is suprathreshold [200, 201].

Few studies have so far obtained a large part or the full spectrum of LEs of balanced spiking networks. They considered a single homogeneous or an excitatory and an inhibitory neuron population [65, 78, 83, 202–204]. We analytically approximate and numerically compute the full spectrum for mixed networks of inhibitory LIF and XIF neurons. Interestingly, it separates into two parts, in contrast to the ones reported previously, including those for networks with separate excitatory and inhibitory populations. Our mean-field approximation states that the number of negative (positive) LEs is equal to the number of LIF (XIF) neurons in the network. Since the underlying mechanism does not depend on details of the neuron models, we expect this to hold for any types of neurons with purely concave or convex rise functions [200].

The mean-field analysis further indicates that the LEs are approximately determined by the neurons' leak or anti-leak rates and the ratios of their actual firing rates to their freely oscillating spike frequencies. In our networks, the LEs are thus largely independent of the collective dynamics and rather reflect properties of individual neurons. In particular, this implies that the growth rates of perturbations do not change with network size. It further implies that in the dynamically balanced state, where the actual firing rates are much smaller than the corresponding free rates, the LEs are mainly determined by the single neuron leak rates; Monteforte and Wolf [83] report a similar finding in large balanced networks of LIF neurons with high indegree. The result is a consequence of the linear subthreshold dynamics of the neurons, which means that the growth or shrinkage of perturbations between spikes is independent of the neuron's membrane potential. We note that Coombes [205] defined the Lyapunov spectrum as consisting of mean-field LEs in a numerical study on networks of LIF neurons.

The presence of discrete events in the dynamics and the potentially large impact of changing their order could, in principle, make it difficult to transfer insights gained from infinitesimal perturbations to finite ones. Jahnke, Memmesheimer, and Timme [77, 82] studied the evolution of finite-size perturbations in pure LIF network models with stable dynamics and showed that small finite-size perturbations decay exponentially fast, while the minimal perturbation leading to a change in event order decreases only algebraically. Thus, for sufficiently small initial perturbations, the probability of a change in event order goes to zero and no difficulties occur. For unstable dynamics, we expect generic changes of event order to be an additional source of deviations between trajectories, so that small finite-size perturbations grow as fast and larger ones at least as fast as their infinitesimal counterparts. We therefore focus mostly on linear stability analysis here.

Our results show that by choosing an appropriate number of LIF and XIF neurons, one can straightforwardly construct spiking balanced networks with a desired number of stable and unstable directions. The associated covariant Lyapunov vectors indicate these directions in the network's phase space. To exploit the stability or instability of the spiking dynamics against perturbations in different directions, we can choose the input weights such that meaningless inputs or perturbations occur in stable directions, while meaningful ones have a component in an unstable direction; the former perturbations are suppressed while the latter are amplified. Our mixed networks thus combine the computational capabilities of purely stable and purely unstable networks. It is tempting to speculate that in the brain, the combination of different neuron types could shape the phase-space structure and lead to combinations of computational capabilities that can be selected with different inputs. Rather than choosing the input weights by hand, plasticity rules for spiking networks in the brain, as well as for future artificial spiking neural networks, may allow them to be found by learning.

CHAPTER 4

Critical avalanches in developing networks of stochastically spiking neurons

This chapter is based on the following published article [206]:

Y. F. Kalle Kossio, **S. Goedeke**, B. van den Akker, B. Ibarz, and R.-M. Memmesheimer
Growing critical: Self-organized criticality in a developing neural system
Phys. Rev. Lett. **121** (2018) 058301 DOI: 10.1103/PhysRevLett.121.058301

In the previous two chapters, we investigated transitions between stable and chaotic collective dynamics in neural networks, where chaos means that the microscopic dynamics exhibit sensitive dependence on the initial conditions. The global network activity, however, remained stable; for example, the response of the network-averaged rate to a small perturbation would be small. In this chapter, we study spiking neural networks close to another transition of their collective dynamics. We consider recurrent networks with excitatory synaptic couplings, in which the spiking activity of an excitatory neuron increases the likelihood of its postsynaptic neurons to be active as well. If a single spike of an excitatory neuron induces on average more than one further (excitatory) spike in the network, the global network activity can become unstable due to positive feedback. Close to this instability, the network dynamics may show critical behavior: bursts or avalanches of activity occurring on many scales, with characteristics governed by power laws. Experiments found such avalanches in various neural systems. According to these findings, the underlying neural network dynamics may operate in a critical regime.

The positive feedback instability happens at a critical value of the overall excitatory coupling strength in the network, which leads to the question how biological neural systems can self-organize into a regime close to this transition point. We therefore turn in this chapter to networks with plastic synaptic connections, meaning that they can change over time. In general, the plasticity of synaptic connections underlies changes that occur during development and learning in neural systems. Here, we investigate a simple plastic network model for developing neural networks and demonstrate how these may grow into a critical regime. The network growth process is guided by homeostatic regulation: Neurons extend and form more excitatory synaptic connections to reach a certain level of spiking activity. If this target level is high compared to the spontaneous activity of the neurons, the network grows towards a stationary

state where most spikes are induced by other spikes in the network. In this case, the network is operating close to the positive feedback instability.

Neuronal avalanches have been observed in several activity modalities. We focus on avalanches of spikes as fluctuations of the ongoing spiking activity. They are initiated by spontaneous spikes of the neurons, which occur randomly at a low rate. To model these fluctuations in a simple manner, we work with networks of stochastically spiking neurons. Each neuron generates spikes with an instantaneous firing rate that is linearly excited by spikes from other neurons in the network. We show that under homeostatic growth of the synaptic connections, such networks develop into a stationary state in which the mostly self-sustained spiking activity consists of avalanches with power-law distributed characteristics. The growth process leads robustly to a point slightly below the positive feedback instability of the linear activity dynamics and does not require fine-tuning.

My first important contribution to the article was the reduction of the network's spiking activity to a scalar self-exciting Hawkes point process. The dimensionless self-excitation coupling strength of this process equals the average number of further spikes that a single spike directly adds to the network activity; this number is known as the "branching parameter". In the stationary state, after average homeostatic network growth has ceased, the self-excitation coupling strength, and hence the branching parameter, is determined only by the spontaneous firing rate and the homeostatic target firing rate. If the target rate is much larger than the spontaneous rate, the coupling strength approaches the critical value 1 and the Hawkes process operates close to critical self-excitation.

Neuronal avalanches of spikes in our network model correspond to clusters of points generated by the self-exciting Hawkes process describing the network activity. My second and crucial contribution was the analytical derivation of the probability distribution of avalanche durations. Together with the distribution of avalanche sizes, the duration distribution is a characteristic often measured in experiments. For the derivation, I expressed the distribution function as the Laplace transform of the instantaneous rate of a transient self-exciting Hawkes process that is initiated by a single spike. This Laplace transform has recently been considered in applied mathematics. Using this connection, I obtained the duration distribution from the solution of a nonlinear ordinary differential equation. The result holds for any value of the self-excitation coupling strength less than or equal to the critical value. Near criticality, perturbation analysis of the differential equation yields the exponent 2 of the power-law tail of the duration distribution. This exponent and the power-law exponent $3/2$ of the avalanche size distribution agree with the corresponding exponents of a critical branching process. These values have also been reported for experimentally observed neuronal avalanches. Since Hawkes processes are widely applied to model endogenous excitation by point-like events, the result generalizes to other fields of science and is one of the main contributions of the article to the scientific progress.

Finally, I wrote to a large extent the parts of the article presenting the above-mentioned results and I contributed to the writing of the other parts. The following sections of this chapter contain the published manuscript with only minor clarifications and changes to adapt the contents to the scope and the format of the thesis.

4.1 Introduction

A hallmark of systems at criticality is the variability of their responses to small perturbations. While small responses are most likely, the probability of large, system-size effects is non-negligible. Various natural and model complex systems show similar behavior [207]. One explanation is that they drive

themselves close to a critical state ["self-organized criticality" [208, 209]]. The dynamics of such systems are characterized by "events" or "avalanches". Their sizes and durations follow power-law distributions, frequently with exponents $3/2$ and 2, respectively, indicating an underlying critical branching process [210–213]. Apparent critical dynamics, "neuronal avalanches", in biological neural networks were first reported by Beggs and Plenz [214, 215]. It has been suggested that they foster information storage and transfer [216, 217]. Experimental studies often report power-law size and duration distributions with exponents $3/2$ and 2. They further indicate that neuronal avalanches emerge during development [218–221], suggesting that neural networks develop into a critical state.

The development of neural networks is determined by an interplay of genetic determinants and environmental influence. Of pivotal importance is neural activity [222, 223]. As a general rule, neurons with low activity level extend their neurites and form more activating connections, while highly active cells reduce these [224–226]. Thereby, neurons maintain their average activity at a particular level by homeostasis [51, 56, 58].

Computational models for avalanches in neural systems rely on static, tuned connectivity [220, 227], on short-term synaptic plasticity [228, 229], or on long-term network changes [230–233]. Here, we propose and analyze a continuous-time spiking neural network model belonging to the third class. The avalanche dynamics follow from a network growth process towards a critical state, which uses local information only [234–236]. The model is rooted in previous models for neural network development [57, 230, 232], but sufficiently simple to be analytically tractable.

4.2 Growing neural network model

We study growing networks of stochastically spiking model neurons. Like biological neurons, our model neurons communicate by sending and receiving spikes in continuous time. Spiking is stochastic, according to an inhomogeneous Poisson point process with instantaneous rate $f_i(t)$ for neuron i [13, 14, 230, 237]. In isolation, neurons have a low spontaneous rate f_0, e.g., due to spontaneous synaptic release or channel fluctuations [238, 239]. A spike from neuron j increases $f_i(t)$ by the time-dependent coupling strength $gA_{ij}(t^-) = \lim_{\varepsilon \searrow 0} gA_{ij}(t-\varepsilon)$, where, in our simple network growth model, we use its value just before the spike has occurred. The increment decays exponentially with time constant τ, which accounts for relaxation due to leak currents. The couplings are excitatory; this is the dominant connection type in developing neural systems [57]. Taken together, the instantaneous rates follow the linear dynamics

$$\tau \dot{f}_i(t) = -f_i(t) + f_0 + \tau g \sum_j A_{ij}(t^-) \sum_{\hat{t}_j} \delta\left(t - \hat{t}_j\right), \tag{4.1}$$

where $\hat{t}_j$ denotes the spike times of neuron j (δ is the Dirac delta distribution). For simplicity, we assume that all neurons have the same parameters. For constant coupling strengths, the network dynamics form a multivariate Hawkes point process [10, 237, 240].

Since neurons are commonly arranged in single or stacked layers, we represent neurite extents by disks with radii $R_i(t)$, with centers, representing cell somas, randomly and uniformly distributed in a planar area [57, 230, 241]. Neurons with more neurite overlap can grow more synaptic connections [242, 243]. Therefore, coupling strengths are set proportional to the overlap areas $A_{ij}(t)$ of the disks, with

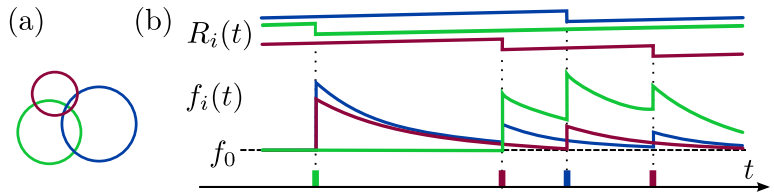

Figure 4.1: Neuron dynamics. (a) Neurons' somas and neurite extents are represented by disks with evolving radii. Coupling strengths are proportional to neurite overlap areas. (b) Neurite radii increase linearly (upper traces), own spiking (lower trace) results in instantaneous shrinkage. Spike arrivals increase the instantaneous firing rate by the coupling strength, it decays exponentially in between (middle trace).

proportionality constant g. We incorporate homeostatic neurite growth by evolving the radii as

$$\dot{R}_i(t) = K\left(1 - \frac{1}{f_{\text{sat}}}\sum_{\hat{t}_i}\delta(t-\hat{t}_i)\right), \tag{4.2}$$

which is illustrated in Fig. 4.1. Between spikes of neuron i, the neurite radius $R_i(t)$ grows linearly with rate K. When the neuron spikes, $R_i(t)$ shrinks by a constant amount K/f_{sat}, which determines the saturation rate f_{sat} at which growth and shrinkage equilibrate, i.e., the homeostatic target level. There are no self-connections. Growth takes much longer than decay of activity, $1/K \gg \tau$ (spatial scales of the network are of order one). Furthermore, we assume $f_{\text{sat}} \gg f_0$, in agreement with experiments [221, 239]. Neurons that are not spontaneously active would reduce the relevant average f_0. The growth model is biologically inspired; it is a simplification of previous growth models [56, 57, 230, 232]. However, many slow homeostatic processes [51, 58, 244] with $f_{\text{sat}} \gg f_0$ will yield similar results.

Figure 4.2 presents a simulation of the network growth dynamics. Initially, the neurons are mostly isolated. Over time, they extend their neurites, form connections, and develop a network. At intermediate stages, neurites and overlaps can overshoot [57, 221, 232]. Finally, the network grows into a stationary state in which average growth ceases. In the following, we investigate this state.

4.3 Stationary state dynamics

When neuron i's time-averaged firing rate $\bar{f}_i$ reaches f_{sat}, its average growth stops [Eq. (4.2)]. Our networks grow into a stationary state, where $\bar{f}_i = f_{\text{sat}}$ for all i. We first compute the average number of spikes that a spike directly causes: Identical $\bar{f}_i$ imply identical time-averaged total overlaps $\sum_j \bar{A}_{ij} = \bar{A}_i = \bar{A}$ and input coupling strengths. Time averaging Eq. (4.1), $\bar{f}_i = f_0 + \tau g \sum_j \bar{A}_{ij}\bar{f}_j$ [here and henceforth we neglect the small fluctuations of $A_{ij}(t)$ around $\bar{A}_{ij}$], and inserting f_{sat} yields $\tau g\bar{A} = 1 - f_0/f_{\text{sat}}$. A spike of neuron j at time $\hat{t}_j$ adds $gA_{ij}(\hat{t}_j^-)e^{-(t-\hat{t}_j)/\tau}\Theta(t-\hat{t}_j)$ to $f_i(t)$ [Eq. (4.1), Θ is the Heaviside step function], such that the number of additionally induced spikes in neuron i is Poisson distributed with mean $\tau gA_{ij}(\hat{t}_j^-)$. Averaged over the stochasticity of spiking, each spike thus generates in total

$$\sigma = \tau g\sum_i \bar{A}_{ij} = \tau g\bar{A} = 1 - \frac{f_0}{f_{\text{sat}}} \tag{4.3}$$

spikes in the network, where we used the symmetry of overlaps, $A_{ij} = A_{ji}$, $\sum_i \bar{A}_{ij} = \sum_j \bar{A}_{ij} = \bar{A}$.

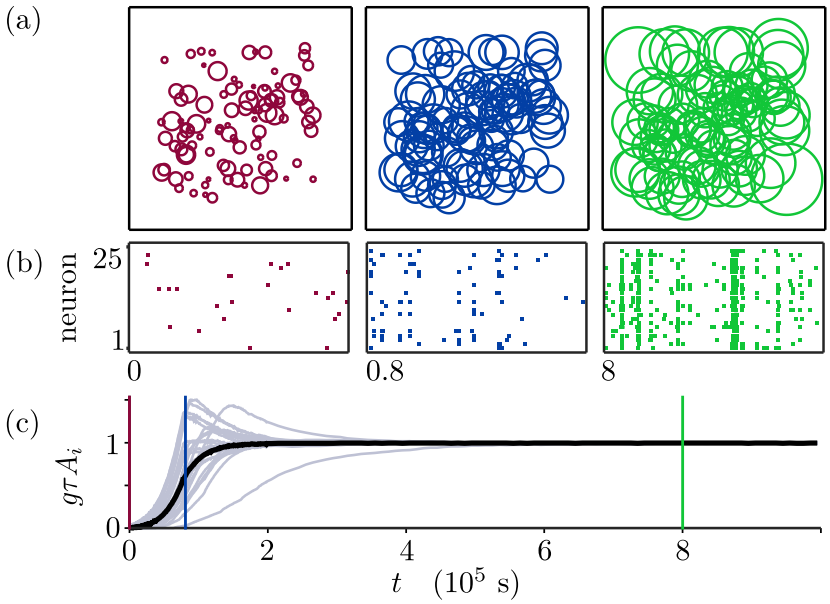

Figure 4.2: Network dynamics. (a) Extents of neurites. (b) Spikes generated by 25 sample neurons (100 s windows). (c) Scaled total overlaps of 25 sample neurons (gray) and the population average (black) as a function of time. For (a) and (b) from left to right: initial state (red), state with growth on average (blue), stationary state (green). Color coded vertical lines in (c) indicate the three different time points in (a) and (b).

Equation (4.3) for the branching parameter σ holds independently of network activity and neuron identity, due to the linearity of Eq. (4.1) and the homogeneity of parameters. In particular, σ equals also the time and population average number of induced spikes ($\propto f_{\text{sat}} - f_0$) per spike ($\propto f_{\text{sat}}$).

The independence of spike offspring generation from other spikes allows us to understand the dynamics as a branching process with branching parameter σ. More specifically, we have an age-dependent or Crump-Mode-Jagers branching process [245]: Individuals (spikes) generate offspring at an age-dependent rate. Neuronal avalanches are trees of offspring, started by a spontaneous spike. For their overall size only the distribution of the single spike offsprings matters. It is Poissonian with parameter σ. The avalanche sizes s therefore follow the Borel distribution [246],

$$P(s) = \frac{(s\sigma)^{s-1} e^{-s\sigma}}{s!}. \tag{4.4}$$

We apply Stirling's approximation to obtain

$$P_{\text{appr}}(s) = \frac{1}{\sqrt{2\pi\sigma}} s^{-3/2} e^{-(\sigma - \ln\sigma - 1)s}, \tag{4.5}$$

explicitly highlighting the power-law tail with exponent $3/2$ of a critical branching process for $\sigma = 1$ [210–213]. The distribution is shown in Fig. 4.3(a). For a subcritical process ($\sigma < 1$), Eq. (4.5) is a power law with exponential cutoff around $s_c(\sigma) = (\sigma - \ln\sigma - 1)^{-1}$. It signifies subcritical dynamics [210, 211, 247], not a finite size effect [209, 211]; the size distribution is independent of the number of neurons. Equation (4.5) inherits the good quality of Stirling's approximation [248], with a relative error of about $1/(12s)$.

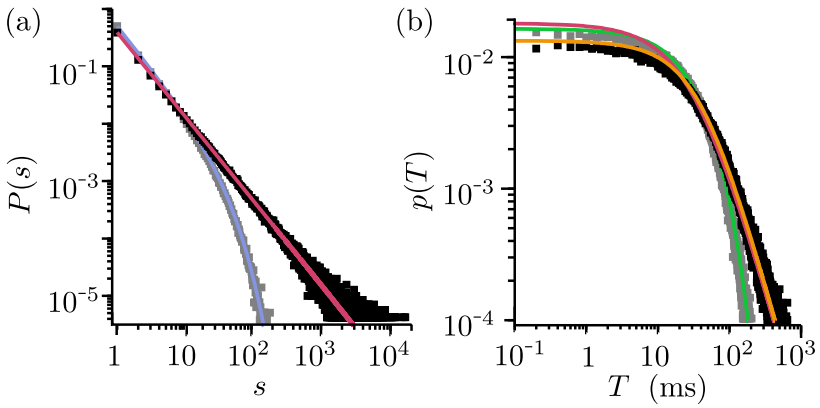

Figure 4.3: Avalanche sizes and durations. (a) Analytical size distributions Eq. (4.5) (discrete points connected) and simulation results for subcritical ($f_{\text{sat}} = 0.04\,\text{Hz}$, $\sigma = 0.75$, $t_{\text{bin}} = 30\,\text{ms}$, blue and gray) and near-critical ($f_{\text{sat}} = 2\,\text{Hz}$, $\sigma = 0.995$, $t_{\text{bin}} = 45\,\text{ms}$, red and black) states. Equation (4.4) yields visually indistinguishable analytics. (b) Analytical duration distributions Eq. (4.10) and simulation results, for subcritical (green and gray) and near-critical (orange and black) states, and closed-form approximation Eq. (4.12) (red).

The heights of Crump-Mode-Jagers trees, i.e., the temporal differences T between their first and last individuals, represent the durations of the corresponding neuronal avalanches. In the following, we derive their probability density $p(T)$. Because of the additivity of Poisson processes, the superposition of all neurons' spike trains can be described as an inhomogeneous Poisson process with rate $f(t) = \sum_i f_i(t)$. Summing Eq. (4.1) over i and inserting $\bar{A} = \sigma/(g\tau)$ [Eq. (4.3)] yields $\tau\dot{f}(t) = Nf_0 - f(t) + \sigma\sum_{\hat{t}}\delta(t-\hat{t})$ with the number of neurons N. $\hat{t}$ are the neurons' spike times; they occur with instantaneous rate $f(t)$. The spiking dynamics may thus be interpreted as a self-exciting Hawkes process. It is Markovian due to the exponentially decaying impact kernel [249, 250]. The spontaneous background rate Nf_0 initiates avalanches. To determine their durations, we consider an analogous process with instantaneous rate $f_a(t)$ and without spontaneous spiking, which is initiated at $t = 0$ by a single spike,

$$\tau\dot{f}_a(t) = -f_a(t) + \sigma\sum_{\hat{t}_a}\delta(t-\hat{t}_a),\;\; f_a(0) = \frac{\sigma}{\tau}. \tag{4.6}$$

The duration of an avalanche is the time T of this process' last spike. The probability that it has occurred before t gives the distribution function $P(T \le t)$ of durations. We first compute this probability conditioned on the instantaneous rate $f_a(t)$ at the end of the considered interval:

$$\begin{aligned} P(T \le t \mid f_a(t)) &= P(\text{no spike in } (t,\infty) \mid f_a(t)) \\ &= e^{-\int_t^\infty f_a(t')dt'} = e^{-\tau f_a(t)}, \end{aligned} \tag{4.7}$$

where we use that the process behaves like a Poisson process with exponentially decaying rate, if no spike is generated. Averaging over $f_a(t)$ yields

$$\begin{aligned} P(T \le t) &= \int_0^\infty P\left(T \le t \mid f_a(t)\right) p\left(f_a(t)\right) df_a(t) \\ &= E\left(e^{-\tau f_a(t)}\right). \end{aligned} \tag{4.8}$$

$E(\cdot)$ denotes the expectation value over the process, Eq. (4.6). Importantly, Eq. (4.8) shows that $P(T \leq t)$ equals the Laplace transform of the random variable $f_a(t)$, evaluated at the decay time constant τ. This Laplace transform has recently been derived [251–253]. Inserting our parameters yields $E(e^{-\tau f_a(t)}) = e^{\sigma a(t)/\tau}$, where $a(t)$ satisfies the ordinary differential equation

$$\dot{a}(t) = -a(t)/\tau + e^{\sigma a(t)/\tau} - 1, \; a(0) = -\tau. \tag{4.9}$$

The resulting $P(T \leq t) = e^{\sigma a(t)/\tau}\Theta(t)$ with $\Theta(0) = 1$ has the density

$$p(T) = \sigma \dot{a}(T) e^{\sigma a(T)/\tau}\Theta(T)/\tau + e^{-\sigma}\delta(T)\,, \tag{4.10}$$

which is displayed in Fig. 4.3(b). We can generalize Eq. (4.9) to Hawkes processes with different kernels using the integral equation for cluster duration distributions [254, 255].

We finally approximate $p(T)$ by closed-form expressions with a focus on its tail near criticality. For large t, $P(T \leq t)$ approaches 1, so $a(t)$ approaches 0. Generally, $\sigma a(t)/\tau$ stays between -1 and 0. Expanding $e^{\sigma a(t)/\tau}$ in Eq. (4.9) around $\sigma a(t)/\tau = 0$ to second order,

$$\dot{a}(t) \approx (\sigma - 1)a(t)/\tau + \sigma^2 a(t)^2/(2\tau^2), \; a(0) = -\tau, \tag{4.11}$$

yields closed-form approximations for $a(t)$. In particular, for nearly critical systems with $\sigma \approx 1$, if we neglect the first term on the right-hand side, the solution becomes $a_{\text{appr}}(t) = -2\tau^2/(2\tau + t)$, leading to a probability density

$$p_{\text{appr}}(T) = 2\tau(2\tau + T)^{-2} e^{a_{\text{appr}}(T)/\tau}\Theta(T) + e^{-1}\delta(T), \tag{4.12}$$

which approaches for large T a power law with critical exponent 2. For large t the error in the expansion Eq. (4.11) becomes negligible, $a_{\text{appr}}(t)$ thus has the right slope and $p_{\text{appr}}(T)$ equals $p(T)$ up to a factor [see Fig. 4.3(b)]. We conclude that the duration distribution has a power-law tail with critical exponent 2. Taking the first term in Eq. (4.11) into account, we expect an exponential cutoff around $T_c(\sigma) = \tau/(1 - \sigma)$. Expanding the exponential to third order yields a closed-form distribution that is a good approximation also for small T.

4.4 Simulations

We complement our analytics with simulations to (i) compare the avalanche distributions, (ii) exemplify the irrelevance of connectivity fluctuations, (iii) investigate the spatial spread of avalanches, and (iv) address the robustness of the results. If not stated otherwise, $N = 100$, $\tau = 10\,\text{ms}$ [4], $g = 500\,\text{Hz}$, $f_0 = 0.01\,\text{Hz}$, $f_{\text{sat}} = 2\,\text{Hz}$ [221, 239], somas are placed on unit square, $K^{-1} = 10^6$ s (fast growth to accelerate simulations) [51, 58, 221, 232]. The simulations use an event-based algorithm. Next spike times are determined using inverse transform sampling of the interspike-interval distribution; we avoid nonelementary functions by splitting each neuron's Poisson process into a homogeneous (with rate f_0) and an inhomogeneous one.

An avalanche should be a sequence of offspring spikes of one spontaneous progenitor. To keep contact with the experimental literature, we analyze numerical data by binning time and considering spike sequences that are not separated by an empty bin as one avalanche [214, 256–258]. Our model yields analytical estimates for the probabilities that binning splits the first avalanche spikes or merges them with the next avalanche, as well as for splitting or merging an average avalanche. Keeping them

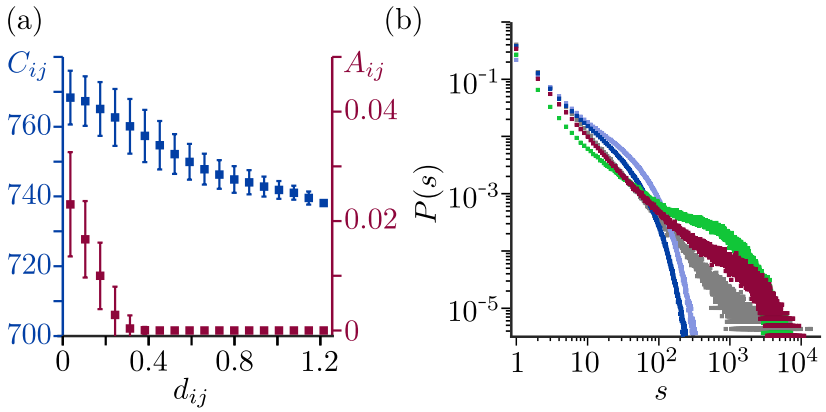

Figure 4.4: (a) Pairwise spike number covariances and overlaps as functions of the intersomatic distances d_{ij} (averages around a particular intersomatic distance, bars: standard deviations), $\sigma = 0.995$, $t_{\text{bin}} = 45\,\text{ms}$. (b) Avalanche size distributions for the same model as in Fig. 3, but with absolute refractory period $\tau_{\text{ref}} = \tau$ (red), $\tau_{\text{ref}} = 4\tau$ (green), and $\tau_{\text{ref}} = 0\,\text{ms}$ (gray) for reference, $t_{\text{bin}} = 45\,\text{ms}$. Size distributions with parameters $\tau = 5\,\text{ms}$, $\tau_{\text{ref}} = 4\tau$, $f_0 = 0.1\,\text{Hz}$, $f_{\text{sat}} = 0.8\,\text{Hz}$, $t_{\text{bin}} = 10\,\text{ms}$ (blue), and $t_{\text{bin}} = 45\,\text{ms}$ (light blue) adapted from Ref. [230].

moderate provides our bin sizes t_{bin} in terms of experimentally accessible quantities ($f_0, \tau, N, f_{\text{sat}}$); see [206, Supplemental Material]. Results are robust against changing t_{bin}.

(i) In all simulations the model reaches a stationary state. The avalanche distributions agree well with the analytically derived ones (Fig. 4.3); the effects of binning and avalanche overlaps are small. We quantitatively test this agreement using the method described in [259]. For both size and duration distributions a pure power law is ruled out, as expected. For the size distribution, a power law with exponential cutoff, see Eq. (4.5), yields a good fit. The analytical values of the power-law exponent, the cutoff $s_c(\sigma)$, and the resulting branching parameter σ are closely matched.

(ii) The fluctuations of $\sum_j A_{ij}(t)$ and the deviations of $\bar{f}_i$ from f_{sat} are small [less than 1%, Fig. 4.2(c)]. Freezing the network ($K = 0$) in the stationary state has no effect on the avalanche statistics: Neuronal growth carries the system close to a critical point, but is not required later on. This is in agreement with self-organized criticality and excludes other mechanisms [260–262].

(iii) To investigate spatial spread near criticality, we compute covariances $C_{ij} = \langle n_i n_j \rangle - \langle n_i \rangle \langle n_j \rangle$ between numbers n_i, n_j of spikes contributed to single avalanches by neurons i, j with various distances. Covariances decay only slowly with distance [Fig. 4.4(a)]. In particular, covariances and thus avalanches spread beyond direct connections.

(iv) To test robustness, we first freeze the network in the stationary state and shuffle the output vectors (columns of the coupling matrix) between neurons. While this alters the network topology and breaks coupling symmetry, it leaves the essential total coupling strengths unchanged. Indeed, we observe little effect on avalanche sizes and durations. Second, we consider moderate nonadditivity of spike impacts. We introduce an absolute refractory period τ_{ref} after a sent spike, during which the neuron cannot spike again. We observe that although the refractory period limits the firing rate, the network reaches a stationary state with the same average individual rate f_{sat} as before: larger overlaps compensate refractoriness. For a refractory period of about τ, which is often biologically plausible [4, 238], the statistics resemble the original one for small and medium size avalanches [Fig. 4.4(b), red vs. gray]. Larger couplings and stop of avalanches lacking available neurons cause an excess of larger avalanches,

followed by a strong reduction. Neurons still frequently contribute several spikes to avalanches. With long refractoriness, little similarity remains [Fig. 4.4(b), green and blue; blue: our model with parameters adapted from Abbott and Rohrkemper [230], calcium variable present in their model does not affect distribution shape].

4.5 Discussion

We suggest an analytically tractable model for neural network growth, which may explain the emergence of subcritical and critical avalanche dynamics. It covers essential features of biological neurons such as operation in continuous time, spiking, leak currents, and network growth. Still, it allows the analytical computation of the avalanche size and duration distributions for subcritical and critical stationary states. Our numerical analysis confirms their validity and robustness and yields additional insight.

Two features are responsible for the emergence of the (near-)critical state (Fig. 4.3): (i) homeostatic growth to attain a saturation rate f_{sat} that is high compared to the spontaneous one f_0 (precise values of f_{sat} and f_0 are irrelevant), and (ii) linear summation of spike impacts. (i) implies that in the stationary state on average each spike generates nearly one successor. This holds for all networks with largely self-sustained activity. Usually, however, branching parameters vary, for example at high network activity spikes generate less offspring. This drives activity excursions back and generates non-power-law distributions [263, 264]. In our networks, (ii) implies that the branching parameter is the same for each spike. Small saturation rates yield subcritical dynamics (Fig. 4.3), strong nonlinearities other deviations [Fig. 4.4(b)]. Our model thus predicts that neural networks may develop criticality already due to their growth, that spontaneous spiking in such networks is low compared to saturation, and that spike effects add rather linearly and are independent of activity level. For example, starburst amacrine cells have radial dendritic trees, interact during development via dendro-dendritic excitatory connections and are reported to generate critical avalanches [220]. Our model predicts that higher precision measurements will reveal deviations as in Fig. 4.4(b), due to the cells' long refractory periods.

Our network model is based on the neurobiologically more detailed ones [57, 230, 232]. Motivated by experiments, van Ooyen and van Pelt [57] propose radial activity-dependent neurite outgrowth steered by calcium dynamics and find convergence to a stationary state for certain parameter ranges. To study avalanches, Abbott and Rohrkemper [230] add stochastically spiking neurons, albeit with long refractoriness and larger f_0/f_{sat}, impeding analytical treatment and causing large deviations from criticality [Fig. 4.4(b)]; Tetzlaff et al. [232] assume antagonistic growth of axons and dendrites and find criticality if a certain fraction of neurons becomes inhibitory; other studies consider more abstract homeostasis and neuron models [244, 265, 266].

Usually, models for neuronal avalanches only allow to estimate size and duration distributions numerically [220, 228, 230, 232, 233]. Eurich, Herrmann, and Ernst [227] obtain an analytical expression of the size distribution for a discrete-time network. In this chapter, we derive size and duration distributions for a continuous-time spiking network model after self-organization. These distributions depend only on the experimentally accessible parameters f_0/f_{sat} and τ, where τ scales the duration distribution. The power-law exponents agree with experimentally found ones and those of simple branching processes [210, 212, 214, 220]. The duration distribution has power-law scaling at the tail [210, 212, 220], a fit to short avalanches [214] would yield different results. Our analytical expressions allow fast parameter scans, delineations of the (near-)critical regime and parameter estimations.

From a general perspective, avalanches in our model are clusters of a self-exciting Hawkes process.

While their size distribution can be straightforwardly computed [Eq. (4.4)], their duration distribution generally requires solving a nonlinear integral equation [254, 255]. Here, we show that for Markovian Hawkes processes it follows from the solution of an ordinary differential equation [Eqs. (4.9) and (4.10)] and we give closed-form approximations. This result may find straightforward application in further fields of science where these processes are employed, for example, to characterize durations of financial market fluctuations [251], earthquakes [267], violence [268], and epidemics [269, 270].

CHAPTER 5

Drifting assemblies in plastic networks of spiking neurons

This chapter is based on the following published article [271]:

Y. F. Kalle Kossio, **S. Goedeke***, C. Klos* and R.-M. Memmesheimer
(* equal contribution)
Drifting assemblies for persistent memory: Neuron transitions and unsupervised compensation
Proc. Natl. Acad. Sci. U.S.A. **118** (2021) e2023832118 DOI: 10.1073/pnas.2023832118

As we have already seen in this thesis, an important determinant of neural network dynamics and function is the synaptic connectivity between the neurons. In the previous chapters, we studied recurrent networks whose connectivity is random or has developed via self-organization without external information. Nevertheless, the study of their dynamics was also motivated from a functional perspective. In this chapter, we consider recurrent neural networks that perform a specific function, which is implemented through their activity dynamics: the representation of associative long-term memories related to information from the external world. For this function, we assume that the synaptic connections in the network have been adapted by means of synaptic plasticity during learning.

Standard models for the neuronal implementation of associative memories are based on so-called Hebbian cell assemblies, ensembles of strongly interconnected neurons that can excite each other. The positive feedback in such assemblies amplifies and completes partial excitation of their neurons. Hence, their activity dynamics can give rise to pattern completion. These models typically assume that stably stored long-term memories are represented by the same neurons over time and that the underlying cell assemblies are therefore static. However, recent experimental results challenge this model. Both synaptic connections and neural representations of external information show apparently undirected changes and fluctuations over time, while learned behaviors and memories remain stable. Such undirected changes in neural representations are also referred to as representational drift. Motivated by these experimental observations, we propose an alternative model for the neuronal implementation of associative memories. It is based on cell assemblies whose constituent neurons can change completely over time. They evolve gradually as coherent structures in the space of neurons of plastic recurrent networks and we call them drifting assemblies. The gradual change allows synaptic plasticity mechanisms to preserve the network

function in a self-organized and unsupervised manner, i.e. without explicit error signals.

In our model, the drift of representations arises through the exchange of neurons that form the assemblies. Over time, a single neuron leaves an assembly by weakening its overall excitatory synaptic coupling strength with the other assembly neurons. If the coupling strength is sufficiently weakened, the neuron will no longer be co-excited when the assembly is active and, therefore, moves out of the associative memory representation implemented by that assembly. Conversely, a neuron that was previously not part of the assembly can newly join it by strengthening the overall coupling between the neuron and the assembly. In this case, the neuron enters the memory representation realized by this assembly. If the exchange of neurons occurs gradually over time and is undirected, the assembly drifts among the available neurons in the network. The continuing drift can lead to complete remodeling of the assembly structure in the network. Thus, after some time, the same memory content may be represented by completely different neurons. Nevertheless, the overall network function and memory are preserved via stable input and output neurons that continue to represent the memory content. The input and output neurons compensate for the drift by adapting their synaptic connections such that each of them maintains its increased excitatory coupling strength to the same drifting assembly, which consists of different neurons over time.

We develop and analyze plastic recurrent network models of spiking neurons that demonstrate the feasibility of drifting cell assemblies and the proposed memory model. The synaptic coupling strengths (weights) of the connections between the excitatory neurons change according to simple biologically plausible spike-timing-dependent plasticity (STDP) rules, which depend on the pre- and postsynaptic activity. The plasticity is Hebbian, meaning that correlated activity leads to weight strengthening. In addition, homeostatic plasticity normalizes each neuron's synaptic weights and thereby introduces competition between them.

In networks of leaky integrate-and-fire (LIF) model neurons, we show that synaptic plasticity in the presence of noisy autonomous network activity can give rise to drifting assemblies while memories persist. We investigate and explain the assembly drift. From the perspective of a single neuron, assemblies act as potential and noise strength wells in a dynamical landscape: When the neuron is part of an assembly, there is a restoring force that reverses perturbations of the overall coupling strength between the neuron and the assembly and thus stabilizes the neuron's assembly membership. Furthermore, the strength of the perturbations is much smaller when the neuron is part of an assembly. In rare events, perturbations are so large that the neuron leaves its assembly well and switches to another assembly. We quantitatively model the dynamics of neuron transitions in a random walk picture. For this, we derive the statistics of different contributions to the weight changes due to STDP induced by the spiking activity in the network. The results explain the random walk's dynamical landscape. We find that strongly correlated spiking in the form of synchronous assembly reactivations plays an important role in the switching dynamics. The weight changes resulting from these assembly reactivations lead both to the potential wells and to the stronger noise in the transition zone between the wells.

Mechanistically, the gradual exchange of neurons underlying assembly drift can be understood as noise-activated switching between metastable states. This suggests that there is a transition from static to drifting assemblies in the network's synaptic weight dynamics. Indeed, we observe in our LIF network model that neuron switching and hence assembly drift ceases when the STDP strength is sufficiently reduced. In this case, the perturbations are too weak to generate neuron switching.

In a modified model, we show that spontaneous (activity-independent) changes of synapses, often referred to as synaptic turnover, can also lead to assembly drift. As a simple model for such turnover, we include spontaneous disappearance and reappearance of excitatory synaptic connections in the network. In particular, since newly appearing synapses have vanishing initial weight, this turnover erases the existing weights over time. Therefore, further ongoing synaptic plasticity is required to maintain the assembly structure. We demonstrate that activity-dependent and homeostatic plasticity is able to preserve the network's assembly structure in the presence of spontaneous synaptic turnover. Simultaneously, we observe assembly drift that is driven by the synaptic turnover. In its absence, the assemblies remain static as in the above-mentioned case with reduced STDP strength.

My first important contribution to the article was the original development and implementation of the LIF network model with STDP, homeostatic normalization and spontaneous synaptic turnover. The neurons receive additional noise inputs, which may originate from the asynchronous irregular activity of neurons that are not explicitly modeled. The noise inputs evoke ongoing spiking activity in the network. As part of this noisy autonomous activity, the assemblies occasionally reactivate by synchronous spiking of their member neurons. The mechanism underlying these assembly reactivations is in fact a positive feedback instability of the excitatory spiking activity in an assembly, similar to the instability discussed in the previous chapter. However, due to the nonlinear spiking neuron model considered here, the assemblies reactivate in a synchronous all-or-none manner. Recurrent inhibition prevents the spreading of excitation in the network. The developed LIF network model exhibits activity dynamics including assembly reactivations that are similar to the dynamics in a network model with much simpler binary neurons operating in discrete time. This allowed us to transfer insights about drifting assemblies obtained with the binary model.

Furthermore, I established drifting assemblies in plastic networks of stochastically spiking linear Poisson neurons in the presence of synaptic turnover. This is important because assembly drift in the other network models relies on assembly reactivations as described above. Hence, the question arises if assemblies can persist and drift without this kind of memory replay, only due to the weak correlations reflecting the assembly structure in the noisy autonomous network activity. To address this question, I investigated pairwise spike correlations and putative assembly reactivations in our different network models. For the reactivations, I developed a method to detect them. The analysis reveals that the spike correlations in the Poisson model are much weaker than in the binary and the LIF models, where the correlations between neurons belonging to the same assembly are quite strong. Moreover, in the Poisson model, we do not observe full assembly reactivations on relevant time scales.

On the theoretical side, I derived from first principles the effective random walk model for neuron transitions between assemblies in the LIF model. The details are presented in Appendix C. In addition to calculating the statistics of weight changes due to STDP during assembly reactivations, this required a model for the probability of a neuron to spike in a reactivation depending on its input coupling strength from the assembly. I could make use of our previous work on propagating synchronous spiking in neural networks, where this probability has been considered. The random walk model allowed us to gain mechanistic understanding of assembly drift in the LIF network model.

The following sections of this chapter contain the parts of the published manuscript that include significant contributions from me. I rearranged and partly modified them to adapt the contents to the scope and the format of the thesis.

5.1 Introduction

Organisms change over time, on many different levels. This holds in particular for the synapses in neural networks [272]: They change their impact and also appear and vanish. On the one hand, this weight and structural plasticity is activity-dependent. Such forms of plasticity have been argued and directly shown to be crucial for learning [273]. On the other hand, weight changes and turnover of connections with similar magnitude occur spontaneously, in excitatory and inhibitory synapses, independent of previous spiking activity and in its absence [33, 35, 274–276]. A similar dichotomy exists for neural representations: They change due to adaptive learning in order to improve task performance, but also spontaneously, often without affecting behavior. The latter has been observed in areas storing long-term memories [277], in sensory areas, for place cells, location and goal-selective cells, and in motor areas [278, 279]. The changes over the durations of the experiments were mostly only partial.

Environments change as well. To flexibly adapt, higher animals acquire information and retain it by forming memories in the brain. In a widely used model, a memory is represented by one or several (depending on their definition) neuronal assemblies, ensembles of strongly interconnected neurons [5, 280]. If an assembly is partially excited, for example by an external input, the remainder of the neurons follow, leading to associative memory recall. For faithful memory storage the ensemble of neurons forming an assembly is assumed to remain the same [34]. Previous theoretical analysis has carefully studied the formation and maintenance of such static neuronal assemblies [41, 281–288]. In particular, it has been suggested that in presence of noisy autonomous (without receiving external stimulation or feedback) network activity [41, 282–284] and spontaneous (activity-independent) synaptic changes [287, 289], assemblies are preserved with the help of activity-dependent synaptic plasticity.

Based on the experimentally observed changes of synaptic weights and connections and of neural representations, we develop a contrasting associative memory model where assemblies are ever and completely changing; they drift or "swim". This happens gradually, by successive exchange of individual neurons. The neuron ensembles forming the same assembly at distant times are not directly related, but indirectly via the ensembles forming the assembly at the times in between. Using an analogy of Wittgenstein [290], this is comparable to a thread, which consists of many rather short overlapping fibers; the ensembles of fibers in spatially distant parts are not directly related. We illustrate our memory model and the analogy in Fig. 5.1. In our model, the participation of single neurons in the memory representation overlaps [Fig. 5.1(a)], like the participation of fibers in the thread. As a consequence, viewed over time the representation looks like a continuous thread [Fig. 5.1(b)]. While fibers adhere together due to the friction between them, neurons in the assembly adhere due to increased synaptic weights. The inputs and outputs track the course of the "assembly thread" to keep behavior and memory stable [Fig. 5.1(c)]; they connect at each time to the correct ensemble of neurons that currently forms the required neural representation. Stable input neurons may be located in the sensory periphery, but also within the brain [278], for example in the primary visual cortex and the dentate gyrus, the input area of the hippocampus [291]; motor neurons are candidates for stable output neurons. We will refer to both input and output neurons [Fig. 5.1(c)] as periphery neurons and to the assembly-forming neurons [Fig. 5.1(a)] as interior ones.

We demonstrate the feasibility of our memory model using spiking neural networks with different types of dynamics. Numerical simulations and theoretical analysis reveal that assembly drift can be driven by synaptic weight fluctuations due to noisy autonomous activity or by spontaneous synaptic turnover (activity-independent appearance and disappearance of synaptic connections). The overall representational structure and memory are maintained by activity-dependent and homeostatic synaptic

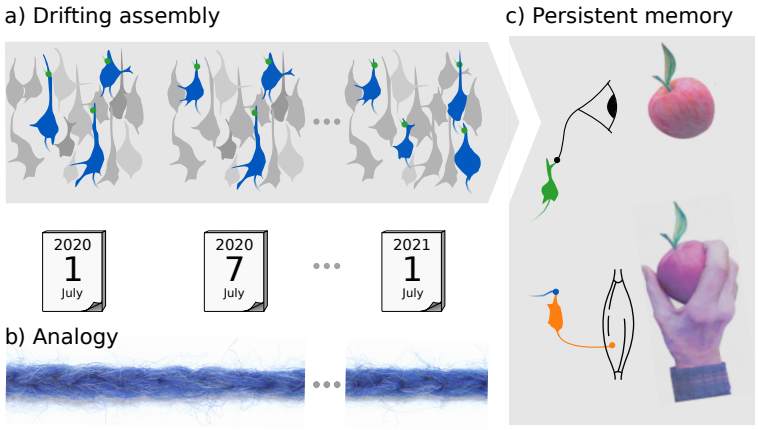

Figure 5.1: Assembly drift and persistent memory. (a) At two nearby times a similar ensemble of neurons forms the neural representation of, for example, "apple" (compare the blue-colored assembly neurons at the first and the second time point). At distant times the representation consists of completely different ensembles (blue-colored assembly neurons at the first and the third time point). Due to their gradual change, temporally distant representations are indirectly related via ensembles in the time period between them. (b) Parts of a thread possess the same form of indirect relation: Nearby parts are composed of similar ensembles of fibers, while distant ones consist of different ensembles, which are connected by those in between. (c) The complete change of memory representations still allows for stable behavior. In the figure, a tasty apple is perceived. At different times, this triggers different ensembles that presently form the representation of "apple"; see panel (a). Assembly activation initiates a reaching movement towards the apple, despite the dissimilarity of the activated neuron ensembles. Memory and behavior are conserved because the gradual change of assembly neurons enables the inputs (green) and outputs (orange) to track the neural representation.

plasticity. In our article [271], we further consider binary neural network models. Using these, we find that assembly drift can be directly related to the evolution of fear memory representations uncovered in recent experiments [277] and that drifting assemblies are suitable for computation.

5.2 Neuron and network models

We consider two types of plastic networks of spiking model neurons. We mainly use networks of noise-driven leaky integrate-and-fire (LIF) neurons, without and with spontaneous synaptic turnover. Further on, we use simplified models of Poisson spiking neurons [13, 41, 206, 237]. The long-term simulations require that we consider networks of medium size, of the order of a hundred neurons. The simplest networks where we can observe neuron switching between assemblies store two assemblies. In networks with three or more assemblies, however, switching includes an additional feature, choosing an assembly to switch to. We thus use networks with three assemblies for our proof-of-principle simulations. Since the memories in our model are stored in the synapses between excitatory neurons, only these are plastic. The remaining synapses are either homogeneous (LIF model) or not explicitly modeled (Poisson model), such that there is no static network weight structure that could contribute to the maintenance of network functionality.

Previous theoretical work has shown that learning [283, 284] or spontaneous emergence [41, 288] of static assemblies as well as their maintenance can be enabled by a combination of spike-timing-dependent plasticity (STDP) and homeostatic plasticity. A recent experimental study investigated STDP in the

recurrent excitatory synapses of the hippocampal region CA3 [38], a region that is assumed to serve as an associative memory network and store assemblies [280]. It found an STDP learning window with long-term potentiation (LTP) for temporally close pre- and postsynaptic spikes, irrespective of their ordering and the stronger the closer they are. Based on the theoretical and experimental results, we choose in our models an STDP rule with a symmetric learning window with centrally peaked LTP [Fig. 5.2(a)]. LTP is compensated by long-term depression (LTD) for temporally more distant spikes, see also [41]. In our networks of Poisson neurons, we use an STDP rule that is similar to voltage- or calcium-based rules [292, 293]. In both models, the synapses further undergo homeostatic normalization. Experiments indicate that the total synaptic input strength [48, 51, 294] to a neuron may be conserved. Further, there is evidence for output normalization [295]. Both input and output normalization may be realized by competition for synaptic resources [48]. Following [41, 48, 296, 297], we thus introduce normalizations of the model neuron's input and output synaptic weights. Finally, we incorporate spontaneous synaptic turnover in some of the models.

5.2.1 LIF networks with STDP

We use a LIF neuron model with current-based synapses and additional noise input. The membrane potential or voltage $V_i(t)$ of neuron i obeys

$$\tau_m \frac{dV_i}{dt} = V_{\text{rest}} - V_i + RI_i^E(t) + RI_i^I(t) + \sqrt{2\tau_m}\sigma\xi_i(t)\,, \tag{5.1}$$

$$RI_i^E(t) = \sum_{j\in M_E}\sum_{t_j\le t} w_{ij}\, e^{-\frac{t-t_j}{\tau_E}},\; RI_i^I(t) = \sum_{j\in M_I}\sum_{t_j\le t} w_{ij}\, e^{-\frac{t-t_j}{\tau_I}}\,. \tag{5.2}$$

Here, $\tau_m = 10\,\text{ms}$ is the membrane time constant, V_{rest} the resting membrane potential, R the input resistance, $I_i^E(t)$ and $I_i^I(t)$ are the total currents generated by the excitatory and inhibitory synapses from the other modeled neurons, and $\xi_i(t)$ is standard Gaussian white noise. The parameter σ equals the standard deviation of the membrane potential's stationary distribution in absence of a threshold and synaptic input currents from the other modeled neurons; the membrane potential then follows an Ornstein-Uhlenbeck process. When the voltage $V_i(t)$ reaches the spike threshold $V_\theta = 20\,\text{mV}$, a spike is emitted and the neuron is reset to $V_0 = 0\,\text{mV}$, where it stays for a refractory period $\tau_{\text{ref}} = 5\,\text{ms}$. V_{rest} is halfway between threshold and reset. A generated spike propagates to postsynaptic neurons, where it generates changes in the synaptic currents. A spike of an excitatory neuron j evokes in the input $RI_i^E(t)$ a jump-like increase of size w_{ij}, the weight of the synaptic connection; thereafter, $RI_i^E(t)$ decays exponentially with time constant $\tau_E = 2\,\text{ms}$. We measure excitatory synaptic weights in terms of the peak of the resulting excitatory postsynaptic potential (EPSP) in units of mV. The factor translating weight w_{ij} into the corresponding peak EPSP can be expressed as $\exp(\log(\tau_E/\tau_m)\,/\,(1-\tau_E/\tau_m)) \approx 0.134$. The decay time constant of inhibitory input currents is $\tau_I = 5\,\text{ms}$. t_j are the spike times of neuron j, M_E is the set of all N_E excitatory and M_I that of all N_I inhibitory neurons modeled in the network.

At each excitatory spike time, the synaptic weights between excitatory neurons are updated according to a pair-based STDP rule, e.g. [5], with symmetric window [Fig. 5.2(a)]:

$$\Delta w_{ij}(\Delta t) = \frac{\eta}{a - b(1+\delta)}\left[a\exp(-a\,|\Delta t|) - b(1+\delta)\exp(-b\,|\Delta t|)\right]\,,$$

where $\Delta t = t_i - t_j$ is the time difference between the postsynaptic and the presynaptic spike. Each side

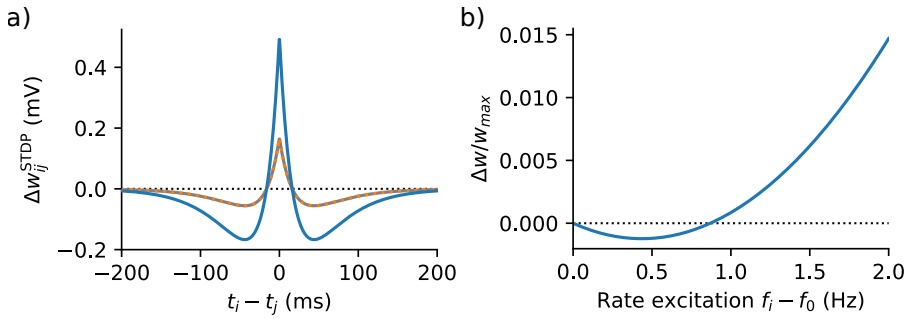

Figure 5.2: Synaptic plasticity rules. (a) STDP windows used in the LIF network model for synapses between interior neurons (blue) and for synapses between interior and periphery neurons (orange); the latter STDP window is used for all synapses (dashed gray) in the LIF network model with spontaneous synaptic turnover. The weight update is given in terms of the change of the peak excitatory postsynaptic potential in units of mV. (b) Dependence of the weight update on the excitation level in the Poisson model. Black dotted lines in (a) and (b) indicate the border between depression and potentiation.

of the window is the difference of two exponentials with decay time constants $a^{-1} = \tau_{\text{LTP}} = 20\,\text{ms}$ and $b^{-1} = \tau_{\text{LTD}} = 40\,\text{ms}$. η is the window amplitude and $1 + \delta$ describes the ratio of integrated LTD and LTP. In terms of the induced change in peak EPSP, peak potentiation is 0.50 mV (0.17 mV), at 0 ms time delay, peak depression is −0.17 mV (−0.06 mV), at ±44 ms, for connections between interior neurons (from or to periphery neurons) in the network model of Sec. 5.3.1; the values in the brackets are also used for all existing connections in the network model of Sec. 5.3.4, i.e. STDP is weaker in the network with spontaneous synaptic turnover. Homeostatic plasticity normalizes the total excitatory input and output weight strength of an interior neuron i to $\sum_{j\in M_E} w_{ij} = \sum_{j\in M_E} w_{ji} = w_{\text{sum}}$; for a periphery neuron the total input and output weight strength is $w_{\text{sum,peri}}$. The normalization is approximated by divisively normalizing after each excitatory spike the columns and the rows of the synaptic weight matrix between the excitatory neurons. The weights between interior neurons are bounded by 0 and w_{max}, for weights from or to periphery neurons the upper bound is $w_{\text{max,peri}}$. These bounds are enforced by clipping weights before and after homeostatic normalization. All possible synapses between excitatory and inhibitory and between inhibitory neurons are present. There are no self-connections. The complete model parameters are given in App. C.1. To simulate our networks, we use Python and the Brian simulator for spiking neural networks [298]; see [299] for example code.

5.2.2 Assemblies

The memories are encoded in a binary, non-graded manner [300]. In particular, most neurons are strongly activated when their assembly is active. Persistent memory manifests itself as a stable behavioral input-output relation: The input neurons that activate the different ensembles forming an assembly are not exchanged over time, and neither are the output neurons that should be activated by them. For simplicity, we assume that memories and the related behaviors are unchanged over time. We further assume that each input and output neuron is specific to one representation and we set up the assemblies without shared neurons. Networks are initialized by setting existing weights between neurons within an assembly and between an assembly and its periphery neurons to 1 and all others to 0; then homeostatic normalization and clipping are applied as described above.

5.2.3 Spontaneous synaptic turnover

A synapse between two excitatory neurons in our networks with spontaneous synaptic turnover has a finite expected lifetime L. It vanishes with rate $1/L$, i.e. in a simulation step of duration Δt with probability $\Delta t/L$, independent of activity and of its current weight. Similarly, if the synapse is absent, it has an average absence time A; it appears in a simulation step with probability $\Delta t/A$. Thus, on average the synapse is present a fraction $L/(L + A)$ of the time; the probability that it is present at a certain time point is $p = L/(L + A)$, the density of synapses of a given type. The spontaneous turnover in our networks switches entries of the binary connectivity matrix between 1 and 0; this matrix indicates the presence or absence of the corresponding synapse. When a synapse vanishes, its weight becomes 0. Newly appearing synapses have weight 0 as well [276].

5.2.4 Simulation of LIF networks and analysis of assembly drift

The networks in Figs. 5.3 and 5.6 consist of $N_E = 102$ excitatory and $N_I = 20$ inhibitory neurons. $N_{\text{int}} = 90$ of the excitatory neurons are interior neurons. These are initiated such that there are three assemblies of $N_{\text{asbly}}(0) = 30$ interior neurons. Each assembly has 4 periphery neurons. The first two periphery neurons are designated input, the second two output neurons. Connectivity in the network of Fig. 5.3 is all-to-all; the connection density is high to compensate the small network size. In Fig. 5.6 the connection density is $p_{\text{int}} = 0.6$ between interior neurons and $p_{\text{peri}} = 0.8$ between interior and periphery neurons. The corresponding synaptic lifetimes and absence times are $L_{\text{int}} = 2000\,\text{s}$ and $A_{\text{int}} = 1333.3\,\text{s}$ for synapses between interior neurons and $L_{\text{peri}} = 2000\,\text{s}$ and $A_{\text{peri}} = 500\,\text{s}$ for those between interior and periphery neurons. We list all model parameters in App. C.1. Four periphery neuron weights exceed the range of the colorbar in Fig. 5.6; the largest weight would evoke a peak EPSP of 3.9 mV. In five alike simulations each with total simulation time 75 hours (100 hours for the network in Fig. 5.6) and different random realizations of networks and noise the representational structure is conserved over time, i.e. the assemblies continuously drift and their periphery neurons faithfully follow them. Clustering of the weight matrix between the excitatory neurons allows us to detect the assemblies and is obtained with the Louvain clustering algorithm [301] as implemented in [302].

Figure 5.4(a) shows the sum of the weights between interior neuron 2 [indexed 6 in Fig. 5.3(d)] and assemblies 1, 2 and 3, normalized by $2w_{\text{sum}}$ (total input plus total output weight). Analogously, panel (b) shows the sum of the weights between input neuron 1 and assemblies 1, 2 and 3, normalized by $2w_{\text{sum,peri}}$. To quantify the similarity of the synaptic weights in the network as a function of time, in panel (e) we consider the Pearson correlation between weight matrices at times 0 and t computed as

$$\text{Corr}(t) = \frac{\sum_{ij} \tilde{w}_{ij}(0)\tilde{w}_{ij}(t)}{\sqrt{\sum_{ij} \tilde{w}_{ij}(0)^2}\sqrt{\sum_{ij} \tilde{w}_{ij}(t)^2}},$$

where $\tilde{w}_{ij} = w_{ij} - \bar{w}$ are the matrix entries centered by the average entry size $\bar{w}$. It holds that $\text{Corr}(0) = 1$ and $|\text{Corr}(t)| \leq 1$; a value close to 1 indicates high similarity while a value around 0 corresponds to chance level similarity. Figure 5.4(f) displays the sum of the weights between the neurons initially forming each of the assemblies, normalized by their maximal sum $N_{\text{asbly}}(0)w_{\text{sum}}$. Further displayed is the sum of weights between the neurons initially forming the first and second, first and third, second and third assembly, normalized by their maximal sum $2N_{\text{asbly}}(0)w_{\text{sum}}$. The inset shows the analogous quantities for the current assemblies at each time, normalized using the current assembly sizes.

By clustering the network at subsequent times, we identify which neurons currently form the different assemblies. The detected assembly structure at a given time allows us to quantify assembly drift in terms of the normalized neuron overlap as a function of time [Fig. 5.4(g)]. We compute this overlap as the number of neurons that the assembly shares with itself at a reference time; it is normalized by the size of the assembly at the reference time. As a first reference, we consider the initialized network with the assembly structure described above. The decay of these overlaps to chance level then demonstrates complete remodeling and defines the remodeling time.

5.2.5 Linear Poisson model

In the linear Poisson (or "Hawkes") neuron model, the impact of each input spike is independent of the current state of the neuron. This fits to background activity, where the state of a neuron stays close to a baseline all the time. Linear Poisson neurons are stochastically spiking neurons with instantaneous rates $f_i(t)$, $i = 1, \ldots, N$, evolving in continuous time. These rates change in response to spikes from the other (explicitly modeled) neurons in the network and follow the linear dynamics

$$\tau \frac{d}{dt} f_i(t) = -f_i(t) + f_0 + \tau \sum_{j=1}^{N} w_{ij} \sum_{t_j} \delta\left(t - t_j\right), \tag{5.3}$$

where f_0 is a constant spontaneous rate due to the assumed embedding in a fluctuation-driven asynchronous irregular activity state. A spike from neuron j increases $f_i(t)$ in a jump-like manner by a nonnegative synaptic weight w_{ij}. Between input spikes $f_i(t)$ decays exponentially with time constant $\tau = 10\,\mathrm{ms}$ to the spontaneous rate $f_0 = 0.75\,\mathrm{Hz}$. A spike of neuron j induces on average τw_{ij} additional spikes in neuron i. Global or explicitly modeled inhibition is assumed to be implicitly contained in the model; it contributes to the randomness of spike generation, both of spontaneous and of excited activity. The network dynamics can be solved in an event-based manner allowing for very long simulation times; see [299] for example code using Python.

The synapses in our Poisson spiking neural networks change according to the following STDP rule. When a neuron spikes, its existing input and output synaptic weights are updated depending on the current level of excitation of the corresponding partner neurons, which we measure by the instantaneous rate above baseline, $f_i(t) - f_0$ for partner neuron i. The dependence is given by a function $\Delta w(f_i - f_0)$, which is negative (giving rise to LTD) for small and average values of the excitation and positive (giving rise to LTP) for larger values [Fig. 5.2(b)]; for simplicity we use a quadratic function with $\Delta w(0) = 0$:

$$\Delta w(\Delta f_i) = 0.01/w_{\max}\,(\Delta f_i - 0.87\,\mathrm{Hz})\,\Delta f_i$$

with $\Delta f_i = f_i(t) - f_0$. The weights are bounded by 0 and $w_{\max}$. We model homeostatic plasticity by input and output normalization of summed synaptic weights as in the LIF models (see Sec. 5.2.1). Here, this implies that the average number of additional spikes in the network induced by a spike of neuron j (also referred to as "branching parameter") remains constant: $\tau \sum_i w_{ij} = \tau w_{\mathrm{sum}} = 0.25$. The synaptic connections in our Poisson network model also turn over spontaneously as described in Sec. 5.2.3. The complete model parameters can again be found in App. C.1.

5.3 Results

5.3.1 A spiking neural network model for drifting memory representations

We show the feasibility of drifting assemblies and our memory model [Fig. 5.1 and Fig. 5.3(a)] using plastic networks of LIF model neurons (see Sec. 5.2.1 and Sec. 5.2.4). Figure 5.3(b) displays the matrix of synaptic weights between the excitatory neurons (henceforth, simply, weight matrix) of such a network with 90 interior and 12 periphery neurons. The network is initialized with three assemblies; each has two input and two output neurons. The strength of a synapse from neuron j to neuron i is given by the entry w_{ij} of the weight matrix. The synaptic weights change due to STDP with a symmetric window and due to divisive homeostatic normalization, which ensures that both input and output weights sum to w_{sum} for each neuron. In addition, they are restricted by a maximal possible weight $w_{\max}$. Periphery neurons differ from interior neurons on the "physiological level" only by a larger $w_{\max}$, smaller w_{sum} and weaker STDP (smaller window amplitude). We further assume that periphery neurons do not connect to each other, because they might lie in distinct brain areas with little interconnectivity.

In Fig. 5.3(b) the neurons initially forming assembly 1 are displayed with lowest indices, in the upper left corner of the weight matrix. Periphery neurons with strong input from and output to assembly 1 therefore have strong weights at the left part of the horizontal and at the upper part of the vertical weight matrices. The interior neurons forming the assembly then gradually change, but the assembly is preserved: It drifts freely in the network. Furthermore, its input and output neurons stay the same [Fig. 5.3(c)]. This holds for all assemblies.

As part of noisy autonomous network activity, neurons forming an assembly occasionally spike synchronously [Fig. 5.3(d)]. Such reactivations appear dispersed over the neuron indices at later times, since the indexing does not fit the assembly structure anymore. Background spiking is irregular and asynchronous; the membrane potentials of the neurons fluctuate irregularly [Fig. 5.3(d)].

We verified that the drifting assemblies have the associative memory property of activating after being partially excited and that they are functional in the sense that they mediate an input-output association: after a sufficiently strong stimulation of its input neurons, an assembly activates and stimulates its output neurons to generate increased spiking activity [271, Supplementary Information, Fig. S2].

5.3.2 Noisy autonomous activity gives rise to drifting assemblies

The assembly drift is reflected by characteristic dynamics on the level of single interior neurons: comparatively long times of stable assembly membership, which are interspersed with fast switches between them [Fig. 5.4(a)]. Periphery neurons do not switch assemblies [Fig. 5.4(b)].

What is the mechanism underlying switching? A neuron generally spikes when its assembly reactivates, resulting in strengthened synaptic coupling between them due to LTP in the STDP rule. However, coincident spiking of an interior neuron together with reactivations of another assembly can increase the synaptic weights between that assembly and the neuron via STDP [Fig. 5.4(c)]. The homeostatic competition between synapses then leads to weakening of the synapses between the neuron and its current assembly. If the weight perturbation is not eliminated (for example, by spiking of the neuron together with its current assembly), the neuron has an increased probability of spiking synchronously together with the other assembly, which will result in even stronger binding. Therefore, a neuron sometimes leaves its current assembly and switches to another one. There are further mechanisms that contribute to the switching in our networks: If an assembly reactivates without near simultaneous spiking of the

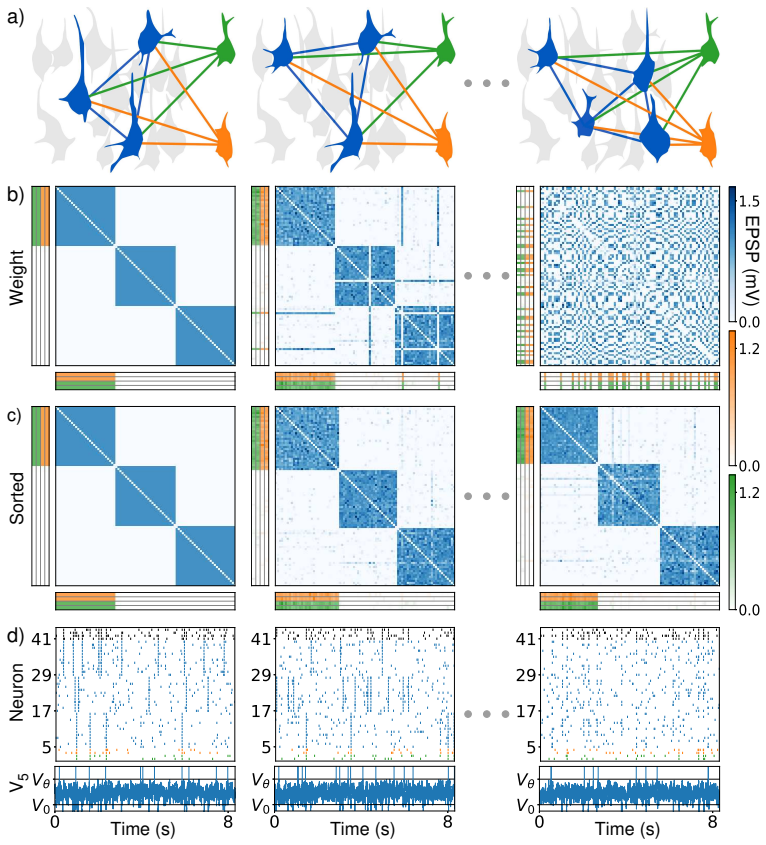

Figure 5.3: Drifting assemblies in spiking neural networks. (a) Schematics emphasizing strong synaptic coupling. While an assembly drifts freely (blue-colored assembly neurons) within the interior neurons, its input and output neurons (green and orange) follow it by adapting their synaptic weights. (b) Weights between interior neurons (blue weight matrix), from input and output neurons to the interior neurons (green and orange vertical weight matrices) and from the interior to the input and output neurons (green and orange horizontal weight matrices). Input (output) weights of neuron i are displayed as the ith row (column). Only weights of the four periphery neurons initially (and thus for all times) attached to assembly 1 are shown for clarity. First column: Network initialization with three assemblies. Second column, after 27 min: Noisy autonomous spiking activity has already driven several interior neurons to attach to a new assembly (blue weight matrix, horizontal and vertical "lines" indicating the changed input and output preference). Third column, after 30 h: The assemblies have drifted away, the weight matrix is completely remodeled. (c) Like (b) but with neurons reordered according to assemblies that they belong to, using a clustering algorithm (Sec. 5.2.4). The assemblies remain intact and the periphery neurons stay strongly coupled to assembly 1. (d, upper) Spike trains of the input (green) and output (orange) neurons of assembly 1, of 12 neurons from each of the ensembles that initially form assembly 1 (5 to 16), 2 (17 to 28) and 3 (29 to 40) and of four inhibitory neurons (black). (d, lower) Membrane potential of the first interior neuron fluctuates irregularly. Spikes are marked by vertical lines above threshold V_θ; reset is to V_0.

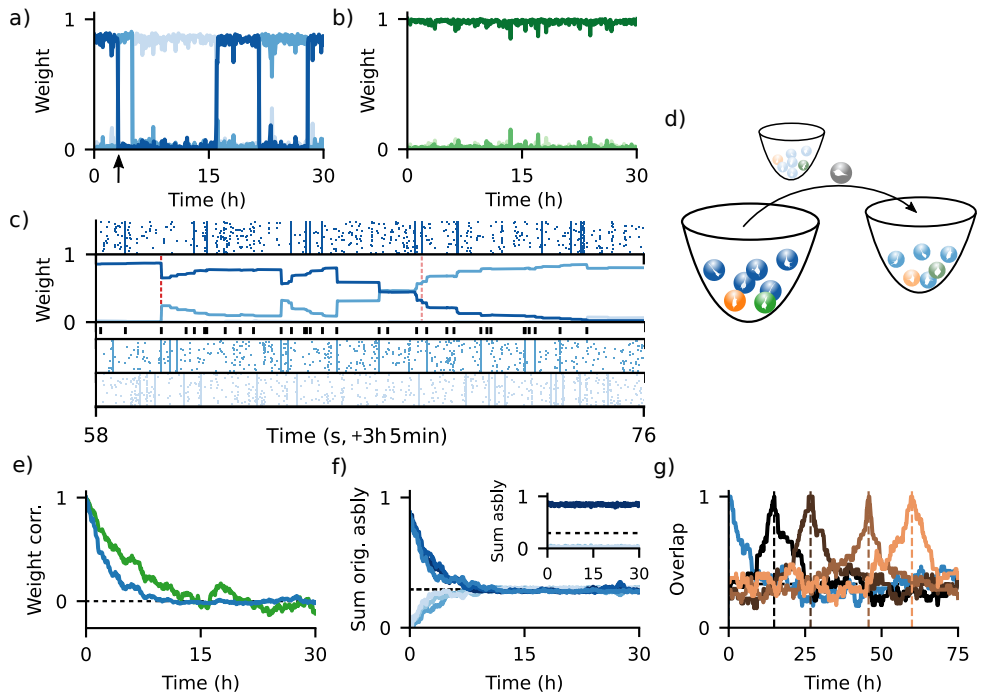

Figure 5.4: Analysis of drifting assemblies and their periphery neurons. (a) Switching of interior neurons in a simulation. Normalized summed weights between a neuron and current assemblies 1, 2 and 3 (dark to light blue) show temporary membership and fast transitions. (b) Periphery neurons stay attached to their assembly. Display is like (a), with greens indicating periphery neuron-assembly weights. (c) Closeup of the transition indicated in (a) by an arrow. Raster plots: spikes of neurons in the three assemblies. The switching neuron is in assembly 1 (first subpanel) before and in assembly 2 (fourth subpanel) after its transition. Second subpanel: summed weight between the neuron and each assembly. Red dashed line: switching neuron spikes together with reactivation of assembly 2; light red dashed line: failure to spike with assembly 1. Third subpanel: spikes of the switching neuron. (d) Schematic illustration of the mechanism underlying assembly drift. Noise drives balls (neurons) out of wells, which are generated by the different assemblies. They move to other wells (neurons switch assemblies). Periphery neurons (green, orange) experience too weak noise to be pushed out of the wells; they stay with their assemblies. (e) Complete network weight remodeling. Pearson correlation between initial and later weights of interior (blue) and periphery-interior synapses (green) converge to chance level. (f) Complete assembly remodeling. Summed weights within and between (darker and lighter blues) the three initial assemblies decay to chance level (1/3 of recurrent interior coupling, black dashed line). Inset shows maintenance of representational structure: sum of weights within current assembly 1 (dark blue) and between it and the current other two assemblies (light blues). (g) Assembly drift continues over time. Overlap between the neuron ensemble forming assembly 1 at a reference time with the neuron ensembles forming assembly 1 at other times. Reference times (dashed vertical lines) are initialization (blue) and first to fourth complete remodeling time (dark to light brown). Chance level is 1/3 (black, mostly covered).

neuron, the homeostatic normalization together with the strengthening of the synapses between the reactivated neurons leads to weakening of the synapses between the neuron and the assembly. In addition, moderately asynchronous spiking of the neuron with respect to an assembly reactivation can result in weakening of the interconnecting synapses due to LTD. Finally, smaller weight changes are induced by the background spiking activity. In the transition phase, when input synapses from both assemblies are impactful, the weights are most volatile. For one of its neurons, an assembly can thus be thought of

acting like a potential and noise well [Fig. 5.4(d)]; the latter refers to a local decrease in noise strength. A neuron tends to jump back towards the bottom after a perturbation of its weights, but several strong perturbations in a short period of time can induce a transition from one well to another, as shown in Fig. 5.4(c). Due to weaker STDP of their synapses, periphery neurons experience weaker perturbations, which are insufficient to escape the well.

Neuron switching and thus assembly drift continues over time. Weights and assemblies completely and continuously remodel, see Fig. 5.4 (e)–(g). We note that neurons often generate transiently strong connections to another than their current assembly [down- and upstrokes in Fig. 5.4(a)]. At these times and during transitions, neurons are to some degree shared between two assemblies and generate an overlap between them.

The mechanisms of assembly stabilization that ensure long times of assembly membership in our networks have been described previously for static assemblies: Correlated activity, for example due to reactivations, leads to strengthening of internal and weakening of inter-assembly synaptic weights via LTP and LTD [38, 41, 283–285, 289]. Additionally, homeostatic plasticity introduces competition, which weakens inter-assembly synapses, as they usually get potentiated less by activity-dependent plasticity [48, 296, 303].

While in our networks the interior neurons switch between assemblies [Fig. 5.4(a)], no assembly vanishes. Responsible for this is likely a higher reactivation rate of smaller assemblies due to stronger internal synapses, as more frequent reactivation leads to stronger binding and recruitment of neurons. Since the assemblies compete for neurons [304], smaller than average assemblies have the tendency to grow and larger ones to shrink.

5.3.3 Simplified models of neuron switching and assembly drift

To further elucidate the mechanism of the neuron transitions between assemblies and as a prerequisite for an effective model of neuron switching, we examine the change Δw_1 of the summed input weight w_1 from assembly 1 to an interior neuron, see Fig. 5.5(a). For simplicity, we focus on a network with two assemblies and without periphery neurons. We consider only the input weights, since the influence of a single neuron's output on an assembly is small. The weights are normalized, such that w_1 is between 0 and 1. We bin this range into 50 bins of size 0.02 and measure for each bin the average $\overline{\Delta w_1}$ and standard deviation $\mathrm{Std}(\Delta w_1)$ of the ensuing weight changes Δw_1 during time intervals whose length equals the average single-neuron interspike interval. As a result, we obtain the average change $\overline{\Delta w_1}(w_1)$ and its fluctuations $\mathrm{Std}(\Delta w_1)(w_1)$ as a function of the summed input weight w_1. If $\overline{\Delta w_1}(w_1)$ is greater than zero, the input strength from assembly 1 on average increases and the neuron is "drawn towards" this assembly. This happens for w_1 close to 1 [Fig. 5.5(b) left], i.e. when the neuron is part of assembly 1. We note that the actual changes are finite size jumps. Since we consider two assemblies, the cases $w_1 \approx 0$ and $w_1 \approx 1$ are symmetric. $w_1 \approx 0$ means that the neuron belongs to assembly 2 and tends to stay away from assembly 1.

The analysis shows that we can understand the dynamics of neuron transitions in the networks of LIF neurons as a random walk between wells [Fig. 5.4(d)]: In each step, there is a deterministic change $\overline{\Delta w_1}(w_1)$ of w_1 and a fluctuation of size $\mathrm{Std}(\Delta w_1)(w_1)$ around it. The deterministic dynamics may be visualized by a potential $U(w_1)$ with $\overline{\Delta w_1}(w_1) = -dU(w_1)/dw_1$, where the summed input weight w_1 of the neuron behaves roughly like an (overdamped) object jumping down the potential's slope. The average weight change alone lets the neuron stay in the potential wells near $w_1 \approx 1$ or $w_1 \approx 0$ [Fig. 5.5(c) left]. The noise in the weight changes, quantified by $\mathrm{Std}(\Delta w_1)(w_1)$ in Fig. 5.5(d) left, is thus crucial for

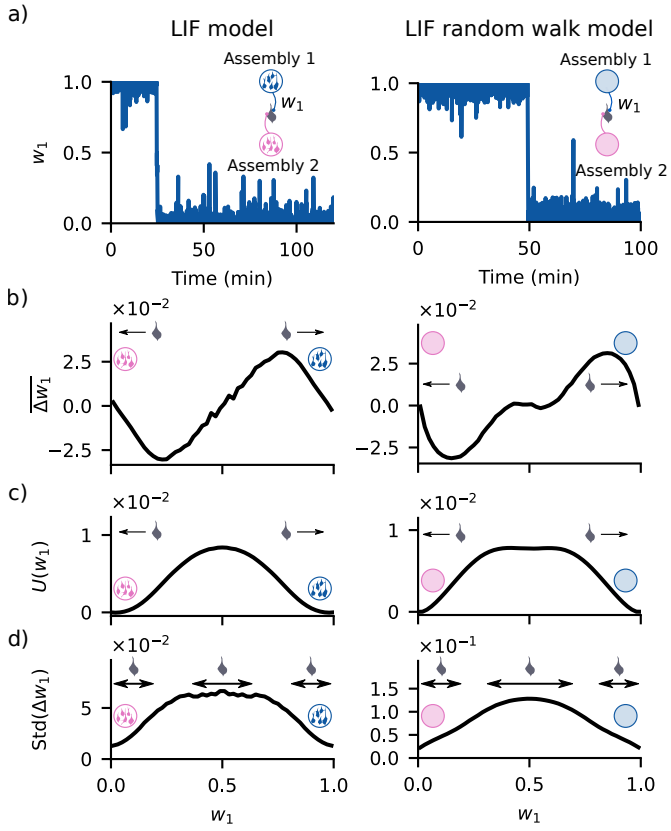

Figure 5.5: Mechanisms of neuron transitions between assemblies driven by weight changes due to noisy autonomous activity. The mechanisms for the LIF network model (left) are well captured by an effective random walk model (right). (a) Long-term membership of a neuron in assemblies and fast transition between them. Large summed input weight w_1 from assembly 1 to the neuron reflects membership in this assembly. (b) Average change $\overline{\Delta w_1}$ of the summed input weight w_1 from assembly 1, as a function of the current weight value. (c) Corresponding potential $U(w_1)$ for the average weight changes. The average weight changes are roughly similar to the displacement of an object jumping down the landscape given by $U(w_1)$. (d) Standard deviation Std(Δw_1) of the change of the summed input weight w_1 from assembly 1, as a function of the current weight value. The two wells near $w_1 \approx 1$ and $w_1 \approx 0$ in both panels (c) and (d) induce the metastable states corresponding to the different assembly memberships.

transitions, as in noise-activated transitions between metastable states [6]. A modified Markovian random walk model without the deterministic part shows that already the substantially stronger noise in the transition zone prevents the neuron from lingering there and drives it to one of the assemblies (App. C.2). In other words, already the inhomogeneous noise alone would lead to the observed metastable states, as in noise-induced multistability [103].

In App. C.3 we also derive an effective random walk model for neuron switching in LIF networks from first principles. It assumes that neurons spike and assemblies reactivate according to Poisson processes with given rates. A "test neuron" further spikes with input weight-dependent probability during assembly reactivations. We can then compute the changes of the test neuron's summed input weights resulting from coactivated synchronous spiking with its assembly, coincident spiking with the other assembly and asynchronous background spiking (App. C.3). This simplified model generates switching dynamics, potentials and noise similar to those obtained in LIF networks (Fig. 5.5 right), confirming our previous explanations for them. In particular, the model confirms that coactivated spiking with the neuron's current assembly generates potential wells and shows that the stronger fluctuations in the transition zone result from opposing weight changes evoked by similarly frequent coactivated spiking with both assemblies.

5.3.4 Spontaneous synaptic turnover also gives rise to drifting assemblies

Assembly drift can also be driven by activity-independent synaptic changes. To show this, we introduce spontaneous turnover of connections; i.e. synapses appear and disappear [33]. For simplicity, we assume a uniform turnover rate that does not depend on the current synaptic weight. A synaptic connection between two excitatory neurons in our networks has a finite expected lifetime (Sec. 5.2.3). Similarly, if the synapse is absent, it has a finite average absence time until it reappears, with initial weight zero. The excitatory connectivity thus completely remodels. The average lifetime of a synapse is about 0.5 h. Experiments indicate average lifetimes in absence of activity that range from minutes for immature synapses to two months for mature ones with large weights [33, 276]. We choose shorter lifetimes to reduce the simulation time to a feasible amount and to roughly include activity-independent weight fluctuations that also occur in biological networks [275]. The presence or absence of the synapse from neuron j to i is indicated by a one or zero in the entry c_{ij} of the connectivity matrix of the network.

Figure 5.6 displays assembly drift in such a network of LIF neurons (Sec. 5.2.1 and 5.2.4). For the employed parameters, the spontaneous synaptic turnover drives the drift: Without it, assemblies stay invariant after a short time of adaptation to the fixed connectivity matrix (Fig. 5.7), because the weight changes due to STDP are weaker than in the network of Fig. 5.3 and noisy network activity alone is insufficient to drive switching of neurons.

5.3.5 Drifting assemblies in networks without spontaneous assembly reactivation

In the LIF neural network models considered so far, the assemblies undergo spontaneous reactivations: events during which practically all neurons in the assembly are active [see Fig. 5.3(d)]. Reactivations have been observed in various cortical areas, from visual and auditory cortex to the hippocampus [305]. To address their necessity for drifting assemblies, we use networks of linear Poisson (or "Hawkes") model neurons [13, 41, 206, 237], see Sec. 5.2.5. The synapses in our Poisson networks change according to a plasticity rule that does not require pre- and postsynaptic spiking, but already generates potentiation if

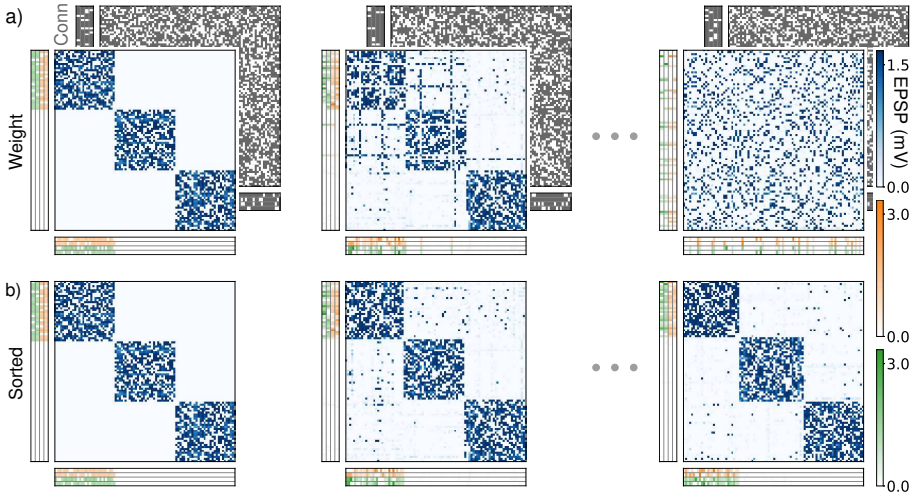

Figure 5.6: Spontaneous synaptic turnover gives rise to assembly drift. (a) The matrix with entries c_{ij} in the background shows the incomplete and spontaneously changing connectivity that drives the drift (present synapses shown in gray). First column shows the weight and connectivity matrices after initialization, second column after 1 h and third column after 37 h. Otherwise the display is like in Fig. 5.3(b). (b) like Fig. 5.3(c).

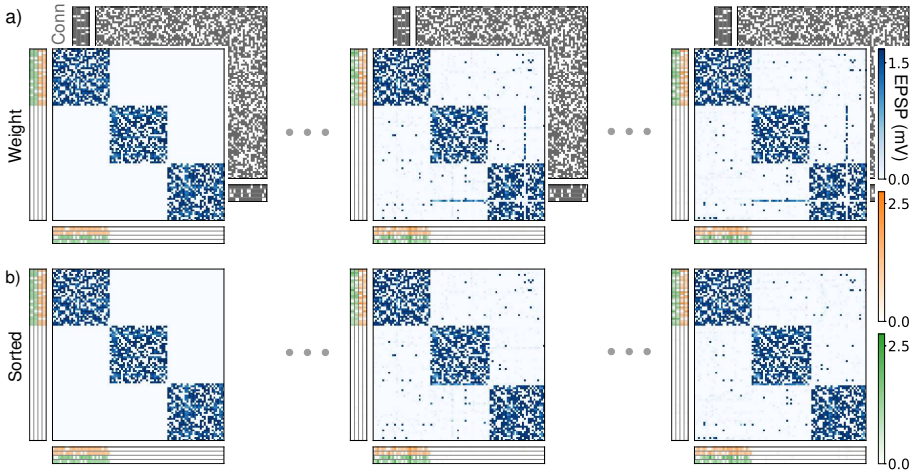

Figure 5.7: LIF network as in Fig. 5.6 but without spontaneous synaptic turnover. The connectivity matrix in the background (gray) is kept constant. First column shows the weight and connectivity matrices after initialization, second column after 2 h and third column after 99 h. After an initial phase of adaptation, no more neurons switch between assemblies.

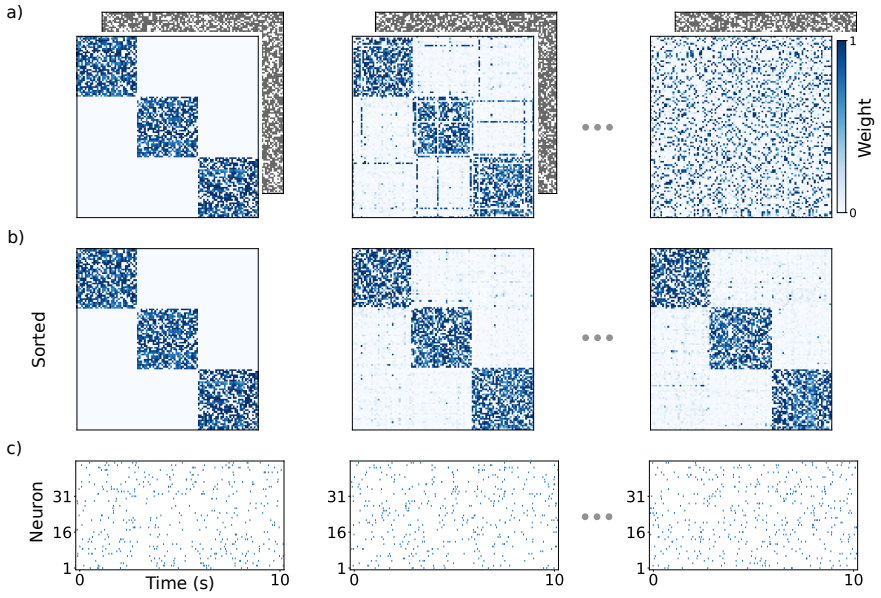

Figure 5.8: Drifting assemblies in a network of linear Poisson spiking neurons with spontaneous synaptic turnover and without periphery neurons. (a) and (b): Display is like panels (a) and (b) in Fig. 5.6 but for 0 h, 72 h and 1440 h hours of simulated time, i.e. the second column shows the network after three days and the third column after 60 days. The underlying synaptic connections spontaneously turn over driving the drift. The weights are divided by their maximal possible value $w_{\max}$. (c) Spike trains of 15 neurons from each of the ensembles that initially form assembly 1 (1-15), 2 (16-30) and 3 (31-45). Spiking activity is asynchronous and irregular without visible assembly reactivations.

one neuron spikes and its partner is more excited than on average [292, 293] (Sec. 5.2.5). Furthermore, the synapses spontaneously turn over. This drives assembly drift as in the LIF network of Fig. 5.6.

Figure 5.8 shows assembly drift in such a network of $N_E = 105$ linear Poisson neurons without periphery neurons. Assembly structure and drift are similar to those observed in the LIF models. However, for the chosen parameters, the drift occurs on a much longer time scale. The autonomous spiking activity in the network shows no visible assembly reactivations.

5.3.6 Spike correlations and reactivation amplitudes

We quantify the pairwise spike correlations in our network models in Fig. 5.9(a) and (b) using the Pearson correlation coefficients of spike counts [306] in time bins of length 150 ms. We measure the spike counts in simulations of static networks with connectivity and weights taken from simulations of networks with drifting assemblies, after the first complete remodeling. The measurement time is 2 h for the LIF and 10 h for the Poisson network. The neurons of the static networks are sorted according to their detected assembly membership as in the figures demonstrating assembly drift. This allows us to partition the neuron pairs into intra-assembly pairs (where both neurons are in the same assembly) and inter-assembly pairs (where each neuron is in a different assembly) and show their contributions to the distributions of correlation coefficients in the networks [Fig. 5.9(b)]. The large pairwise correlations in the LIF model are mainly due to the frequent synchronous reactivation of individual assemblies. In contrast, the pairwise correlations in the Poisson model are much smaller. They are also smaller than

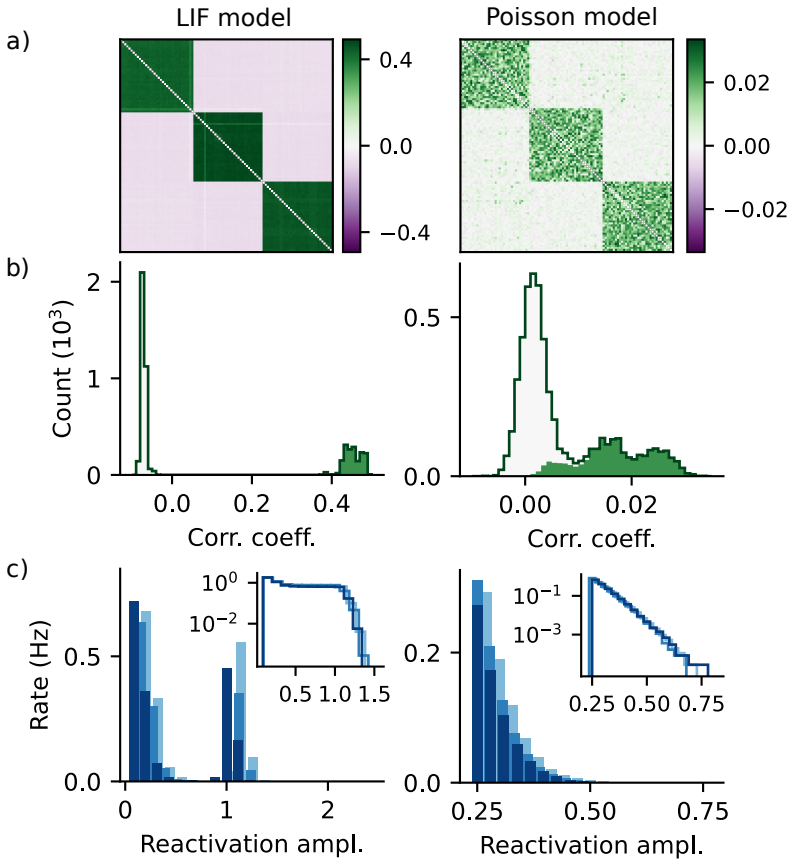

Figure 5.9: Spike correlations and assembly reactivations in different network models. Left: LIF network of Fig. 5.3. Right: Poisson network of Fig. 5.8. The networks have frozen connectivity and weights, which are obtained by fixing plastic networks after the first complete remodeling. (a) and (b): Correlations between Poisson neurons (right) are much smaller than those in the LIF network (left). (a) Matrices of measured Pearson correlation coefficients of spike counts with neurons sorted according to their assembly membership. Diagonal entries (equal to one) are blanked in white. (b) Histograms of correlation coefficients. Solid line gives the full histogram over all neuron pairs, green shading indicates the contribution from intra-assembly neuron pairs, light shading the remaining contribution from pairs of neurons belonging to different assemblies. (c) Assembly reactivation is absent in Poisson networks. Main panels show how often different maxima of the number of spikes in a moving time window are detected within an assembly. Occurrence rate is plotted against number of spikes divided by the assembly size (different blues for the assemblies). Events of spontaneous assembly reactivation are reflected by a second peak, separate from the background continuum near zero. Insets show the complementary cumulative distributions with logarithmic rate scale.

observed in the brain [306].

To further investigate assembly reactivation, in Fig. 5.9(c) we consider the summed spiking activity of all neurons in an assembly. This assembly spiking activity is then temporally filtered with a moving time window of size 15 ms for the LIF and 100 ms for the Poisson model. Reactivation events in our LIF networks are equal to or shorter than the corresponding window size. Further, the analytical duration distribution [206] shows that for our Poisson networks' time constant and branching parameter $\tau w_{\text{sum}} = 0.25$ (Sec. 5.2.5) only 0.013% of single avalanches are longer than 100 ms, justifying the latter

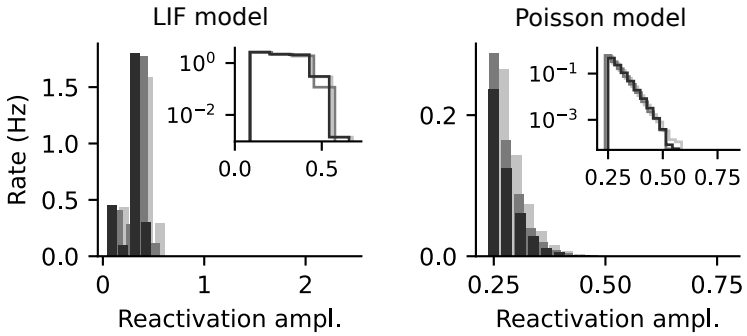

Figure 5.10: Distributions of reactivation amplitudes as in Fig. 5.9(c), but for randomly selected neuron ensembles of the same sizes as the three assemblies in the networks. For the LIF model (left) the histograms exhibit a peak around $1/3$ of the typical assembly reactivation amplitudes, which agrees with the expected neuron overlap of a random ensemble with the assemblies. For the linear Poisson model (right) the amplitude distribution appears to exponentially decay in a similar fashion as for the assemblies in the network [Fig. 5.9(c) right]. The semi-logarithmic plot of the complementary cumulative distribution in the inset reveals that for the randomly selected ensembles, the exponential decay is slightly faster.

window size. The filtering results in a time series, which gives at each time the number of spikes that have occurred in an assembly during the preceding time window. We locate local maxima of this time series to detect putative assembly reactivation. If two local maxima are found with a temporal distance less than the filter window size, only the larger one is kept. The heights of the local maxima are initially numbers of spikes; we normalize them by the corresponding assembly size and call them (relative) reactivation amplitudes. We only consider amplitudes above or equal to a minimal value (0.125 for LIF and 0.25 for Poisson networks). Figure 5.9(c) shows histograms of the amplitudes obtained from the same simulations as the correlation measurements. Dividing the counts of different amplitudes by the measurement time gives the displayed occurrence rate. The associated complementary cumulative distributions [Fig. 5.9(c) insets] indicate the rates at which putative assembly reactivation with sizes exceeding the given amplitude occur. For comparison, in Fig. 5.10 we also show analogous histograms and cumulative distributions for a random partition of the network into neuron ensembles of the same sizes as the assemblies. For the LIF model, the result is consistent with synchronous reactivations that are confined to the assembly and during which most assembly neurons spike once.

Ignition-like reactivations as in the LIF model can in principle not occur in networks of linear Poisson neurons, since all spike impacts superpose linearly. In contrast, the autonomous activity consists of overlapping “avalanches”, transient sequences of spikes evoked by single spontaneous ones [206]. For small summed synaptic weights w_{sum} of the neurons, these are short-lived and small. An avalanche is usually confined to a part of a single assembly, since the interconnections there are still relatively strong. Avalanches involving entire assemblies do not occur on relevant time scales [Fig. 5.9(c) right]. The slightly slower decay of the distribution of assembly reactivation amplitudes compared to the analogous one for randomly selected neuron ensembles may reflect this confinement [insets in Fig. 5.9(c) right and Fig. 5.10 right]. Therefore, the occurring partial activation already suffices to sustain a drifting assembly: Its inactive neurons also have an increased level of excitation, due to the received spikes. The resulting potentiation between inactive and active neurons adds to that between the active ones. Both contributions together keep the assemblies intact and let neurons switch in conjunction with spontaneous synaptic turnover.

5.4 Discussion

We propose that associative memories can be represented by drifting assemblies. The drift results from fast transitions of neurons between assemblies and can be driven by noisy autonomous activity or spontaneous synaptic turnover. The represented memories nevertheless persist, because activity-dependent and homeostatic synaptic plasticity are able to keep the representational structure intact and inputs, outputs and the assemblies consistent. This happens without requiring teaching signals or behavioral feedback, even though the network connectivity, the weight matrices and the ensembles of neurons forming the assemblies change completely.

Several general suggestions have been raised on how neural network functionality may be maintained despite spontaneous synaptic turnover and unstable neural representations: First, the changes might have no impact on the relevant part of network dynamics because they are too weak or because they are eliminated by downstream attractor dynamics [272, 278, 307]. Second, ongoing remodeling may generate transitions between redundant networks that are equally well suited for the required tasks [272, 307, 308]. Third, sensory feedback or feedback from other brain areas may lead to the correction of remodeling that is not in redundant directions [279, 307]. Previous computational studies addressing the impact of intrinsic synaptic changes often focused on the question how neural representations can nevertheless be stable. They proposed that a preserved core structure keeps neuronal activity stable [278, 309–311], that the networks are retrained to counteract degradation [308, 312, 313] or that the ensembles of neurons storing memories are kept invariant via reactivations and unsupervised plasticity [287, 289, 314]. A few computational studies addressed how representations that change may emerge and partially also how the change's impact may be attenuated. Rokni et al. [308] find that synaptic changes let the preferred directions of hidden layer model neurons fluctuate similar to preferred directions in some recordings of motor cortex neurons. Rule et al. [315] demonstrate that modest retraining of linear readouts by supervised learning allows to detect location, speed and head direction from changing representations in the posterior parietal cortex. Gillett, Pereira, and Brunel [311] show that if a fluctuating part is added to otherwise stable synaptic weights storing a sequence, individual neurons partially and temporarily change their dynamics, but the original sequence can still be detected.

In our networks, there is no preserved core in the network structure or memory representation. Only the full circuit consisting of inputs, assemblies and outputs preserves memory and behavior. Interior neurons switch from representing one memory to representing another, only periphery neurons are stable. The inputs and outputs therefore have to constantly compensate for remodeling. Our results demonstrate that this is possible without external feedback or exchange of error signals within the brain. Collective autonomous network activity guides synaptic plasticity and tells the neurons where to attach. Such unsupervised, self-organized compensation may occur for other types of representational drift as well. Further it may compensate other perturbations like death of neurons. On the level of the synapses there is constant change in both periphery and interior neurons in our model; no synapse is persistent. This predicts that in memory areas weight remodeling will turn out to be complete [35].

The developed models combine established model neurons [5, 13], simple STDP rules [5, 38, 41] and homeostatic normalization [41, 48, 296, 297]. Robustness is indicated by the occurrence of drifting assemblies with different neuron models, the simple mechanisms underlying the drift and the broad range of spike correlation strengths that can sustain drifting assemblies; this range includes values found in biological neural networks [306]. The number of assemblies in the investigated networks is largely fixed by the network parameters. The approximately constant number of assemblies renders our memory networks less flexible. Versatile learning could still be implemented by input-dependent imprinting

of new assemblies. This would lead to the vanishing or merging of existing assemblies and, if these already represent memories, to forgetting or generalization. It is not known whether the number of assemblies changes in the brain; it might be that assemblies form during development [316]. Preexisting assemblies could become meaningful when specific inputs and outputs are assigned to them through learning, similar to preexisting sequential structures in a model of episodic memory [317]. Our models can generate very different drifting time scales from hours to months, which depend on underlying parameters like STDP strength and synaptic lifetime.

We find that noisy autonomous activity and the resulting synaptic weight changes can cause transitions of neurons between assemblies. This is due to noise-activated switching between metastable states and inhomogeneous noise that prevents neurons from staying between assemblies. Furthermore, our findings suggest a specific link between spontaneous synaptic changes and the remodeling of neural representations: Synaptic connectivity turnover pushes individual neurons out of their assembly and lets them switch to other ones. Since in the brain spontaneous synaptic change is significant [33, 35, 275] and activity is noisy [5, 306], we suggest that both together may drive the drifting of assemblies in biological neural networks. The strength of a neuron's input synapses thereby decides whether it continues to stay with its assembly, suggesting that input neurons rely on feedback from their assembly to be able to follow it. Our model thus provides an additional explanation for the abundance of feedback connections throughout cortical areas, including early sensory areas.

Our model suggests that the changes through assembly drift are per se functionally neutral. Modification and acquisition of memories will happen on top. The drift could nevertheless help the storage of memories. First, it may contribute to solving the stability–plasticity dilemma: New memories can be imprinted in the same highly plastic (sub-)region [see also the two-stage model for memory [280]]. Over time, their representations drift away together with their input and output connections, giving way to further ones and joining older representations in less plastic regions. Second, in our networks, inputs, outputs and assemblies need to stay consistent for persistent memory. Having static assemblies is an additional requirement. In the brain, this may be difficult to realize due to noisy activity, ongoing plasticity and spontaneous synaptic change. The proposed self-organized compensation could have abolished the evolutionary pressure to develop static assemblies; the synaptic plasticity rules were less constrained and free to satisfy other requirements.

The results make specific and general predictions for memory systems. In particular, they predict a continued drift, reduction of overlap to chance level and gradual but complete remodeling of synaptic weights, as well as the mechanisms underlying the drift and ways to interfere with it. On a more general level, our results indicate that impairments of synaptic plasticity do not only lead to problems in acquiring new memories but also to forgetting, either because assemblies break down or because the improperly compensating inputs and outputs disconnect from their drifting assemblies. Changes in synaptic plasticity as observed in dementia may thus directly lead to the symptomatic forgetting. Further, unsupervised compensation of drifting representations could be a common theme in the brain. Finally, our work suggests that Heraclitus of Ephesus' 2500 year old idea is also true for memory representation, namely that everything flows or, in other words, drifts.

CHAPTER 6

Summary and conclusion

In this thesis, we have studied transition phenomena that emerge in the dynamics of recurrent neural networks. The explored transitions represent important qualitative changes in the dynamical and functional behavior of the considered neural systems. We used concepts and methods from dynamical systems theory, stochastic processes, and statistical physics to address questions arising in neuroscience. Specifically, we investigated the effect of a generic stochastic drive on the transition to chaos in rate-based random neural networks. In balanced networks of spiking neurons, we discovered a transition from stable to chaotic dynamics when a single neuron whose membrane potential accelerates towards spike generation is inserted into the network. Furthermore, in developing networks of stochastically spiking neurons undergoing homeostatic growth, we showed the emergence of critical avalanche dynamics near the transition from stable to unstable recurrent excitation. Finally, we explained the drift of cell assemblies in plastic neural networks as a result of noise-activated transitions between metastable states. Overall, the results presented in this thesis provide a theoretical understanding of dynamical transitions in tractable but also biologically relevant neural network models. Given the generality of such transition phenomena, our insights are not limited to the specific models and may be applicable to other complex dynamical systems.

Time-varying input shapes the transition to chaos in driven random neural networks. In Chapter 2, we investigated how a time-varying input interacts with the nonlinear recurrent dynamics of large rate-based neural networks with random synaptic coupling weights. We modeled the input as a generic stochastic drive and studied the resulting nonautonomous dynamical system. The central goal of this chapter was to theoretically understand the impact of the stochastic drive on the transition to chaos that occurs in random neural networks when the overall recurrent coupling strength is increased [24, 29, 73]. To achieve this goal, we developed the dynamic mean-field theory (DMFT) for stochastically driven networks based on a path-integral formulation. In addition to the self-consistent activity statistics, the theory provides the maximum Lyapunov exponent (LE), which quantifies the dynamical stability against generic infinitesimal perturbations, and the memory capacity for reconstructing past inputs. Using the maximum LE, we determined the critical coupling strength for the transition from stable to chaotic dynamics in stochastically driven networks. For Gaussian white noise input, we derived an exact condition for the transition, which allowed us to delineate the phase diagram as a function of recurrent coupling strength and input amplitude. Beyond a previously known static effect [131], the input suppresses chaos by a dynamic mechanism, shifting the transition to significantly higher coupling

strengths than predicted by time-local linear stability analysis. This dynamic mechanism gives rise to a regime with locally expansive yet asymptotically stable, nonchaotic dynamics. Our analytical results for the memory capacity show that this dynamical regime yields optimal memory of past inputs.

Chaotic dynamics are characterized by their sensitive dependence on the initial conditions. Consequently, we used the maximum LE to identify the onset of chaos, in contrast to previous studies that used a decaying activity autocorrelation function as an indicator [24, 126, 127]. Due to the stochastic drive, this indicator is not applicable to the nonautonomous dynamics studied here. Instead, we treated the stochastically driven network as a random dynamical system (see also Section 1.2.4). This means that the convergence or divergence of trajectories with nearby initial conditions is evaluated for the identical stochastic realization of the time-varying input; in other words, the noise is regarded as "frozen". A positive maximum LE indicates sensitive dependence on the initial conditions for a given input realization. From a neuroscientific perspective, stable dynamics, indicated by a negative maximum LE, corresponds to the reliability of the network response to the same input across different trials [90, 91].

The development of the DMFT approach for the maximum LE of random networks driven by a time-varying input constitutes a significant advance. Following an idea of Derrida and Pomeau [139], we developed this approach by considering the self-consistent joint DMFT of two copies or replicas of the network. This provides a mathematically clear derivation of the Schrödinger eigenvalue equation that determines the maximum LE. While this equation was previously stated for the original autonomous random neural network model [29], no details were given. By analyzing this equation, we gained further insight into how time-varying inputs affect the maximum LE and the transition to chaos.

Since the transition to chaos in the autonomous dynamics of random neural networks has received a lot of attention [24, 29, 30, 73, 74, 127, 148, 174, 318], the work presented in this chapter addressed an important open question in theoretical neuroscience. Nonautonomous dynamics in such networks, due to time-varying inputs or noise, both of which are biologically relevant, has been studied by only a few previous works. For a simpler discrete-time random network model, Molgedey, Schuchhardt, and Schuster [131] showed that a white noise input suppresses chaos. The suppression of chaos can be fully explained by the effect of the input on the spectral radius of the time-local Jacobian matrix [131]. In particular, a random static input ("quenched noise") should have the same effect [319]. For continuous-time network dynamics, Rajan, Abbott, and Sompolinsky [126] found that the suppression of chaos by sinusoidal periodic inputs depends on the frequency of the input. This may be partly due to the dynamic mechanism uncovered here. The biologically relevant setting of continuous-time dynamics driven by aperiodic time-varying inputs or by noise was still open. Our work has closed this gap.

With the transition, we have identified the "edge of chaos" of the driven random neural network. Dynamics near the edge of chaos may give rise to optimal information processing and computational capabilities [117, 164–167, 320]. The edge of chaos likely represents a balance where a system generates sufficiently rich internal dynamics while remaining responsive to external input signals. Therefore, it is necessary to consider driven dynamical systems where time-varying inputs interact with the internal dynamics. To study information processing, we analyzed the memory capacity for reconstructing past input signals and found optimal performance slightly below the transition to chaos. While we demonstrated this for linearly decodable memory, future work could investigate nonlinear computational capabilities, e.g., [30, 156], and ask how they relate to the uncovered regime with locally expansive but asymptotically stable dynamics.

An important extension would be to investigate the case of structured time-varying inputs, where the input components received by different neurons are correlated. Furthermore, such structured inputs could be considered together with additional structure in the recurrent connectivity [21, 174, 321]. The

assumption of uncorrelated input components across neurons enabled the DMFT analysis because it is consistent with the effective stochastic description of the recurrent input to the neurons. This setting also allowed us to obtain the memory capacity based on our DMFT analysis. Low-dimensional structured inputs can lead to correlations between the activity of the neurons. For simpler discrete-time dynamics, this was partially considered in [163, 322]. Building on the results presented here, we have developed and solved a nonstationary DMFT that can account for correlations between neurons due to a common time-dependent input to the network [323]. Remarkably, even in this nonstationary setting, the DMFT yields the maximum Lyapunov exponent, as validated by numerical simulations. This demonstrates the viability of extending our work.

Our DMFT approach for determining dynamical stability using replicas has recently been generalized in a model-independent framework and applied to random networks of binary neuron-like units [324]; see also [325]. Therefore, it may be possible to combine the DMFT approach with the heuristic mean-field approach for approximating Lyapunov exponents in networks of spiking neurons introduced in Chapter 3.

Neurons that accelerate towards spiking induce chaos in inhibitory balanced networks. Turning to spiking neural networks, in Chapter 3 we studied the collective dynamics in inhibitory balanced random networks composed of two types of integrate-and-fire model neurons with different intrinsic membrane potential dynamics. In addition to conventional leaky integrate-and-fire (LIF) neurons, we introduced "anti-leaky" integrate-and-fire neurons, abbreviated as XIF neurons, which model that the membrane potential of biological neurons accelerates towards spike generation. The main goal of this chapter was to elucidate the interplay between the neurons' intrinsic dynamics and the recurrent network behavior, with a focus on the dynamical stability properties. Confirming that the mechanism of dynamical balance is quite general [26, 27, 65, 76, 77, 79, 326], we first demonstrated that the autonomous deterministic dynamics of mixed networks containing both LIF and XIF neurons with inhibitory pulse coupling can generate asynchronous irregular spiking activity. To address the detailed stability properties of the underlying microscopic dynamics in such mixed networks, we examined in this chapter not only the maximum LE, but the full spectrum of LEs. While the dynamics of networks consisting only of LIF neurons is stable against infinitesimal and small perturbations [76, 77, 82, 83, 179, 327], we found that the spectrum of LEs in our mixed networks separates into a negative and a positive part, with approximately one positive LE per XIF neuron in the network. In particular, inserting even a single XIF neuron into a network of LIF neurons leads to a transition from stable to chaotic dynamics.

The study was guided by two analytical mean-field approaches: one for the activity statistics and one for the LEs. In the statistical mean-field approach, assuming asynchronous irregular spiking, we approximated the recurrent input to a neuron with Poisson shot noise and determined self-consistent firing rates and membrane potential probability distributions of both LIF and XIF neurons. Unlike the usual diffusion approximation [85, 86], the shot-noise approach takes into account the finite-amplitude impact of individual input spikes [183, 189]; hence, it also applies to networks in which the neurons receive recurrent input from only a small number of other neurons. Using this approach, we obtained mixed networks that exhibit self-consistent asynchronous irregular spiking activity with homogeneous firing rates for any ratio of the numbers of LIF and XIF neurons. In line with this, there is no obvious change in the statistics of the spiking activity at the transition to chaos induced by inserting a single XIF neuron into a network of LIF neurons.

In the mean-field approach for the LEs, we neglected the transmission of perturbations between the neurons and considered the growth or shrinkage of infinitesimal perturbations during the dynamics of

individual neurons. The resulting mean-field LE of a single LIF or XIF neuron depends only on the neuron's intrinsic leak or anti-leak rate and its firing rate relative to the "free" firing rate due to the constant excitatory input alone. In the presence of inhibitory input spikes, the mean-field LE becomes negative for LIF neurons and positive for XIF neurons. This is because inhibitory input spikes prolong the time between threshold crossings, during which infinitesimal perturbations shrink or grow according to the intrinsic membrane potential dynamics. If the actual firing rate is much smaller than the free firing rate, the mean-field LE will be close to the negative (anti-)leak rate of the neuron. Surprisingly, this simple mean-field approach explains salient properties of the spectrum of LEs in our mixed networks, namely that the number of positive and negative LEs is approximately equal to the number of XIF and LIF neurons, respectively, and that their values are mostly close to the negative anti-leak or leak rate.

The remarkable agreement of the mean-field approach with the numerically computed spectrum of LEs of the recurrent network dynamics indicates that the spectrum of LEs is largely determined by the dynamical stability of the individual neurons. The transmission of perturbations between neurons seems to be a secondary effect that modulates the shape of the spectrum. Since inhibitory input spikes stabilize the dynamics of LIF neurons, our mean-field approach provides an additional explanation for the known stability of inhibitory balanced LIF networks [76, 77, 82, 83, 179, 327]. For the dynamics of XIF neurons, on the other hand, inhibitory input spikes lead to the growth of infinitesimal perturbations. To counteract this growth through the recurrent network dynamics, perturbations transmitted from other neurons must have the opposite sign. Such an arrangement seems implausible for generic infinitesimal perturbations, especially during asynchronous irregular activity. Therefore, inhibitory input spikes also destabilize the dynamics of XIF neurons when the full network dynamics is considered; see also [200]. The numerically computed spectra of LEs confirm that each XIF neuron in our mixed networks acts as a source of instability, amplifying small perturbations and driving chaotic dynamics in the network.

Neural computations involving temporal spike patterns may require both dynamical convergence ("pattern completion") and divergence ("pattern separation") of small perturbations [83, 178, 324, 327]. Our work suggests that this could be realized with different classes of inhibitory neurons if they operate in different stability regimes; see also [182, 198, 199]. This functional separation would have the advantage that inputs to different populations of neurons could arise through development and learning. Furthermore, neurons that are intrinsically prone to instability, such as our XIF neurons in the balanced state, could act as a source of deterministic noise through chaotic dynamics, which can be useful for probabilistic computations and sampling [328, 329].

Our mean-field approach provides a novel perspective on the stability properties of spiking networks and could be extended to more complex integrate-and-fire models that incorporate both crucial aspects of the intrinsic dynamics in biological neurons: dissipation (stability) and dynamic spike generation (instability). For example, the well-known quadratic and exponential integrate-and-fire neuron models, e.g. [5], incorporate these two opposing mechanisms. For such models with nonlinear membrane potential dynamics, stability will depend on the time spent in stable and unstable phases of the intrinsic dynamics and thus the probability distribution of the membrane potential; see also [330]. The employed shot-noise approach for the probability distribution can also be extended to nonlinear integrate-and-fire neuron models [331–333]. An important line of work suggests that inhibitory inputs that prolong the acceleration phase intensify chaos in balanced-state networks [65, 78, 333]. With nonlinear intrinsic dynamics, however, the transmission of perturbations between the neurons may contribute more significantly. Understanding how the stability properties of the recurrent network dynamics emerge from the stability properties of individual neurons and the transmission of perturbations remains an interesting direction for future research [333].

In Chapters 2 and 3, we studied transitions in recurrent neural networks with random but fixed synaptic connectivity. The transitions occurred either by changing the overall recurrent coupling strength or by modifying the intrinsic neuron dynamics, both of which remained fixed during the activity dynamics. In Chapters 4 and 5, we shifted our focus to neural networks with plastic synaptic connections. There, activity-dependent forms of synaptic plasticity introduce a dynamic interplay between the evolving connectivity and the collective activity, which can lead to the self-organization of structure in the connectivity. In both studies, we observed such self-organization involving dynamical transition phenomena.

Critical avalanche dynamics emerge in developing neural networks. In Chapter 4, we investigated how neuronal avalanches, bursts of activity with power-law distributed sizes and durations frequently observed in neural systems, might arise during the development of neural networks. We studied this phenomenon by implementing a homeostatically regulated neuronal growth process (structural plasticity) in a biologically plausible yet analytically tractable spiking network model based on linear Poisson neurons [13, 14, 230, 237]. These neurons incorporate essential biological features such as continuous-time operation, leaky integration, and stochastic variability in their spike responses. The goal was to understand how developing neural networks can self-organize into a regime exhibiting critical avalanche dynamics without requiring fine-tuning. We showed that networks of excitatory neurons can develop, through homeostatic regulation alone, into a configuration close to the transition where recurrent excitation becomes unstable. In this near-critical regime, the network generates avalanches of spikes with statistical properties similar to those of critical branching processes, providing a mechanistic model at the level of spiking neurons for the experimentally observed avalanche criticality.

We identified two key mechanisms that underlie this self-organization towards criticality: (i) homeostatic growth that saturates at a firing rate much higher than the spontaneous rate of the neurons, and (ii) linear superposition of postsynaptic rate responses in the network. Under these conditions, a stationary state develops in which each spike generates nearly one successor spike on average. Due to the linearity of the network's rate response, this branching parameter [210, 212, 216] remains constant across spikes regardless of the activity level, enabling the emergence of scale-free avalanches.

A crucial step for our analysis was the exact reduction of the network's spiking dynamics to a scalar self-exciting Hawkes point process [10, 249, 254, 255]. This approach allowed us to derive the distributions of avalanche sizes and durations analytically, rather than relying on numerical simulations. For avalanche sizes, the underlying branching process of spikes initiated by a spontaneous spike has a Poisson offspring distribution, resulting in an explicit power-law distribution with exponent 3/2 near criticality. For avalanche durations, we developed a novel derivation showing that the distribution can be obtained from the solution of a nonlinear ordinary differential equation, yielding a power-law tail with exponent 2 at criticality. Both exponents match those of critical branching processes [210, 212, 213] and are consistent with experimental observations of neuronal avalanches. These mathematical results establish a link between continuous-time spiking neural network dynamics and more abstract branching process models often used to investigate criticality [97, 216, 244, 334]. Importantly, the results can be extended to Hawkes processes with more general kernels [335], offering broader applicability beyond neuroscience.

The critical transition in our model occurs at the point where the stationary state of the activity dynamics becomes unstable. If we treat spontaneous spiking as an external perturbation, it is the state of vanishing activity that becomes unstable. Since this state represents an absorbing state of the stochastic

network dynamics in the absence of spontaneous spiking, the transition is similar to a non-equilibrium phase transition with an absorbing state [95, 97]. In the linear model with spontaneous spiking, the fixed point of the average firing rate moves to infinity as the recurrent excitatory coupling approaches the instability point. Although this does not constitute a bifurcation in the strict sense [59, 93], we found critical (avalanche) dynamics near the instability.

Linear Poisson neurons provide a tractable stochastic model for studying avalanche dynamics, but they still represent a linearized approximation for the irregular spiking activity of nonlinear neurons at an operating point. During critical avalanches, however, the activity of individual neurons can vary greatly, potentially invalidating the linearization at a given operating point. For nonlinear neuron dynamics, the branching parameter may change with the activity level, leading to deviations from ideal scale-free behavior. Our simulations with a long refractory period demonstrated this effect. While the homeostatic growth process compensates by increasing the excitatory coupling between the neurons, the resulting avalanches show rather strong deviations from power-law scaling. In our networks, strong reciprocal connections between neighboring neurons may amplify these deviations; see also [336].

Our work has contributed to the continuing research on homeostatic plasticity and criticality in neural systems, e.g., [109, 230, 232, 337–340]. Zierenberg, Wilting, and Priesemann [244] showed that homeostatic plasticity in a branching network model gives rise to different dynamical regimes depending on the strength of the external input, which may explain differences between in vitro and in vivo experimental observations [341]. In our model, spontaneous spiking plays a similar role to an external input, and, accordingly, the emerging dynamical regime depends on the spontaneous firing rate relative to the homeostatic target rate. When spontaneous spiking or external input is weak, such as during development or under in vitro conditions, the networks adapt to generate "missing" activity, bringing them closer to instability. Thus, the near-critical regime in our model aligns with observed signatures of criticality in developing neural systems [219–221, 232, 342]. Priesemann and colleagues [247, 334, 341] have proposed that cortical networks in vivo operate in a slightly subcritical, reverberating regime, which may have functional advantages. Dahmen et al. [343] have further enriched this perspective by identifying a second type of criticality in cortical networks, characterized by weak population-level fluctuations but a large dispersion of correlations across neurons, which emerges from inhibition-dominated dynamics with heterogeneous connectivity. Investigating these functionally relevant regimes in our spiking network model, especially after incorporating inhibition, could be an interesting avenue for future research.

Noise-activated transitions give rise to representational drift in memory networks. Motivated by experimentally observed continual changes in synaptic connections and neural representations, in Chapter 5 we showed that cell assemblies of strongly interconnected neurons representing associative memories can drift over time while preserving their memory function. The neurons that form these drifting assemblies are gradually exchanged, allowing stable coding input and output neurons to compensate for the drift by adapting their synaptic connections. As a result, the same memory content can be represented by completely different neurons at different points in time, consistent with temporally evolving memory representations [277] and, more generally, with representational drift [279, 344]. Our study had two intertwined goals: to explore how memories can be maintained despite the dynamic nature of neural circuits, and to understand the mechanisms that enable and regulate the drift of cell assemblies.

We established the feasibility of drifting assemblies in several models of plastic spiking neural networks implementing simple, biologically plausible spike-timing-dependent and homeostatic synaptic plasticity. The drift results from noisy but also directed synaptic weight changes, either due to plasticity

driven by noisy spiking activity alone or in combination with spontaneous synaptic turnover [33, 35, 272]. Since the employed activity-dependent plasticity models (rules) are deterministic, stochastic synaptic weight changes must originate from noisy activity. Crucially, the coordinated drift that maintains the overall assembly structure can occur "offline" during ongoing autonomous activity, without structured inputs or teaching signals.

The stable coding input and output neurons track the assembly drift by adjusting their plastic synaptic connections. Their synaptic connections with the assembly neurons follow the same plasticity rules as the connections between the assembly neurons, just with a lower learning rate or, equivalently, a lower overall amplitude of weight changes. Importantly, this lower learning rate results in reduced weight fluctuations. In conjunction with the noisy ongoing activity, Hebbian spike-timing dependent plasticity specifically strengthens connections between input or output neurons and assembly neurons that tend to be active together. This enables self-organized tracking of the drift without explicit error signals or supervision. It may be fruitful to synthesize this mechanism with other models of representational drift and tracking [345, 346].

The drift in our network models arises from the switching of neurons between assemblies. We conceptualize this switching as noise-activated transitions between metastable states in a dynamical landscape, where these states are formed by the assembly structure. We support this conceptual picture by constructing and analyzing effective stochastic models of the switching dynamics. Since the time scale of such noise-activated transitions can vary on an exponential scale, a prediction of our model is that the drift time scales can vary widely: small changes in synaptic plasticity strength or network activity statistics can dramatically alter drift time scales. This also provides a possible mechanism for modulating the drift speed; see, e.g., [347, 348]. Furthermore, if the fluctuations are too weak to drive the transitions, the drift ceases and the assemblies become static.

Drift in our model occurs between functionally equivalent network configurations, namely a given number of assemblies consisting of an approximately constant number of neurons. To maintain functionality in the presence of ongoing changes, these configurations should be stable. Using the conceptual picture of a dynamical landscape, there should be a "restoring force" that drives back fluctuations, especially those that would deteriorate functionality. However, if different functionally equivalent configurations exist, as in our model, a restoring force for fluctuations, or drift, between these configurations is not required for stable function; see also [346, 349]. In our network models, functionality is invariant with respect to the assignment of the neurons to the assemblies. For symmetry reasons, configurations with permuted neurons are equally stable. Notably, drift occurs through the "smallest" configurational change that maintains functionality: the exchange of a single neuron in an assembly. Hence, the restoring force in the dynamical landscape may be weakest for such changes between equally stable and nearby configurations. This is somewhat reminiscent of dynamics on a line or continuous attractor, e.g., [350], but in the space of synaptic weight configurations. Generalizing this abstract picture to other forms of drifting representations, especially those involving continuous variables, provides an interesting opportunity for future work. As a starting point, one could consider relevant neural network models with invariances or symmetries [350, 351].

Previous theoretical and computational studies on assemblies in plastic neural networks have addressed two related questions [352]: 1) Can assemblies or similar structures form, for instance during development, through the interplay of synaptic plasticity with the activity in the network? Several mechanisms have been identified that support the self-organization of such structures [40, 41, 48, 288, 297]. 2) Can assemblies in plastic neural networks persist in the presence of fluctuations during ongoing activity, or do they fade away? A prominent idea is that autonomously generated activity reflects the assembly

structure so that activity-dependent plasticity mechanisms can reinforce it [283, 285, 287, 289, 353]. Beyond that, we explored the possibility that assemblies can continuously remodel while maintaining their functional identity. Since drift involves fluctuations that generally lead to a weakening of the assembly structure in the network, self-organized reinforcement of assemblies may be a prerequisite for gradual drift. Future theoretical work could investigate the fluctuations in synaptic weight dynamics due to synaptic plasticity, possibly building on the derivation of the random walk model presented in this chapter and the pioneering work of Kempter, Gerstner, and van Hemmen [13].

Our model of drifting assemblies implies a balance between stability and plasticity. To be stable, the assembly structure in the synaptic connectivity should be an attracting configuration of the plasticity dynamics driven by the network activity. As mentioned, such stability can be achieved during ongoing autonomous activity, even in the presence of spontaneous synaptic changes [283, 285, 287, 289, 353]. However, both the reinforcement of the assembly structure and the learning of new memories require plasticity, i.e. changes in synaptic connections. Therefore, the required flexibility introduces a degree of instability into the plasticity dynamics. This balance between stability and plasticity is reflected by the qualitative change from static to drifting assemblies depending on the strength of plasticity, which represents another example of a transition in neural network dynamics—the central theme of this thesis.

APPENDIX A

Details of field-theoretical dynamic mean-field calculations in Chapter 2

A.1 Derivation of mean-field equation

The generating functional $Z[\mathbf{I}](\mathbf{J})$ in Eq. (2.2) is properly normalized independent of the realization of $\mathbf{J}$. This property allows us to follow De Dominicis and Peliti [136] and to introduce the disorder-averaged generating functional

$$\begin{aligned} \bar{Z}[\mathbf{I}] &= \langle Z[\mathbf{I}](\mathbf{J})\rangle_{\mathbf{J}} \\ &= \int \Pi_{ij} dJ_{ij}\, \mathcal{N}\left(0, \frac{g^2}{N}, J_{ij}\right) Z[\mathbf{I}](\mathbf{J})\,. \end{aligned} \tag{A.1}$$

The coupling term $\exp\left(-\sum_{i\neq j} J_{ij}\tilde{x}_i^{\mathrm{T}}\phi(x_j)\right)$ in Eq. (2.2) factorizes over neuron indices i, j, and the random weights J_{ij} appear linear in the exponent. Thus, we can separately integrate over the independently and identically distributed J_{ij}, $i \neq j$, by completing the square and obtain

$$\int dJ_{ij}\, \mathcal{N}\left(0, \frac{g^2}{N}, J_{ij}\right) \exp\left(-J_{ij}\tilde{x}_i^{\mathrm{T}}\phi(x_j)\right) = \exp\left(\frac{g^2}{2N}\left(\tilde{x}_i^{\mathrm{T}}\phi(x_j)\right)^2\right)\,. \tag{A.2}$$

We reorganize the resulting sum in the exponent of the coupling term as

$$\begin{aligned} \frac{g^2}{2N}\sum_{i\neq j}\left(\int \tilde{x}_i(t)\phi(x_j(t))\,dt\right)^2 &= \frac{g^2}{2N}\sum_{i\neq j}\iint \tilde{x}_i(t)\phi(x_j(t))\,\tilde{x}_i(t')\phi(x_j(t'))\,dt\,dt' \\ &= \frac{1}{2}\sum_i \iint \tilde{x}_i(t)\tilde{x}_i(t')\left(\frac{g^2}{N}\sum_j \phi(x_j(t))\phi(x_j(t'))\right) dt\,dt' \\ &\quad - \frac{1}{2}\sum_i \iint \tilde{x}_i(t)\tilde{x}_i(t')\,\frac{g^2}{N}\phi(x_i(t))\phi(x_i(t'))\,dt\,dt'\,, \end{aligned} \tag{A.3}$$

where we use $\left(\int f(t)\,dt\right)^2 = \iint f(t)f(t')\,dt\,dt'$ in the first step and $\sum_{ij} x_i y_j = \sum_i x_i \sum_j y_j$ in the second. The last line is the diagonal (self-coupling) to be skipped in the double sum. It is a correction of

order N^{-1} and will be neglected in the following. The disorder-averaged generating functional (A.1) therefore takes the form

$$\bar{Z}[\mathbf{l}] = \int \mathcal{D}\mathbf{x} \int \mathcal{D}\tilde{\mathbf{x}} \, \exp\left(S_0[\mathbf{x}, \tilde{\mathbf{x}}] + \mathbf{l}^{\mathrm{T}}\mathbf{x}\right) \times \exp\left(\frac{1}{2}\tilde{\mathbf{x}}^{\mathrm{T}} Q_1 \tilde{\mathbf{x}}\right), \tag{A.4}$$

where we extend our notation with $x^{\mathrm{T}}Ay = \iint x(t)\, A(t,t')\, y(t')\, dt\, dt'$ to bilinear forms and define

$$Q_1(t,t') = \frac{g^2}{N} \sum_j \phi(x_j(t))\phi(x_j(t')) \,. \tag{A.5}$$

The field Q_1 is an empirical average over N contributions, which, by the law of large numbers and in the case of weak correlations, will converge to its expectation value for large N. This heuristic argument is validated in the following more formally: A saddle-point approximation leads to the replacement of Q_1 by its (self-consistent) expectation value. To this end, we first decouple the interaction term by inserting the Fourier representation of the Dirac delta functional:

$$\delta\left[-\frac{N}{g^2}Q_1(t,s) + \phi(\mathbf{x}(t))^{\mathrm{T}}\,\phi(\mathbf{x}(s))\right] = \int \mathcal{D}Q_2 \, \exp\left(-\frac{N}{g^2}Q_1^{\mathrm{T}}Q_2 + \phi(\mathbf{x})^{\mathrm{T}}Q_2\phi(\mathbf{x})\right), \tag{A.6}$$

where we further extend our notation with $Q_1^{\mathrm{T}}Q_2 = \iint Q_1(t,s)\, Q_2(t,s)\, dt\, ds$ and $\phi(\mathbf{x}(t))^{\mathrm{T}}\phi(\mathbf{x}(s)) = \sum_{i=1}^{N} \phi(x_i(t))\phi(x_i(s))$. We note that the conjugate field Q_2 is purely imaginary. We hence rewrite Eq. (A.4) as

$$\bar{Z}[j,\tilde{j}] = \int \mathcal{D}Q_1 \int \mathcal{D}Q_2 \exp\left(-\frac{N}{g^2}Q_1^{\mathrm{T}}Q_2 + N\,\ln\,\Omega[Q_1,Q_2] + j^{\mathrm{T}}Q_1 + \tilde{j}^{\mathrm{T}}Q_2\right), \tag{A.7}$$

$$\Omega[Q_1,Q_2] = \int \mathcal{D}x \int \mathcal{D}\tilde{x} \, \exp\left(S_0[x,\tilde{x}] + \frac{1}{2}\tilde{x}^{\mathrm{T}}Q_1\tilde{x} + \phi(x)^{\mathrm{T}}Q_2\phi(x)\right),$$

where we introduce source terms $j, \tilde{j}$ for the auxiliary fields and drop the original source terms $\mathbf{l}^{\mathrm{T}}\mathbf{x}$. The integral measures $\mathcal{D}Q_{1,2}$ must be defined suitably. By writing $N\,\ln\,\Omega[Q_1,Q_2]$, we use the fact that the auxiliary fields couple only to sums of fields $\sum_i \phi^2(x_i)$ and $\sum_i \tilde{x}_i^2$, so the generating functional for the fields $\mathbf{x}$ and $\tilde{\mathbf{x}}$ factorizes into a product of N identical factors $\Omega[Q_1,Q_2]$.

The remaining problem can be considered as a field theory for the auxiliary fields Q_1 and Q_2. The form (A.7) clearly exposes the N-dependence of the action for these latter fields: It is of the form $\int dQ\, \exp(Nf(Q))$, which, for large N, suggests a saddle-point approximation. This approximation neglects fluctuations in the auxiliary fields and hence sets them equal to their expectation value; this point is the dominant contribution to the probability mass. To obtain the saddle-point equations, we consider the Legendre-Fenchel transform of $\ln\,\bar{Z}$ as

$$\Gamma[q_1,q_2] = \sup_{j,\tilde{j}} \left\{ j^{\mathrm{T}}q_1 + \tilde{j}^{\mathrm{T}}q_2 - \ln\bar{Z}[j,\tilde{j}] \right\}, \tag{A.8}$$

which is called the vertex-generating functional or effective action [144, 354]. It holds that $\frac{\delta\Gamma}{\delta q_1} = j$

and $\frac{\delta\Gamma}{\delta q_2} = \tilde{j}$, the so-called equations of state. The leading-order mean-field or tree-level approximation amounts to the approximation $\Gamma[q_1, q_2] \simeq -S[q_1, q_2]$, where $S[Q_1, Q_2] = -\frac{N}{g^2} Q_1^{\mathrm{T}} Q_2 + N \ln \Omega[Q_1, Q_2]$ is the action for the auxiliary fields Q_1 and Q_2. We insert this tree-level approximation into the equations of state and further set $j = \tilde{j} = 0$ since the source fields have no physical meaning and thus must vanish. We get the saddle-point equations

$$\begin{aligned} 0 &= \frac{\delta S[Q_1, Q_2]}{\delta Q_{\{1,2\}}} \\ &= \frac{\delta}{\delta Q_{\{1,2\}}} \left(-\frac{N}{g^2} Q_1^{\mathrm{T}} Q_2 + N \ln \Omega[Q_1, Q_2] \right), \end{aligned} \tag{A.9}$$

from which we obtain a pair of equations

$$\begin{aligned} 0 &= -\frac{N}{g^2} Q_1^*(t,s) + \frac{N}{\Omega} \left. \frac{\delta \Omega[Q_1, Q_2]}{\delta Q_2(t,s)} \right|_{Q^*} \\ &\leftrightarrow Q_1^*(t,s) = g^2 \left\langle \phi(x(t)) \phi(x(s)) \right\rangle_{Q^*} = g^2 C_{\phi(x)\phi(x)}(t,s) \\ 0 &= -\frac{N}{g^2} Q_2^*(t,s) + \frac{N}{\Omega} \left. \frac{\delta \Omega[Q_1, Q_2]}{\delta Q_1(t,s)} \right|_{Q^*} \\ &\leftrightarrow Q_2^*(t,s) = \frac{g^2}{2} \langle \tilde{x}(t) \tilde{x}(s) \rangle_{Q^*} = 0, \end{aligned} \tag{A.10}$$

where we define the average autocorrelation function $C_{\phi(x)\phi(x)}(t,s)$ of the nonlinearly transformed activity of the units. The second saddle point $Q_2^* = 0$ vanishes because the field was introduced to represent a Dirac delta constraint in the Fourier domain. One can show that, consequently, $\int \mathcal{D}Q \, \exp(S[Q_1, Q_2]) Q_2 = 0$, which is the true mean value $Q_2^* = \langle Q_2 \rangle = 0$.

Here $\langle \circ \rangle_{Q^*}$ denotes the expectation value with respect to realizations of x evaluated at the saddle point Q^*. The expectation value must be computed self-consistently, since the values of the saddle points, by Eq. (A.7), influence the statistics of the fields $\mathbf{x}$, which in turn determines the function Q_1^* by Eq. (A.10). Inserting the saddle-point solution into the generating functional (A.7) we get Eq. (2.4):

$$\bar{Z}^* \propto \int \mathcal{D}x \int \mathcal{D}\tilde{x} \, \exp\left(S_0[x, \tilde{x}] + \frac{g^2}{2} \tilde{x}^{\mathrm{T}} C_{\phi(x)\phi(x)} \tilde{x} \right).$$

The action has the important property that it decomposes into a sum of actions for individual, non-interacting units which are driven by a field with common, self-consistently determined statistics, characterized by the second cumulant $C_{\phi(x)\phi(x)}$. Prior to the saddle-point approximation (A.7), the fluctuations in the field Q_1 are common to all of the single units effectly coupling them. The saddle-point approximation replaces the fluctuating field Q_1 by its mean (A.10), which reduces the network to N noninteracting units, or, equivalently, a single-unit system. The second term in Eq. (2.4) represents a Gaussian noise with a correlation function $g^2 C_{\phi(x)\phi(x)}(t,s)$. Physically, it corresponds to the fluctuating input each unit receives from the $N-1$ other units. Its autocorrelation function is given by the summed autocorrelation functions of the output activities $\phi(x_i(t))$ weighted by g^2/N, which incorporates the Gaussian distribution of the couplings.

The interpretation of the noise can be appreciated by explicitly considering the moment-generating functional of a Gaussian stochastic process with a given autocorrelation function $C(t,s)$, which leads to

the cumulant-generating functional $\ln Z_\eta[\tilde{x}]$ that appears in the exponent of Eq. (2.4) and has the form

$$\begin{aligned} \ln\, Z_\eta[-\tilde{x}] &= \ln \left\langle \exp\left(-\tilde{x}^T \eta\right)\right\rangle \\ &= \frac{1}{2}\tilde{x}^{\mathrm{T}}\, C\, \tilde{x}. \end{aligned}$$

Note that the only nonvanishing cumulant of the effective noise is the second cumulant; the cumulant-generating functional is quadratic in $\tilde{x}$. This means that the effective noise is Gaussian and only couples pairs of time points according to the correlation function.

A.2 Stationary process

We rewrite Eq. (2.5) as

$$(\partial_t + 1)\, x(t) = \tilde{\eta}(t), \tag{A.11}$$

where we combine the two independent Gaussian processes $\eta(t)$ and $\xi(t)$ appearing in Eq. (2.5) into $\tilde{\eta}(t)$. We then multiply Eq. (A.11) for time points t and s and take the expectation value over realizations of the noise $\tilde{\eta}$ on both sides, which leads to

$$(\partial_t + 1)\,(\partial_s + 1)\; C_{xx}(t,s) = g^2\, C_{\phi(x)\phi(x)}(t,s) + 2\sigma^2 \delta(t-s)\,, \tag{A.12}$$

where we define the autocorrelation function of the activities $C_{xx}(t,s) = \langle x(t)x(s)\rangle$. We are now interested in a stationary state of the system with $C_{xx}(t,s) = c(t-s)$. The inhomogeneity in Eq. (A.12) is then also time-translation invariant since $C_{\phi(x)\phi(x)}(t,s)$ is only a function of $\tau = t - s$. Therefore, the differential operator $(\partial_t + 1)\,(\partial_s + 1)\, c(t-s)$, with $\tau = t - s$, simplifies to $(-\partial_\tau^2 + 1)\, c(\tau)$, and we get

$$(-\partial_\tau^2 + 1)\, c(\tau) = g^2\, C_{\phi(x)\phi(x)}(\tau, 0) + 2\sigma^2\, \delta(\tau),$$

given as Eq. (2.7) in Section 2.3 of the main text.

A.3 Replica calculation for the maximum Lyapunov exponent

We start from the generating functional for the pair of systems (2.13) and perform the average over realizations of the coupling matrix $\mathbf{J}$, as in Eq. (A.2). We therefore need to evaluate the Gaussian integral

$$\begin{aligned} \int dJ_{ij}\, \mathcal{N}\left(0, \frac{g^2}{N}, J_{ij}\right) \exp\left(-J_{ij} \sum_{\alpha=1}^{2} \tilde{x}_i^{\alpha\mathrm{T}} \phi(x_j^\alpha)\right) \\ = \exp\left(\frac{g^2}{2N} \sum_{\alpha=1}^{2} \left[\tilde{x}_i^{\alpha\mathrm{T}} \phi(x_j^\alpha)\right]^2\right) \exp\left(\frac{g^2}{N} \left[\tilde{x}_i^{1\mathrm{T}} \phi(x_j^1)\right] \left[\tilde{x}_i^{2\mathrm{T}} \phi(x_j^2)\right]\right). \end{aligned} \tag{A.13}$$

The first exponential factor only includes variables of a single subsystem and is identical to the term appearing in Eq. (A.2). The second exponential factor is a coupling term between the two systems arising from the identical matrix $\mathbf{J}$ in the two replicas in each realization that enters the expectation value.

We treat the former terms as before and concentrate on the mixed coupling term here. Analogous to Eq. (A.3), the exponent of the mixed coupling term can be rewritten as

$$\frac{g^2}{N}\sum_{i\neq j}\left[\tilde{x}_i^{1\mathrm{T}}\phi(x_j^1)\right]\left[\tilde{x}_i^{2\mathrm{T}}\phi(x_j^2)\right] = \iint\sum_i \tilde{x}_i^1(t)\tilde{x}_i^2(s)\frac{g^2}{N}\sum_j \phi(x_j^1(t))\,\phi(x_j^2(s))\,dt\,ds, \tag{A.14}$$

where we included the self-coupling term $i = j$, which is again a subleading correction of order N^{-1}.

We now follow the steps in Appendix A.1 and introduce three pairs of auxiliary variables. The pairs Q_1^α, Q_2^α are defined as before in Eqs. (A.5) and (A.6), but for each subsystem, and the pair T_1, T_2 decouples the mixed term (A.14) by defining

$$T_1(t,s) = \frac{g^2}{N}\sum_j \phi(x_j^1(t))\,\phi(x_j^2(s)).$$

Taken together, we can therefore rewrite the generating functional (2.13) averaged over the couplings as

$$\begin{aligned}\bar{Z} &= \langle Z(\mathbf{J})\rangle_{\mathbf{J}} \\ &= \left\{\Pi_{\alpha=1}^2\int\mathcal{D}Q_1^\alpha\int\mathcal{D}Q_2^\alpha\right\}\int\mathcal{D}T_1\int\mathcal{D}T_2\,\exp\left(S[\{Q_1^\alpha,Q_2^\alpha\}_{\alpha\in\{1,2\}},T_1,T_2]\right),\end{aligned} \tag{A.15}$$

with

$$S[\{Q_1^\alpha,Q_2^\alpha\}_{\alpha\in\{1,2\}},T_1,T_2] = -\sum_{\alpha=1}^2\frac{N}{g^2}Q_1^{\alpha\mathrm{T}}Q_2^\alpha - \frac{N}{g^2}T_1^{\mathrm{T}}T_2 + N\,\ln\,\Omega^{12}[\{Q_1^\alpha,Q_2^\alpha\}_{\alpha\in\{1,2\}},T_1,T_2]$$

and

$$\begin{aligned}&\Omega^{12}[\{Q_1^\alpha,Q_2^\alpha\}_{\alpha\in\{1,2\}},T_1,T_2] \\ &\quad=\left\{\Pi_{\alpha=1}^2\int\mathcal{D}x^\alpha\int\mathcal{D}\tilde{x}^\alpha\,\exp\left(S_0[x^\alpha,\tilde{x}^\alpha]+\frac{1}{2}\tilde{x}^{\alpha\mathrm{T}}Q_1^\alpha\tilde{x}^\alpha+\phi(x^\alpha)^{\mathrm{T}}Q_2^\alpha\phi(x^\alpha)\right)\right\} \\ &\qquad\times\exp\left(\tilde{x}^{1\mathrm{T}}\left(T_1+2\sigma^2\right)\tilde{x}^2+\phi(x^1)^{\mathrm{T}}T_2\phi(x^2)\right)\Big).\end{aligned}$$

Analogously to Appendix A.1, we can introduce sources for the auxiliary fields Q_1^α, Q_2^α, T_1, T_2. Then, the equations of state are obtained from the vertex-generating functional Γ as before, which, in the tree-level approximation is given by $\Gamma = -S$ and, for vanishing sources, leads to the saddle-point equations $\frac{\delta S}{\delta Q_{1,2}^\alpha} = \frac{\delta S}{\delta T_{1,2}} = 0$. From the latter, we obtain the set of equations

$$\begin{aligned}Q_1^{\alpha*}(t,s) &= g^2\,\frac{1}{\Omega^{12}}\,\frac{\delta\Omega^{12}}{\delta Q_2^\alpha(t,s)} = g^2\,\langle\phi(x^\alpha(t))\phi(x^\alpha(s))\rangle_{Q^*,T^*}\,,\\ Q_2^{\alpha*}(t,s) &= 0\,,\\ T_1^*(t,s) &= g^2\,\frac{1}{\Omega^{12}}\,\frac{\delta\Omega^{12}}{\delta T_2(t,s)} = g^2\,\langle\phi(x^1(t))\phi(x^2(s))\rangle_{Q^*,T^*}\,,\\ T_2^*(t,s) &= 0\,.\end{aligned} \tag{A.16}$$

Therefore, the generating functional $\bar{Z}^*$ at the saddle point reads

$$\left\{\Pi_{\alpha=1}^{2} \iint \mathcal{D}x^\alpha \mathcal{D}\tilde{x}^\alpha \exp\left(S_0[x^\alpha, \tilde{x}^\alpha] + \frac{1}{2}\tilde{x}^{\alpha\mathrm{T}} Q_1^{\alpha*} \tilde{x}^\alpha\right)\right\} \exp\left(\tilde{x}^{1\mathrm{T}}\left(T_1^* + 2\sigma^2\right)\tilde{x}^2\right). \tag{A.17}$$

We make the following observations:

1. The two marginal subsystems $\alpha = 1, 2$ in Eq. (A.17) have the same form as in Eq. (2.4). This consistency is expected because there is no physical coupling between the two systems. This implies that the marginal statistics of the activity in one system cannot be affected by the mere presence of the other. Hence, in particular, the saddle points $Q_{1,2}^{\alpha*}$ must be the same as in Eq. (2.4).
2. If the last factor in Eq. (A.17) was absent, the two systems would be independent. Two sources, however, contribute to correlations between the systems: the common Gaussian white noise that gives rise to the term proportional to $2\sigma^2$ and the effective Gaussian noise due to a nonzero value of the auxiliary field $T_1^*(t, s)$.
3. Only products of pairs of fields appear in Eq. (A.17), so x^α are Gaussian.

From Eqs. (A.17) and (A.16) we can read off the pair of effective dynamical equations (2.14) with self-consistent statistics (2.15).

A.3.1 Derivation of the variational equation

We multiply Eq. (2.14) with itself for equal or different values of α and take the expectation value on both sides. Thus, we get, for $\alpha, \beta \in \{1, 2\}$,

$$(\partial_t + 1)(\partial_s + 1)\, c^{\alpha\beta}(t, s) = 2\sigma^2\delta(t - s) + g^2 F_\phi\left(c^{\alpha\beta}(t, s), c^{\alpha\alpha}(t, t), c^{\beta\beta}(s, s)\right), \tag{A.18}$$

where the function F_ϕ is defined as the expectation value

$$F_\phi(c^{12}, c^{11}, c^{22}) = \left\langle \phi(x^1)\phi(x^2)\right\rangle$$

for the centered bivariate Gaussian distribution

$$\begin{pmatrix} x^1 \\ x^2 \end{pmatrix} \sim \mathcal{N}_2\left(0, \begin{pmatrix} c^{11} & c^{12} \\ c^{12} & c^{22} \end{pmatrix}\right).$$

First, we observe that the equations for the autocorrelation functions $c^{\alpha\alpha}(t, s)$ decouple and can each be solved separately, leading to the same equation (2.9) as before. This formal result could have been anticipated because the marginal statistics of each subsystem cannot be affected by the mere presence of the other system. Their stationary solutions

$$c^{11}(t, s) = c^{22}(t, s) = c(t - s)$$

then provide the "background" for the equation for the cross-correlation function between the two copies; they fix the second and third arguments of the function F_ϕ on the right-hand side of Eq. (A.18). It remains to determine the equation of motion for $c^{12}(t, s)$.

We first determine the stationary solution $c^{12}(t,s) = k(t-s)$. We see immediately from Eq. (A.18) that $k(\tau)$ obeys the same equation of motion as $c(\tau)$, so $k(\tau) = c(\tau)$ is a solution. The distance (2.12) between replicas for this solution therefore vanishes; the dynamics in both replicas follows identical trajectories. Let us now study the stability of this solution. We hence need to expand c^{12} around the stationary solution

$$c^{12}(t,s) = c(t-s) + \epsilon\, k^{(1)}(t,s)\,,\ \epsilon \ll 1\,. \tag{A.19}$$

We expand the right-hand side of Eq. (A.18) to linear order in ϵ using Price's theorem and Eq. (2.8):

$$\begin{aligned} F_\phi\left(c^{12}(t,s), c_0, c_0\right) &= f_\phi\left(c^{12}(t,s), c_0\right) \\ &\simeq f_\phi\left(c(t-s), c_0\right) \\ &\quad + \epsilon\, f_{\phi'}\left(c(t-s), c_0\right)\, k^{(1)}(t,s)\,. \end{aligned} \tag{A.20}$$

Inserting the expansion into Eq. (A.18) and using the fact that $c(\tau)$ solves the zeroth-order equation, we get the equation of motion for the linear deflection (2.16). By Eq. (2.12), the deflection $k^{(1)}(t,s)$ determines the squared distance between the two subsystems as

$$\begin{aligned} d(t) &= \underbrace{c^{11}(t,t)}_{=c_0} + \underbrace{c^{22}(t,t)}_{=c_0} \underbrace{-c^{12}(t,t) - c^{21}(t,t)}_{=-2c_0-2\epsilon\, k^{(1)}(t,t)} \\ &= -2\epsilon\, k^{(1)}(t,t)\,. \end{aligned} \tag{A.21}$$

A.3.2 Schrödinger equation for the maximum Lyapunov exponent

We here reduce the equation for the variation of the cross-correlation (2.16) into a Schrödinger equation, as in the original work [29, Eq. 10].

First, noting that $C_{\phi'\phi'}(t,s) = f_{\phi'}\left(c(t-s), c_0\right)$ is time-translation invariant, it is advantageous to introduce the coordinates $T = t+s$ and $\tau = t-s$ and write $k^{(1)}(t,s)$ as $k(T,\tau)$ with $k^{(1)}(t,s) = k(t+s, t-s)$. The differential operator $(\partial_t + 1)\,(\partial_s + 1)$ with the chain rule $\partial_t \to \partial_T + \partial_\tau$ and $\partial_s \to \partial_T - \partial_\tau$ in the new coordinates is $(\partial_T + 1)^2 - \partial_\tau^2$. A separation ansatz $k(T,\tau) = \mathrm{e}^{\kappa T/2}\,\psi(\tau)$ then yields the eigenvalue equation

$$\left(\frac{\kappa}{2} + 1\right)^2 \psi(\tau) - \partial_\tau^2 \psi(\tau) = g^2 f_{\phi'}\left(c(\tau), c_0\right)\psi(\tau)$$

for the growth rates κ of $d(t) = -2\epsilon k^{(1)}(t,t) = -2\epsilon k(2t,0)$. We can express the right-hand side by the second derivative of the potential (2.10) so that with

$$V''(c(\tau); c_0) = -1 + g^2 f_{\phi'}\left(c(\tau), c_0\right) \tag{A.22}$$

we get the time-independent Schrödinger equation

$$\left(-\partial_\tau^2 - V''(c(\tau); c_0)\right)\,\psi(\tau) = \underbrace{\left(1 - \left(\frac{\kappa}{2} + 1\right)^2\right)}_{=E}\,\psi(\tau), \tag{A.23}$$

where the time lag τ plays the role of a spatial coordinate for the Schrödinger equation. The eigenvalues ("energies") E_n determine the exponential growth rates κ_n of the solutions $k(2t,0) = \mathrm{e}^{\kappa_n t}\,\psi_n(0)$ at $\tau = 0$ with

$$\kappa_n^{\pm} = 2\left(-1 \pm \sqrt{1 - E_n}\right). \tag{A.24}$$

We can therefore determine the growth rate of the mean-square distance of the two subsystems by Eq. (A.21). The fastest-growing mode of the distance is hence given by the ground-state energy E_0 and the plus in Eq. (A.24). The deflection between the two subsystems therefore grows with the rate

$$\begin{aligned}\lambda_{\max} &= \frac{1}{2}\kappa_0^{+} \\ &= -1 + \sqrt{1 - E_0}\,,\end{aligned} \tag{A.25}$$

where the factor $1/2$ in the first line follows from d being the squared distance.

A.4 Colored and quenched-noise input

A.4.1 General case of colored Gaussian noise

In this appendix, we consider a network driven by N independent colored Gaussian noise inputs $\xi_i(t)$ with zero mean and stationary covariance functions $\langle \xi_i(t+\tau)\xi_j(t)\rangle = \delta_{ij} c_{\mathrm{cn}}(\tau)$. For concreteness, we use Ornstein-Uhlenbeck processes with autocorrelation function

$$c_{\mathrm{cn}}(\tau) = \sigma^2 \left(1 + \frac{1}{\tau_{\mathrm{n}}}\right) \mathrm{e}^{-|\tau|/\tau_{\mathrm{n}}}\,, \tag{A.26}$$

where the correlation time $0 < \tau_{\mathrm{n}} \leq \infty$ determines how fast the input varies in time. The limit $\tau_{\mathrm{n}} \to \infty$ leads to quenched-noise input, which is randomly Gaussian distributed but constant in time, while the limit $\tau_{\mathrm{n}} \to 0$ corresponds to Gaussian white-noise input as in Eq. (2.1). The cumulant-generating functional of the noise takes the form

$$W_{\xi}[j] = \frac{1}{2}\, j^{\mathrm{T}}\, c_{\mathrm{cn}}\, j\,.$$

It appears in the action of the disorder-averaged moment-generating functional, analogous to Eq. (A.4), as

$$\begin{aligned}\bar{Z}[\mathbf{l}] &= \langle Z[\mathbf{l}](\mathbf{J})\rangle_{\mathbf{J}} \\ &= \int \mathcal{D}\mathbf{x} \int \mathcal{D}\tilde{\mathbf{x}}\, \exp\left(\tilde{\mathbf{x}}^{\mathrm{T}}(\partial_t + 1)\,\mathbf{x} + \frac{1}{2}\,\tilde{\mathbf{x}}^{\mathrm{T}}\, c_{\mathrm{cn}}\,\tilde{\mathbf{x}} + \mathbf{l}^{\mathrm{T}}\mathbf{x}\right) \exp\left(\frac{1}{2}\tilde{\mathbf{x}}^{\mathrm{T}}\, Q_1 \tilde{\mathbf{x}}\right),\end{aligned} \tag{A.27}$$

where Q_1 is defined as before in Eq. (A.5). Introducing a second auxiliary field Q_2 and performing the saddle-point approximation analogously to Appendix A.1, we obtain from the stationary point for the auxiliary field Q_1 the differential equation for the stationary autocorrelation function $c(\tau)$ corresponding

to Eq. (2.7):

$$(-\partial_\tau^2 + 1)\, c(\tau) = g^2\, f_\phi(c(\tau), c_0) + c_{\mathrm{cn}}(\tau)\,. \tag{A.28}$$

The factor $1 + 1/\tau_\mathrm{n}$ in Eq. (A.26) has been chosen such that, for $g = 0$, the solution of Eq. (A.28) yields the same variance $c_0 = c(0) = \sigma^2$ for all correlation times τ_n and is identical to the case of white-noise input with autocorrelation function $\langle \xi_i(t)\xi_i(s)\rangle = 2\sigma^2\delta(t-s)$.

Except in the case of quenched input, the second term in Eq. (A.28) depends on τ. If we rewrite this equation as a motion of a classical particle, this term corresponds to an additional, time-dependent force:

$$\ddot{c}(\tau) \;=\; -V'(c(\tau)) - c_{\mathrm{cn}}(\tau)\,, \tag{A.29}$$

where the potential $V(c) = V(c; c_0)$ is defined as in Eq. (2.10) and has to be determined self-consistently. Because of the time-dependent force, however, we cannot use energy conservation anymore to obtain c_0.

We may obtain a self-consistent solution numerically: Choosing an initial value $c(0) = c_0 \geq \sigma^2$ arbitrarily, we integrate the differential equation (A.29) forward in time $\tau \geq 0$ based on the system

$$\frac{d}{dt}\begin{pmatrix} c \\ \dot{c} \end{pmatrix} \;=\; \begin{pmatrix} \dot{c} \\ -V'(c; c_0) - c_{\mathrm{cn}}(\tau) \end{pmatrix}$$

with the initial conditions $c(0) = c_0$, $\dot{c}(0) = 0$. To determine the correct initial value c_0, we apply the shooting method to find the initial value c_0 at which $\lim_{\tau\to\infty} c(\tau) = 0$. It leads to a self-consistent autocorrelation function $c(\tau)$ which monotonically decreases to zero.

A.4.2 Quenched Gaussian noise

The case of quenched Gaussian noise follows from the above by considering the limit $\tau_n \to \infty$ in Eq. (A.26). We hence obtain from Eq. (A.28) the differential equation for the autocorrelation function:

$$(-\partial_\tau^2 + 1)\, c(\tau) = g^2\, f_\phi(c(\tau), c_0) + \sigma_\mathrm{q}^2\,, \tag{A.30}$$

where σ_q^2 denotes the variance of the static quenched inputs with mean zero. Since the additional term in Eq. (A.30) does not depend on τ, we can include it into a potential and write

$$\ddot{c}(\tau) = -V_\mathrm{q}'(c(\tau))$$

with the modified potential $V_\mathrm{q}(c; c_0) = V(c; c_0) + \sigma_\mathrm{q}^2 c$. Again the potential has to be determined self-consistently through the initial value $c(0) = c_0$. Without time-dependent input, a constant autocorrelation function $c(\tau) = c_0$ can be a solution; it is determined by $V_\mathrm{q}'(c_0; c_0) = 0$, which has a solution for all values of g. Depending on the coupling strength g, however, other time-dependent self-consistent solutions exist as in the case without any input [29], which corresponds to $\sigma_\mathrm{q}^2 = 0$ here.

We only consider a self-consistent autocorrelation function $c(\tau)$ that monotonically decays to an asymptotic value c_∞, which is related to the variance of the quenched input σ_q^2. In Section 2.5, we describe how to construct this solution. Here, we briefly discuss its existence: Such a solution can only exist if the self-consistent potential $V_\mathrm{q}(c; c_0)$ has both a maximum (at c_∞) and a minimum (between c_∞ and c_0) because the classical motion of $c(\tau)$ starts with zero kinetic energy, i.e., $\dot{c}(0) = 0$. There exists a critical value g_nec at which the second derivative at the endpoint of the potential $V_\mathrm{q}''(c_0; c_0) = -1 + g^2\langle[\phi'(x)]^2\rangle$

changes sign from negative to positive; here, c_0 is determined for each g by $V_\mathrm{q}'(c_0; c_0) = 0$. This implies that for $g > g_\mathrm{nec}$, the self-consistent potential $V_\mathrm{q}(c; c_0)$ for a constant autocorrelation function $c(\tau) = c_0$ ends in a minimum and has an intermediate maximum. As a maximum followed by a minimum of the potential is necessary for the time-dependent, monotonically decaying, self-consistent, autocorrelation function, we conjecture that such a solution only exists for $g > g_\mathrm{nec}$. This result agrees with the condition found for networks with internally generated quenched noise [24].

A.4.3 Lyapunov exponent for colored and quenched noise input

The Lyapunov exponent for quenched or colored noise input can be obtained from the same Schrödinger eigenvalue equation (2.17) as in the white-noise case. This result follows from a replica calculation analogous to the white-noise case (Appendix A.3), where the only difference is the form of the noise term. The saddle-point equation leads to a set of self-consistency equations for the correlations within and between the replicas, analogous to Eq. (A.18). The explicit appearance of the noise term on the right-hand side vanishes in the equation for the distance, Eq. (2.12), after the linearization (A.20); hence, the correlation function of the noise enters only implicitly by determining the quantum potential (A.22) in the Schrödinger equation (A.23), where we need to insert the self-consistently determined autocorrelation function $c(\tau)$ obeying Eq. (A.28) for colored noise or Eq. (A.30) for quenched noise.

A.5 Dynamic mean-field theory for nonvanishing mean coupling

In this appendix, we treat the case of random couplings with nonzero mean. For concreteness, we assume here that the coupling weights are independently and identically Gaussian distributed according to

$$J_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\frac{\bar{g}}{N}, \frac{g^2}{N}\right) . \tag{A.31}$$

Alternatively, we assume a connectivity $J_{ij} \overset{\text{i.i.d.}}{\sim} J_0/\sqrt{N}\,\mathcal{B}(p)$, where $\mathcal{B}(p)$ denotes the Bernoulli distribution with success probability p. This assumption leads to a directed Erdős–Rényi network with connection probability p and nonzero synaptic weights $J_0/\sqrt{N}$. We show below how the latter parameters, J_0, p, determine the former ones, $\bar{g}, g$, and that both network variants behave similar in the limit of large N. The network dynamics again follows Eq. (2.1). The moment-generating functional takes the same form as for the case of vanishing mean coupling, Eq. (2.2). Taking the disorder average, the calculation proceeds analogously to Eq. (A.1). Introducing the abbreviation $y_{ij} = -\int \tilde{x}_i(t)\,\phi\big(x_j(t)\big)\,dt$, we may write the term in the action that is affected as

$$\begin{aligned}
\left\langle \exp\left(-\tilde{x}^\mathrm{T} J \phi\right)\right\rangle_J &= \left\langle \exp\left(\sum_{ij} J_{ij} y_{ij}\right)\right\rangle_J \\
&= \exp\left(\frac{\bar{g}}{N}\sum_{ij} y_{ij} + \frac{1}{2}\frac{g^2}{N}\sum_{ij} y_{ij}^2\right) \\
&= \exp\left(\sum_i R_1^T \tilde{x}_i + \frac{1}{2}\sum_i \tilde{x}_i{}^\mathrm{T} Q_1 \tilde{x}_i\right) ,
\end{aligned} \tag{A.32}$$

where we introduce the auxiliary fields

$$\begin{aligned} R_1(t) &= \frac{\bar{g}}{N} \sum_j \phi(x_j(t)) \,, \\ Q_1(t,s) &= \frac{g^2}{N} \sum_j \phi(x_j(t))\, \phi(x_j(s)) \,. \end{aligned} \tag{A.33}$$

Analogously, the disorder average over an Erdős–Rényi network can be expressed by noting that its cumulant-generating function for a single coupling weight reads

$$\begin{aligned} \varphi_{J_{ij}}(y) &= \ln \left\langle \mathrm{e}^{y \frac{J_0}{\sqrt{N}} s} \right\rangle_{s\sim\mathcal{B}(p)} \\ &= \ln\!\left(p\, \mathrm{e}^{\frac{J_0}{\sqrt{N}} y} - 1 \right) \\ &= \frac{J_0 p}{\sqrt{N}}\, y + \frac{p(1-p)\, J_0^2}{2N}\, y^2 + O(N^{-\frac{3}{2}}) \,, \end{aligned}$$

where the first two cumulants $pJ_0/\sqrt{N}$ and $p(1-p)J_0^2/N$ appear as coefficients in the polynomial in y.

We therefore get the same form for the disorder-averaged action as Eq. (A.32), but with different scalar factors in the definition of the auxiliary fields as $R_1(t) = J_0 p/\sqrt{N} \sum_j \phi(x_j(t))$ and $Q_1(t,s) = p(1-p)J_0^2/(2N) \sum_j \phi(x_j(t))\, \phi(x_j(s))$. The terms that originate from cumulants of order higher than two are suppressed for large network size with at least $O(N^{-3/2})$, so that they do not contribute in the limit of large N. Comparing to Eq. (A.33), we may hence treat the Erdős–Rényi network on the same footing as the network with Gaussian connectivity by identifying the parameters as

$$\begin{aligned} \bar{g} &= \sqrt{N}\, J_0 p \,, \\ g^2 &= p(1-p) J_0^2 \,. \end{aligned}$$

We note that the effective mean coupling strength $\bar{g}$ of the Erdős–Rényi network grows with network size N; to describe large but finite-sized networks in dynamic mean-field theory, one has to use the corresponding value.

We now derive the dynamic mean-field equations for the system. The four-point coupling term involving Q_1 is handled in the same manner as before by introducing an auxiliary field Q_2 that serves to represent a delta functional as in Eq. (A.6). Analogously, we perform a second Hubbard-Stratonovich transform to write

$$\delta\left[-(N/\bar{g})\, R_1 + \sum_i \phi(x_i) \right] \propto \int \mathcal{D}R_2 \, \exp\left(-(N/\bar{g})\, R_2^{\mathrm{T}} R_1 + R_2^{\mathrm{T}} \sum_i \phi(x_i) \right)$$

to obtain the disorder-averaged moment-generating functional

$$\begin{aligned} \bar{Z}[k,\tilde{k},j,\tilde{j}] = \iint \{\mathcal{D}R_{1,2}\} \iint \{\mathcal{D}Q_{1,2}\} \exp\left(-\frac{N}{\bar{g}} R_1^{\mathrm{T}} R_2 - \frac{N}{g^2} Q_1^{\mathrm{T}} Q_2 + N \ln \Omega[R_1, R_2, Q_1, Q_2] \right) \\ \times \exp\left(k^{\mathrm{T}} R_1 + \tilde{k}^{\mathrm{T}} R_2 + j^{\mathrm{T}} Q_1 + \tilde{j}^{\mathrm{T}} Q_2 \right) \,, \end{aligned} \tag{A.34}$$

with

$$\Omega[R_1, R_2, Q_1, Q_2] = \int \mathcal{D}x \int \mathcal{D}\tilde{x} \, \exp\left(S_0[x, \tilde{x}] - R_1^{\mathrm{T}}\tilde{x} + R_2^{\mathrm{T}}\phi(x) + \frac{1}{2}\tilde{x}^{\mathrm{T}}Q_1\tilde{x} + \phi(x)^{\mathrm{T}}Q_2\phi(x)\right), \tag{A.35}$$

where again the generating functional factorizes into independent, identical single-unit contributions Ω giving rise to the factor N in front of $\ln\Omega$.

To formally derive a self-consistent approximation, it is standard to define the vertex-generating functional or effective action as

$$\Gamma[r_1, r_2, q_1, q_2] = \sup_{\{j,\tilde{j},k,\tilde{k}\}} \left\{k^{\mathrm{T}}r_1 + \tilde{k}^T r_2 + j^{\mathrm{T}}q_1 + \tilde{j}^{\mathrm{T}}q_2 - \ln\, Z[k, \tilde{k}, j, \tilde{j}]\right\} .$$

The equations of state

$$\frac{\delta\Gamma}{\delta\{r_1, r_2, q_1, q_2\}} \quad = \quad 0\,, \tag{A.36}$$

solved for the auxiliary fields, then yield a coupled set of self-consistency equations for these fields in the absence of external sources.

In the lowest order, the tree-level approximation [144], the effective action Γ is approximated by the negative action:

$$\Gamma[r_1, r_r, q_1, q_2] \simeq -S[r_1, r_2, q_1, q_2]\,, \tag{A.37}$$

$$S[r_1, r_2, q_1, q_2] = -\frac{N}{\bar{g}}\, r_1^{\mathrm{T}} r_2 - \frac{N}{g^2}\, q_1^{\mathrm{T}} q_2 + N\, \ln\Omega[r_1, r_2, q_1, q_2]\,. \tag{A.38}$$

The self-consistency equations (A.36) for the auxiliary fields then take the explicit form

$$\begin{aligned}
0 &= \frac{\delta\Gamma}{\delta r_1} \rightarrow R_2^*(t) = \bar{g}\,\langle\tilde{x}(t)\rangle_\Omega = 0\,,\\
0 &= \frac{\delta\Gamma}{\delta r_2} \rightarrow R_1^*(t) = \bar{g}\,\langle\phi(x(t))\rangle_\Omega\,,\\
0 &= \frac{\delta\Gamma}{\delta q_1} \rightarrow Q_2^*(s,t) = \frac{g^2}{2}\,\langle\tilde{x}(s)\tilde{x}(t)\rangle_\Omega = 0\,,\\
0 &= \frac{\delta\Gamma}{\delta q_2} \rightarrow Q_1^*(s,t) = g^2\,\langle\phi(x(s))\phi(x(t))\rangle_\Omega\,,
\end{aligned} \tag{A.39}$$

where $\langle\circ\rangle_\Omega$ is the expectation value with respect to the statistics of the fields x and $\tilde{x}$ given by Ω in Eq. (A.35) with the saddle-point values for the auxiliary fields inserted. The effective equation of motion follows from inserting the saddle-point values into the action as

$$S_0[x, \tilde{x}] - R_1^{*\mathrm{T}}\tilde{x} + \frac{1}{2}\tilde{x}^{\mathrm{T}}Q_1^*\tilde{x}\,,$$

which demonstrates that we get a mean input proportional to R_1^* and Gaussian fluctuations with

covariance Q_1^*, analogous to Appendix A.2:

$$\begin{aligned}(\partial_t + 1)\; x(t) &= \bar{g}\,\langle \phi(x(t))\rangle + \eta(t)\,, \\ \langle \eta(t)\,\eta(s)\rangle &= Q_1^*(t,s) + 2\sigma^2\,\delta(t-s)\,,\end{aligned} \tag{A.40}$$

where $\eta(t)$ is again a Gaussian process with zero mean. Hence, the stationary solution for the mean satisfies the nonlinear self-consistency equation

$$\langle x\rangle = \bar{g}\,\langle \phi(x)\rangle\,.$$

Defining the deflection from the mean as $\delta x(t) = x(t) - \langle x\rangle$, we get a differential equation for the stationary covariance function $c(\tau) = \langle \delta x(t+\tau)\delta x(t)\rangle$ [using the equation of motion (A.40) with the mean $\langle x\rangle$ subtracted on both sides]

$$(-\partial_\tau^2 + 1)\, c(\tau) = g^2\, C_{\phi(x)\phi(x)}(t+\tau, t) + 2\sigma^2\,\delta(\tau),$$

which is identical to Eq. (2.7).

The new term is therefore the mean input $\bar{g}\,\langle \phi(x)\rangle$ from the network. We can obtain the joint solution of the mean $\langle x\rangle$ and the covariance function $c(\tau)$ using

$$\begin{aligned}\langle \phi(x)\rangle &= \int Dz\; \phi(\langle x\rangle + \sqrt{c_0}\, z)\,, \\ C_{\phi\phi}(\tau) &= f_{\phi(\circ + \langle x\rangle)}(c_0, c(\tau))\,,\end{aligned}$$

where $f_u(c_0, c_1)$ is defined in Eq. (2.8) and Dz denotes the standard Gaussian measure. Here, we need to insert the function $\phi(\circ + \langle x\rangle)$ to capture the nonzero mean of x; the Gaussian fluctuating part is contained in the definition of f.

Note that for $\langle \phi(x)\rangle \neq 0$, $C_{\phi\phi}(\tau)$ does not decay to 0 as $\tau \to \infty$, but it approaches a nonzero asymptotic value $C_{\phi\phi}(\infty) = f_{\phi(\circ+\langle x\rangle)}(c_0, c_\infty)$. This signifies that the disorder in the connectivity contributes to a distribution of average activities across units in the population.

A.6 Lyapunov exponent for nonvanishing mean coupling

The calculation of the maximum Lyapunov exponent in the case of nonvanishing mean of the connections proceeds along the same line as in Appendix A.3. We therefore only highlight the steps in which the calculation differs from the former case.

As shown in Appendix A.5, Erdős–Rényi networks and networks with Gaussian connectivity lead to the same mean-field equations in the limit of large N. Without loss of generality, we keep the notation of Gaussian connectivity here. To assess the Lyapunov exponent, one needs to compute the disorder-average

over a pair of systems (2.13). The only term affected takes the form

$$\int dJ_{ij}\mathcal{N}\left(\frac{\bar{g}}{N},\frac{g^2}{N},J_{ij}\right)\exp\left(-J_{ij}\sum_{\alpha=1}^{2}\tilde{x}_i^{\alpha\mathrm{T}}\phi(x_j^\alpha)\right)=\exp\left(-\frac{\bar{g}}{N}\sum_{\alpha=1}^{2}\tilde{x}_i^{\alpha\mathrm{T}}\phi(x_j^\alpha)\right)$$
$$\times\exp\left(\frac{g^2}{2N}\sum_{\alpha=1}^{2}\left[\tilde{x}_i^{\alpha\mathrm{T}}\phi(x_j^\alpha)\right]^2\right)$$
$$\times\exp\left(\frac{g^2}{N}\left[\tilde{x}_i^{1\mathrm{T}}\phi(x_j^1)\right]\left[\tilde{x}_i^{2\mathrm{T}}\phi(x_j^2)\right]\right), \qquad (A.41)$$

where the only difference from Eq. (A.13) is the additional first factor on the right-hand side. The important property of this additional term is that it does not couple variables belonging to different replicas; hence, it does not contribute to the correlation between the two replicas. As we demonstrate below, the analysis of the transition to chaos therefore remains unchanged.

As a direct consequence, the corresponding saddle-point equations for the auxiliary variables $R_{1,2}^\alpha$ are identical to Eq. (A.39) and hence read

$$R_2^{*\alpha}(t)=\bar{g}\,\langle\tilde{x}^\alpha(t)\rangle_\Omega=0\,,$$
$$R_1^{*\alpha}(t)=\bar{g}\,\langle\phi(x^\alpha(t))\rangle_\Omega\,.$$

Performing the same steps as in Appendix A.3 leads to the same form of the saddle-point equations (A.16) for the remaining fields, so the generating functional at the saddle point takes the form

$$\bar{Z}^*\propto\left\{\Pi_{\alpha=1}^2\iint\mathcal{D}x^\alpha\mathcal{D}\tilde{x}^\alpha\exp\left(S_0[x^\alpha,\tilde{x}^\alpha]-R_1^{*\alpha\mathrm{T}}\tilde{x}^\alpha+\frac{1}{2}\tilde{x}^{\alpha\mathrm{T}}Q_1^{\alpha*}\tilde{x}^\alpha\right)\right\}\exp\left(\tilde{x}^{1\mathrm{T}}\left(T_1^*+2\sigma^2\right)\tilde{x}^2\right)$$

analogous to Eq. (A.17). The only differences are the terms $-R_1^{*\alpha\mathrm{T}}\tilde{x}^\alpha$, which correspond to the mean input received by the neuron in each replicon. The corresponding set of effective equations can be read off in the same manner as before, yielding Eq. (2.24) given in the main text.

A.7 Memory curve

To evaluate Eq. (2.27) we need to determine the disorder-averaged sum of squared correlation functions

$$\sum_{i=1}^{K}\overline{\langle x_i(t)z(t_0)\rangle^2}\,, \qquad (A.42)$$

with $t_0=t-\tau$ and K denoting the number of neurons available to the readout, which we initially leave as a free parameter. Here, $\langle\circ\rangle$ denotes the average over realizations of the inputs $\xi_i(t)$ (or, alternatively, over time), and the overbar is the average over realizations of the connectivity $\mathbf{J}$ as in Eq. (A.1). Moreover, here we examine a more general signal $z(t)=\sum_{j=1}^N v_j\xi_j(t)$, a linear combination with weights v_i of the inputs to different neurons.

We pick two points in time $t, s \geq t_0$ and define

$$h^K(t,s) = \sum_{i=1}^{K} \overline{\langle x_i(t) z(t_0)\rangle \langle x_i(s) z(t_0)\rangle} \,. \tag{A.43}$$

The quantity of interest (A.42) then follows for $t = s$. The key idea is to express the correlation function $\langle x_i(t) z(t_0)\rangle$ as a weighted sum of response functions $\langle x_i(t)\tilde{x}(t_0)\rangle$, which we show in the following. We introduce a scalar source term $k(t)$ for the signal $z(t)$ and average over the inputs. This leads to the generating functional

$$Z[\mathbf{l}, k] = \int \mathcal{D}\mathbf{x} \int \mathcal{D}\tilde{\mathbf{x}} \, \exp\left(S_0[\mathbf{x}, \tilde{\mathbf{x}}] - \tilde{\mathbf{x}}^{\mathrm{T}} \mathbf{J} \boldsymbol{\phi}(\mathbf{x}) + \mathbf{l}^{\mathrm{T}} \mathbf{x} - \sum_i 2\sigma^2 v_i \, k^{\mathrm{T}} \tilde{x}_i \right). \tag{A.44}$$

Evaluating the correlation yields

$$\langle x_i(t) z(t_0)\rangle = \frac{\delta^2 Z}{\delta l_i(t) \delta k(t_0)}\Big|_{l_i = k = 0} = -2\sigma^2 \sum_{j=1}^{N} v_j \langle x_i(t) \tilde{x}_j(t_0)\rangle \,. \tag{A.45}$$

We next consider a pair of systems (replicas) as in Appendix A.3 with the difference, however, that the two systems receive independent realizations of the inputs. We need two independently driven systems to express the product of the two correlation functions in Eq. (A.43). By independence, the corresponding average factorizes,

$$\begin{aligned} \langle x_i(t) z(t_0)\rangle \langle x_i(s) z(t_0)\rangle &= \left\langle x_i^1(t) z^1(t_0) \, x_i^2(s) z^2(t_0) \right\rangle \\ &= h_i(t, s, t_0) \,, \end{aligned} \tag{A.46}$$

where the superscript denotes the replica index as before. It is now sufficient to introduce a single source term

$$4\sigma^4 \sum_{j,l} v_j v_l \int dt \epsilon(t) \tilde{x}_j^1(t) \tilde{x}_l^2(t) \,, \tag{A.47}$$

with source $\epsilon(t)$, to the corresponding generating functional of the pair, which allows us to obtain $\langle x_i(t) z(t_0)\rangle$ with Eqs. (A.45) and (A.46) as

$$\langle x_i(t) z(t_0)\rangle \langle x_i(s) z(t_0)\rangle = \frac{\delta}{\delta l_i^1(t)} \frac{\delta}{\delta l_i^2(s)} \frac{\delta}{\delta \epsilon(t_0)} Z\big|_{l_i^1 = l_i^2 = \epsilon = 0} \,.$$

The additional source term (A.47) has the physical interpretation of a common input with time-dependent covariance $\epsilon(t)$ injected into pairs of units in the two replicas. The absence of quadratic terms $(\tilde{x}^\alpha)^2$ shows that this common input does not affect the marginal statistics of the two systems alone. This interpretation is only mentioned here for illustrative purposes; the derivation does not rely on it. Because of the weight $v_j v_l$ for different neuron pairs (j, l), we keep the neuron index in the following.

To compute Eq. (A.43), we need the disorder average of Eq. (A.46), $\overline{h_i(t, s, t_0)}$. Similar to Appendix A.3.1, we proceed by deriving a differential equation for this function.

First, after averaging over the disorder, completely analogous to Appendix A.3, we can read off effective equations for the single units,

$$(\partial_t + 1)\, x_i^\alpha(t) = \xi_i^\alpha(t) + \eta_i^\alpha(t) + \rho_i^\alpha(t)\,, \tag{A.48}$$

$\alpha \in \{1,2\}$, $i \in \{1,...,N\}$, together with a set of self-consistency equations for the correlations of the noises:

$$\begin{aligned}
\langle \xi_i^\alpha(t)\, \xi_j^\beta(s)\rangle &= 2\sigma^2 \delta_{\alpha\beta}\delta_{ij}\delta(t-s)\,,\\
\langle \eta_i^\alpha(t)\, \eta_j^\beta(s)\rangle &= \delta_{ij}\frac{g^2}{N}\sum_{k=1}^{N}\langle \phi(x_k^\alpha(t))\phi(x_k^\beta(s))\rangle\,,\\
\langle \rho_i^\alpha(t)\, \rho_j^\beta(s)\rangle &= 4\sigma^4(1-\delta_{\alpha\beta})\, v_i v_j \epsilon(t)\, \delta(t-s)\,.
\end{aligned} \tag{A.49}$$

The first line in Eq. (A.49) represents the independent noise driving the two systems, the second line the common connectivity, and the third line the common noise component we introduced in Eq. (A.47) to generate the function (A.46).

Second, we obtain $\overline{h_i(t,s,t_0)}$ by a functional derivative with respect to $l_i^1(t)$, $l_i^2(s)$, and $\epsilon(t_0)$, which follows from the representation (A.46). Writing the functional derivative with respect to ϵ explicitly as a limit, we can express it by taking a limit of the cross-correlation between the pair of systems as

$$\overline{h_i(t,s,t_0)} = \lim_{\iota\to 0}\frac{1}{\iota}\langle x_i^1(t) x_i^2(s)\rangle\Big|_{\epsilon=\iota\,\delta(\circ-t_0)]}\,, \tag{A.50}$$

where we use that for $\epsilon = 0$ the two systems are uncorrelated.

We now combine the effective equations (A.48) and (A.50) to obtain a partial differential equation:

$$\begin{aligned}
(\partial_t + 1)\,(\partial_s + 1)\,\overline{h_i(t,s,t_0)} = \frac{g^2}{N}\,\lim_{\iota\to 0}\frac{1}{\iota}\Big[\sum_{j=1}^{N}\langle \phi(x_j^1(t))\phi(x_j^2(s))\rangle\Big]\Big|_{\epsilon=\iota\,\delta(\circ-t_0)}\\
+ 4\sigma^4 v_i^2 \delta(t-s)\,\delta(t-t_0)\,.
\end{aligned} \tag{A.51}$$

Since we are interested in the limit $\iota \to 0$, we expand around the uncorrelated state to linear order as follows

$$\begin{aligned}
\langle \phi(x_i^1(t))\phi(x_i^2(s))\rangle &= f_\phi(0,c_0) + \partial_c f_\phi(0,c_0)\,\langle x_i^1(t) x_i^2(s)\rangle\\
&= \iota\overline{\langle\phi'(x_i)\rangle}^2\,\overline{h_i(t,s,t_0)}\,,
\end{aligned} \tag{A.52}$$

where $c_0 = \overline{\langle (x_i(t))^2\rangle}$ is the stationary variance and the first term vanishes as $\overline{\langle\phi(x_i)\rangle} = 0$. Inserting Eq. (A.52) into Eq. (A.51), we arrive at

$$\begin{aligned}
(\partial_t + 1)\,(\partial_s + 1)\,\overline{h_i(t,s,t_0)} = \frac{g^2}{N}\overline{\langle\phi'(x_j)\rangle}^2\sum_{j=1}^{N}\overline{h_j(t,s,t_0)}\\
+ 4\sigma^4\, v_i^2\,\delta(t-t_0)\,\delta(s-t_0)\,.
\end{aligned}$$

Here, $\overline{\langle \phi'(x_j)\rangle}$ is independent of j because the expectation value is taken with respect to the disorder-averaged unperturbed system; thus, we use a representative unit j as the index. Taking the sum over $i = 1, ..., K$ yields

$$(\partial_t + 1)\,(\partial_s + 1)\; h^K(t,s) = g^2 \overline{\left\langle \phi'\left(x_j\right)\right\rangle}^2 \frac{K}{N}\, h^N(t,s) + 4\sigma^4 \|\boldsymbol{v}_K\|^2 \delta(t-t_0)\delta(s-t_0) \tag{A.53}$$

with $\|\boldsymbol{v}_K\|^2 = \sum_{i=1}^K v_i^2$. For the complete sum of squared response functions, $h^N(t,s)$, the following closed linear partial differential equation holds:

$$(\partial_t + 1)\,(\partial_s + 1)\, h^N(t,s) = g^2 \overline{\left\langle \phi'\left(x_j\right)\right\rangle}^2 h^N(t,s) + 4\sigma^4 \delta(t-t_0)\delta(s-t_0)\,, \tag{A.54}$$

where we set $\|\boldsymbol{v}_N\|^2 = \|\boldsymbol{v}\|^2 = 1$ without loss of generality. In the main text Section 2.7, we consider $\|\boldsymbol{v}\|^2 = 1/N$, which only leads to an additional factor in the following and cancels in the formula for the memory curve. The solution to this equation describes the shape of the memory curve if the readout has access to the states of all neurons.

To determine $h^K(t,s)$, we note that the difference $h^K(t,s) - (K/N)\, h^N(t,s)$ is proportional to the solution of

$$(\partial_t + 1)\,(\partial_s + 1)\; h^{(0)}(t,s) = \delta(t-t_0)\delta(s-t_0)\,, \tag{A.55}$$

which by direct integration yields

$$h^{(0)}(t,s) = \mathrm{e}^{-(t-t_0)}\Theta(t-t_0)\mathrm{e}^{-(s-t_0)}\Theta(s-t_0)\,. \tag{A.56}$$

Thus, $h^K(t,s)$ is given by

$$\begin{aligned} h^K(t,s) &= \frac{K}{N} h^N(t,s) + 4\sigma^4 \left(\|\boldsymbol{v}_K\|^2 - \frac{K}{N}\right) h^{(0)}(t,s) \\ &= 4\sigma^4 \frac{K}{N} h^{(1)}(t,s) + 4\sigma^4 \left(\|\boldsymbol{v}_K\|^2 - \frac{K}{N}\right) h^{(0)}(t,s)\,, \end{aligned} \tag{A.57}$$

where $h^{(1)}(t,s)$ solves

$$(\partial_t + 1)\,(\partial_s + 1)\, h^{(1)}(t,s) = a^2 h^{(1)}(t,s) + \delta(t-t_0)\delta(s-t_0) \tag{A.58}$$

with parameter $a^2 = g^2 \overline{\left\langle \phi'(x_j)\right\rangle}^2 = 1 - 1/\tau_\infty^2$. Here, τ_∞ is the asymptotic decay time of the autocorrelation function relating the shape of the memory curve to the decay of the autocorrelations of the network dynamics.

As in Appendix A.3.2, it is useful to change coordinates to $T = t + s - 2t_0$ and $\tau = t - s$. In these coordinates, Eq. (A.58) takes the form

$$(\partial_T + 1)^2 h^{(1)}(T,\tau) - \partial_\tau^2 h^{(1)}(T,\tau) = a^2 h^{(1)}(T,\tau) + 2\,\delta(T)\delta(\tau)$$

and setting $h^{(1)}(T,\tau) = \mathrm{e}^{-T} u(T,\tau)$ simplifies the equation further to

$$\partial_T^2 u(T,\tau) - \partial_\tau^2 u(T,\tau) = a^2 u(T,\tau) + 2\,\delta(T)\delta(\tau)\,, \tag{A.59}$$

a Klein-Gordon wave equation with temporal coordinate T and spatial coordinate τ (and negative squared mass $-a^2$). We are looking for the solution $u(T,\tau)$ in $T \geq 0, \tau \in \mathbb{R}$. To this end, we consider the temporal Laplace and the spatial Fourier transform of Eq. (A.59). Fourier transformation in τ yields

$$\partial_T^2 \hat{u}(T,k) + \left(k^2 - a^2\right)\hat{u}(T,k) = 2\,\delta(T) \tag{A.60}$$

with $k \in \mathbb{R}$ and the Fourier representation

$$u(T,\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{ik\tau} \hat{u}(T,k)\, dk\,.$$

For each $k \in \mathbb{R}$, the Laplace transformation in T,

$$\tilde{u}(p,k) = \int_0^\infty e^{-pT} \hat{u}(T,k)\, dT\,,$$

of Eq. (A.60) reads

$$p^2 \tilde{u}(p,k) - p \underbrace{\hat{u}(0,k)}_{=0} - \underbrace{\partial_T \hat{u}(0,k)}_{=0} + \left(k^2 - a^2\right)\tilde{u}(p,k) = 2\,.$$

Hence, in the Laplace-Fourier domain, we obtain the solution

$$\tilde{u}(p,k) = \frac{2}{p^2 + k^2 - a^2}\,.$$

For the memory curve, we only need the solution $u(T,\tau)$ for $\tau = 0$, the diagonal $s = t$ in the original coordinates. Setting $\tau = 0$ in the Fourier representation gives the Laplace transform of $u(T,\tau=0)$:

$$\begin{aligned} \tilde{u}(p) = \tilde{u}(p,\tau=0) &= \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{1}{k^2 + p^2 - a^2}\, dk \\ &= \frac{1}{\sqrt{p^2 - a^2}} \end{aligned} \tag{A.61}$$

with $p \in \mathbb{C}$ such that $\mathrm{Re}\left(p^2\right) > a^2$. The function on the right is the Laplace transform of the modified Bessel function of the first kind $I_0(aT)$ [248]. Together with $h^{(1)}(T,\tau) = \mathrm{e}^{-T} u(T,\tau)$, we therefore obtain the shape of the memory curve as

$$h^{(1)}(T) = h^{(1)}(T,\tau=0) = \mathrm{e}^{-T}\, I_0(aT)\, \Theta(T)\,. \tag{A.62}$$

Finally, using Eqs. (A.56) and (A.62) in Eq. (A.57) gives the following explicit expression (setting $t_0 = 0$) for the disorder-averaged sum of squared correlation functions, Eq. (A.42):

$$\begin{aligned} h_K(t) = h_K(t,t) &= 4\sigma^4 \frac{K}{N} \mathrm{e}^{-2t}\, I_0(a2t)\, \Theta(t) \\ &\quad + 4\sigma^4 \left(\|\boldsymbol{v}_K\|^2 - \frac{K}{N} \right) \mathrm{e}^{-2t}\, \Theta(t) \\ &= 4\sigma^4 \frac{K}{N} \mathrm{e}^{-2t}\, (I_0(a2t) - 1)\, \Theta(t) \\ &\quad + 4\sigma^4 \|\boldsymbol{v}_K\|^2 \mathrm{e}^{-2t}\, \Theta(t)\,. \end{aligned} \tag{A.63}$$

In Eq. (A.63), we split $h_K(t)$ into two contributions: a network contribution $h_K^{\mathrm{net}}(t)$ proportional to K/N with shape $\mathrm{e}^{-2t}\, (I_0(a2t) - 1)\, \Theta(t)$ and a local contribution proportional to $\|\boldsymbol{v}_K\|^2$ with shape $\mathrm{e}^{-2t}\, \Theta(t)$. The latter is just the memory of the signal due to the leaky integration of the single units, while the former describes the memory due to the collective network dynamics; only this contribution is affected by the network coupling strength.

We can evaluate Eq. (2.27) using Eq. (A.63) with the choice $v_i = 1/N$, which leads to Eq. (2.28) and to the network memory (2.29).

APPENDIX B

Lyapunov exponents for integrate-and-fire neuron dynamics in Chapter 3

B.1 Mean-field Lyapunov exponents

In this appendix, we compute the mean-field LEs for LIF and XIF neurons in Chapter 3. For this, we describe the dynamics by a sequence of discrete maps from the state at a time shortly after a spike to the state at a time shortly after the next spike. We take a stroboscopic perspective, i.e., the times are fixed by the reference trajectory. In our mean-field approach, we focus on the evolution of small perturbations in the effective single neuron dynamics and neglect perturbations of the input spike times. To this end, we compute the scalar "single spike Jacobian" factor [65, 83] $J_{\mathrm{mf}}(t_{k+1})$ that describes the linear evolution of infinitesimal perturbations of the neuron's membrane potential from time $t_k^+ = t_k + \varepsilon$ to time $t_{k+1}^+ = t_{k+1} + \varepsilon$ (with arbitrarily small $\varepsilon > 0$), where t_k, $k \in \mathbb{N}$, are the times at which the considered neuron spikes and resets. In this approximation, the additive inhibitory input spikes only prolong the duration $t_{k+1} - t_k$ between two consecutive spike times of the neuron; they do not affect perturbations directly.

To compute the Jacobian factor

$$J_{\mathrm{mf}}(t_{k+1}) = \lim_{\varepsilon \searrow 0} \frac{\partial V(t_{k+1} + \epsilon)}{\partial V(t_k + \epsilon)}$$

it is helpful to note that when the neuron spikes and resets, say at time t_k, an infinitesimal voltage perturbation $\delta V(t) = \tilde{V}(t) - V(t)$ gets translated into an infinitesimal perturbation of the spike time, $\delta t_k = \tilde{t}_k - t_k$. Hence, at their respective spike times, the perturbed and the unperturbed membrane potentials are reset, i.e. $\tilde{V}(\tilde{t}_k) = V(t_k) = V_{\mathrm{re}} = 0$. To obtain the resulting infinitesimal voltage perturbation $\delta V(t_k + \epsilon)$ shortly after the spike, we evolve the perturbed and the unperturbed membrane potentials using the autonomous subthreshold dynamics to the time $t_k + \epsilon$ with $\max(\tilde{t}_k, t_k) < t_k + \epsilon$. For infinitesimal δt_k, we can take the limit $\varepsilon \searrow 0$ in the end. Linear Taylor expansion yield

$$\begin{aligned}\tilde{V}(t_k + \epsilon) &= \tilde{V}(\tilde{t}_k) + \dot{V}\left(\tilde{V}(\tilde{t}_k)\right)(t_k + \epsilon - \tilde{t}_k) + \cdots \\ &= V_{\mathrm{re}} + \dot{V}(V_{\mathrm{re}})(\epsilon - \delta t_k) + \cdots\end{aligned}$$

and

$$\begin{aligned} V(t_k + \epsilon) &= V(t_k) + \dot{V}\left(V(t_k)\right)(t_k + \epsilon - t_k) + \cdots \\ &= V_{\mathrm{re}} + \dot{V}\left(V_{\mathrm{re}}\right)\epsilon + \cdots \end{aligned}$$

with the subthreshold dynamics' vector field $\dot{V}(V) = -\gamma V + I^{\mathrm{ext}}$. Therefore, the infinitesimal voltage perturbation immediately after the spike reads

$$\delta V(t_k^+) = \lim_{\varepsilon \searrow 0} \delta V(t_k + \epsilon) = -\dot{V}\left(V_{\mathrm{re}}\right)\delta t_k \,. \tag{B.1}$$

An initial infinitesimal voltage perturbation $\delta V(t_k^+)$ at an equal time t_k^+ immediately after a spike evolves according to the linearization of the subthreshold dynamics, $\dot{\delta V} = -\gamma\,\delta V$. This evolution can be directly integrated up to a time $t_{k+1} - \varepsilon$ shortly before the next spike, which leads to

$$\delta V(t_{k+1} - \varepsilon) = \mathrm{e}^{-\gamma(t_{k+1} - \epsilon - t_k)}\,\delta V(t_k^+) \,. \tag{B.2}$$

Here, we point out again that inhibitory input spikes occurring at identical times for the perturbed and the unperturbed neuron dynamics do not directly change voltage perturbations because they impact the membrane potential additively. We further assume that the spike times are generic: input spike times do not exactly coincide with the neuron's spike times [and for a XIF neuron also not with instances at which $V(t) = V_{\mathrm{cutoff}}$, the cutoff of the step-function voltage dependence (3.3)]. As mentioned above, an infinitesimal voltage perturbation $\delta V(t_{k+1} - \varepsilon)$ gets translated into an infinitesimal perturbation $\delta t_{k+1} = \tilde{t}_{k+1} - t_{k+1}$ of the next spike time. To relate these, we evolve the perturbed and the unperturbed potential backwards in time from their respective spike times using again linear Taylor expansion:

$$\begin{aligned} \tilde{V}(t_{k+1} - \varepsilon) &= \tilde{V}(\tilde{t}_{k+1}^-) + \dot{V}\left(\tilde{V}(\tilde{t}_{k+1}^-)\right)(t_{k+1} - \varepsilon - \tilde{t}_{k+1}) + \cdots \\ &= V_{\mathrm{th}} - \dot{V}\left(V_{\mathrm{th}}\right)(\epsilon + \delta t_{k+1}) + \cdots \end{aligned}$$

and

$$\begin{aligned} V(t_{k+1} - \varepsilon) &= V(t_{k+1}^-) + \dot{V}\left(V(t_{k+1}^-)\right)(t_{k+1} - \varepsilon - t_{k+1}) + \cdots \\ &= V_{\mathrm{th}} - \dot{V}\left(V_{\mathrm{th}}\right)\epsilon + \cdots \end{aligned}$$

with $t_{k+1} - \varepsilon < \min(\tilde{t}_{k+1}, t_{k+1})$. The relationship between the infinitesimal voltage perturbation immediately before the next spike and the infinitesimal perturbation of the corresponding spike time is thus

$$\delta V(t_{k+1}^-) = \lim_{\varepsilon \searrow 0} \delta V(t_{k+1} - \varepsilon) = -\dot{V}\left(V_{\mathrm{th}}\right)\delta t_{k+1} \,. \tag{B.3}$$

Taking the limit $\varepsilon \searrow 0$ in Eq. (B.2) gives this voltage perturbation in terms of the perturbation immediately after the previous spike, $\delta V(t_{k+1}^-) = \mathrm{e}^{-\gamma(t_{k+1} - t_k)}\,\delta V(t_k^+)$.

From Eq. (B.3) together with Eq. (B.1) evaluated at t_{k+1}, we obtain the infinitesimal voltage perturbation

immediately after the next spike

$$\delta V(t_{k+1}^+) = -\dot{V}\left(V_{\mathrm{re}}\right)\delta t_{k+1} = \frac{\dot{V}\left(V_{\mathrm{re}}\right)}{\dot{V}\left(V_{\mathrm{th}}\right)}\delta V(t_{k+1}^-) \tag{B.4}$$

$$= \frac{I^{\mathrm{ext}}}{-\gamma V_{\mathrm{th}} + I^{\mathrm{ext}}}\, \mathrm{e}^{-\gamma(t_{k+1}-t_k)}\, \delta V(t_k^+)\,, \tag{B.5}$$

where we inserted the subthreshold vector field $\dot{V}(V) = -\gamma V + I^{\mathrm{ext}}$ (with $V_{\mathrm{re}} = 0$). The "saltation" factor $\dot{V}\left(V_{\mathrm{re}}\right)/\dot{V}\left(V_{\mathrm{th}}\right)$ in Eq. (B.4) describes the effect of the neuron's spike and reset on infinitesimal voltage perturbations [64, 355, 356]. It can be expressed using the spike frequency $\rho_{\mathrm{free}} = 1/T_{\mathrm{free}}$ of the freely oscillating neuron dynamics in the absence of recurrent inhibition as

$$\frac{I^{\mathrm{ext}}}{-\gamma V_{\mathrm{th}} + I^{\mathrm{ext}}} = \mathrm{e}^{\gamma T_{\mathrm{free}}} = \mathrm{e}^{\gamma/\rho_{\mathrm{free}}}\,,$$

which holds for both LIF and XIF neurons, see also Eqs. (3.4) and (3.5). With this expression, Eq. (B.5) yields the mean-field single neuron Jacobian factor $J_{\mathrm{mf}}(t_{k+1}) = \exp(\gamma/\rho_{\mathrm{free}} - \gamma(t_{k+1} - t_k))$ stated in Eq. (3.20).

The asymptotic rate of exponential growth or shrinkage of infinitesimal perturbations is then given by the long-term average of the logarithms of the Jacobian factors resulting in the mean-field LE

$$\begin{aligned}
\lambda_{\mathrm{mf}} &= \lim_{L\to\infty} \frac{1}{t_L} \sum_{k=1}^{L} \ln\left|J_{\mathrm{mf}}(t_k)\right| \\
&= \lim_{L\to\infty} \frac{1}{t_L} \sum_{k=1}^{L} \left(\frac{\gamma}{\rho_{\mathrm{free}}} - \gamma(t_k - t_{k-1})\right) \\
&= \gamma \lim_{L\to\infty} \left(\frac{1}{\rho_{\mathrm{free}}}\frac{L}{t_L} - \frac{(t_L - t_0)}{t_L}\right) \\
&= \gamma\left(\frac{\rho}{\rho_{\mathrm{free}}} - 1\right)
\end{aligned} \tag{B.6}$$

with $t_0 = 0$ and the long-term average firing rate $\rho = \lim_{L\to\infty} L/t_L$.

B.2 Single spike Jacobian matrix of the recurrent network dynamics

Here, we derive an expression for the single spike Jacobian matrix [65, 83] of the full recurrent network dynamics (3.1). Therefore, we consider the vector $\delta V_i(t) = \tilde{V}_i(t) - V_i(t)$, $i \in \{1, \ldots, N\}$, of infinitesimal membrane potential perturbations. During the networks dynamics, the neurons only interact at spike events and the single spike Jacobian matrix $J_{ij}(t_{k+1})$, $i, j \in \{1, \ldots, N\}$, describes the linear evolution of infinitesimal perturbations of the network's state between subsequent spike times. We denote the sequence of all spike times in the network by t_k, $k \in \mathbb{N}$. We assume that there are no exact coincident spike times of different neurons, which we expect to hold for generic trajectories of our autonomous

inhibitory network dynamics. As in Appendix B.1 we take a stroboscopic perspective, i.e., the spike times are fixed by the reference trajectory.

Our goal is to obtain the single spike Jacobian matrix with components

$$J_{ij}(t_{k+1}) = \lim_{\varepsilon \searrow 0} \frac{\partial V_i(t_{k+1} + \epsilon)}{\partial V_j(t_k + \epsilon)} .$$

The initial infinitesimal voltage perturbations $\delta V_i(t_k^+) = \lim_{\varepsilon \searrow 0} \delta V_i(t_k + \epsilon)$ immediately after the spike at time t_k evolve for each neuron i separately according to the linear subthreshold dynamics, $\delta \dot{V}_i = -\gamma_i \, \delta V_i$ [see Eq. (3.1)]. This evolution yields the vector of perturbations at a time $t_{k+1} - \epsilon$ shortly before the next spike

$$\delta V_i(t_{k+1} - \varepsilon) = \mathrm{e}^{-\gamma_i (t_{k+1} - \epsilon - t_k)} \delta V_i(t_k^+) . \tag{B.7}$$

At time t_{k+1}, a neuron in the network emits a spike and is reset; we denote its index by $l \in \{1, \ldots, N\}$. The effect of the spike and reset mechanism on the neuron's own voltage perturbation $\delta V_l(t_{k+1}^+)$ is given by Eq. (B.5) in Appendix B.1 above. Importantly, an infinitesimal voltage perturbation $\delta V_l(t_{k+1}^-) = \lim_{\varepsilon \searrow 0} \delta V_l(t_{k+1} - \epsilon)$ of the neuron immediately before the spike gets translated into an infinitesimal perturbation $\delta t_{k+1} = \tilde{t}_{k+1} - t_{k+1}$ of the spike time [Eq. (B.3)]:

$$\begin{aligned} \delta t_{k+1} &= -\frac{1}{\dot{V}_l \, (V_\mathrm{th})} \delta V_l(t_{k+1}^-) \\ &= -\frac{1}{-\gamma_l V_\mathrm{th} + I_l^\mathrm{ext}} \delta V_l(t_{k+1}^-) . \end{aligned} \tag{B.8}$$

Here and in the following, we use the subthreshold dynamics' vector fields $\dot{V}_i(V) = -\gamma_i V + I_i^\mathrm{ext}$.

The infinitesimal perturbation δt_{k+1} of the spike time t_{k+1} leads to an additional perturbation of the membrane potentials of the postsynaptic neurons, because the perturbed and the unperturbed potentials, $\tilde{V}_i(t) = V_i(t) + \delta V_i(t)$ and $V_i(t)$, receive the resulting jump-like decreases at infinitesimally different times, $\tilde{t}_{k+1} = t_{k+1} + \delta t_{k+1}$ and t_{k+1}. Following a similar strategy as in Appendix B.1, we evolve the perturbed and the unperturbed potentials from the times $\tilde{t}_{k+1}$ and t_{k+1} backwards in time to an equal time $t_{k+1} - \epsilon$ with $t_{k+1} - \varepsilon < \min(\tilde{t}_{k+1}, t_{k+1})$ using linear Taylor expansion:

$$\tilde{V}_i(t_{k+1} - \varepsilon) = \tilde{V}_i(\tilde{t}_{k+1}^-) - \dot{V}_i \left(\tilde{V}_i(\tilde{t}_{k+1}^-) \right) (\epsilon + \delta t_{k+1}) + \cdots$$

and

$$V_i(t_{k+1} - \varepsilon) = V_i(t_{k+1}^-) - \dot{V}_i \left(V_i(t_{k+1}^-) \right) \epsilon + \cdots ,$$

where $\tilde{V}_i(\tilde{t}_{k+1}^-)$ and $V_i(t_{k+1}^-)$ are the left-hand side limits at the respective spike times $\tilde{t}_{k+1}$ and t_{k+1}. Hence, we can relate the infinitesimal voltage perturbations at an equal time shortly before the spike to the infinitesimal difference $\tilde{V}_i(\tilde{t}_{k+1}^-) - V_i(t_{k+1}^-)$ of the perturbed and the unperturbed membrane potentials when the spike is received:

$$\delta V_i(t_{k+1}^-) = \lim_{\varepsilon \searrow 0} \delta V_i(t_{k+1} - \varepsilon) = \tilde{V}_i(\tilde{t}_{k+1}^-) - V_i(t_{k+1}^-) - \dot{V}_i \left(\tilde{V}_i(\tilde{t}_{k+1}^-) \right) \delta t_{k+1} . \tag{B.9}$$

The receiving of the spike emitted by neuron l results in the jump-like decrease of the perturbed and

unperturbed membrane potentials according to

$$\tilde{V}_i(\tilde{t}_{k+1}^+) = \tilde{V}_i(\tilde{t}_{k+1}^-) + h_i\left(\tilde{V}_i(\tilde{t}_{k+1}^-)\right) C_{il}$$

and

$$V_i(t_{k+1}^+) = V_i(t_{k+1}^-) + h_i\left(V_i(t_{k+1}^-)\right) C_{il}\,,$$

where $\tilde{V}_i(\tilde{t}_{k+1}^+)$ and $V_i(t_{k+1}^+)$ are the right-hand side limits at the respective spike times $\tilde{t}_{k+1}$ and t_{k+1}. For the step-function voltage dependency $h_i(V_i)$ given by Eq. (3.3), the impact is additive and the infinitesimal difference $\tilde{V}_i(\tilde{t}_{k+1}^+) - V_i(t_{k+1}^+)$ remains unchanged [neglecting exceptional trajectories with $V_i(t_{k+1}^-) = V_{\text{cutoff}}$, which we expect to not be generic]. Together with Eq. (B.9), we therefore obtain

$$\tilde{V}_i(\tilde{t}_{k+1}^+) - V_i(t_{k+1}^+) = \delta V_i(t_{k+1}^-) + \dot{V}_i\left(\tilde{V}_i(\tilde{t}_{k+1}^-)\right)\delta t_{k+1}\,. \tag{B.10}$$

As a final step, we need to evolve the perturbed and the unperturbed membrane potentials to an equal time $t_{k+1}+\epsilon$ with $\max(\tilde{t}_{k+1}, t_{k+1}) < t_{k+1}+\epsilon$ shortly after the spike. The differences of the linear Taylor expansions

$$\tilde{V}_i(t_{k+1}+\varepsilon) = \tilde{V}_i(\tilde{t}_{k+1}^+) + \dot{V}_i\left(\tilde{V}_i(\tilde{t}_{k+1}^+)\right)(\epsilon - \delta t_{k+1}) + \cdots$$

and

$$V_i(t_{k+1}+\varepsilon) = V_i(t_{k+1}^+) + \dot{V}_i\left(V_i(t_{k+1}^+)\right)\epsilon + \cdots$$

yield the infinitesimal voltage perturbations $\delta V_i(t_{k+1}^+) = \lim_{\varepsilon\searrow 0}\delta V_i(t_{k+1}+\varepsilon)$ of the postsynaptic neurons $i \neq l$ immediately after the spike of neuron l as follows:

$$\begin{aligned}
\delta V_i(t_{k+1}^+) &= \tilde{V}_i(\tilde{t}_{k+1}^+) - V_i(t_{k+1}^+) - \dot{V}_i\left(\tilde{V}_i(\tilde{t}_{k+1}^+)\right)\delta t_{k+1} \\
&= \delta V_i(t_{k+1}^-) - \left[\dot{V}_i\left(\tilde{V}_i(\tilde{t}_{k+1}^+)\right) - \dot{V}_i\left(\tilde{V}_i(\tilde{t}_{k+1}^-)\right)\right]\delta t_{k+1} \\
&= \delta V_i(t_{k+1}^-) + \gamma_i\left[\tilde{V}_i(\tilde{t}_{k+1}^+) - \tilde{V}_i(\tilde{t}_{k+1}^-)\right]\delta t_{k+1} \\
&= \delta V_i(t_{k+1}^-) + \gamma_i h_i\left(V_i(t_{k+1}^-)\right) C_{il}\,\delta t_{k+1}\,,
\end{aligned} \tag{B.11}$$

where we used Eq. (B.10) in the second line and $\dot{V}_i(V) = -\gamma_i V + I_i^{\text{ext}}$ in the third. The second term in the last line of Eq. (B.11) describes how an infinitesimal perturbation δt_{k+1} of the spike time of neuron l is added to the infinitesimal voltage perturbations of the postsynaptic neurons. This term depends linearly on the coupling sizes $h_i\left(V_i(t_{k+1}^-)\right) C_{il}$ of the jump-like network interactions because of the linearity of the subthreshold vector fields $\dot{V}_i(V)$.

If we write δt_{k+1} in Eq. (B.11) in terms of the infinitesimal voltage perturbation $\delta V_l(t_{k+1}^-)$ of the spiking neuron l using Eq. (B.8), we obtain the linear "saltation" map [64, 355, 356] of infinitesimal voltage perturbations (at equal times) after a single spike in the network,

$$\delta V_i(t_{k+1}^+) = \begin{cases} \delta V_i(t_{k+1}^-) + \frac{-\gamma_i h_i\left(V_i(t_{k+1}^-)\right) C_{il}}{I_l^{\text{ext}} - \gamma_l V_{\text{th}}}\delta V_l(t_{k+1}^-) & \text{for } i \neq l \\ \frac{I_l^{\text{ext}}}{I_l^{\text{ext}} - \gamma_l V_{\text{th}}}\delta V_l(t_{k+1}^-) & \text{for } i = l \end{cases}, \tag{B.12}$$

with the "saltation" factor from Eq. (B.11) for $i = l$, which describes the effect of the spike and reset mechanism. The uncoupled subthreshold evolution of infinitesimal voltage perturbations yields $\delta V_i(t_{k+1}^-) = \mathrm{e}^{-\gamma_i(t_{k+1}-t_k)}\delta V_i(t_k^+)$ for all neurons i. Hence, the components $\delta V_i(t_{k+1}^+)/\delta V_j(t_k^+)$ of the

single spike Jacobian matrix can be read off from Eq. (B.12) as

$$J_{ij}(t_{k+1}) = \begin{cases} \mathrm{e}^{-\gamma_i(t_{k+1}-t_k)} & \text{for } i = j \neq l \\ \frac{-\gamma_i h_i\left(V_i(t_{k+1}^-)\right)C_{il}}{I_l^{\mathrm{ext}}-\gamma_l V_{\mathrm{th}}}\mathrm{e}^{-\gamma_l(t_{k+1}-t_k)} & \text{for } i \neq l\,,\ j = l \\ \mathrm{e}^{-\gamma_i(t_{k+1}-t_k)} + \frac{\gamma_l V_{\mathrm{th}}}{I_l^{\mathrm{ext}}-\gamma_l V_{\mathrm{th}}}\mathrm{e}^{-\gamma_l(t_{k+1}-t_k)} & \text{for } i = j = l \\ 0 & \text{otherwise} \end{cases} \tag{B.13}$$

$$= \mathrm{e}^{-\gamma_i(t_{k+1}-t_k)}\delta_{ij} + \frac{\gamma_l V_{\mathrm{th}}\delta_{il} - \gamma_i h_i\left(V_i(t_{k+1}^-)\right)C_{il}}{I_l^{\mathrm{ext}}-\gamma_l V_{\mathrm{th}}}\mathrm{e}^{-\gamma_l(t_{k+1}-t_k)}\delta_{jl}\,.$$

The second expression using Kronecker deltas exposes that the single spike Jacobian matrix (B.13) equals a sum of a diagonal matrix and a matrix with a single nonzero column, i.e. a rank-one matrix.

We finish this appendix with computing the determinant of the single spike Jacobian matrix (B.13). This determinant equals the change of an infinitesimal N-dimensional volume in the voltage phase space of the recurrent network dynamics between subsequent spike events. We employ the matrix determinant lemma, $\det\left(\mathbf{A} + \mathbf{a}\,\mathbf{b}^{\mathrm{T}}\right) = \left(1 + \mathbf{b}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{a}\right)\det(\mathbf{A})$, for the sum of an invertible matrix $\mathbf{A}$ and a rank-one matrix $\mathbf{a}\,\mathbf{b}^{\mathrm{T}}$. It yields

$$\begin{aligned} \det\left(J_{ij}(t_{k+1})\right) &= \left(1 + \frac{\gamma_l V_{\mathrm{th}}}{I_l^{\mathrm{ext}}-\gamma_l V_{\mathrm{th}}}\right)\exp\left(-\sum_{i=1}^{N}\gamma_i(t_{k+1}-t_k)\right) \\ &= \exp\left(\frac{\gamma_l}{\rho_{\mathrm{free},l}} - \sum_{i=1}^{N}\gamma_i(t_{k+1}-t_k)\right), \end{aligned} \tag{B.14}$$

with the free spike frequency $\rho_{\mathrm{free},l}$ of neuron l in the absence of recurrent input, since

$$\frac{I_l^{\mathrm{ext}}}{I_l^{\mathrm{ext}}-\gamma_l V_{\mathrm{th}}} = \mathrm{e}^{\gamma/\rho_{\mathrm{free},l}}$$

holds for both LIF and XIF neurons.

APPENDIX C

Parameters and details on the random walk models in Chapter 5

C.1 Parameters of models used in the simulations

LIF model in Figs. 5.3 and 5.4

Neuron numbers: excitatory neurons: $N_E = 102$; interior neurons: $N_{\text{int}} = 90$; periphery neurons: 12; inhibitory neurons: $N_I = 20$.

Network structure: connection probability between interior neurons: $p_{\text{int}} = 1$; connection probability between interior and periphery neurons: $p_{\text{peri}} = 1$; periphery neurons have no connections between each other; connection probability between excitatory and inhibitory neurons and between inhibitory and inhibitory neurons: 1; there are no self-connections of neurons.

Neuron parameters: spike threshold: $V_\theta = 20\,\text{mV}$; reset potential $V_0 = 0\,\text{mV}$; resting potential: $V_{\text{rest}} = 10\,\text{mV}$; membrane time constant: $\tau_m = 10\,\text{ms}$; absolute refractory period: $\tau_{\text{ref}} = 5\,\text{ms}$; sum of input and sum of output weights of an interior neuron: $w_{\text{sum}} = 256.25\,\text{mV} = \frac{1}{2}\left[\left(N_{\text{asbly}} - 1\right) p_{\text{int}} w_{\text{max}} + N_{\text{peri}} p_{\text{peri}} w_{\text{max,peri}}\right]$, the square bracketed term is the expected input of an interior neuron from a typical size assembly and its periphery neurons, if all weights were at their individual maximum; sum of input and sum of output weights of a periphery neuron: $w_{\text{sum,peri}} = 225.0\,\text{mV} = \frac{1}{5}\left[N_{\text{asbly}} p_{\text{peri}} w_{\text{max,peri}}\right]$, the square bracketed term is the expected input of a periphery neuron from a typical size assembly, if all weights are at their individual maximum; noise strength: $\sigma = 3.5\,\text{mV}$.

Excitatory synapses: time constant: $\tau_E = 2\,\text{ms}$; maximal synaptic strength of synapses between interior neurons: $w_{\text{max}} = 12.5\,\text{mV}$, evoking a peak EPSP of 1.67 mV in a resting postsynaptic neuron; maximal synaptic strength of synapses between interior and periphery neurons: $w_{\text{max,peri}} = 37.5\,\text{mV}$, evoking a peak EPSP of 5.02 mV in a resting postsynaptic neuron; strength of synapses to inhibitory neurons: $w_{\text{E}\to\text{I}} = 5.02\,\text{mV}$, evoking a peak EPSP of 0.67 mV in a resting postsynaptic neuron.

Inhibitory synapses: time constant: $\tau_I = 5\,\text{ms}$; strength of synapses to excitatory neurons: $w_{\text{I}\to\text{E}} = -5.13\,\text{mV}$, evoking a peak inhibitory postsynaptic potential (IPSP) of $-1.28\,\text{mV}$ in a resting postsynaptic neuron; strength of synapses to inhibitory neurons: $w_{\text{I}\to\text{I}} = -5.39\,\text{mV}$ evoking a peak IPSP of $-1.35\,\text{mV}$ in a resting postsynaptic neuron.

STDP window: $\Delta w_{ij}(\Delta t) = \frac{\eta}{a - b(1+\delta)} \left[a \exp(-a\,|\Delta t|) - b(1+\delta) \exp(-b\,|\Delta t|)\right]$, where $\Delta t = t_i - t_j$

is the time difference between the postsynaptic and the presynaptic spike; window amplitude for connections between interior neurons: $\eta = 3.75\,\mathrm{mV}$; window amplitude for connections between interior and periphery neurons: $\eta = 1.25\,\mathrm{mV}$; LTP peak at 0 ms; LTP decay rate: $a = 1/\tau_{\mathrm{LTP}} = 1/(20\,\mathrm{ms})$; LTD decay rate: $b = 1/\tau_{\mathrm{LTD}} = 1/(40\,\mathrm{ms})$; ratio of integrated LTD and LTP: $1 + \delta = 1 + 1/3$.

Spontaneous synaptic turnover: none.

Memory representation: number of assemblies: $n_{\mathrm{asbly}} = 3$; initial number of interior neurons per assembly: $N_{\mathrm{asbly}}(0) = 30$; periphery neurons per assembly: $N_{\mathrm{peri}} = 4$.

Simulation: time step: 0.25 ms; total simulated time: 75 hours.

LIF network model for transition mechanism analysis in Fig. 5.5

Same parameters as above with the following exceptions:

Neuron numbers: excitatory neurons: $N_E = 68$; interior neurons: $N_{\mathrm{int}} = 68$; periphery neurons: 0; inhibitory neurons: $N_I = 13$.

Neuron parameters: sum of input and sum of output weights of an interior neuron: $w_{\mathrm{sum}} = 309.375\mathrm{mV} = \frac{3}{4}\left(N_{\mathrm{asbly}} - 1\right) p_{\mathrm{int}} w_{\mathrm{max}}$.

Excitatory synapses: strength of synapses to inhibitory neurons: $w_{\mathrm{E}\to\mathrm{I}} = 9.10\,\mathrm{mV}$, evoking a peak EPSP of 1.22 mV in a resting postsynaptic neuron.

Inhibitory synapses: strength of synapses to excitatory neurons: $w_{\mathrm{I}\to\mathrm{E}} = -9.52\,\mathrm{mV}$, evoking a peak IPSP of −2.38 mV in a resting postsynaptic neuron; strength of synapses to inhibitory neurons: $w_{\mathrm{I}\to\mathrm{I}} = -10.31\,\mathrm{mV}$ evoking a peak IPSP of −2.58 mV in a resting postsynaptic neuron.

STDP window: window amplitude: $\eta = 5\,\mathrm{mV}$.

Spontaneous synaptic turnover: none.

Memory representation: number of assemblies: $n_{\mathrm{asbly}} = 2$; initial number of interior neurons per assembly: $N_{\mathrm{asbly}}(0) = 34$.

Simulation: total simulated time: 50 hours.

LIF random walk model in Fig. 5.5

Same parameters as in the LIF network model for the transition mechanism analysis above where applicable. Additional parameters:

Assemblies: number of assemblies: $n_{\mathrm{asbly}} = 2$; number of neurons per assembly: $N_{\mathrm{asbly}} = 33$; assembly reactivation rate: $r_{\mathrm{asbly}} = 0.75\,\mathrm{Hz}$; assembly reactivation duration: $\tau_{\mathrm{react}} = 15\,\mathrm{ms}$; average waiting time between assembly reactivations: $1/\left(n_{\mathrm{asbly}} r_{\mathrm{asbly}}\right) = 0.66\,\mathrm{s}$.

Neuron parameters: background spike rate: $r_{\mathrm{bg}} = 0.75\,\mathrm{Hz}$; coactivation probability: Eq. (C.23) with $\mu_V = V_{\mathrm{rest}} = 10\,\mathrm{mV}$ and $\sigma_V = \sigma = 3.5\,\mathrm{mV}$; sum of input weights: $w_{\mathrm{sum}} = 1$; weight translation and attenuation factor: $\gamma = 0.73 \cdot 41.4\,\mathrm{mV} = 30.2\,\mathrm{mV}$, equal to the peak compound post-synaptic potential evoked by an assembly reactivation for input weight 1.

STDP window: window shape: $h(\Delta t) = \left[a \exp(-a\,|\Delta t|) - b(1+\delta)\exp(-b\,|\Delta t|)\right] / (a - b(1+\delta))$ with integral $\int_{-\infty}^{+\infty} h(\Delta t)\, d\Delta t = -40\,\mathrm{ms}$; window amplitude: $\eta = 0.01616$; positive zero-crossing: $\tau_{\mathrm{pot}} = 16.2\,\mathrm{ms}$

Weight changes: Equations (C.4) and (C.11): $\overline{\Delta w_a^{\mathrm{bg,react}}} = -\eta\, N_{\mathrm{asbly}}\, 0.03 = -0.016$ and $\mathrm{Std}\left(\Delta w_a^{\mathrm{bg,react}}\right) = 0.062$; Eqs. (C.13) and (C.17): $\overline{\Delta w_a^{\mathrm{coact}}} = \eta\, N_{\mathrm{asbly}}\, 0.69 = 0.37$ and $\mathrm{Std}\left(\Delta w_a^{\mathrm{coact}}\right) = 0.04$; Eqs. (C.19) and (C.21): $\overline{\Delta w_b^{\mathrm{coact,bg}}} = -\eta\, N_{\mathrm{asbly}}\, 0.03 = -0.016$ and $\mathrm{Std}\left(\Delta w_b^{\mathrm{coact,bg}}\right) = 0.011$;

Eq. (C.22): $\overline{\Delta w_b^{\mathrm{bg,bg}}}(T) = -\eta\, N_{\mathrm{asbly}}\, T\, 0.0225\,\mathrm{s}^{-1} = -T\, 0.012\,\mathrm{s}^{-1}$ with time interval length T measured in seconds.

Simulation: event-based; total simulated time: 1000 hrs.

LIF model in Figs. 5.6 and 5.7

Same parameters as in the LIF model of Fig. 5.3 with the following exceptions:

Network structure: connection probability between interior neurons: $p_{\mathrm{int}} = 0.6$; connection probability between interior and periphery neurons: $p_{\mathrm{peri}} = 0.8$;

Neuron parameters: sum of input and sum of output weights of an interior neuron: $w_{\mathrm{sum}} = 253.125\,\mathrm{mV} = \frac{3}{4}\left[\left(N_{\mathrm{asbly}} - 1\right) p_{\mathrm{int}} w_{\mathrm{max}} + N_{\mathrm{peri}} p_{\mathrm{peri}} w_{\mathrm{max,peri}}\right]$; sum of input and sum of output weights of a periphery neuron: $w_{\mathrm{sum,peri}} = 225.0\,\mathrm{mV} = \frac{1}{4}\left[N_{\mathrm{asbly}} p_{\mathrm{peri}} w_{\mathrm{max,peri}}\right]$.

Excitatory synapses: strength of synapses to inhibitory neurons: $w_{\mathrm{E}\to\mathrm{I}} = 4.96\,\mathrm{mV}$, evoking a peak EPSP of 0.66 mV in a resting postsynaptic neuron.

Inhibitory synapses: time constant: $\tau_I = 5\,\mathrm{ms}$; strength of synapses to excitatory neurons: $w_{\mathrm{I}\to\mathrm{E}} = -5.06\,\mathrm{mV}$, evoking a peak IPSP of $-1.27\,\mathrm{mV}$ in a resting postsynaptic neuron; strength of synapses to inhibitory neurons: $w_{\mathrm{I}\to\mathrm{I}} = -5.33\,\mathrm{mV}$ evoking a peak IPSP of $-1.33\,\mathrm{mV}$ in a resting postsynaptic neuron.

STDP window: window amplitude for all connections: $\eta = 1.25\,\mathrm{mV}$.

Spontaneous synaptic turnover: lifetime and absence time of synapses between interior neurons: $L_{\mathrm{int}} = 2000\,\mathrm{s}$ and $A_{\mathrm{int}} = 1333.3\,\mathrm{s}$; lifetime and absence time of synapses between interior and periphery neurons: $L_{\mathrm{peri}} = 2000\,\mathrm{s}$ and $A_{\mathrm{peri}} = 500\,\mathrm{s}$. No synaptic turnover in Fig. 5.7.

Simulation: total simulated time: 100 hours.

Linear Poisson model

Neuron numbers: $N_E = 105$; interior neurons: $N_{\mathrm{int}} = 105$; periphery neurons: 0.

Network structure: connection probability between neurons: $p_{\mathrm{int}} = 0.65$; there are no self-connections of neurons.

Neuron parameters: time constant: $\tau = 10\,\mathrm{ms}$; spontaneous rate: $f_0 = 0.75\,\mathrm{Hz}$; sum of input and sum of output weights of a neuron: $\tau w_{\mathrm{sum}} = 0.25$, maximal synaptic weight: $\tau w_{\mathrm{max}} = 0.0124$.

Plasticity rule: symmetric weight change upon a spike of neuron j: $\Delta w(\Delta f_i) = 0.01/w_{\mathrm{max}}(\Delta f_i - 0.87\,\mathrm{Hz})\Delta f_i$, where $\Delta f_i = f_i(t) - f_0$ is the current level of excitation of the post- or presynaptic partner neuron.

Spontaneous synaptic turnover: lifetime and absence time of synapses: $L_{\mathrm{int}} = 6\,\mathrm{h}$ and $A_{\mathrm{int}} = 3.23\,\mathrm{h}$.

Memory representation: number of assemblies: $n_{\mathrm{asbly}} = 3$; initial number of interior neurons per assembly: $N_{\mathrm{asbly}}(0) = 35$.

Simulation: event-based; simulated time: 125 days.

C.2 Random walk models based on statistics of weight changes

In this section, we construct and analyze Markovian random walk models for the dynamics of w_1, the summed input weight to a neuron from assembly 1, based on the average $\overline{\Delta w_1}$ and standard deviation $\mathrm{Std}(\Delta w_1)$ of the ensuing weight changes Δw_1 during short time intervals (Sec. 5.3.3). We also

approximate the random walk by a diffusion process [6]. The obtained models allow us to numerically and analytically compute stationary probability densities and thereby detect metastable states. By removing the average change or homogenizing the fluctuations, we can selectively study the impact of the inhomogeneous noise or the deterministic part of the weight changes on the switching dynamics.

We use the observed mean $\overline{\Delta w_1}(w_1)$ and standard deviation $\mathrm{Std}(\Delta w_1)(w_1)$ of the weight changes to construct a first random walk model. We make the Markov assumption that the change in w_1 depends only on its previous value, which is reasonable for sufficiently long time steps. For simplicity, we further assume that the noise is normally distributed. We therefore consider the discrete-time stochastic dynamics

$$w_1(t+1) = w_1(t) + \overline{\Delta w_1}(w_1(t)) + \mathrm{Std}(\Delta w_1)(w_1(t))\,\xi(t)\,, \tag{C.1}$$

where $\xi(t) \sim \mathcal{N}(0,1)$ independently in each step, and w_1 is clipped to the interval $[0,1]$ after each step. The resulting stationary probability density of w_1 of the Markovian random walk model (C.1) is in acceptable agreement with the density obtained in simulations of the LIF network model [Fig. C.1(a) right]. A likely source of deviations is that the weight changes in the network are not Gaussian distributed.

To study the effect of the inhomogeneous noise, we consider a modified random walk model without the deterministic part, i.e. $w_1(t+1) = w_1(t) + \mathrm{Std}(\Delta w_1)(w_1(t))\,\xi(t)$. We find that this dynamics still gives rise to a bimodal density with peaks corresponding to assembly membership [Fig. C.1(b) right]. This is similar to noise-induced multistability, which has been described in different models and systems, e.g. for electrical and chemical oscillations, populations dynamics and foraging behavior [103, 105, 110, 357]. To also examine the impact of the deterministic part of the weight change, we consider a model with homogeneous noise:

$$w_1(t+1) = w_1(t) + \overline{\Delta w_1}(w_1(t)) + \langle \mathrm{Std}(\Delta w_1)\rangle\,\xi(t)\,,$$

where $\langle \mathrm{Std}(\Delta w_1)\rangle = \int_0^1 \mathrm{Std}(\Delta w_1)(w_1)\,p_{\mathrm{stat}}(w_1)\,dw_1$ and $p_{\mathrm{stat}}(w_1)$ is the stationary probability density of w_1 in simulations of the LIF network model. In simulations of this random walk model, the neuron's summed input weight w_1 stays close to $w_1 \approx 0$ or $w_1 \approx 1$ during the whole simulation time depending on the initial condition [Fig. C.1(c) right]. Therefore, with this noise, the neuron does not leave the potential wells shown in Fig. 5.5(c). Taken together, we conclude that both the inhomogeneous noise and the deterministic weight changes contribute to the observed switching dynamics. In particular, the stronger noise in the transition zone appears to be crucial for neuron transitions and contributes to the metastable states of long-term assembly membership.

We can analytically obtain the stationary probability densities for the Markovian random walk models using a diffusion approximation. For this we interpret the stochastic dynamics (C.1) as the Euler–Maruyama discretization of an Itô stochastic differential equation [6]. With a time step equal to 1, we can directly read off the coefficients specifying the drift and the noise strength in the stochastic differential equation: the mean $\overline{\Delta w_1}(w_1)$ and the standard deviation $\mathrm{Std}(\Delta w_1)(w_1)$ of the weight updates, which both depend on the weight w_1. The stationary solution of the Fokker-Planck equation associated to the stochastic differential equation is given by

$$p_{\mathrm{FP}}(w_1) \propto \frac{1}{\mathrm{Std}(\Delta w_1)^2(w_1)} \exp\left[2\int_0^{w_1} \frac{\overline{\Delta w_1}(u)}{\mathrm{Std}(\Delta w_1)^2(u)}\,du\right] \tag{C.2}$$

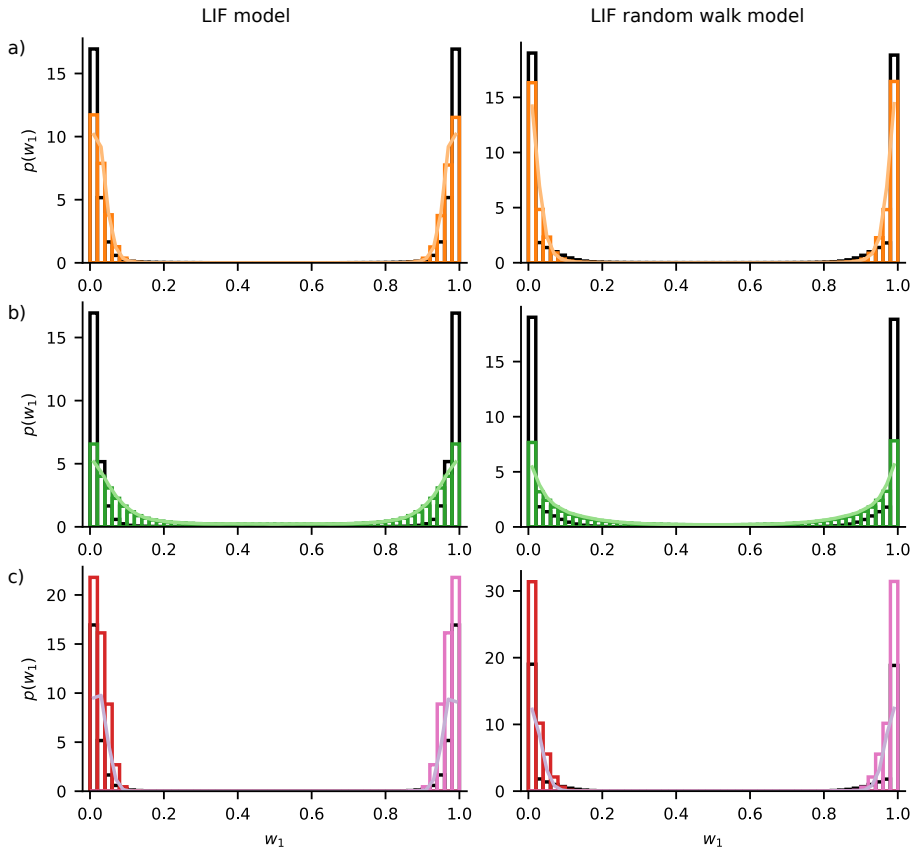

Figure C.1: Stationary distributions of w_1 for the LIF network model with two assemblies in Fig. 5.5 (left, black histograms) and for the corresponding effective random walk model (right, black histograms) compared to stationary distributions of related Markovian random walk models (colored histograms) and analytical diffusion approximations (colored curves). (a) Random walk model (C.1) based on mean and standard deviation of weight changes (orange). (b) Random walk model without deterministic weight changes, accounting only for inhomogeneous noise (green). (c) Random walk model with deterministic part but homogeneous noise. In this case, the distribution obtained in simulations of the random walk model depends on the initial condition, $w_1(0) = 0$ (red) or $w_1(0) = 1$ (pink), since switching does not occur within the simulated time. Each of the numerically estimated distributions is normalized to have unit integral. The analytical probability density (light purple) accounts for both distributions.

for $w_1 \in [0, 1]$; the normalization $\int_0^1 dw_1\, p_{\mathrm{FP}}(w_1) = 1$ yields the stationary probability density. Here, we use reflecting boundary conditions to keep the probability density in the interval $[0, 1]$. For the modified random walk models, the expression holds with drift or noise strength replaced accordingly. The resulting analytical probability densities agree with the ones of the random walk models except at the boundaries (Fig. C.1). The deviations at the boundaries originate from the clipping of w_1 in the random walk model; if we replace the clipping by reflecting the increments at the boundaries, the densities also agree there.

C.3 Random walk model from first principles for neuron switching in LIF networks

We derive an effective random walk model from first principles for the neuron switching dynamics in LIF networks where noisy autonomous activity drives the assembly drift (Figs. 5.3 and 5.4). In the random walk model, a single "test neuron" spikes at a fixed background rate and with an input weight-dependent probability when its current or another assembly reactivates. The different assemblies reactivate randomly at a fixed rate and their neurons generate asynchronous background spiking as well. The resulting STDP and homeostatic normalization lead to a random walk of the neuron's summed input weights from the different assemblies, which determine its assembly membership and its probability to be activated by a reactivating assembly. Model parameters like the background spike rate and the reactivation rate of the assemblies will be chosen to agree with the LIF network simulations for the transition mechanism analysis in Fig. 5.5 to render the random walk model comparable to the LIF model.

We focus on the test neuron's summed input weights $w_a \geq 0$ from the different assemblies $a = 1, \ldots, n_{\mathrm{asbly}}$, where n_{asbly} is the number of assemblies in the network. The small effect of the neuron on the assemblies and their activities is neglected. Homeostatic plasticity normalizes the sum of the w_a to w_{sum}, which we set to one, thus $\sum_a w_a = 1$. We further use that $w_{\mathrm{sum}}/N_a \leq w_{\max}$ for the sizes N_a of the assemblies. Thus, on the level of the summed input weights from the assemblies, the constraint set by homeostatic normalization is stronger than the upper bound set by the individual maximal synaptic weight. For simplicity, we use a fixed and identical size N_{asbly} for all assemblies. From the perspective of the neuron, the assemblies are regarded as static except for the plasticity of the summed input weights w_a from them. In our LIF networks, weight changes are driven by STDP combined with homeostatic normalization. Pairs of spikes of the test neuron and an assembly neuron contribute to changes of the summed synaptic input weight w_a from the assembly to the test neuron according to the STDP rule (Sec. 5.2.1). Spontaneous assembly reactivations occur at rates r_a for the different assemblies. Again for simplicity, we use an identical reactivation rate r_{asbly} for all assemblies and measure it in our simulations. The activity of the test neuron consists of irregular background spiking, which is modeled as an independent Poisson process with rate r_{bg}. Furthermore, the neuron can be coactivated by an assembly reactivation with a probability $p_{\mathrm{coact}}(w_a)$ depending on the momentary input weight w_a from the assembly. In addition to the spontaneous assembly reactivations, the assembly neurons exhibit asynchronous irregular background spiking as well. As for the test neuron, we model this background spiking activity by independent Poisson processes with identical rates r_{bg}.

To specify our random walk model, we estimate the weight changes induced by STDP around a single assembly reactivation. Motivated by our network simulations [Fig. 5.9(c) left], we assume that each assembly neuron spikes exactly once during a single reactivation and that these spikes occur independently of each other with a given temporal distribution $\rho_{\mathrm{react}}(t)$. The assembly reactivations

in our LIF networks are synchronous with a temporal duration of at most 15 ms, and, hence, one may consider approximating the temporal distribution by a Dirac delta function. We nevertheless choose to include the finite temporal duration here as it has a considerable effect on the estimated weight changes and on the coactivation probability $p_{\text{coact}}(w_a)$. For the explicit formulas below, we coarsely approximate this temporal distribution $\rho_{\text{react}}(t)$ by a uniform density of duration τ_{react}. We write the STDP window as $\Delta w_{ij}(t_i - t_j) = \eta\, h(t_i - t_j)$, where the amplitude η has the units of the weights and, thus, is given as a fraction of w_{sum}. The dimensionless function $h(\Delta t)$ describes the shape of the STDP window, see Fig. 5.2(a). By separating the spiking activity into reactivation and background spikes of the assembly neurons and of the test neuron, we obtain the following four contributions.

First, the reactivation of, say, assembly a can coincide to some degree with the random background spiking of the test neuron. The corresponding STDP-induced weight change due to a single reactivation event takes the form

$$\Delta w_a^{\text{bg,react}} = \eta \sum_{t^a_{\text{react}}} \sum_{t_{\text{bg}}} h\left(t_{\text{bg}} - t^a_{\text{react}}\right) , \tag{C.3}$$

where t_{bg} indicate the test neuron's background spike times, which follow an independent Poisson process with rate r_{bg}, while t^a_{react} denote the (by assumption exactly N_a) spike times of the neurons in the assembly. This contribution to the weight change is a random quantity because the spike times are distributed randomly. Its mean and standard deviation under the given assumptions can be computed by taking averages over the different sets of spike times.

We start with the mean:

$$\begin{aligned}
\overline{\Delta w_a^{\text{bg,react}}} &= \eta \left\langle \sum_{t^a_{\text{react}}} \sum_{t_{\text{bg}}} h\left(t_{\text{bg}} - t^a_{\text{react}}\right) \right\rangle_{t_{\text{bg}}, t^a_{\text{react}}} \\
&= \eta\, N_a \int_{-\infty}^{+\infty} \left(r_{\text{bg}} \int_{-\infty}^{+\infty} h(s-t)\, ds \right) \rho_{\text{react}}(t)\, dt \\
&= \eta\, N_a\, r_{\text{bg}} \int_{-\infty}^{+\infty} h(\Delta t)\, d\Delta t\,,
\end{aligned} \tag{C.4}$$

where we use that the average of the summed values of a function $f(t)$ at the random points t_{Poi} of a homogeneous Poisson point process with rate r_{Poi} is given by

$$\left\langle \sum_{t_{\text{Poi}}} f\left(t_{\text{Poi}}\right) \right\rangle_{t_{\text{Poi}}} = r_{\text{Poi}} \int_{-\infty}^{\infty} f(s)\, ds \tag{C.5}$$

[Campbell's theorem for the mean, e.g. [358]]. We note that the mean, Eq. (C.4), does not depend on the temporal distribution $\rho_{\text{react}}(t)$ of the assembly reactivation spikes.

To obtain the variance of the weight change contribution, we use the law of total variance,

$$\text{Var}\left(\Delta w_a^{\text{bg,react}}\right) = \left\langle \text{Var}\left(\Delta w_a^{\text{bg,react}} \middle| t^a_{\text{react}}\right) \right\rangle_{t^a_{\text{react}}} + \text{Var}\left(\text{E}\left(\Delta w_a^{\text{bg,react}} \middle| t^a_{\text{react}}\right) \right)_{t^a_{\text{react}}} , \tag{C.6}$$

where $\text{E}\left(\Delta w_a^{\text{bg,react}} \middle| t^a_{\text{react}}\right)$ is the conditional mean (or expectation value). We first compute the conditional

variance given the reactivation spike times:

$$\begin{aligned}\mathrm{Var}\left(\Delta w_a^{\mathrm{bg,react}}\middle| t^a_{\mathrm{react}}\right) &= \eta^2\,\mathrm{Var}\left(\sum_{t_{\mathrm{bg}}}\sum_{t^a_{\mathrm{react}}} h\left(t_{\mathrm{bg}}-t^a_{\mathrm{react}}\right)\middle| t^a_{\mathrm{react}}\right)\\ &= \eta^2 r_{\mathrm{bg}}\int_{-\infty}^{+\infty}\left(\sum_{t^a_{\mathrm{react}}} h\left(s-t^a_{\mathrm{react}}\right)\right)^2 ds\,, \end{aligned} \tag{C.7}$$

where we use that for a homogeneous Poisson process with the notation of Eq. (C.5) the variance of the random sum reads

$$\mathrm{Var}\left(\sum_{t_{\mathrm{Poi}}} f\left(t_{\mathrm{Poi}}\right)\right) = r_{\mathrm{Poi}}\int_{-\infty}^{\infty}\left(f(s)\right)^2\,ds \tag{C.8}$$

[Campbell's theorem for the variance, e.g. [358]]. The conditional mean given the reactivation spike times equals the unconditional mean,

$$\begin{aligned}\mathrm{E}\left(\Delta w_a^{\mathrm{bg,react}}\middle| t^a_{\mathrm{react}}\right) &= \eta\sum_{t^a_{\mathrm{react}}}\left\langle\sum_{t_{\mathrm{bg}}} h\left(t_{\mathrm{bg}}-t^a_{\mathrm{react}}\right)\right\rangle_{t_{\mathrm{bg}}}\\ &= \eta\, r_{\mathrm{bg}}\sum_{t^a_{\mathrm{react}}}\int_{-\infty}^{+\infty} h\left(s-t^a_{\mathrm{react}}\right)\,ds\\ &= \eta\, N_a\, r_{\mathrm{bg}}\int_{-\infty}^{+\infty} h(\Delta t)\,d\Delta t\,, \end{aligned} \tag{C.9}$$

where we used the substitution $\Delta t = s - t^a_{\mathrm{react}}$ in the integral of the second line. The conditional mean's independence of t_{react} implies that its variance $\mathrm{Var}\left(\mathrm{E}\left(\Delta w_a^{\mathrm{bg,react}}\middle| t^a_{\mathrm{react}}\right)\right)_{t^a_{\mathrm{react}}}$ vanishes. Equation (C.6) then yields the unconditional variance

$$\begin{aligned}\mathrm{Var}\left(\Delta w_a^{\mathrm{bg,react}}\right) &= \left\langle\mathrm{Var}\left(\Delta w_a^{\mathrm{bg,react}}\middle| t^a_{\mathrm{react}}\right)\right\rangle_{t^a_{\mathrm{react}}}\\ &= \eta^2 r_{\mathrm{bg}}\int_{-\infty}^{+\infty}\left\langle\left(\sum_{t^a_{\mathrm{react}}} h\left(s-t^a_{\mathrm{react}}\right)\right)^2\right\rangle_{t^a_{\mathrm{react}}} ds\\ &= \eta^2 r_{\mathrm{bg}}\left[N_a(N_a-1)\int_{-\infty}^{+\infty}\tilde{h}^2(s)\,ds + N_a\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h^2(s-t)\,\rho_{\mathrm{react}}(t)\,dt\,ds\right]\\ &= \eta^2 r_{\mathrm{bg}}\left[N_a^2\int_{-\infty}^{+\infty}\tilde{h}^2(s)\,ds + N_a\left(\int_{-\infty}^{+\infty} h^2(s)\,ds - \int_{-\infty}^{+\infty}\tilde{h}^2(s)\,ds\right)\right], \end{aligned} \tag{C.10}$$

where $\tilde{h}(s) = \int_{-\infty}^{+\infty} h(s-t)\rho_{\mathrm{react}}(t)\,dt$ is the convolution of the STDP window with the reactivation spike density. To compute the average over the reactivation spike times t^a_{react} in the second line of Eq. (C.10),

we expanded the square and used our assumption that these times are independently and identically distributed with the density $\rho_{\text{react}}(t)$. For the parameter regime of our LIF networks, we can neglect the term proportional to N_a in the last line of Eq. (C.10) and obtain a compact expression for the standard deviation

$$\text{Std}\left(\Delta w_a^{\text{bg,react}}\right) \approx \eta\, N_a \left(r_{\text{bg}} \int_{-\infty}^{+\infty} \tilde{h}^2(s)\, ds\right)^{1/2} . \tag{C.11}$$

For the STDP window function employed in the LIF network models and a uniform reactivation spike density, the integrals in Eqs. (C.4) and (C.11) can be computed analytically.

The second contribution to the STDP-induced weight change arises because an assembly can coactivate the test neuron during a reactivation event. This happens in particular if the test neuron is part of the assembly. We describe our model for the coactivation probability $p_{\text{coact}}(w_a)$ below. We assume that the coactivated neuron independently emits one spike with the same temporal distribution $\rho_{\text{react}}(t)$ as the (other) assembly reactivation spikes such that the resulting weight change can be written as

$$\Delta w_a^{\text{coact}} = \eta \sum_{t_{\text{react}}^a} h\left(t_{\text{coact}} - t_{\text{react}}^a\right) . \tag{C.12}$$

Here, t_{coact} is the coactivation spike time while t_{react}^a are the N_a assembly reactivation spike times as before. Under our assumptions, the differences $t_{\text{coact}} - t_{\text{react}}^a$ are identically distributed with probability density $\hat{\rho}(s) = \int_{-\infty}^{+\infty} \rho_{\text{react}}(s+u)\rho_{\text{react}}(u)\, du$, which results in a symmetric triangular (or "hat") density $\max(\tau_{\text{react}} - |s|, 0)/\tau_{\text{react}}^2$ with one-sided temporal width τ_{react} for the uniform approximation of the spike density. Therefore, the mean of this weight change contribution is given by

$$\begin{aligned}
\overline{\Delta w_a^{\text{coact}}} &= \eta \left\langle \sum_{t_{\text{react}}^a} h\left(t_{\text{coact}} - t_{\text{react}}^a\right) \right\rangle_{t_{\text{coact}}, t_{\text{react}}^a} \\
&= \eta\, N_a \int_{-\infty}^{+\infty} h(s)\hat{\rho}(s)\, ds \\
&\approx \eta\, N_a \left(1 - \frac{1}{3}\frac{\tau_{\text{react}}}{\tau_{\text{pot}}}\right) ,
\end{aligned} \tag{C.13}$$

where the third line is a useful approximation appropriate for $\tau_{\text{react}} \leq \tau_{\text{pot}}$; it follows from approximating the central, positive part of the STDP window by another symmetric triangular function $\eta\left(1 - |\Delta t|/\tau_{\text{pot}}\right)$ with peak value η and zero-crossings at $\pm\tau_{\text{pot}}$. The condition $\tau_{\text{react}} \leq \tau_{\text{pot}}$ also implies that this weight change consists only of potentiation (LTP). We present the approximation here because it compactly exposes the effect of the temporal distribution of the reactivation spike times on the STDP-induced weight change originating from the test neuron's coactivated spiking with the assembly. For the variance $\text{Var}\left(\Delta w_a^{\text{coact}}\right)$ of this weight change contribution, we note that the differences $t_{\text{coact}} - t_{\text{react}}^a$ are not independently distributed since they all contain the same random time t_{coact}. We thus proceed by

computing the second moment as follows:

$$\begin{aligned}
\mathrm{Var}\left(\Delta w_a^{\mathrm{coact}}\right) &= \eta^2 \left\langle \left(\sum_{t_{\mathrm{react}}^a} h\left(t_{\mathrm{coact}} - t_{\mathrm{react}}^a\right) \right)^2 \right\rangle_{t_{\mathrm{coact}}, t_{\mathrm{react}}^a} - \left(\overline{\Delta w_a^{\mathrm{coact}}}\right)^2 \\
&= \eta^2 \int_{-\infty}^{+\infty} \left\langle \left(\sum_{t_{\mathrm{react}}^a} h\left(s - t_{\mathrm{react}}^a\right) \right)^2 \right\rangle_{t_{\mathrm{react}}^a} \rho_{\mathrm{react}}(s)\, ds - \left(\overline{\Delta w_a^{\mathrm{coact}}}\right)^2 \\
&= \eta^2 \left[N_a(N_a-1) \int_{-\infty}^{+\infty} \tilde{h}^2(s)\, \rho_{\mathrm{react}}(s)\, ds \right. \\
&\qquad \left. + N_a \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h^2\left(s-t\right)\, \rho_{\mathrm{react}}(t)\, \rho_{\mathrm{react}}(s)\, dt\, ds \right] - \left(\overline{\Delta w_a^{\mathrm{coact}}}\right)^2 \\
&= \eta^2 \left[N_a(N_a-1) \int_{-\infty}^{+\infty} \tilde{h}^2(s)\, \rho_{\mathrm{react}}(s)\, ds + N_a \int_{-\infty}^{+\infty} h^2\left(s\right)\, \hat{\rho}(s)\, ds \right. \\
&\qquad \left. - N_a^2 \left(\int_{-\infty}^{+\infty} h(s) \hat{\rho}(s)\, ds \right)^2 \right],
\end{aligned} \tag{C.14}$$

where $\tilde{h}(s) = \int_{-\infty}^{+\infty} h(s-t)\rho_{\mathrm{react}}(t)\, dt$ is the convolution of the STDP window with the reactivation spike density as in Eq. (C.10). For $\tau_{\mathrm{react}} \leq \tau_{\mathrm{pot}}$ we can obtain a compact formula for the variance using the same approximations as for the mean weight change in Eq. (C.13). To evaluate the first integral in the last line of Eq. (C.14) we only need to know $\tilde{h}(s)$ for values of s where $\rho_{\mathrm{react}}(s)$ is nonzero. For these values of s, we approximate $\tilde{h}(s)$ by convolving the uniform spike density of duration τ_{react} (for concreteness centered at zero) with the central, positive part of the STDP window shape, which we again replace by the symmetric triangular function $\left(1 - |s|/\tau_{\mathrm{pot}}\right)$,

$$\tilde{h}(s) \approx 1 - \frac{1}{4}\frac{\tau_{\mathrm{react}}}{\tau_{\mathrm{pot}}} - \frac{s^2}{\tau_{\mathrm{react}}\tau_{\mathrm{pot}}}, \ |s| \leq \frac{\tau_{\mathrm{react}}}{2}. \tag{C.15}$$

For the second integral in the last line of Eq. (C.14), we integrate $\left(1 - |s|/\tau_{\mathrm{pot}}\right)^2$ with respect to the symmetric triangular density $\max(\tau_{\mathrm{react}} - |s|, 0)/\tau_{\mathrm{react}}^2$. Together with the approximation of the mean in Eq. (C.13), we therefore obtain

$$\begin{aligned}
\mathrm{Var}\left(\Delta w_a^{\mathrm{coact}}\right) &\approx \eta^2 \left[N_a(N_a-1) \left(1 - \frac{2}{3}\frac{\tau_{\mathrm{react}}}{\tau_{\mathrm{pot}}} + \frac{7}{60}\left(\frac{\tau_{\mathrm{react}}}{\tau_{\mathrm{pot}}}\right)^2 \right) + N_a \left(1 - \frac{2}{3}\frac{\tau_{\mathrm{react}}}{\tau_{\mathrm{pot}}} + \frac{1}{6}\left(\frac{\tau_{\mathrm{react}}}{\tau_{\mathrm{pot}}}\right)^2 \right) \right. \\
&\qquad \left. - N_a^2 \left(1 - \frac{2}{3}\frac{\tau_{\mathrm{react}}}{\tau_{\mathrm{pot}}} + \frac{1}{9}\left(\frac{\tau_{\mathrm{react}}}{\tau_{\mathrm{pot}}}\right)^2 \right) \right] \\
&= \eta^2 \left(\frac{N_a^2}{180} + \frac{N_a}{20} \right) \left(\frac{\tau_{\mathrm{react}}}{\tau_{\mathrm{pot}}} \right)^2,
\end{aligned} \tag{C.16}$$

which gives the standard deviation of the weight change

$$\mathrm{Std}\left(\Delta w_a^{\mathrm{coact}}\right) \approx \eta\, N_a \left(\frac{1}{180} + \frac{1}{20N_a}\right)^{1/2} \frac{\tau_{\mathrm{react}}}{\tau_{\mathrm{pot}}}\,. \tag{C.17}$$

We note that the variability of this weight change contribution solely originates from the variability of the spike times within an assembly reactivation and, thus, vanishes if its temporal duration approaches zero, i.e. for $\tau_{\mathrm{react}} \to 0$.

The third and fourth contributions to the STDP-induced weight change arise from the asynchronous irregular background spiking of the assembly neurons, which we model as independent Poisson processes with identical rates r_{bg}. These weight change contributions affect the summed input weights w_b from all assemblies $b = 1, \ldots, n_{\mathrm{asbly}}$ in the network, while the first and the second contributions only affect the summed input weight w_a from the assembly a that has currently reactivated. The third contribution is only present when the test neuron is coactived by an assembly reactivation. In this case, it emits a spike, which can coincide to some degree with the background spiking of the assembly neurons. The corresponding STDP-induced changes of the summed input weights w_b from the different assemblies $b = 1, \ldots, n_{\mathrm{asbly}}$ read

$$\Delta w_b^{\mathrm{coact,bg}} = \eta \sum_{t_{\mathrm{bg}}^b} h\left(t_{\mathrm{coact}} - t_{\mathrm{bg}}^b\right)\,, \tag{C.18}$$

where t_{bg}^b indicate the background spike times of all neurons in assembly b while t_{coact} is the coactivation spike time as before. The superposition of independent Poisson processes forms another Poisson process whose rate equals the sum of the individual rates. Therefore, the background spike times t_{bg}^b of all neurons in assembly b follow a Poisson process with rate $N_b\, r_{\mathrm{bg}}$, and we can compute the mean and standard deviation of the third weight change contribution in a similar manner to the first contribution above. Campbell's theorem for the mean, Eq. (C.5), yields

$$\begin{aligned}
\overline{\Delta w_b^{\mathrm{coact,bg}}} &= \eta \left\langle \sum_{t_{\mathrm{bg}}^b} h\left(t_{\mathrm{coact}} - t_{\mathrm{bg}}^b\right) \right\rangle_{t_{\mathrm{coact}}, t_{\mathrm{bg}}^b} \\
&= \eta \int_{-\infty}^{+\infty} \left(N_b\, r_{\mathrm{bg}} \int_{-\infty}^{+\infty} h(t-s)\, ds \right) \rho_{\mathrm{react}}(t)\, dt \\
&= \eta\, N_b\, r_{\mathrm{bg}} \int_{-\infty}^{+\infty} h(\Delta t)\, d\Delta t\,.
\end{aligned} \tag{C.19}$$

To obtain the variance, we employ the law of total variance together with Campbell's theorem for the

variance, Eq. (C.8), and use that the variance of the conditional mean vanishes as in Eq. (C.10):

$$\begin{aligned}\mathrm{Var}\left(\Delta w_b^{\mathrm{coact,bg}}\right) &= \left\langle \mathrm{Var}\left(\Delta w_b^{\mathrm{coact,bg}}\middle| t_{\mathrm{coact}}\right)\right\rangle_{t_{\mathrm{coact}}} + \mathrm{Var}\left(\mathrm{E}\left(\Delta w_b^{\mathrm{coact,bg}}\middle| t_{\mathrm{coact}}\right)\right)_{t_{\mathrm{coact}}} \\ &= \eta^2 N_b\, r_{\mathrm{bg}} \left\langle \int_{-\infty}^{+\infty} h^2(t_{\mathrm{coact}} - s)\, ds \right\rangle_{t_{\mathrm{coact}}} \\ &= \eta^2 N_b\, r_{\mathrm{bg}} \int_{-\infty}^{+\infty} h^2(\Delta t)\, d\Delta t\,. \end{aligned} \tag{C.20}$$

Hence, in contrast to the standard deviation of the first contribution, Eq. (C.11), the resulting standard deviation

$$\mathrm{Std}\left(\Delta w_b^{\mathrm{coact,bg}}\right) = \eta \left(N_b\, r_{\mathrm{bg}} \int_{-\infty}^{+\infty} h^2(\Delta t)\, d\Delta t\right)^{1/2} \tag{C.21}$$

is proportional to the square root of the number of neurons in the assembly, making it significantly smaller than the former.

Finally, also the background spiking of the test neuron can coincide with the background spiking of the assembly neurons, which leads to STDP-induced weight changes that occur independently of the assembly reactivations. Here, we only include the mean of this fourth weight change contribution as it is non-zero for the employed STDP window. We expect that the fluctuations are proportional to the square root of the number of neurons in the assembly as in Eq. (C.21) and, therefore, small compared to the fluctuations of the first weight change contribution, Eq. (C.11). It is, however, difficult to estimate the standard deviation of this weight change contribution because of the ongoing homeostatic weight normalization. We therefore consider the mean weight change over a not too long time interval without ongoing weight normalization and apply it only it at the end of the interval. As before, we model the background spiking of the test neuron and of the assembly neurons as independent Poisson processes with rates r_{bg} for the test neuron and $N_b\, r_{\mathrm{bg}}$ for assembly b. The average of the resulting (additive) STDP-induced changes of the summed input weights w_b during a time interval of length T is then proportional to the product of the rates [13]:

$$\overline{\Delta w_b^{\mathrm{bg,bg}}}(T) = \eta\, N_b\, r_{\mathrm{bg}}^2\, T \int_{-\infty}^{+\infty} h(\Delta t)\, d\Delta t\,, \tag{C.22}$$

where T should be at least several times longer than the temporal extent of the STDP window $h(\Delta t)$. On the other hand, to approximately implement the ongoing weight normalization, the length T should not be too long. In our random walk model, we use the mean weight change Eq. (C.22) for the time intervals between successive assembly reactivations. Their average length is $1/\left(n_{\mathrm{asbly}} r_{\mathrm{asbly}}\right)$.

The test neuron's probability $p_{\mathrm{coact}}(w_a)$ of being coactivated by the synchronous assembly reactivation depends on the momentary input weight w_a from the assembly. Because the second weight change contribution from above is only present when the test neuron is coactive with the assembly and also only affects, specifically strengthens, the summed input weight from this assembly, the coactivation probability implicitly gives rise to a weight-dependence of the STDP-induced weight changes. It is thus an important part of the random walk model. The probability that a noisy LIF neuron emits a spike in response to synchronous spiking of excitatory inputs has been previously investigated, see e.g. [359, 360]. As the assembly reactivations in our LIF networks consist of synchronous spiking with approximately

one spike per assembly neuron, the response probability is well described by the probability of finding the neuron's fluctuating membrane potential V in the interval $[V_\theta - \gamma\, w_a, V_\theta]$, where V_θ is the spike threshold and $\gamma\, w_a$ is the expected peak value of the compound post-synaptic potential generated by the synchronous inputs. The factor γ thus translates the dimensionless input weight w_a we use in our random walk model to synaptic input weights in terms of their peak EPSPs, as we display them in Fig. 5.3, see also Sec. 5.2.1. It also includes an attenuation of the peak value due to the finite temporal duration of the reactivation spike density $\rho_{\text{react}}(t)$; for the uniform density of duration τ_{react}, we have $\gamma = \gamma(\tau_{\text{react}})$. To obtain an explicit formula for the coactivation probability, we approximate the membrane potential distribution $p_V(v)$ of the neuron before a reactivation event by a Gaussian density with mean μ_V and standard deviation σ_V [359, 360] leading to

$$\begin{aligned} p_{\text{coact}}(w_a) &= \int_{V_\theta - \gamma\, w_a}^{V_\theta} p_V(v)\, dv \\ &= \frac{1}{2}\left[\text{erf}\left(\frac{\gamma\, w_a}{\sqrt{2}\sigma_V} - \frac{V_\theta - \mu_V}{\sqrt{2}\sigma_V}\right) + \text{erf}\left(\frac{V_\theta - \mu_V}{\sqrt{2}\sigma_V}\right)\right], \end{aligned} \tag{C.23}$$

which has the natural property $p_{\text{coact}}(0) = 0$ and increases in a sigmoidal manner to values close to 1 for $\gamma w_a - (V_\theta - \mu_V) \gg \sqrt{2}\sigma_V$. For simplicity, we use $\mu_V = V_{\text{rest}}$ and $\sigma_V = \sigma$, the mean and the standard deviation of the stationary membrane potential distribution of our LIF model neurons in absence of a threshold and synaptic inputs from other modeled neurons (Sec. 5.2.1). We expect that the asynchronous background spiking at low rates of the other neurons does not significantly affects the membrane potential distribution. Alternatively, instead of the Gaussian density in Eq. (C.23) one could use the known stationary density of the LIF neuron with white noise input, see e.g. [27]. Here, we opt for the more explicit approximation.

The obtained STDP-induced weight change contributions around a single assembly reactivation together with the test neuron's coactivation probability enable us to specify our random walk model describing the dynamics of the test neuron's summed input weights w_a from the assemblies. To model the dynamics in continuous time, we assume that the times of the spontaneous assembly reactivations follow independent Poisson processes with identical reactivation rates $r_a = r_{\text{asbly}}$. In the random walk model we move along the reactivation events. At the reactivation of an assembly, we apply the different weight change contributions as follows.

We start with the fourth contribution resulting from the background spiking of both the test neuron and the assembly neurons. All the weights change according to Eq. (C.22) with T given by the time interval since the previous assembly reactivation, which on average equals $1/\left(n_{\text{asbly}} r_{\text{asbly}}\right)$. After the application of the STDP-induced additive changes to the weights, we clip them to $w_a \geq 0$ and employ homeostatic divisive normalization by updating the weights to

$$w_a = \frac{w'_a}{\sum_{b=1}^{n_{\text{asbly}}} w'_b}, \tag{C.24}$$

where w'_a are the changed weights after clipping. We use this updated weight w_a, i.e. including clipping and normalization after the update, in the coactivation probability $p_{\text{coact}}(w_a)$ to randomly decide if the test neuron is coactived by the assembly reactivation. Next, we apply the first weight change contribution, which originates from the coincidence of the assembly reactivation with the background spiking of the test neuron. To include the important fluctuations of this contribution, we randomly draw

the (additive) weight change from a normal distribution with mean Eq. (C.4) and standard deviation Eq. (C.11). Afterwards, we again clip the changed weights to $w_a \geq 0$ and then normalize them according to Eq. (C.24). In case the test neuron is coactive with the assembly, the neuron's coactivated spike leads to further STDP-induced weight changes, which are characterized above as the second and third contributions. We also randomly draw them from normal distributions with corresponding means and standard deviations, Eqs. (C.13) and (C.17) for the second and Eqs. (C.19) and (C.21) for the third contribution. Of particular importance is the second contribution because it describes how coactivated synchronous spiking of the neuron reinforces the weight from the reactivating assembly. As before, the changed weights are clipped and normalized, where we employ Eq. (C.4) only after the application of the second and third contribution together. Following all the weight updates, we move forward to the next time of an assembly reactivation, which in our model occurs after an exponentially distributed waiting time with mean $1/\left(n_{\text{asbly}} r_{\text{asbly}}\right)$. These steps completely define our effective random walk model of a neuron's input weights from the assemblies in our LIF networks. We note that the random walk model does not incorporate the indirect weight decrease via homeostatic normalization when an assembly reactivates and the neuron does not spike, because this would require the modeling of weight changes within the assembly.

Figure 5.5(a), right, shows a neuron transition between two assemblies generated by the random walk model for the dynamics of w_1 whereas panels (b) and (d), right, displays the mean and standard deviations of the resulting weight changes of w_1 including normalization. These statistics describe the weight changes between two consecutive assembly reactivations. In contrast, the statistics in Fig. 5.5(b) and (d), left, of a LIF network (with two assemblies and without periphery neurons) are obtained by sampling the weight changes during average single-neuron interspike intervals. In our network simulations, the two used time intervals, however, approximately agree. This is also reflected in the relation $r_{\text{asbly}} + r_{\text{bg}} = n_{\text{asbly}} r_{\text{asbly}}$ of the chosen random walk model parameters (App. C.1), since the average single neuron rate is approximately $r_{\text{asbly}} + r_{\text{bg}}$ (neurons belonging to an assembly usually reliably spike with it). Therefore, the statistics in Fig. 5.5(b) and (d), left and right, agree well even quantitatively. Based on the statistics of weight changes between two consecutive assembly reactivations, we repeat the analysis using Markovian random walk models in App. C.2 for the random walk model derived in this section and find similar results as for the LIF network model (Fig. C.1 right).

Bibliography

[1] M. O'Shea, *The Brain: A Very Short Introduction*, Oxford University Press, 2005, ISBN: 9780192853929.

[2] J. C. Eccles, P. Fatt, and K. Koketsu,
Cholinergic and inhibitory synapses in a pathway from motor-axon collaterals to motoneurones,
J. Physiol. (Lond.) **126** (1954) 524.

[3] P. Strata and R. Harvey, *Dale's principle*, Brain Research Bulletin **50** (1999) 349.

[4] P. Dayan and L. F. Abbott,
Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems,
MIT Press, 2001, ISBN: 9780262041997.

[5] W. Gerstner, W. M. Kistler, R. Naud, and L. Paninski,
Neuronal Dynamics: From Single Neurons to Networks and Models of Cognition,
Cambridge University Press, 2014, ISBN: 9781107060838.

[6] C. W. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, 4th ed.,
Springer, 2009, ISBN: 9783540707134.

[7] W. R. Softky and C. Koch, *The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs*, J. Neurosci. **13** (1993) 334.

[8] D. Heeger, *Poisson model of spike generation*, Handout, 2000,
URL: `https://www.cns.nyu.edu/~david/handouts/poisson.pdf`.

[9] D. R. Cox and V. Isham, *Point Processes*, Chapman & Hall/CRC, 1980, ISBN: 9780412219108.

[10] A. G. Hawkes, *Spectra of some self-exciting and mutually exciting point processes*,
Biometrika **58** (1971) 83.

[11] A. G. Hawkes, *Point spectra of some mutually exciting point processes*,
Journal of the Royal Statistical Society: Series B (Methodological) **33** (1971) 438.

[12] D. H. Johnson, *Point process models of single-neuron discharges*,
J. Comput. Neurosci. **3** (1996) 275.

[13] R. Kempter, W. Gerstner, and J. L. van Hemmen, *Hebbian learning and spiking neurons*,
Phys. Rev. E **59** (1999) 4498.

[14] W. Gerstner and W. M. Kistler, *Spiking Neuron Models: Single Neurons, Populations, Plasticity*,
Cambridge University Press, 2002.

[15] N. Brunel, F. S. Chance, N. Fourcaud, and L. F. Abbott,
Effects of synaptic noise and filtering on the frequency response of spiking neurons,
Phys. Rev. Lett. **86** (2001) 2186.

[16] B. Lindner and L. Schimansky-Geier,
Transmission of noise coded versus additive signals through a neuronal ensemble,
Phys. Rev. Lett. **86** (2001) 2934.

[17] S. Ostojic and N. Brunel, *From spiking neuron models to linear-nonlinear models*,
PLOS Comput. Biol. **7** (2011) e1001056.

[18] A. Renart, N. Brunel, and X.-J. Wang, "Mean-field theory of irregularly spiking neuronal populations and working memory in recurrent cortical networks,"
Computational Neuroscience: A Comprehensive Approach, ed. by J. Feng,
Chapman & Hall/CRC, 2004 431.

[19] D. Sussillo, *Neural circuits as computational dynamical systems*,
Current Opinion in Neurobiology **25** (2014) 156.

[20] O. Barak, *Recurrent neural networks as versatile tools of neuroscience research*,
Curr. Opin. Neurobiol. **46** (2017) 1.

[21] F. Mastrogiuseppe and S. Ostojic,
Linking connectivity, dynamics, and computations in low-rank recurrent neural networks,
Neuron **99** (2018) 609.

[22] G. B. Ermentrout and D. H. Terman, *Mathematical Foundations of Neuroscience*,
Springer, 2010, ISBN: 9780387877075.

[23] O. Shriki, D. Hansel, and H. Sompolinsky,
Rate models for conductance-based cortical neuronal networks,
Neural Comput. **15** (2003) 1809.

[24] O. Harish and D. Hansel, *Asynchronous rate chaos in spiking neuronal circuits*,
PLOS Comput. Biol. **11** (2015) e1004266.

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

[26] C. van Vreeswijk and H. Sompolinsky,
Chaos in neuronal networks with balanced excitatory and inhibitory activity,
Science **274** (1996) 1724.

[27] N. Brunel,
Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons,
J. Comput. Neurosci. **8** (2000) 183.

[28] B. Barbour, N. Brunel, V. Hakim, and J.-P. Nadal,
What can we learn from synaptic weight distributions? Trends Neurosci. **30** (2007) 622.

[29] H. Sompolinsky, A. Crisanti, and H. J. Sommers, *Chaos in random neural networks*,
Phys. Rev. Lett. **61** (3 1988) 259.

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

[31] M. F. Bear, B. W. Connors, and M. A. Paradiso, *Neuroscience: Exploring the Brain*, 4th ed., Wolters Kluwer, 2016, ISBN: 9780781778176.

[32] L. Luo, *Principles of Neurobiology*, 2nd ed., CRC Press, 2021, ISBN: 0815346050.

[33] N. Yasumatsu, M. Matsuzaki, T. Miyazaki, J. Noguchi, and H. Kasai,
Principles of long-term dynamics of dendritic spines, J. Neurosci. **28** (2008) 13592.

[34] G. Mongillo, S. Rumpel, and Y. Loewenstein,
Intrinsic volatility of synaptic connections — a challenge to the synaptic trace theory of memory,
Curr. Opin. Neurobiol. **46** (2017) 7.

[35] N. E. Ziv and N. Brenner,
Synaptic tenacity or lack thereof: Spontaneous remodeling of synapses,
Trends Neurosci. **41** (2018) 89.

[36] J. Sjöström and W. Gerstner, *Spike-timing dependent plasticity*,
Scholarpedia **5** (2010) 1362, revision #184913.

[37] D. E. Feldman, *The spike-timing dependence of plasticity*, Neuron **75** (2012) 556.

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

[39] D. O. Hebb, *The organization of behavior: a neuropsychological theory*, Wiley, 1949.

[40] G. K. Ocker, A. Litwin-Kumar, and B. Doiron,
Self-organization of microcircuits in networks of spiking neurons with plastic synapses,
PLOS Comput. Biol. **11** (2015) e1004458.

[41] N. Ravid Tannenbaum and Y. Burak, *Shaping neural circuits by high order synaptic interactions*,
PLOS Comp. Biol. **12** (2016) e1005056.

[42] E. Oja, *Simplified neuron model as a principal component analyzer*,
Journal of Mathematical Biology **15** (1982) 267.

[43] L. F. Abbott and S. B. Nelson, *Synaptic plasticity: taming the beast*,
Nat. Neurosci. **3** (2000) 1178.

[44] F. Zenke and W. Gerstner,
Hebbian plasticity requires compensatory processes on multiple timescales,
Phil. Trans. R. Soc. B **372** (2017) 20160259.

[45] M. Chistiakova, N. M. Bannon, J.-Y. Chen, M. Bazhenov, and M. Volgushev,
Homeostatic role of heterosynaptic plasticity: models and experiments,
Front. Comput. Neurosci. **9** (2015) 89.

[46] K. D. Miller and D. J. C. MacKay, *The role of constraints in Hebbian learning*,
Neural Comput. **6** (1994) 100.

[47] K. D. Miller, *Synaptic economics: competition and cooperation in synaptic plasticity*,
Neuron **17** (1996) 371.

[48] I. R. Fiete, W. Senn, C. Z. H. Wang, and R. H. R. Hahnloser,
Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity, Neuron **65** (2010) 563.

[49] S. Song, K. D. Miller, and L. F. Abbott,
Competitive Hebbian learning through spike-timing-dependent synaptic plasticity,
Nat. Neurosci. **3** (2000) 919.

[50] R. Kempter, W. Gerstner, and J. L. van Hemmen,
Intrinsic stabilization of output rates by spike-based Hebbian learning,
Neural Comput. **13** (2001) 2709.

[51] G. Turrigiano,
Homeostatic synaptic plasticity: Local and global mechanisms for stabilizing neuronal function,
Cold Spring Harb. Perspect. Biol. **4** (2012) a005736.

[52] G. G. Turrigiano and S. B. Nelson, *Hebb and homeostasis in neuronal plasticity*,
Curr. Opin. Neurobiol. **10** (2000) 358.

[53] G. G. Turrigiano, *The self-tuning neuron: Synaptic scaling of excitatory synapses*,
Cell **135** (2008) 422.

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

[55] M. Butz, F. Wörgötter, and A. van Ooyen, *Activity-dependent structural plasticity*,
Brain Research Reviews **60** (2009) 287.

[56] A. van Ooyen and M. Butz-Ostendorf, eds., *The Rewiring Brain*, Academic Press, 2017,
ISBN: 9780128037843.

[57] A. van Ooyen and J. van Pelt, *Activity-dependent outgrowth of neurons and overshoot phenomena in developing neural networks*, J. Theor. Biol. **167** (1994) 27.

[58] A. van Ooyen, *Using theoretical models to analyse neural development*,
Nat. Rev. Neurosci. **12** (2011) 311.

[59] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, 2nd ed., CRC Press, 2015, ISBN: 9780813349107.

[60] A. Pikovsky and A. Politi, *Lyapunov Exponents: A Tool to Explore Complex Dynamics*,
Cambridge University Press, 2016, ISBN: 9781107030428.

[61] J.-P. Eckmann and D. Ruelle, *Ergodic theory of chaos and strange attractors*,
Rev. Mod. Phys. **57** (1985) 617.

[62] P. Holmes and E. T. Shea-Brown, *Stability*, Scholarpedia **1** (2006) 1838, revision #137538.

[63] G. Teschl, *Ordinary Differential Equations and Dynamical Systems*,
American Mathematical Society, 2012, ISBN: 978-0-8218-8328-0.

[64] P. C. Müller, *Calculation of Lyapunov exponents for dynamic systems with discontinuities*,
Chaos, Solitons & Fractals **5** (1995) 1671.

[65] M. Monteforte and F. Wolf,
Dynamical entropy production in spiking neuron networks in the balanced state,
Phys. Rev. Lett. **105** (2010) 268104.

[66] Y. Fyodorov, *Random matrix theory*, Scholarpedia **6** (2011) 9886, revision #90306.

[67] J. Ginibre, *Statistical ensembles of complex, quaternion, and real matrices*,
J. Math. Phys. **6** (1965) 440.

[68] V. L. Girko, *Circular law*, Theory of Probability and Its Applications **29** (1985) 694.

[69] H. J. Sommers, A. Crisanti, H. Sompolinsky, and Y. Stein,
Spectrum of large random asymmetric matrices, Phys. Rev. Lett. **60** (1988) 1895.

[70] Z. D. Bai, *Circular law*, Ann. Probab. **25** (1997) 494.

[71] R. M. May, *Will a large complex system be stable?* Nature **238** (1972) 413.

[72] S. Allesina and S. Tang,
The stability–complexity relationship at age 40: a random matrix perspective,
Population Ecology **57** (2015) 63.

[73] J. Kadmon and H. Sompolinsky, *Transition to chaos in random neuronal networks*,
Phys. Rev. X **5** (2015) 041030.

[74] S. Ostojic,
Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons,
Nat. Neurosci. **17** (2014) 594.

[75] R. Engelken, F. Farkhooi, D. Hansel, C. van Vreeswijk, and F. Wolf, *A reanalysis of "Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons"*,
F1000Research **5** (2016) 2043, [version 1; peer review: 2 approved].

[76] R. Zillmer, N. Brunel, and D. Hansel, *Very long transients, irregular firing, and chaotic dynamics in networks of randomly connected inhibitory integrate-and-fire neurons*,
Phys. Rev. E **79** (2009) 031909.

[77] S. Jahnke, R.-M. Memmesheimer, and M. Timme, *How chaotic is the balanced state?*
Front. Comput. Neurosci. **3** (2009) 13.

[78] R. Engelken, *Chaotic Neural Circuit Dynamics*,
PhD thesis: Georg-August-University Göttingen, 2017,
URL: `http://dx.doi.org/10.53846/goediss-6718`.

[79] C. van Vreeswijk and H. Sompolinsky, *Chaotic balanced state in a model of cortical circuits*,
Neural Comput. **10** (1998) 1321.

[80] M. Timme, F. Wolf, and T. Geisel,
Coexistence of regular and irregular dynamics in complex networks of pulse-coupled oscillators,
Phys. Rev. Lett. **89** (2002) 258701.

[81] C. van Vreeswijk and H. Sompolinsky,
"Course 9 - Irregular activity in large networks of neurons,"
Methods and Models in Neurophysics, Elsevier, 2005 341.

[82] S. Jahnke, R.-M. Memmesheimer, and M. Timme,
Stable irregular dynamics in complex neural networks, Phys. Rev. Lett. **100** (2008) 048102.

[83] M. Monteforte and F. Wolf, *Dynamic flux tubes form reservoirs of stability in neuronal circuits*, Phys. Rev. X **2** (2012) 041007.

[84] J. Honerkamp, *Stochastische Dynamische Systeme*, German, Wiley-VCH, 1990, ISBN: 3527279458.

[85] H. C. Tuckwell, *Introduction to theoretical neurobiology, Volume 2: Nonlinear and stochastic theories*, Cambridge University Press, 1988, ISBN: 9780521019323.

[86] A. N. Burkitt, *A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input*, Biol. Cybern. **95** (2006) 1.

[87] M. J. E. Richardson, *Firing-rate response of linear and nonlinear integrate-and-fire neurons to modulated current-based and conductance-based synaptic drive*, Phys. Rev. E **76** (2007) 021919.

[88] M. J. E. Richardson, *Spike-train spectra and network response functions for non-linear integrate-and-fire neurons*, Biol. Cybern. **99** (2008) 381.

[89] L. Arnold, *Random Dynamical Systems*, Springer, 1998, ISBN: 3540637583.

[90] K. K. Lin, E. Shea-Brown, and L.-S. Young, *Spike-time reliability of layered neural oscillator networks*, J. Comput. Neurosci. **27** (2009) 135.

[91] G. Lajoie, K. K. Lin, and E. Shea-Brown, *Chaos and reliability in balanced spiking networks with temporal drive*, Phys. Rev. E **87** (2013) 052901.

[92] Y. A. Kuznetsov, *Elements of Applied Bifurcation Theory*, Springer, 2004, ISBN: 978-0-387-21906-6.

[93] T. Gross, *Not one, but many critical states: a dynamical systems perspective*, Front. Neural Circuits **15** (2021) 614268.

[94] M. C. Cross and P. C. Hohenberg, *Pattern formation outside of equilibrium*, Rev. Mod. Phys. **65** (1993) 851.

[95] M. Henkel, H. Hinrichsen, and S. Lübeck, *Non-Equilibrium Phase Transitions: Volume 1: Absorbing Phase Transitions*, Springer, 2008, ISBN: 9781402087653.

[96] R. V. Solé, *Phase Transitions*, Primers in Complex Systems, Princeton University Press, 2011, ISBN: 9780691150758.

[97] M. A. Muñoz, *Colloquium: Criticality and dynamical scaling in living systems*, Rev. Mod. Phys. **90** (2018) 031001.

[98] M. Fruchart, R. Hanai, P. B. Littlewood, and V. Vitelli, *Non-reciprocal phase transitions*, Nature **592** (2021) 363.

[99] U. Feudel, *Complex dynamics in multistable systems*, Int. J. Bifurcation Chaos **18** (2008) 1607.

[100] U. Feudel, A. N. Pisarchik, and K. Showalter, *Multistability and tipping: From mathematics and physics to climate and brain—Minireview and preface to the focus issue*, Chaos **28** (2018) 033501.

[101] E. Forgoston and R. O. Moore,
A primer on noise-induced transitions in applied dynamical systems,
SIAM Review **60** (2018) 969.

[102] B. A. W. Brinkman et al., *Metastable dynamics of neural circuits and networks*,
Appl. Phys. Rev. **9** (2022) 011313.

[103] W. Horsthemke and R. Lefever,
Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology,
Springer, 1984, ISBN: 9783540113591.

[104] C. Van den Broeck, J. M. R. Parrondo, and R. Toral,
Noise-induced nonequilibrium phase transition, Phys. Rev. Lett. **73** (1994) 3395.

[105] T. Biancalani, L. Dyson, and A. J. McKane,
Noise-induced bistable states and their mean switching time in foraging colonies,
Phys. Rev. Lett. **112** (2014) 038101.

[106] P. Glendinning, *Stability, Instability and Chaos: An Introduction to the Theory of Nonlinear Differential Equations*, Cambridge University Press, 1994, ISBN: 9780521425667.

[107] J. J. Binney, N. J. Dowrick, A. J. Fisher, and M. E. J. Newman,
The Theory of Critical Phenomena: An Introduction to the Renormalization Group,
Oxford University Press, 1992, ISBN: 9780198513933.

[108] J. P. Sethna, *Statistical Mechanics: Entropy, Order Parameters, and Complexity*, 2nd ed.,
Oxford University Press, 2021, ISBN: 9780198865254.

[109] J. Hesse and T. Gross, *Self-organized criticality as a fundamental property of neural systems*,
Front. Syst. Neurosci. **8** (2014).

[110] G. Jetschke, *Mathematik der Selbstorganisation*, German, 2. korrigierte und erw. Aufl,
Verlag Harri Deutsch, 2008, ISBN: 3817118252.

[111] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory*, Springer, 1999, ISBN: 0387989315.

[112] P. Hänggi, P. Talkner, and M. Borkovec, *Reaction-rate theory: fifty years after Kramers*,
Rev. Mod. Phys. **62** (1990) 251.

[113] J. Schuecker, S. Goedeke, and M. Helias, *Optimal sequence memory in driven random networks*,
Phys. Rev. X **8** (2018) 041029.

[114] W. Maass, T. Natschläger, and H. Markram, *Real-time computing without stable states: a new framework for neural computation based on perturbations*, Neural Comput. **14** (2002) 2531.

[115] H. Jaeger and H. Haas, *Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication*, Science **304** (2004) 78.

[116] R. Legenstein and W. Maass,
Edge of chaos and prediction of computational performance for neural circuit models,
Neural Networks **20** (2007) 323.

[117] T. Toyoizumi and L. F. Abbott, *Beyond the edge of chaos: Amplification and temporal integration by recurrent networks in the chaotic regime*, Phys. Rev. E **84** (2011) 051908.

[118] O. Barak, D. Sussillo, R. Romo, M. Tsodyks, and L. F. Abbott,
From fixed points to chaos: Three models of delayed discrimination,
Prog. Neurobiol. **103** (2013) 214.

[119] K. Rajan, C. D. Harvey, and D. W. Tank,
Recurrent network models of sequence generation and memory, Neuron **90** (2016) 128.

[120] N. Li, K. Daie, K. Svoboda, and S. Druckmann,
Robust neuronal dynamics in premotor cortex during motor planning, Nature **532** (2016) 459.

[121] J. D. Murray et al., *Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex*, Proc. Natl. Acad. Sci. U.S.A. **114** (2016) 394.

[122] R. Laje and D. V. Buonomano,
Robust timing and motor patterns by taming chaos in recurrent neural networks,
Nat. Neurosci. **16** (2013) 925.

[123] V. Mante, D. Sussillo, K. V. Shenoy, and W. T. Newsome,
Context-dependent computation by recurrent dynamics in prefrontal cortex,
Nature **503** (2013) 78.

[124] H. Sompolinsky and A. Zippelius, *Dynamic theory of the spin-glass phase*,
Phys. Rev. Lett. **47** (1981) 359.

[125] H. Sompolinsky and A. Zippelius, *Relaxational dynamics of the Edwards-Anderson model and the mean-field theory of spin-glasses*, Phys. Rev. B **25** (1982) 6860.

[126] K. Rajan, L. F. Abbott, and H. Sompolinsky,
Stimulus-dependent suppression of chaos in recurrent neural networks,
Phys. Rev. E **82** (2010) 011903.

[127] J. Aljadeff, M. Stern, and T. Sharpee,
Transition to chaos in random networks with cell-type-specific connectivity,
Phys. Rev. Lett. **114** (2015) 088101.

[128] K. Rajan and L. F. Abbott, *Eigenvalue spectra of random matrices for neural networks*,
Phys. Rev. Lett. **97** (2006) 188104.

[129] M. Helias and D. Dahmen, *Statistical Field Theory for Neural Networks*, Springer, 2020,
ISBN: 9783030464448.

[130] https://doi.org/10.5281/zenodo.3754062, This archive contains the source code to reproduce most figures from the book "Statistical field theory for neural networks" by Moritz Helias and David Dahmen published 2020 with Springer Lecture Notes in Physics.

[131] L. Molgedey, J. Schuchhardt, and H. G. Schuster,
Suppressing chaos in neural networks by noise, Phys. Rev. Lett. **69** (1992) 3717.

[132] C. Zhou and J. Kurths,
Noise-induced phase synchronization and synchronization transitions in chaotic oscillators,
Phys. Rev. Lett. **88** (2002) 230602.

[133] P. C. Martin, E. D. Siggia, and H. A. Rose, *Statistical dynamics of classical systems*,
Phys. Rev. A **8** (1973) 423.

[134] H.-K. Janssen, *On a Lagrangean for classical field dynamics and renormalization group calculations of dynamical critical properties*, Z. Physik B: Cond. Mat. **23** (1976) 377.

[135] C. De Dominicis,
Techniques de renormalisation de la théorie des champs es dynamique des phénomènes critiques,
J. Phys. Colloques **37** (1976) C1-247-C1.

[136] C. De Dominicis and L. Peliti, *Field-theory renormalization and critical dynamics above T_c: Helium, antiferromagnets, and liquid-gas systems*, Phys. Rev. B **18** (1978) 353.

[137] J. Schuecker, S. Goedeke, D. Dahmen, and M. Helias,
Functional methods for disordered neural networks, 2016,
arXiv: 1605.06758 [cond-mat.dis-nn], URL: https://arxiv.org/abs/1605.06758.

[138] M. Moshe and J. Zinn-Justin, *Quantum field theory in the large N limit: a review*,
Phys. Rep. **385** (2003) 69.

[139] B. Derrida and Y. Pomeau, *Random networks of automata: a simple annealed approximation*,
Europhys. Lett. **1** (1986) 45.

[140] H. Jaeger, *Short term memory in echo state networks*, GMD Report 152,
GMD – Forschungszentrum Informationstechnik, 2001,
URL: https://doi.org/10.24406/publica-fhg-291107.

[141] K. H. Fischer and J. A. Hertz, *Spin Glasses*, Cambridge University Press, 1991.

[142] S.-I. Amari, *Characteristics of random nets of analog neuron-like elements*,
IEEE Transactions on Systems, Man, and Cybernetics **2** (1972) 643.

[143] A. Crisanti and H. Sompolinsky, *Dynamics of spin systems with randomly asymmetric bonds: Langevin dynamics and a spherical model*, Phys. Rev. A **36** (1987) 4922.

[144] J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena*, 4th ed.,
Oxford University Press, 2002.

[145] T. Cabana and J. Touboul, *Large deviations, dynamics and phase transitions in large stochastic and disordered neural networks*, J. Stat. Phys. **153** (2013) 211.

[146] A. Altland and B. Simons, *Condensed Matter Field Theory*, 2nd ed.,
Cambridge University Press, 2010, ISBN: 9780521769754.

[147] C. C. Chow and M. A. Buice, *Path integral methods for stochastic differential equations*,
The Journal of Mathematical Neuroscience **5** (2015) 8.

[148] F. Mastrogiuseppe and S. Ostojic,
Intrinsically-generated fluctuating activity in excitatory-inhibitory networks,
PLOS Comput. Biol. **13** (2017) e1005498.

[149] R. Price, *A useful theorem for nonlinear devices having Gaussian inputs*,
IRE Transactions on Information Theory **4** (1958) 69.

[150] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed.,
McGraw-Hill, 1991, ISBN: 9780070484771.

[151] S. Kirkpatrick and D. Sherrington, *Infinite-ranged models of spin-glasses*,
Phys. Rev. B **17** (1978) 4384.

[152] J. Hahne et al., *Integration of continuous-time dynamics in a spiking neural network simulator*, Front. Neuroinf. **11** (2017) 34.

[153] A. Peyser et al., *NEST 2.14.0*, 2017, URL: https://doi.org/10.5281/zenodo.882971.

[154] N. Fourcaud and N. Brunel,
Dynamics of the firing probability of noisy integrate-and-fire neurons,
Neural Comput. **14** (2002) 2057.

[155] M. Helias, T. Tetzlaff, and M. Diesmann, *Echoes in correlated neural systems*,
New J. Phys. **15** (2013) 023002.

[156] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar,
Information processing capacity of dynamical systems, Sci. Rep. **2** (2012) 514.

[157] M. Hermans and B. Schrauwen, "Memory in reservoirs for high dimensional input,"
The 2010 International Joint Conference on Neural Networks (IJCNN), IEEE, 2010 1.

[158] S. Ostojic, *Response to Comment on "Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons"*, 2015,
URL: https://www.biorxiv.org/content/early/2015/06/05/020354.

[159] J. Trousdale, Y. Hu, E. Shea-Brown, and K. Josić,
Impact of network structure and cellular response on spike time correlations,
PLOS Comput. Biol. **8** (2012) e1002408.

[160] B. Dummer, S. Wieland, and B. Lindner, *Self-consistent determination of the spike-train power spectrum in a neural network with sparse connectivity*, Front. Comput. Neurosci. **8** (2014) 104.

[161] S. Wieland, D. Bernardi, T. Schwalger, and B. Lindner,
Slow fluctuations in recurrent networks of spiking neurons, Phys. Rev. E **92** (2015) 040901.

[162] G. Wainrib and J. Touboul, *Topological and dynamical complexity of random neural networks*,
Phys. Rev. Lett. **110** (2013) 118101.

[163] G. Wainrib and M. N. Galtier,
A local Echo State Property through the largest Lyapunov exponent,
Neural Networks **76** (2016) 39.

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

[165] R. Legenstein and W. Maass, "What makes a dynamical system computationally powerful?"
New Directions in Statistical Signal Processing: From Systems to Brains,
ed. by S. Haykin, J. C. Principe, T. Sejnowski, and J. McWhirter, MIT Press, 2007 127,
URL: https://igi-web.tugraz.at/people/Abstracts/LegensteinMaass:05/.

[166] L. Büsing, B. Schrauwen, and R. Legenstein,
Connectivity, dynamics, and memory in reservoir computing with binary and analog neurons,
Neural Comput. **22** (2010) 1272.

[167] J. Boedecker, O. Obst, J. T. Lizier, N. M. Mayer, and M. Asada,
Information processing in echo state networks at the edge of chaos,
Theory in Biosciences **131** (2012) 205.

[168] L. Livi, F. M. Bianchi, and C. Alippi, *Determination of the edge of criticality in echo state networks through Fisher information maximization*,
IEEE Transactions on Neural Networks and Learning Systems **29** (2017) 706.

[169] P. C. Hohenberg and B. I. Halperin, *Theory of dynamic critical phenomena*,
Rev. Mod. Phys. **49** (1977) 435.

[170] M. Prokopenko, J. T. Lizier, O. Obst, and X. R. Wang,
Relating Fisher information to order parameters, Phys. Rev. E **84** (2011) 041116.

[171] L. Mazzucato, A. Fontanini, and G. La Camera,
Stimuli reduce the dimensionality of cortical activity, Front. Syst. Neurosci. **10** (2016) 11.

[172] J. A. Hertz, Y. Roudi, and P. Sollich,
Path integral methods for the dynamics of stochastic and disordered systems,
J. Phys. A: Math. Theor. **50** (2017) 033001.

[173] S. Song, P. J. Sjöström, M. Reigl, S. Nelson, and D. B. Chklovskii,
Highly nonrandom features of synaptic connectivity in local cortical circuits,
PLOS Biol. **3** (2005) e68.

[174] D. Martí, N. Brunel, and S. Ostojic, *Correlations between synapses in pairs of neurons slow down dynamics in randomly connected neural networks*, Phys. Rev. E **97** (2018) 062314.

[175] F. A. Rodrigues, T. K. D. Peron, P. Ji, and J. Kurths, *The Kuramoto model in complex networks*, Phys. Rep. **610** (2016) 1.

[176] T. Nishikawa and A. E. Motter,
Comparative analysis of existing models for power-grid synchronization,
New J. Phys. **17** (2015) 015012.

[177] A. Pomerance, E. Ott, M. Girvan, and W. Losert,
The effect of network topology on the stability of discrete state models of genetic control,
Proc. Natl. Acad. Sci. U.S.A. **106** (2009) 8209.

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

[179] R. Zillmer, R. Livi, A. Politi, and A. Torcini, *Desynchronization in diluted neural networks*, Phys. Rev. E **74** (2006) 036203.

[180] A. Politi, R. Livi, G.-L. Oppo, and R. Kapral, *Unpredictable behaviour in stable systems*, Europhys. Lett. **22** (1993) 571.

[181] A. Politi and A. Torcini, "Stable chaos,"
Nonlinear Dynamics and Chaos: Advances and Perspectives,
ed. by M. Thiel, J. Kurths, M. Romano, G. Károlyi, and A. Moura, Springer, 2010 103.

[182] R. Tremblay, S. Lee, and B. Rudy,
GABAergic interneurons in the neocortex: From cellular properties to circuits,
Neuron **91** (2016) 260.

[183] S. Olmi, D. Angulo-Garcia, A. Imparato, and A. Torcini,
Exact firing time statistics of neurons driven by discrete inhibitory noise,
Sci. Rep. **7** (2017) 1577.

[184] R. E. Mirollo and S. H. Strogatz, *Synchronization of pulse-coupled biological oscillators*, SIAM J. Appl. Math. **50** (1990) 1645.

[185] R.-M. Memmesheimer and M. Timme, *Designing complex networks*, Physica D **224** (2006) 182.

[186] P. Goel and B. Ermentrout, *Synchrony, stability, and firing patterns in pulse-coupled oscillators*, Physica D **163** (2002) 191.

[187] R. M. Smeal, G. B. Ermentrout, and J. A. White, *Phase-response curves and synchronized neural networks*, Phil. Trans. R. Soc. B **365** (2010) 2407.

[188] A. Viriyopase, R.-M. Memmesheimer, and S. Gielen, *Analyzing the competition of gamma rhythms with delayed pulse-coupled oscillators in phase representation*, Phys. Rev. E **98** (2018) 022217.

[189] M. J. E. Richardson and R. Swarbrick, *Firing-rate response of a neuron receiving excitatory and inhibitory synaptic shot noise*, Phys. Rev. Lett. **105** (2010) 178102.

[190] D. Angulo-Garcia, S. Luccioli, S. Olmi, and A. Torcini, *Death and rebirth of neural activity in sparse inhibitory networks*, New J. Phys. **19** (2017) 053011.

[191] J. K. Hale and S. M. Verduyn Lunel, *Introduction to Functional Differential Equations*, Springer, 1993.

[192] F. Ginelli et al., *Characterizing dynamics with covariant Lyapunov vectors*, Phys. Rev. Lett. **99** (2007) 130601.

[193] P. V. Kuptsov and U. Parlitz, *Theory and computation of covariant Lyapunov vectors*, J. Nonlinear Sci. **22** (2012) 727.

[194] B. Kriener, T. Tetzlaff, A. Aertsen, M. Diesmann, and S. Rotter, *Correlations and population dynamics in cortical networks*, Neural Comput. **20** (2008) 2185.

[195] S. Denève and C. K. Machens, *Efficient codes and balanced networks*, Nat. Neurosci. **19** (2016) 375.

[196] A. A. Prinz, *Insights from models of rhythmic motor systems*, Curr. Opin. Neurobiol. **16** (2006) 615.

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

[198] X.-J. Wang, J. Tegnér, C. Constantinidis, and P. S. Goldman-Rakic, *Division of labor among distinct subtypes of inhibitory neurons in a cortical microcircuit of working memory*, Proc. Natl. Acad. Sci. U.S.A. **101** (2004) 1368.

[199] A. Litwin-Kumar, R. Rosenbaum, and B. Doiron, *Inhibitory stabilization and visual coding in cortical circuits with multiple interneuron subtypes*, J. Neurophysiol. **115** (2016) 1399.

[200] R.-M. Memmesheimer and M. Timme,
Stable and unstable periodic orbits in complex networks of spiking neurons with delays,
Discr. Cont. Dyn. Syst. **28** (2010) 1555.

[201] Q. L. Gu, Z. K. Tian, G. Kovačič, D. Zhou, and D. Cai,
The dynamics of balanced spiking neuronal networks under poisson drive is not chaotic,
Front. Comput. Neurosci. **12** (2018) 47.

[202] S. Luccioli, S. Olmi, A. Politi, and A. Torcini, *Collective dynamics in sparse networks*,
Phys. Rev. Lett. **109** (2012) 138103.

[203] G. Lajoie, J.-P. Thivierge, and E. Shea-Brown,
Structured chaos shapes spike-response noise entropy in balanced neural networks,
Front. Comput. Neurosci. **8** (2014) 123.

[204] E. Ullner and A. Politi, *Self-sustained irregular activity in an ensemble of neural oscillators*,
Phys. Rev. X **6** (2016) 011015.

[205] S. Coombes, "Chaos in integrate-and-fire dynamical systems," *AIP Conference Proceedings*,
2000 88.

[206] F. Y. Kalle Kossio, S. Goedeke, B. van den Akker, B. Ibarz, and R.-M. Memmesheimer,
Growing critical: Self-organized criticality in a developing neural system,
Phys. Rev. Lett. **121** (2018) 058301.

[207] D. Marković and C. Gros, *Power laws and self-organized criticality in theory and nature*,
Phys. Rep. **536** (2014) 41.

[208] P. Bak, C. Tang, and K. Wiesenfeld, *Self-organized criticality: An explanation of the 1/f noise*,
Phys. Rev. Lett. **59** (1987) 381.

[209] P. Bak, C. Tang, and K. Wiesenfeld, *Self-organized criticality*, Phys. Rev. A **38** (1988) 364.

[210] S. Zapperi, K. B. Lauritsen, and H. E. Stanley,
Self-organized branching processes: Mean-field theory for avalanches,
Phys. Rev. Lett. **75** (1995) 4071.

[211] H. J. Jensen, *Self-Organized Criticality*, Cambridge University Press, 1998.

[212] T. E. Harris, *The Theory of Branching Processes*, Dover Publications, 2002.

[213] S. di Santo, P. Villegas, R. Burioni, and M. A. Muñoz, *Simple unified view of branching process statistics: Random walks in balanced logarithmic potentials*, Phys. Rev. E **95** (2017) 032115.

[214] J. M. Beggs and D. Plenz, *Neuronal avalanches in neocortical circuits*,
J. Neurosci. **23** (2003) 11167.

[215] J. M. Beggs and D. Plenz, *Neuronal avalanches are diverse and precise activity patterns that are stable for many hours in cortical slice cultures*, J. Neurosci. **24** (2004) 5216.

[216] C. Haldeman and J. M. Beggs, *Critical branching captures activity in living neural networks and maximizes the number of metastable states*, Phys. Rev. Lett. **94** (2005) 058101.

[217] W. L. Shew and D. Plenz, *The functional benefits of criticality in the cortex*,
The Neuroscientist **19** (2013) 88.

[218] A. Mazzoni et al., *On the dynamics of the spontaneous activity in neuronal networks*, PLOS ONE **2** (2007) e439.

[219] E. D. Gireesh and D. Plenz, *Neuronal avalanches organize as nested theta- and beta/gamma-oscillations during development of cortical layer 2/3*, Proc. Natl. Acad. Sci. U.S.A. **105** (2008) 7576.

[220] M. H. Hennig, C. Adams, D. Willshaw, and E. Sernagor, *Early-stage waves in the retinal network emerge close to a critical state transition between local and global functional connectivity*, J. Neurosci. **29** (2009) 1077.

[221] Y. Yada et al., *Development of neural population activity toward self-organized criticality*, Neuroscience **343** (2017) 55.

[222] S. B. Kater, M. P. Mattson, C. Cohan, and J. Connor, *Calcium regulation of the neuronal growth cone*, Trends Neurosci. **11** (1988) 315.

[223] S. B. Kater and L. R. Mills, *Regulation of growth cone behavior by calcium*, J. Neurosci. **11** (1991) 891.

[224] C. S. Cohan and S. B. Kater, *Suppression of neurite elongation and growth cone motility by electrical activity*, Science **232** (1986) 1638.

[225] F. Van Huizen and H. J. Romijn, *Tetrodotoxin enhances initial neurite outgrowth from fetal rat cerebral cortex cells in vitro*, Brain Research **408** (1987) 271.

[226] R. D. Fields, E. A. Neale, and P. G. Nelson, *Effects of patterned electrical activity on neurite outgrowth from mouse sensory neurons*, J. Neurosci. **10** (1990) 2950.

[227] C. W. Eurich, J. M. Herrmann, and U. A. Ernst, *Finite-size effects of avalanche dynamics*, Phys. Rev. E **66** (2002) 066137.

[228] A. Levina, J. M. Herrmann, and T. Geisel, *Dynamical synapses causing self-organized criticality in neural networks*, Nat. Phys. **3** (2007) 857.

[229] A. Levina, J. M. Herrmann, and T. Geisel, *Phase transitions towards criticality in a neural system with adaptive interactions*, Phys. Rev. Lett. **102** (2009) 118110.

[230] L. F. Abbott and R. Rohrkemper, *A simple growth model constructs critical avalanche networks*, Progress in Brain Research **165** (2007) 13.

[231] V. Gómez, A. Kaltenbrunner, V. López, and H. Kappen, "Self-organization using synaptic plasticity," *Advances in Neural Information Processing Systems*, ed. by D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, vol. 21, 2008, URL: `https://papers.nips.cc/paper/3394-self-organization-using-synaptic-plasticity`.

[232] C. Tetzlaff, S. Okujeni, U. Egert, F. Wörgötter, and M. Butz, *Self-organized criticality in developing neuronal networks*, PLOS Comput. Biol. **6** (2010) e1001013.

[233] B. D. Papa, V. Priesemann, and J. Triesch,
Criticality meets learning: Criticality signatures in a self-organizing recurrent neural network,
PLOS ONE **12** (2017) e0178683.

[234] S. Bornholdt and T. Rohlf,
Topological evolution of dynamical networks: Global criticality from local dynamics,
Phys. Rev. Lett. **84** (2000) 6114.

[235] S. Bornholdt and T. Röhl, *Self-organized critical neural networks*,
Phys. Rev. E **67** (2003) 066118.

[236] C. Meisel and T. Gross, *Adaptive self-organization in a realistic neural network model*,
Phys. Rev. E **80** (2009) 061917.

[237] V. Pernice, B. Staude, S. Cardanobile, and S. Rotter,
How structure determines correlations in neuronal networks,
PLOS Comput. Biol. **7** (2011) e1002059.

[238] C. Koch, *Biophysics of Computation: Information Processing in Single Neurons*,
Oxford University Press, 1999.

[239] K. J. Ford, A. L. Félix, and M. B. Feller,
Cellular mechanisms underlying spatiotemporal features of cholinergic retinal waves,
J. Neurosci. **32** (2012) 850.

[240] A. N. Burkitt, M. Gilson, and J. L. van Hemmen,
Spike-timing-dependent plasticity for neurons with recurrent connections,
Biol. Cybern. **96** (2007) 533.

[241] J. Barral and A. D. Reyes, *Synaptic scaling rule preserves excitatory–inhibitory balance and salient neuronal network dynamics*, Nat. Neurosci. **19** (2016) 1690.

[242] M. Abeles, *Corticonics: Neural Circuits of the Cerebral Cortex*,
Cambridge University Press, 1991.

[243] A. van Ooyen et al., *Independently outgrowing neurons and geometry-based synapse formation produce networks with realistic synaptic connectivity*, PLOS ONE **9** (2014) e85858.

[244] J. Zierenberg, J. Wilting, and V. Priesemann,
Homeostatic plasticity and external input shape neural network dynamics,
Phys. Rev. X **8** (2018) 031018.

[245] K. S. Crump and C. J. Mode, *A general age-dependent branching process. I*,
J. Math. Anal. Appl. **24** (1968) 494.

[246] J. C. Tanner, *A derivation of the Borel distribution*, Biometrika **48** (1961) 222.

[247] V. Priesemann et al., *Spike avalanches in vivo suggest a driven, slightly subcritical brain state*,
Front. Syst. Neurosci. **8** (2014) 108.

[248] M. Abramowitz and I. A. Stegun,
Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables,
Dover Publications, 1974.

[249] D. Oakes, *The Markovian self-exciting process*, J. Appl. Probab. **12** (1975) 69.

[250] E. Bacry, I. Mastromatteo, and J.-F. Muzy, *Hawkes processes in finance*, Mark. Microstruct. Liquidity **1** (2015) 1550005.

[251] E. Errais, K. Giesecke, and L. R. Goldberg, *Affine point processes and portfolio credit risk*, SIAM J. Finan. Math. **1** (2010) 642.

[252] A. Dassios and H. Zhao, *A dynamic contagion process*, Adv. Appl. Probab. **43** (2011) 814.

[253] X. Gao and L. Zhu, *Limit theorems for Markovian Hawkes processes with a large initial intensity*, Stoch. Process. Their Appl. **128** (2018) 3807.

[254] A. G. Hawkes and D. Oakes, *A cluster process representation of a self-exciting process*, J. Appl. Probab. **11** (1974) 493.

[255] J. Møller and J. G. Rasmussen, *Perfect simulation of Hawkes processes*, Adv. Appl. Probab. **37** (2005) 629.

[256] V. Priesemann, M. H. J. Munk, and M. Wibral, *Subsampling effects in neuronal avalanche distributions recorded in vivo*, BMC Neurosci. **10** (2009) 40.

[257] G. Hahn et al., *Neuronal avalanches in spontaneous activity in vivo*, J. Neurophysiol. **104** (2010) 3312.

[258] A. Levina and V. Priesemann, *Subsampling scaling*, Nat. Commun. **8** (2017) 15140.

[259] A. Clauset, C. R. Shalizi, and M. E. J. Newman, *Power-law distributions in empirical data*, SIAM Review **51** (2009) 661.

[260] D. Sornette, *Sweeping of an instability: an alternative to self-organized criticality to get powerlaws without parameter tuning*, J. Phys. I France **4** (1994) 209.

[261] J. A. Bonachela and M. A. Muñoz, *Self-organization without conservation: true or just apparent scale-invariance?* J. Stat. Mech. **2009** (2009) P09009.

[262] J. A. Bonachela, S. de Franciscis, J. J. Torres, and M. A. Muñoz, *Self-organization without conservation: are neuronal avalanches generically critical?* J. Stat. Mech. **2010** (2010) P02015.

[263] T. P. Vogels and L. F. Abbott, *Signal propagation and logic gating in networks of integrate-and-fire neurons*, J. Neurosci. **25** (2005) 10786.

[264] A. Kumar, S. Schrader, A. Aertsen, and S. Rotter, *The high-conductance state of cortical networks*, Neural Comput. **20** (2008) 1.

[265] A. Levina, U. Ernst, and J. M. Herrmann, *Criticality of avalanche dynamics in adaptive recurrent networks*, Neurocomputing **70** (2007) 1877.

[266] F. Droste, A.-L. Do, and T. Gross, *Analytical investigation of self-organized criticality in neural networks*, J. R. Soc. Interface **10** (2013) 20120558.

[267] T. Wang, M. Bebbington, and D. Harte, *Markov-modulated Hawkes process with stepwise decay*, Ann. Inst. Stat. Math. **64** (2012) 521.

[268] E. Lewis, G. Mohler, P. J. Brantingham, and A. L. Bertozzi, *Self-exciting point process models of civilian deaths in Iraq*, Security Journal **25** (2012) 244.

[269] H. Kim, *Spatio-Temporal Point Process Models for the Spread of Avian Influenza Virus (H5N1)*, PhD thesis: UC Berkeley, 2011, URL: http://www.escholarship.org/uc/item/8nc0r19n.

[270] F. P. Schoenberg, M. Hoffmann, and R. J. Harrigan, *A recursive point process model for infectious diseases*, Ann. Inst. Stat. Math. **71** (2019) 1271.

[271] Y. F. Kalle Kossio, S. Goedeke, C. Klos, and R.-M. Memmesheimer, *Drifting assemblies for persistent memory: Neuron transitions and unsupervised compensation*, Proc. Natl. Acad. Sci. U.S.A. **118** (2021) e2023832118.

[272] S. Rumpel and J. Triesch, *The dynamic connectome*, e-Neuroforum **22** (2016) 48.

[273] Y. Humeau and D. Choquet, *The next generation of approaches to investigate the link between synaptic plasticity and learning*, Nat. Neurosci. **22** (2019) 1536.

[274] A. Rubinski and N. E. Ziv, *Remodeling and tenacity of inhibitory synapses: Relationships with network activity and neighboring excitatory synapses*, PLOS Comput. Biol. **11** (2015) e1004632.

[275] R. Dvorkin and N. E. Ziv, *Relative contributions of specific activity histories and spontaneous processes to size remodeling of glutamatergic synapses*, PLOS Biol. **14** (2016) e1002572.

[276] K. P. Berry and E. Nedivi, *Spine dynamics: Are they all the same?* Neuron **96** (2017) 43.

[277] L. A. DeNardo et al., *Temporal evolution of cortical ensembles promoting remote memory retrieval*, Nat. Neurosci. **22** (2019) 460.

[278] C. Clopath, T. Bonhoeffer, M. Hübener, and T. Rose, *Variance and invariance of neuronal long-term representations*, Phil. Trans. R. Soc. B **372** (2017) 20160161.

[279] M. E. Rule, T. O'Leary, and C. D. Harvey, *Causes and consequences of representational drift*, Curr. Opin. Neurobiol. **58** (2019) 141.

[280] G. Buzsáki, *Neural syntax: Cell assemblies, synapsembles, and readers*, Neuron **68** (2010) 362.

[281] A. Scott, *Neuroscience: A Mathematical Primer*, Springer, 2002, ISBN: 9780387954035.

[282] T. P. Vogels, H. Sprekeler, F. Zenke, C. Clopath, and W. Gerstner, *Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks*, Science **334** (2011) 1569.

[283] A. Litwin-Kumar and B. Doiron, *Formation and maintenance of neuronal assemblies through synaptic plasticity*, Nat. Commun. **5** (2014) 5319.

[284] F. Zenke, E. J. Agnes, and W. Gerstner, *Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks*, Nat. Commun. **6** (2015) 6922.

[285] G. K. Ocker and B. Doiron,
Training and spontaneous reinforcement of neuronal assemblies by spike timing plasticity,
Cerebral Cortex **29** (2018) 937.

[286] J. Herpich and C. Tetzlaff,
Principles underlying the input-dependent formation and organization of memories,
Network Neuroscience **3** (2019) 606.

[287] J. Humble, K. Hiratsuka, H. Kasai, and T. Toyoizumi, *Intrinsic spine dynamics are critical for recurrent network learning in models with and without autism spectrum disorder*,
Front. Comput. Neurosci. **13** (2019) 38.

[288] L. Montangie, C. Miehl, and J. Gjorgjieva,
Autonomous emergence of connectivity assemblies via spike triplet interactions,
PLOS Comput. Biol. **16** (2020) e1007835.

[289] M. J. Fauth and M. C. W. van Rossum,
Self-organized reactivation maintains and reinforces memories despite synaptic turnover,
eLife **8** (2019) e43717.

[290] L. Wittgenstein, *Philosophische Untersuchungen/Philosophical investigations*,
ed. by P. M. S. Hacker and J. Schulte, Wiley-Blackwell, 2009.

[291] T. Hainmueller and M. Bartos,
Parallel emergence of stable and dynamic memory engrams in the hippocampus,
Nature **558** (2018) 292.

[292] D. V. Buonomano,
A learning rule for the emergence of stable dynamics and timing in recurrent networks,
J. Neurophysiol. **94** (2005) 2275.

[293] C. Clopath, L. Büsing, E. Vasilaki, and W. Gerstner,
Connectivity reflects coding: a model of voltage-based STDP with homeostasis,
Nat. Neurosci. **13** (2010) 344.

[294] S. Royer and D. Paré,
Conservation of total synaptic weight through balanced synaptic depression and potentiation,
Nature **422** (2003) 518.

[295] M. Letellier, F. Levet, O. Thoumine, and Y. Goda, *Differential role of pre- and postsynaptic neurons in the activity-dependent control of synaptic strengths across dendrites*,
PLOS Biol. **17** (2019) e2006223.

[296] A. Lazar, G. Pipa, and J. Triesch, *SORN: a self-organizing recurrent neural network*,
Front. Comput. Neurosci. **3** (2009) 23.

[297] M. A. Triplett, L. Avitan, and G. J. Goodhill, *Emergence of spontaneous assembly activity in developing neural networks without afferent input*, PLOS Comput. Biol. **14** (2018) e1006421.

[298] M. Stimberg, R. Brette, and D. F. M. Goodman,
Brian 2, an intuitive and efficient neural simulator, eLife **8** (2019).

[299] https://github.com/fkalle/drifting-assemblies, Code to reproduce the main results of the article.

[300] H. G. Rey et al., *Single neuron coding of identity in the human hippocampal formation*, Current Biology **30** (2020) 1152.

[301] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, *Fast unfolding of communities in large networks*, J. Stat. Mech. **2008** (2008) P10008.

[302] R. LaPlante et al., *bctpy v0.5.2: Brain Connectivity Toolbox for Python*, 2020, URL: `https://github.com/aestrivex/bctpy`.

[303] E. L. Bienenstock, L. N. Cooper, and P. W. Munro, *Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex*, J. Neurosci. **2** (1982) 32.

[304] R. Quian Quiroga, *Plugging in to human memory: advantages, challenges, and insights from human single-neuron recordings*, Cell **179** (2019) 1015.

[305] A. U. Sugden et al., *Cortical reactivations of recent sensory experiences predict bidirectional network changes during learning*, Nat. Neurosci. **23** (2020) 981.

[306] A. Renart et al., *The asynchronous state in cortical circuits*, Science **327** (2010) 587.

[307] A. R. Chambers and S. Rumpel, *A stable brain from unstable components: Emerging concepts and implications for neural computation*, Neuroscience **357** (2017) 172.

[308] U. Rokni, A. G. Richardson, E. Bizzi, and H. S. Seung, *Motor learning with unstable neural representations*, Neuron **54** (2007) 653.

[309] G. Mongillo, S. Rumpel, and Y. Loewenstein, *Inhibitory connectivity defines the realm of excitatory plasticity*, Nat. Neurosci. **21** (2018) 1463.

[310] L. Susman, N. Brenner, and O. Barak, *Stable memory with unstable synapses*, Nat. Commun. **10** (2019) 4441.

[311] M. Gillett, U. Pereira, and N. Brunel, *Characteristics of sequential activity in networks with temporally asymmetric Hebbian learning*, Proc. Natl. Acad. Sci. U.S.A. **117** (2020) 29948.

[312] R. Ajemian, A. D'Ausilio, H. Moorman, and E. Bizzi, *A theory for how sensorimotor skills are learned and retained in noisy and nonstationary neural circuits*, Proc. Natl. Acad. Sci. U.S.A. **110** (2013) E5078.

[313] D. Kappel, R. Legenstein, S. Habenschuss, M. Hsieh, and W. Maass, *A dynamic connectome supports the emergence of stable computational function of neural circuits through reward-based learning*, eNeuro **5** (2018) e0301.

[314] D. Acker, S. Paradis, and P. Miller, *Stable memory and computation in randomly rewiring neural networks*, J. Neurophysiol. **122** (2019) 66.

[315] M. E. Rule et al., *Stable task information from an unstable neural population*, eLife **9** (2020) e51121.

[316] T. Pietri et al., *The emergence of the spatial structure of tectal spontaneous activity is independent of visual inputs*, Cell Reports **19** (2017) 939.

[317] S. Cheng, *The CRISP theory of hippocampal function in episodic memory*, Front. Neural Circuits **7** (2013) 88.

[318] M. Stern, H. Sompolinsky, and L. F. Abbott,
Dynamics of random neural networks with bistable units, Phys. Rev. E **90** (2014) 062710.

[319] B. Cessac, *Increase in complexity in random neural networks*,
Journal de Physique I **5** (1995) 409.

[320] C. G. Langton, *Computation at the edge of chaos: Phase transitions and emergent computation*,
Physica D **42** (1990) 12.

[321] A. Rivkind and O. Barak, *Local dynamics in trained recurrent neural networks*,
Phys. Rev. Lett. **118** (2017) 258101, ISSN: 1079-7114.

[322] M. Massar and S. Massar, *Mean-field theory of echo state networks*,
Phys. Rev. E **87** (2013) 042809.

[323] R. Engelken, A. Ingrosso, R. Khajeh, S. Goedeke, and L. F. Abbott,
Input correlations impede suppression of chaos and learning in balanced firing-rate networks,
PLOS Comput. Biol. **18** (2022) e1010590, ed. by J. Aljadeff.

[324] C. Keup, T. Kühn, D. Dahmen, and M. Helias,
Transient chaotic dimensionality expansion by recurrent networks,
Phys. Rev. X **11** (2021) 021064.

[325] M. Samuelides and B. Cessac, *Random recurrent neural networks dynamics*,
The European Physical Journal Special Topics **142** (2007) 89.

[326] G. Mongillo and D. J. Amit,
Oscillations and irregular emission in networks of linear spiking neurons,
J. Comput. Neurosci. **11** (2001) 249.

[327] M. Puelma Touzel and F. Wolf, *Statistical mechanics of spike events underlying phase space partitioning and sequence codes in large-scale models of neural circuits*,
Phys. Rev. E **99** (2019) 052402.

[328] J. Jordan et al., *Deterministic networks for probabilistic computing*, Sci. Rep. **9** (2019).

[329] Y. Terada and T. Toyoizumi,
Chaotic neural dynamics facilitate probabilistic computations through sampling,
Proc. Natl. Acad. Sci. U.S.A. **121** (2024).

[330] J. Ritt, *Evaluation of entrainment of a nonlinear neural oscillator to white noise*,
Phys. Rev. E **68** (2003) 041915.

[331] M. J. E. Richardson, *Spike shape and synaptic-amplitude distribution interact to set the high-frequency firing-rate response of neuronal populations*, Phys. Rev. E **98** (2018) 042405.

[332] M. J. E. Richardson, *Linear and nonlinear integrate-and-fire neurons driven by synaptic shot noise with reversal potentials*, Phys. Rev. E **109** (2024) 024407.

[333] R. Engelken, M. Monteforte, and F. Wolf, *Sparse chaos in cortical circuits*, 2024,
arXiv: 2412.21188 [q-bio.NC], URL: https://arxiv.org/abs/2412.21188.

[334] J. Wilting et al.,
Operating in a reverberating regime enables rapid tuning of network states to task requirements,
Front. Syst. Neurosci. **12** (2018).

[335] S. Goedeke, F. Y. Kalle Kossio, and R.-M. Memmesheimer,
"The cluster duration distribution of Hawkes processes,"
International Conference on Mathematical NeuroScience (ICMNS), 2019,
URL: https://web.math.ku.dk/~susanne/ICMNS2019/index.html.

[336] J. Zierenberg, J. Wilting, V. Priesemann, and A. Levina,
Description of spreading dynamics by microscopic network models and macroscopic branching processes can differ due to coalescence, Phys. Rev. E **101** (2020) 022301.

[337] A. van Ooyen and M. Butz-Ostendorf,
"Homeostatic structural plasticity can build critical networks,"
The Functional Role of Critical Dynamics in Neural Systems, Springer, 2019 117,
ISBN: 9783030209650.

[338] Z. Ma, G. G. Turrigiano, R. Wessel, and K. B. Hengen,
Cortical circuit dynamics are homeostatically tuned to criticality in vivo,
Neuron **104** (2019) 655.

[339] R. Zeraati, V. Priesemann, and A. Levina,
Self-organization toward criticality by synaptic plasticity, Front. Phys. **9** (2021).

[340] S. Okujeni and U. Egert,
Structural modularity tunes mesoscale criticality in biological neuronal networks,
J. Neurosci. **43** (2023) 2515.

[341] J. Wilting and V. Priesemann,
25 years of criticality in neuroscience – established results, open controversies, novel concepts,
Curr. Opin. Neurobiol. **58** (2019) 105.

[342] V. Pasquale, P. Massobrio, L. L. Bologna, M. Chiappalone, and S. Martinoia,
Self-organization and neuronal avalanches in networks of dissociated cortical neurons,
Neuroscience **153** (2008) 1354.

[343] D. Dahmen, S. Grün, M. Diesmann, and M. Helias,
Second type of criticality in the brain uncovers rich multiple-neuron dynamics,
Proc. Natl. Acad. Sci. U.S.A. **116** (2019) 13051.

[344] L. N. Driscoll, L. Duncker, and C. D. Harvey, *Representational drift: Emerging theories for continual learning and experimental future directions*,
Curr. Opin. Neurobiol. **76** (2022) 102609.

[345] M. E. Rule and T. O'Leary, *Self-healing codes: How stable neural populations can track continually reconfiguring neural representations*, Proc. Natl. Acad. Sci. **119** (2022).

[346] S. Qin et al., *Coordinated drift of receptive fields in Hebbian/anti-Hebbian network models during noisy representation learning*, Nat. Neurosci. **26** (2023) 339.

[347] N. Geva, D. Deitch, A. Rubin, and Y. Ziv,
Time and experience differentially affect distinct aspects of hippocampal representational drift,
Neuron **111** (2023) 2357.

[348] D. Khatib et al.,
Active experience, not time, determines within-day representational drift in dorsal CA1,
Neuron **111** (2023) 2348.

[349] C. Micou and T. O'Leary,
Representational drift as a window into neural and behavioural plasticity,
Curr. Opin. Neurobiol. **81** (2023) 102746.

[350] M. Khona and I. R. Fiete, *Attractor and integrator networks in the brain*,
Nat. Rev. Neurosci. **23** (2022) 744.

[351] C. Langdon, M. Genkin, and T. A. Engel,
A unifying perspective on neural manifolds and circuits for cognition,
Nat. Rev. Neurosci. **24** (2023) 363.

[352] C. Miehl, S. Onasch, D. Festa, and J. Gjorgjieva,
Formation and computational implications of assemblies in neural circuits,
J. Physiol. **601** (2022) 3071.

[353] Y. Wei and A. A. Koulakov, *Long-term memory stabilized by noise-induced rehearsal*,
J. Neurosci. **34** (2014) 15804.

[354] J. W. Negele and H. Orland, *Quantum Many-Particle Systems*, Perseus Books, 1998.

[355] D. Zhou, Y. Sun, A. V. Rangan, and D. Cai,
Spectrum of Lyapunov exponents of non-smooth dynamical systems of integrate-and-fire type,
J. Comput. Neurosci. **28** (2010) 229.

[356] F. Bizzarri, A. Brambilla, and G. S. Gajani,
Lyapunov exponents computation for hybrid neurons, J. Comput. Neurosci. **35** (2013) 201.

[357] L. Arnold, W. Horsthemke, and R. Lefever,
White and coloured external noise and transition phenomena in nonlinear systems,
Z. Physik B: Cond. Mat. **29** (1978) 367.

[358] J. F. C. Kingman, *Poisson Processes*, Oxford University Press, 1993, ISBN: 9780198536932.

[359] S. Goedeke and M. Diesmann,
The mechanism of synchronization in feed-forward neuronal networks,
New J. Phys. **10** (2008) 015007.

[360] S. Jahnke, R.-M. Memmesheimer, and M. Timme,
Propagating synchrony in feed-forward networks, Front. Comput. Neurosci. **7** (2013) 153.

Acknowledgements

First and foremost, I would like to thank my supervisor Raoul-Martin Memmesheimer. I learned immensely from him, and I am grateful for the many opportunities he made possible for me. I am particularly glad to have been part of his group in Bonn from the beginning. His exceptional mentorship, characterized by both empathy and deep scientific understanding, has been invaluable throughout my doctoral journey. I would also like to thank my second reviewer, Moritz Helias. I fondly recall our collaboration and my early experiences in Jülich, which shaped my path towards theoretical neuroscience and fostered my interest in statistical physics. Further, I would like to thank Lena Funcke and Anton Bovier from the University of Bonn for agreeing to be part of my doctoral committee. I extend my gratitude to Markus Diesmann and Tom Tetzlaff for their invaluable advice and support throughout my scientific journey.

I am grateful to my colleagues Felipe Kalle Kossio, Christian Klos, Paul Manz, and Jannis Schücker for their scientific collaboration on the research presented in this thesis. Working with them was a great experience. For proofreading and providing insightful feedback on various parts of this thesis, I would like to thank Simon Altrogge, Rainer Engelken, Kristine Heiney, Philippa Hennessey, Christian Klos, Tobias Kühn, Christian Leibold, Xiao-Xiong Lin, Natalie Schieferstein, Tom Tetzlaff, and Sacha van Albada.

Furthermore, I would like to thank all current and former colleagues of the Neural Network Dynamics and Computation group at the University of Bonn for the pleasant atmosphere, many entertaining moments, and stimulating journal clubs. Special thanks go to Fabian Pallasdies and Luke Eilers, whose master's theses I had the pleasure to co-supervise. I am especially grateful to my friend Wilhelm Braun, with whom I explored Bonn and shared many fun experiences alongside Christian Klos. I would also like to thank Simone Christian, Stefan Klein, and Walter Witke from the Institute of Genetics for their support and assistance.

I would also like to thank Claudia Bachmann, Inga Blundell, Tobias Kühn, and especially Sacha van Albada from my time in Jülich. I am grateful that our paths crossed there.

Many thanks to the Theoretical Systems Neuroscience group at the Bernstein Center Freiburg, where I have already settled in well and made new friends. I am especially grateful to Christian Leibold for his guidance and support. I look forward to future experiences there.

I would also like to thank my long-time scientific and personal friend Rainer Engelken for his collaboration, exchange of ideas, support, and friendship, which have made so much possible for me. I look forward to our future endeavors.

Finally, I would like to express my deepest gratitude to my parents, my sister, her partner, and my nephews Rasmus, Justus, and Titus. They have always been there for me and have made this journey possible.