Pioneer neurons and stimulus-selective sequential recruitment in unstructured networks of spiking neurons *in silico*: phenomenology and mechanisms

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

genehmigt durch die Fakultät für Naturwissenschaften der
Otto-von-Guericke-Universität Magdeburg

von Dipl.-Phys. Christoph Bauermeister
geb. am 28.04.1985 in Dresden

Gutachter: Prof. Jochen Braun, PhD
Prof. Dr. Stefan Rotter

eingereicht am: 23.04.2021
verteidigt am: 30.05.2022

Zusammenfassung

Ich beschreibe unstrukturierte Zufallsnetzwerke *in silico* von (pulsierenden) Neuronen vom "Integrate-and-Fire-Typ" mit Leckstrom, frequenzabhängigen, leitungsbasierten Synapsen und mit einer neuartigen Art von heterogener Topologie, im superkritischen Regime, in dem sich ausgeprägte Synchronitätsereignisse (genannt *Netzwerkpulse*) mit Phasen von vernachlässigbarer Aktivität abwechseln. In Übereinstimmung mit früheren Arbeiten experimenteller und theoretischer Natur finde ich eine Kohorte von priviligierten Pionierneuronen, die durch ihr frühes Feuern die Netzwerkpulse ankündigen. Abgesehen von der Darlegung makroskopischer Aspekte der Aktivität – sowohl in Bezug auf spontanes als auch evoziertes Verhalten – beschreibe ich in großem Detail verschiedene Aspekte der mikroskopischen Aktivität der Pionierneurone, wie Phänomenologie, Ursachen und Mechanismen, die im frühen Feuern der Pionierneurone involviert sind sowie der erfolgreichen Codierung der Lokalisierung schwacher externer Stimuli durch zeitbasierte neuronale Codierschemata. Letzteres bedeutet, dass der Ursprung schwacher externer Stimuli reliabel durch die Rangfolge der ersten Pulse der Pionierneurone decodiert werden kann. Ratenbasierte Codierschemata und die Rangfolge von Nichtpionieren sind im Gegensatz dazu zu dieser Codierung nicht fähig. Ich stelle auch die Frage, ob Pioniere in nicht-heterogenen Netzwerken auftreten können und berichte, dass Heterogenität den Pioniereffekt wesentlich verstärkt. Ich gebe ferner eine Erklärung für diese starke Abhängigkeit der Phänomenologie von dem Grad der Heterogenität des zugrunde liegenden Netzwerkes. Eine Folge davon ist, dass mein unstrukturiertes heterogenes Netzwerk ein *minimales* Modell für die Effekte darstellt, die ich beschreibe, und es tut dies *ohne* heterogene Pulsschwellen, Hintergrundströme oder andere Heterogenität in den neuronalen Parametern.

Als ein Nebenprodukt dieser Bemühungen beschreibe ich auch eine "Bayessche" computationale Methode, die es möglich macht, starke Synapsen und starke kausale Wechselwirkungen zwischen Paaren von Einheiten in neuronalen Systemen zu detektieren. Die Methode wird mit Hilfe meiner *in-silico*-Netzwerke verifiziert, aber sie ist immer anwendbar, wenn superkritisches Verhalten auftritt und kann daher auch auf geeignete neuronale Kulturen *in vitro* angewandt werden, sogar wenn starke Untererprobung die Messungen limitiert. Dies ist wichtig, denn das Problem, die Struktur neuronaler Systeme aus Aktivitätsmessungen herzuleiten, ist ein schwieriges "Reverse-Engineering-Problem".

Abstract

I describe unstructured random networks of (spiking) leaky integrate-and-fire neurons *in silico* with frequency-dependent conductance-based synapses and with a novel type of heterogeneous topology in the super-critical regime, where pronounced synchronisation events (termed 'network spikes,' NS) alternate with phases of near-silence. In agreement with earlier experimental and modelling studies, I observe a cohort of privileged 'pioneer' neurons which herald NS by their early firing. Apart from describing the macroscopic aspects of the activity in the spontaneous and in the evoked case, I elaborate in considerable detail on several aspects related to the microscopic dynamics of pioneer neurons, such as phenomenology, reasons and mechanisms involved in their early firing and such as the successful encoding of locations of weak external stimuli by means of time-based neuronal coding schemes. The latter means, more specifically, that the origin of weak external stimulation can reliably be decoded by the rank-order of the first spikes of pioneer neurons. In contrast, rate-based coding schemes, or rank-order by non-pioneers, are unable to do so. I also consider whether pioneers can arise in *non*-heterogeneous networks and report that heterogeneity considerably enhances the pioneer-effect. I also give an explanation for this strong dependence of the phenomenology on the degree of heterogeneity in the underlying network. As a consequence, my unstructured heterogeneous network provides a *minimal* model for the effects which are described here, and does so *without* heterogeneous firing thresholds or heterogeneous background currents or other heterogeneity in neuronal parameters.

As a by-product of my work, I also describe a special 'Bayesian-like' computational method which makes it possible to infer strong synapses or strong causal interaction between pairs of units in neural systems. The method is verified by using my *in silico* networks; the method, however, applies whenever super-critical behaviour is observed and may thus be applied to appropriate neuronal cultures *in vitro* – and all this even when strong sub-sampling of neurons is at work. This is important, because inferring structure from activity measurements is a difficult 'reverse-engineering' problem in general.

Contents

Chapter 1

Background, previous work and aims

In this chapter, I first give a very concise presentation of neurons and synapses and of the ordinary differential equations which are typically used for their mathematical description. Additionally, I summarise previous studies which are pertinent to my work. Particular emphasis will be given to the experimental observations that motivated this project. Lastly, I introduce the goals and contributions of the thesis.

1.1 Background

None of the material presented in this background section is original. Useful references for the topics described in this section and references for important topics left out are: Kandel et al. (2000), Bear et al. (2007), Thompson (2000), Dayan et al. (2001), Gerstner et al. (2014), Izhikevich (2007), Tuckwell (2005), Rolls (2008), Rolls and Deco (2010), Rolls (2016), Koch (1999), Burkitt (2006a), Burkitt (2006b), Tsodyks et al. (1998), Tsodyks and Markram (1997), Markram et al. (1998), Risken (1996), Honerkamp (1993), Stratonovich (1967), Van Kampen (1992), Gardiner (2009), Pa-

poulis and Pillai (2002). In the sequel, I will *rarely* mention particular references when I introduce a concept. All that I review (and much more, of course) can be found in these textbooks and papers.

1.1.1 Neurons and neuron models

Around 1900, it was discovered (primarily by Santiago Ramón y Cajal) that nervous tissue is composed of *discrete* entities. These discrete units are specialised cells which are referred to as *nerve cells* or *neurons*. Before 1900, it was widely believed that nervous tissue provides an exception to the rule – first proposed by Jacob Schleiden and Theodor Schwann around 1830 – that all tissue consists of single discrete cells. We now know, however, that it also applies to the brain.

Thus, in the very end, the brain is 'simply' an extremely complex network of neurons in which the neurons communicate with each other at sites called *synapses*. It also contains other cells, however, which are required for the correct functioning of neurons, such as cells which comprise the components of blood vessels and so-called *glia cells*, which are believed to have a supporting and modulatory function. In the sequel, only neurons are relevant.

Neurons are able to generate certain 'signals' by means of which they process and transmit information. These are the *action potentials*. Action potentials are often referred to as *spikes*, and I will follow this custom. In this thesis, a striking instance of self-similarity is considered where an entire network of neurons behaves 'like a single neuron,' thus producing a *network spike*. In order to distinguish these two entirely different types of 'spike,' I will sometimes refer to action potentials (as generated by single neurons) by using the term *'individual'* (*individual spike*). When I talk about *spikes* or *action potentials*, I will *always* mean *individual* spikes. In contrast, when I talk about *network* spikes, I will *always* retain the word 'network.'

An action potential manifests itself by a *stereotypical, very short* and *strong, all-or-none* fluctuation of the so-called *membrane voltage* $V(t)$, which is the (time-dependent) difference $V_{inside}(t) - V_{outside}(t)$ of the electrical potentials inside and outside the neuron. Here one makes use of the idealisation that these two potentials

are well-defined, i.e. independent of the specific choices of positions within these two regions, because the electric conductances within these two regions are (ideally) very high,[1] so that electrical fields are immediately neutralised, as, e.g., within a solid body made of metal.

The *shape* of an action potential is so strongly *localised* in time – and believed to be so *inessential* for neuronal coding – that spikes are mathematically modelled as (Dirac) δ-peaks. A *sequence* of spikes is then described by a *sequence* of δ-peaks, a so-called *Dirac comb* (see below).

To a certain approximation – and by employing a couple of idealisations[2] – the behaviour of the membrane voltage $V(t)$ of neurons *between spikes* can phenomenologically be described by the differential equation

$$\frac{\mathrm{d}V}{\mathrm{d}t}(t) = \frac{E_L - V(t)}{\tau_m} + \frac{R_m I_b}{\tau_m} + \frac{R_m I_{input}(t)}{\tau_m}. \tag{1.1}$$

I first want to elaborate on *this* equation, which does *not* yet include *spikes*. E_L, τ_m, R_m and I_b are parameters, whereas $I_{input}(t)$ is a current which describes the part of the 'input' to the neuron different from the *background current* I_b. E_L is the so-called *leak reversal potential*. It is the value to which the voltage relaxes if the input current I_{input} and the constant background current I_b are switched off. τ_m is the time-scale (*membrane time constant*) according to which the relaxation occurs. As far as R_m is concerned (*membrane resistance*), for my purposes, one may consider it simply as a constant of proportionality which translates a current into a voltage. The background current I_b effectively shifts the leak reversal potential E_L to the *effective leak reversal potential* $E_{L,eff} = E_L + R_m I_b$. Thus, if the input current I_{input} (but not the background current) is switched off, the voltage relaxes to the

[1] The fluids inside and outside the neuron consist of water where large amounts of salts are solved. This is the reason for the high conductibility within these two regions. The neuronal membrane, on the other hand, is *not* permeable to most of the ions which are solved in the intra-cellular and in the extra-cellular fluid. Therefore, the membrane voltage can be non-zero.

[2] With this step from neurons and action potentials to the leaky integrate-and-fire neuron model (which I am up to introduce), I make a very long story very short. In reality, this is a very dramatic simplification which ignores many details, such as the spatial extent of neurons, and the intricacies of the precise dynamics of the large family of different ion channels. However, integrate-and-fire neurons are very useful for minimal network models. See the literature mentioned at the beginning of this section for more details on the modelling of single neurons and their spikes.

effective leak reversal potential.

The differential equation just mentioned can be understood in several ways. In the first sense, it is simply a 'machine' which translates the input I_{input} into an output $V(t)$. And from a more specific point of view, it is an input-output relation which performs a certain type of low-pass filtering on the input signal. I want to mention that the ordinary differential equation (ODE) at hand (essentially) gives rise to Ornstein-Uhlenbeck noise if the signal I_{input} is chosen to be Gaussian white noise. Finally, I want to stress that the equation as it stands so far is entirely linear, in the sense that the space of differences of solutions (for fixed I_{input}) is a vector space.

The ODE at hand is *forgetful*. This means that any initial condition is 'forgotten' (i.e. not relevant) after a time which is much larger than the membrane time constant, see Figs. 1.1 and 1.2 below.

Now I take the generation of spikes into account. This is done by the *ad hoc* specification that a spike is generated whenever the voltage $V(t)$ reaches a *threshold* V_{thr} from below. When this happens, the voltage is *immediately* reset to the *reset voltage* V_{res}. Upon adding these two rules, the equation above gives a coherent and (strongly) idealised description of the behaviour of (real) neurons (see footnote 2, however). This equation together with the two additional specifications mentioned (threshold voltage and reset) is referred to as the *leaky integrate-and-fire (LIF) neuron model* (see Fig. 1.1 and Fig. 1.2 for an illustration[3]). The threshold-and-reset rule originates from the fact that for many purposes the spike generation of neurons can be considered threshold-governed, although in principle already this is an idealisation. The ODE at hand, including the *ad hoc* specifications, can still be considered as an input-output relation, but it is *highly non-linear*[4]. When the LIF model is considered in this way (as an input-output correspondence), the output signal is, however, simply the list of spike times, i.e. the list of times of threshold crossings of $V(t)$. In other words, one is usually not interested in the precise behaviour of

[3] In these two figures, the notion of the $*$-voltage is introduced, which plays a central role in this thesis.

[4] The non-linearity of neurons is one of the primary reasons for the complexity and richness of the information processing capabilities of neurons and neural networks.

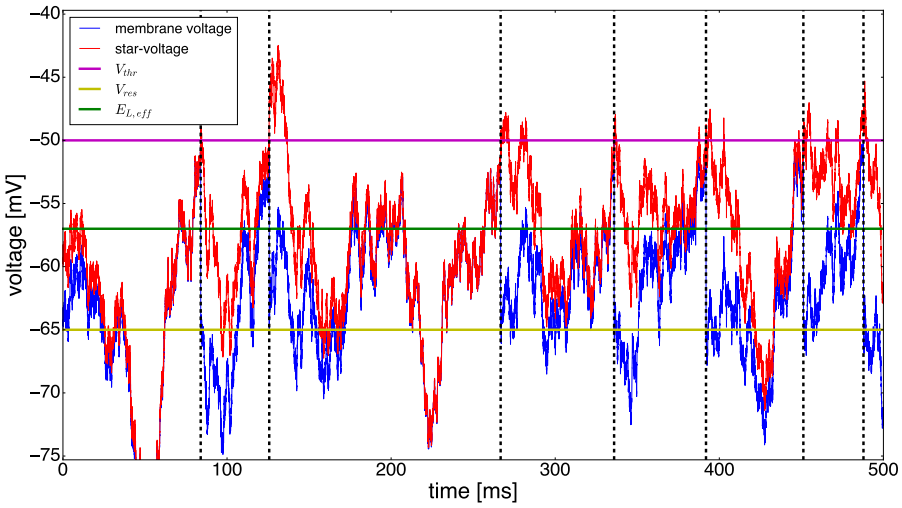

Figure 1.1 | Illustration of the LIF neuron model. The magenta line is the threshold voltage and the yellow line is the reset voltage, which is here identical to the (bare) leak reversal potential. The green line indicates the *effective* reversal potential. The blue curve depicts the *actual* voltage, and the red curve the so-called $*$-*voltage*, which is *not* reset, but otherwise follows the same ODE as the actual voltage. (The $*$-voltage is a *fake* voltage without physiological relevance. However it is very useful for certain theoretical considerations of the behaviour of the model neurons.) This figure is generated with parameters where the neuron is in the *noise-driven* regime, i.e. the (total) input current is highly irregular and its mean is *below* the critical point where spiking occurs even without noise. Therefore, the spikes (marked by vertical, black, dashed lines) occur at *irregular* intervals. Because the ODE at hand is 'forgetful,' and because the average inter-spike interval is large compared to the membrane time constant, both voltages agree when the time since the last spike is relatively large. Parameters: $V_{thr} = -50.0$ mV; $V_{res} = -65.0$ mV; $E_L = -65.0$ mV; $\tau_m = 20.0$ ms; $R_m = 0.4$ GΩ; $I_b = 20.0$ pA. The input current I_{input} was chosen to be Gaussian white noise (with mean zero) with a certain noise strength.

sub-threshold fluctuations of the membrane voltage $V(t)$, but only in the instances where action potentials occur. The list of spike times is usually described by the so-called *neural response function* ρ (a Dirac comb), which is defined as

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

where the sum is over all spikes, and where t_k are the spike times. Here, δ denotes the well-known Dirac δ-function.

I mention at this point that the highly non-linear LIF neuron model can give rise to an effect which is counter-intuitive at first sight but qualitatively easy to

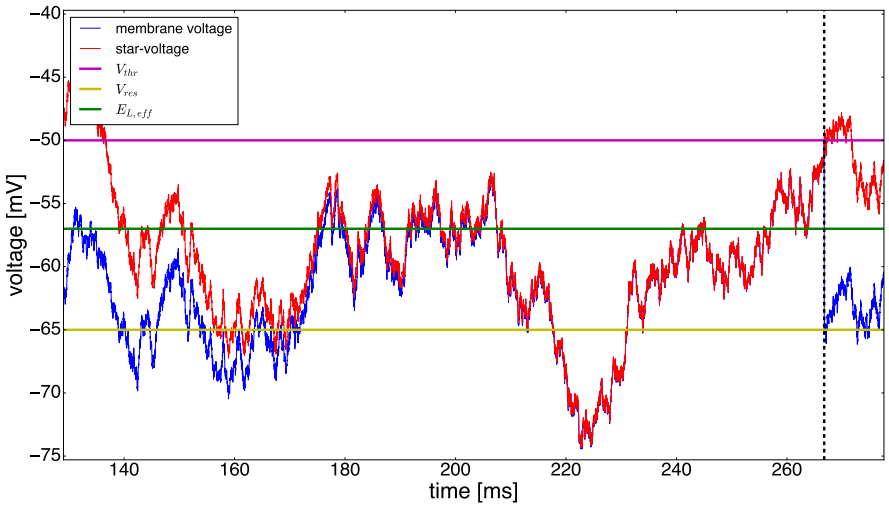

Figure 1.2 | Illustration of the LIF neuron model (continued). Magnification of a region in Fig. 1.1 where a spike occurs (black, dashed, vertical line). Here one can nicely observe the agreement of both voltages when the time since the last spike is large relative to the membrane time constant (20 ms); and one can readily observe that the actual voltage is reset when the spike occurs, whereas the $*$-voltage is not. The agreement of both voltages, when the time since the last spike is large, also demonstrates that initial conditions are 'forgotten' after some time.

understand: coherence resonance, see Lindner (2002). This means that (in a certain regime) the output signal (spike train) can become more *regular* if the noisiness of the input is *increased*.

The LIF model is often extended to include a so-called *absolute refractory period*. This is a short time τ_{ref} after a spike where the voltage is not allowed to evolve, i.e., it is 'frozen' to the reset voltage. This comes from the fact that real neurons are unable to produce a new spike immediately after a previous spike, no matter how large the input is, and in this case the neuron is said to be *refractory*. The period where this applies is, by definition, the *absolute refractory period*[5].

Finally, one may add a mechanism of *adaptation* to the LIF model with refractory period. Adaptation is a mechanism by which the firing rate (the rate of occurrences of spikes) slowly decreases upon constant stimulation, which is often observed in cortical neurons. This means that the neuron becomes less active and thus 'adapts'

[5] In reality, the absolute refractory period is followed by a *relative* refractory period, where it is more difficult, but not impossible, to evoke a spike. In the LIF model, this is to a certain extent incorporated by the reset rule.

to the given stimulus. The LIF model with refractory period and adaptation can quite reliably reproduce[6] actual spiking patterns of individual neurons (referred to as *spike trains*) apart from few missed or additional spikes. In this thesis, I will *exclusively* consider the LIF model *without* adaptation. The refractory period is retained in the simulations. The neglect of adaptation is done for simplicity. The possibility to explain in simple physical terms what I actually observe *in silico* will profit from this neglect, as well as the degree to which the networks are amenable to analytical computation. It is also my aim to find and describe a *minimal* model which gives rise to the class of phenomena I am interested in, and as I will describe, adaptation is not needed if another fatigue mechanism is present, see below. In fact, my *synapses* are *dynamical* and in particular may deplete, and to a certain degree the effects caused by this are qualitatively comparable to the effect of adaptation of neurons (see Figs. 1.3 and 1.4 below). Adaptation (of neurons) and depletion (of synapses, see below) are collectively referred to as *fatigue mechanisms*.

The fact that here I only describe the LIF model (with refractory period and without adaptation) also includes the specification that the LIF neuron model with refractory period (and without adaptation) is the only neuron model used in those network simulations whose phenomenology is described in this thesis.

By fitting the LIF model to the behaviour of *real* neurons, one observes that the physiological range of some of the parameters is typically as described below.

Physiological range of parameters:

- $\tau_m \sim 20$ ms;
- $R \sim \mathrm{G}\Omega$;
- $E_L \sim -65$ mV;
- $V_{thr} \sim -50$ mV;

[6] This applies in experiments where – in the first phase – the model neuron is fitted to a real neuron by stimulating the real neuron with a certain (known) injected input current I_{input}, and where – in the second phase – one attempts to predict (using the model neuron) actual spikes in the real neuron by using another certain (known) input current and the parameters found in the first phase. In contrast, spike trains of neurons which are embedded in real neural networks, and which are *not artificially* stimulated, can usually not be predicted and require a probabilistic description, because the (effective) input currents are not known *a priori*.

- $V_{res} \sim -65$ mV;
- $\tau_{ref} \sim 1$ ms.

1.1.2 Synapses and synapse models

Different neurons communicate with each other by means of *synapses*. This communication is spatially and temporally *directed*: The transfer of information is always from a *pre-synaptic* neuron towards a *post-synaptic* neuron. The biological realisation of this process is chemical[7]: When an all-or-none spike is triggered in the pre-synaptic neuron, and when this spike has propagated to the *synaptic terminals*, a chemical substance referred to as *neurotransmitter*[8] is released by the pre-synaptic side of the synapse, then diffuses through the *synaptic cleft* (which isolates the two neurons from each other), and then causes a small regular voltage fluctuation (the *post-synaptic potential, PSP*) in the post-synaptic neuron. When many spikes – from many pre-synaptic neurons – occur within a sufficiently small time interval, these fluctuations may add up so that the post-synaptic neuron also reaches threshold and by itself produces a spike. Thus, the post-synaptic neuron is essentially a (highly non-linear) 'machine' which produces an all-or-none signal (a spike) when the total input – caused by pre-synaptic neurons – is large enough.

Every neuron produces only PSPs which *decrease* (*increase*[9], respectively) the distances to threshold of the voltages of *all* post-synaptic neurons (Dale's law). They are referred to as *excitatory* (*inhibitory*, respectively) neurons. Essentially, the PSP caused by *excitatory* neurons (*excitatory* PSP, *EPSP*) comprises a *positive* deviation of the membrane voltage $V(t)$ (of the post-synaptic neuron) from its previous level, and the PSP caused by *inhibitory* neurons (*inhibitory* PSP, *IPSP*) manifests itself by a *negative* deviation of the membrane voltage of the post-synaptic neuron from its previous level.

[7] I ignore electrical synapses.

[8] There are many different known neurotransmitters and receptors for these. The latter are proteins, embedded in the post-synaptic membrane, which detect neurotransmitter molecules in a 'lock-and-key' way. I will not describe the precise chemical names of neurotransmitters and receptors, not to mention the agonists and antagonists of the receptors.

[9] Here I slightly idealise matters by ignoring, e.g., the effect called *shunting inhibition*, see below.

As far as the modelling is concerned, two conceptually different manifestations of PSPs are considered. PSPs can be initiated by directly prescribing a certain input current component I_{syn} which effectively stimulates the post-synaptic neuron and adds (linearly) to any other input currents which may be present. Model synapses which make use of this way of initiating the PSP are referred to as *current synapses*. The other, more 'physiological,' but also more 'complicated' possibility found in the modelling literature prescribes the post-synaptic *conductance* changes (which add linearly to any other conductance changes – w.r.t to the same type of ion – which may be present) instead of *current* changes. The post-synaptic current (which induces the PSP) caused by pre-synaptic spikes is then determined by these changing (synaptic) conductances and the running voltage by (essentially) Ohm's law, so that one could say that w.r.t. conductance synapses there are two computational 'steps' instead of one. Conductance synapses could be dubbed voltage-*dependent* and current synapses voltage-*independent*. From a different point of view, one may think of *current* synapses (as described so far) as being *linear* and of conductance synapses as being *non*-linear, because in the latter case the synaptic currents which cause the PSP do not add up linearly in general. This is also the reason why an analytical treatment of conductance synapses is more involved than consideration of current synapses. *Shunting inhibition*, the effect that *strong inhibitory input* to a neuron does *not* directly alter the membrane voltage, but effectively *makes the co-active excitatory input in-effective*, can only occur with *conductance* synapses. Thus the more 'realistic' conductance synapses in principle show a richer phenomenology. These are the reasons why I used them in spite of my aim to provide a minimal model for the effects I am interested in.

Current synapses: mathematical modelling

Assume that i is a certain neuron within a network of LIF neurons, such that these neurons are connected to each other by *current* synapses. Then the membrane

voltage of neuron i (between spikes) behaves as

$$\frac{\mathrm{d}V_i}{\mathrm{d}t}(t) = \frac{E_{L,i} - V_i(t)}{\tau_{m,i}} + \frac{R_{m,i} I_{b,i}}{\tau_{m,i}} + \frac{R_{m,i} I_{syn,exc,i}(t)}{\tau_{m,i}} + \frac{R_{m,i} I_{syn,inh,i}(t)}{\tau_{m,i}}, \tag{1.3}$$

assuming that the *synaptic currents* $I_{syn,exc,i}$ (*excitatory* synaptic current) and $I_{syn,inh,i}$ (*inhibitory* synaptic current) are the *only* components of the input current[10] I_{input}. The synaptic currents are often modelled by letting them evolve according to the ODEs

$$\frac{\mathrm{d}I_{syn,exc,i}}{\mathrm{d}t}(t) = -\frac{I_{syn,exc,i}(t)}{\tau_{syn,exc,i}} + \sum_{j \text{ exc.}} A_{ij}\rho_j(t) \tag{1.4}$$

and

$$\frac{\mathrm{d}I_{syn,i}}{\mathrm{d}t}(t) = -\frac{I_{syn,inh,i}(t)}{\tau_{syn,inh,i}} + \sum_{j \text{ inh.}} A_{ij}\rho_j(t), \tag{1.5}$$

where in the first equation the sum is over all excitatory neurons and in the second equation over all inhibitory neurons in the network. $\tau_{syn,exc,i}$ and $\tau_{syn,inh,i}$ are time constants which may depend on the neuron i. They are called the *excitatory* (*inhibitory*, respectively) *synaptic time constant* of neuron i. A_{ij} is a constant depending on the pair (i,j) with physical dimension of Ampère. It is referred to as the *synaptic weight* of the synapse $j \to i$. ρ_j is the neural response function of neuron j. In model networks like the one which I describe here, if j and i are neurons such that A_{ij} is non-zero, then neuron j is said to be *pre-synaptic* to neuron i and neuron i is said to be *post-synaptic* to neuron j[11]. Obviously, in the above equations, it would be sufficient to sum over all neurons j which are pre-synaptic to neuron i.

Equation 1.4 just says that the excitatory synaptic current of neuron i is elevated by the amplitude A_{ij} if a pre-synaptic spike arrives which originates from the excitatory neuron j. Between spikes, the excitatory synaptic current relaxes to zero with the time-constant $\tau_{syn,exc,i}$. An analogous statement can be made for the inhibitory synaptic current. When j is a pre-synaptic neuron of neuron i, the weight A_{ij} will be positive if neuron j is excitatory, and negative if neuron j is inhibitory.

[10] If, in addition, a certain constant or variable current I_e were artificially *injected* into neuron i, the input current would be $I_{input} = I_e + I_{syn,exc} + I_{syn,inh}$.

[11] An analogous terminology applies in networks of model neurons connected by *conductance* synapses, see below.

Thus, in model networks of neurons connected by *current* synapses, the *signs* of the out-going weights of neuron j (either all non-negative or all non-positive according to Dale's law) determine whether the neuron is excitatory or inhibitory. In my simulations, however, I use *conductance* synapses, which I describe now.

Conductance synapses: mathematical modelling

Now let i be a neuron within a network of LIF neurons connected by *conductance* synapses. Then the membrane voltage of neuron i behaves (between spikes) as (assuming again that no current is artificially injected)

$$\frac{\mathrm{d}V_i}{\mathrm{d}t}(t) = \frac{E_{L,i} - V_i(t)}{\tau_{m,i}} + \frac{R_{m,i} I_{b,i}}{\tau_{m,i}} + \frac{R_{m,i} I_{syn,i}(t)}{\tau_{m,i}}, \tag{1.6}$$

where

$$I_{syn,i}(t) = g_{exc,i}(t)(E_{exc,i} - V_i(t)) + g_{inh,i}(t)(E_{inh,i} - V_i(t)). \tag{1.7}$$

Here, $g_{exc,i}$ is the *excitatory synaptic conductance* of neuron i an $g_{inh,i}$ is the *inhibitory synaptic conductance* of neuron i. These quantities are always non-negative. The idea now is that the *excitatory synaptic reversal potential* $E_{exc,i}$ is *above* threshold, and that an incoming spike originating from an excitatory pre-synaptic neuron j increases $g_{exc,i}$ by a certain (positive) amount Δg_{ij}. Then the change of the first term in the right hand side of equation 1.7 will be positive, and hence neuron i is excited towards threshold. In contrast, the *inhibitory synaptic reversal potential* $E_{inh,i}$ is typically close to the reset voltage – or even more negative; if a spike originating from the *inhibitory* pre-synaptic neuron j arrives at neuron i, the inhibitory synaptic conductance $g_{inh,i}$ is (again) *increased* by the (positive) amount Δg_{ij}. Because the membrane voltage is typically above the reset value, this means that the change of the second term on the right hand side of equation 1.7 will be negative in general, so that neuron i is likely to be inhibited. The Δg_{ij} are referred to as the *weights* in networks of neurons connected by conductance synapses. Note that in networks of neurons connected by *current* synapses, the weights have dimension of *Ampère* and

are positive or negative (or zero). In contrast, in networks of neurons connected by *conductance* synapses, the weights have dimension of *Siemens* and are always *positive* (or zero).

In networks of (non-dynamical) conductance synapses, $g_{exc,i}$ and $g_{inh,i}$ evolve typically according to the ODEs

$$\frac{\mathrm{d}g_{exc,i}}{\mathrm{d}t}(t) = -\frac{g_{exc,i}(t)}{\tau_{syn,exc,i}} + \sum_{j \text{ exc.}} \Delta g_{ij} \rho_j(t) \tag{1.8}$$

and

$$\frac{\mathrm{d}g_{inh,i}}{\mathrm{d}t}(t) = -\frac{g_{inh,i}(t)}{\tau_{syn,inh,i}} + \sum_{j \text{ inh.}} \Delta g_{ij} \rho_j(t), \tag{1.9}$$

where in the first equation the sum is over all excitatory neurons and in the second equation over all inhibitory neurons. $\tau_{syn,exc,i}$ and $\tau_{syn,inh,i}$ are again the synaptic time constants. The above equations just say that a spike originating from a pre-synaptic excitatory neuron j elevates the excitatory synaptic conductance of neuron i by the (positive) amount Δg_{ij}, and likewise a spike originating from an inhibitory pre-synaptic neuron j elevates the inhibitory synaptic conductance of neuron i by the (positive) amount Δg_{ij}. Between spikes, the synaptic conductances relax to zero with respective time constants.

I want to emphasise that all that I have said so far applies when the synapses are *not* dynamical, i.e. *static*. I shall need the extension from (static) conductance synapses to dynamical conductance synapses, and this is the topic of the next subsection.

Dynamical synapses: the quantal model

Synapses are here made *dynamical* by using a model which was first proposed by Tsodyks, Pawelzik and Markram (Tsodyks et al. (1998)). The model assumes that there exist (for each synapse) four time-dependent, non-dimensional quantities $R(t)$, $E(t)$, $I(t)$, and $u(t)$ (with values between zero and unity) which evolve

according to the following equations:

$$\frac{\mathrm{d}R}{\mathrm{d}t}(t) = \frac{I(t)}{\tau_{rec}} - u(t+\epsilon)R(t-\epsilon)\rho(t); \tag{1.10}$$

$$\frac{\mathrm{d}E}{\mathrm{d}t}(t) = -\frac{E(t)}{\tau_I} + u(t+\epsilon)R(t-\epsilon)\rho(t); \tag{1.11}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t}(t) = \frac{E(t)}{\tau_I} - \frac{I(t)}{\tau_{rec}}; \tag{1.12}$$

$$\frac{\mathrm{d}u}{\mathrm{d}t}(t) = \frac{-u(t)}{\tau_{facil}} + U(1-u(t-\epsilon))\rho(t); \tag{1.13}$$

also, the relation

$$R(t)+E(t)+I(t)=1 \tag{1.14}$$

holds. Here, $\rho(t)$ is the neural response function of the neuron which is *pre*-synaptic to the synapse which is being described. τ_{rec}, τ_I, and τ_{facil} are time constants which are referred to as the *recovery time constant*, the *inactivation time constant*, and the *facilitation time constant*, respectively. U is a non-dimensional constant between zero and one. $u(t+\epsilon)$ is just a suggestive notation for the one-sided limit $\lim_{\epsilon \downarrow 0} u(t+\epsilon)$ and similarly for the other expressions of the same type in the above equations. My notation should suggest adding (subtracting, respectively) an infinitesimally small quantity to the argument t. This is needed because the quantities described in general have jumps at the instances where they are evaluated (due to the Dirac functions which appear in the neural response function).

The above equations comprise a *phenomenological* model for the depletion and recovery of synaptic resources. I do not want to elaborate on possible biophysical interpretations of these equations, but I shall follow the simplest (and very suggestive) way of understanding these equations, which says that $R(t)$ is the fraction of neurotransmitter which is ready for release (at time t), that $E(t)$ is the fraction of neurotransmitter in the effective state (at time t), i.e. docking at post-synaptic receptors, and that $I(t)$ is the fraction of neurotransmitter in the inactive state, i.e. non-docking and distributed in the synaptic cleft (again at time t).

Equation 1.14 then says that every molecule of neurotransmitter must be in one of the three states and that it cannot be in more than one state at a time. The

quantity $u(t+\epsilon)$ is the fraction of *ready-to-release* neurotransmitter which becomes active when a pre-synaptic spike arrives at the synapse at time t. This implies that the fraction of the *entire* neurotransmitter which becomes active upon arrival of a pre-synaptic spike is the product $u(t+\epsilon)R(t-\epsilon)$. Here, R is evaluated as a left-sided limit because of causality. It is simplest to explain this assuming that the entire neurotransmitter in the recovered state becomes segregated into the synaptic cleft. In this case, $R(t+\epsilon)$ would be zero, but the fraction of neurotransmitter which changed in the active state is of course non-zero in general. Therefore multiplying $u(t+\epsilon)$ with $R(t+\epsilon)$ would not make sense. It must be multiplied with $R(t-\epsilon)$.

We can now understand equation 1.11: It says that resources in the active state become inactivated (and hence leave the active state) with a rate $\frac{1}{\tau_I}$, and that – upon arrival of a pre-synaptic spike – E is elevated by the fraction of resources which are released into the synaptic cleft. Equation 1.10 then says that R is decreased by the same amount – if a pre-synaptic spike arrives –, and that the neurotransmitter in the inactive state recovers with a rate of $\frac{1}{\tau_{rec}}$. Equation 1.12 is the unique differential equation for I which is implied by the other equations.

Finally, the equation for $u(t)$ says that u relaxes to zero with a time constant of τ_{facil}, and that it is elevated by $U(1-u(t-\epsilon))$ if a pre-synaptic spike arrives at the synapse. U in principle describes how strongly it is elevated, and the actual (more complicated) form of the size of the jump, $U(1-u(t-\epsilon))$, has this particular form in order to make sure that u does not become larger than unity. The entire set of equations stated above phenomenologically describes the (short-term) dynamics of synaptic resources (depression and facilitation), and it has been shown that they can well reproduce observations of the dynamics of synapses in the neocortex. The model with $\tau_{facil}=0$ ($\tau_{facil}>0$) is illustrated in Fig. 1.3 (Fig. 1.4, respectively). In the simulations described in this thesis, I modelled – following Tsodyks and colleagues – synapses where the post-synaptic neuron is excitatory with $\tau_{facil}=0$ and synapses where the post-synaptic neuron is inhibitory with $\tau_{facil}\neq 0$ (see appendix).

Finally, the connection with synaptic *conductances* is provided by the equations

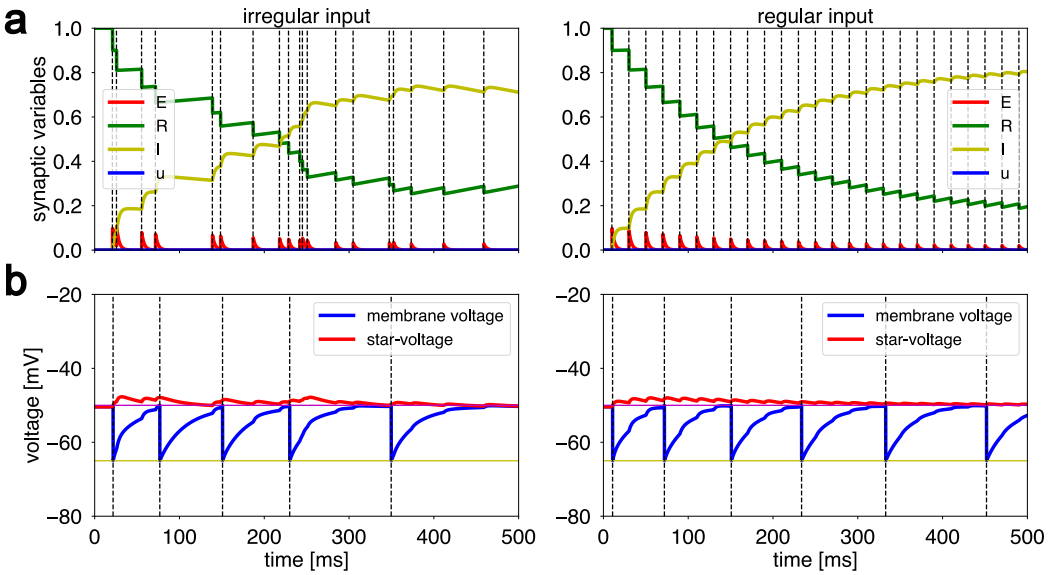

Figure 1.3 | Illustration of the quantal model without facilitation. I considered two neurons 1 and 2 connected by a synapse S from 1 to 2. Neuron 1 (the 'pre-synaptic' neuron) fires according to a homogeneous Poisson process (left figures, entitled 'irregular input') or according to a completely regular spike train (right figures, entitled 'regular input'); rate of neuron 1 is 50 Hz in both cases. The vertical black, dashed lines in the two top figures show the spike times of neuron 1. The synaptic resources of synapse S are shown in colour in the top figures. The bottom figures show the membrane voltages (in colour) and the spike instances (vertical black, dashed lines) of neuron 2. One observes that the inter-spike intervals (the time intervals between consecutive spikes) of neuron 2 slightly increase with time. This shows that the depletion of synaptic resources can have a similar effect as the adaptation of neurons when the parameters are chosen appropriately. Parameters of neuron 2: $\tau_m = 30$ ms, $R_m = 40$ MΩ, $E_L = -70$ mV, $V_{thr} = -50$ mV, $V_{res} = -65$ mV, $E_{exc} = 0$ mV, $I_b = (V_{thr} - E_L) \cdot 0.98/R_m$. Parameters of synapse S: $U = 0.1$, $\tau_{rec} = 1$ s, $\tau_{facil} = 0$ s, $\tau_I = 3$ ms, $\Delta g = 100$ nS.

$$g_{exc,i}(t) = \sum_{j \text{ exc.}} \Delta g_{ij} E_{ij}(t) \tag{1.15}$$

$$g_{inh,i}(t) = \sum_{j \text{ inh.}} \Delta g_{ij} E_{ij}(t), \tag{1.16}$$

which just translate the *fraction* of resources in the active state in the amount of *synaptic conductance*, with the specification that the conductance changes caused by the pre-synaptic neurons add up linearly. These two equations replace equations 1.8 and 1.9 when synapses are dynamical. In this case, the inactivation time constants replace the synaptic time constants and play their role.

Networks with dynamical synapses, i.e. with synapses showing the type of short-term dynamics described here, can show a rich phenomenology and many effects

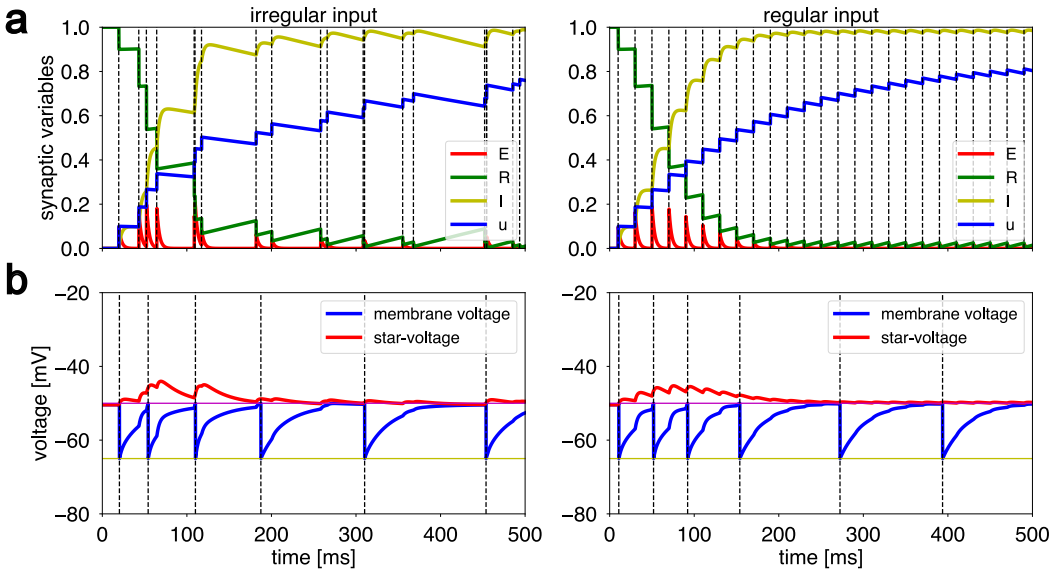

Figure 1.4 | Illustration of the quantal model (continued). Same as Fig. 1.3, but with facilitation switched on (τ_{facil} = 1 s).

not possible with static synapses. For example, when synapses are dynamical, the stationary deviation of the post-synaptic potential from its resting level *saturates* as the pre-synaptic firing rate increases, essentially because the amount of synaptic resources is *finite*. Thus, in this regime, the rate of the post-synaptic neuron is essentially independent of the firing pattern of the pre-synaptic neuron and rate coding becomes impossible. Since rate-coding is perfectly prominent when the pre-synaptic firing rate is small, dynamic synapses are sometimes referred to as *frequency-dependent* synapses. When the pre-synaptic firing rate is large, temporal features of the post-synaptic spike train (instead of its rate) encode the information which is provided by the spiking activity of the pre-synaptic neuron.

Note that, in this thesis, I only include *short-term dynamics* of synapses, but not *long-term effects* such as learning. I do not want the network to be able to 'learn' to be in the dynamical state which I am interested in, because one has to study this state *in isolation* first. Long-term dynamics of synapses, such as *learning*, would be contrary to having a *minimal* model for the effects I am primarily interested in.

1.2 Previous work

In this section, I shall elaborate on research results which have already been obtained by other researches and are relevant for my work. I shall do so by considering several key words of topics in neural science to which my thesis contributes. After each key word, results in the respective field are described. Along the lines, I will establish how the results which I review are related to the *present* work.

1.2.1 Neuronal cultures: important experimental results

Neuronal cultures *in vitro* provide important model systems which can be observed with *multi-electrode arrays*. Also, these preparations can be stimulated electrically and pharmacologically. An important source for general aspects of cultured networks *in vitro*, in particular about how these cultures are obtained, how their activity develops in time, etc., is Marom and Shahaf (2002). Another useful source in this respect is Morin et al. (2005). In addition to this, the following four publications (see below) are very important for the motivation of the present work. Note that the principal mode of operation of neuronal cultures is the *network spike (NS)*, which is a recurring strong all-or-none fluctuation in the population activity (essentially the sum of the activities of all neurons) of the culture, which in turn is interrupted by phases of near-silent, irregular activity (see Fig. 2.2 for a computer simulation of an artificial neural network showing this type of activity).

Eytan and Marom (2006)

The authors ask where the *shape* and *time-scale* ($\mathcal{O}$(100 ms)) of NS comes from and argue that assembly activation is a *threshold-governed* phenomenon. They further conjecture that the effective topology of the neural cultures which they study is *scale-free* (see below), since this would explain why the time-scale of NS is *independent* of the size of the network. They further report that NS are completely *abolished* by blockade of excitation, so that NS are not due to so-called *self-pacing* neurons. They also show that NS can be evoked by *stimulation*, a fact which is very important

for the present work. They report furthermore that the initial phase of NS follows an *exponential growth*, and that the rate of recruitment is *independent* of the size of NS. Next, they demonstrate that the decline of activity *after* the peak of the NS occurs even if inhibition is *blocked*, so the decline is presumably caused by fatigue mechanisms at the *cellular* level. Moreover, they argue that the initial recruitment of neurons is *non-random and hierarchical*. There exists a cohort of *privileged* neurons which reliably increase their firing tens of milliseconds *before* the peak of the NS, and this phenomenon is *stable* over time. This fact is very important to the present work, since this is one of the first experimental evidences for the existence of *'pioneer neurons,'* which are the topic of this thesis. Eytan and colleagues also observe that the same neurons – pioneer neurons – are activated early, independent of whether a NS is *evoked* or *spontaneous*.

The authors also observe that in *coupled* networks of cultured neurons, the temporal relation between 'upstream' NS and evoked 'downstream' NS can be tuned as a function of the *identity* of neurons which are coupled, and argue that this might be the basis for *fast* information processing.

In summary, Eytan and Marom establish that NS are all-or-nothing, threshold-governed events which are terminated mainly by synaptic depression, and that the order of recruitment of individual neurons is non-random and sequential. As already mentioned, this paper comprises the experimental impetus which initiated the current project, which is aimed (among other things) at the minimal modelling of the phenomenon of pioneer neurons.

Shahaf et al. (2008)

This paper provides an extension of the ideas presented in the previously described publication. The authors ask whether neuronal coding by *recruitment order* (i.e. spike order) is applicable for stimuli which are not ordered temporally, in complex recurrent neural networks. They also ask, if the temporal code *is* applicable, how it deals with trial-to-trial variations (they also give evidence for the latter) in the spike times of individual neurons ('time-warping'). They argue that the response of the

network to stimulation is such that directly activated neurons ('receptive sheath') is very reliable (which is somewhat trivial), but they move on to the next step in processing: the response of neurons beyond the receptive sheath. The times of first-spikes of these neurons are highly variable. However, they demonstrate that the *rank-order* (the order of the recruitment of the early neurons which are sampled) is stimulus-site specific. They do this by comparing strings ('order strings') which describe the order of recruitment of individual neurons, so that 'FDA' would mean: first neuron F, then neuron D, then neuron A. In doing so, they compare (using the so-called Levenshtein edit distance) order strings of the privileged neurons during evoked NS which are evoked by stimulation at different sites (or loci). This comparison shows what they want to demonstrate, i.e. that recruitment order is stimulus-site specific, in spite of variability in the *absolute* spike times.

In summary, they provide experimental evidence for rank-order-based coding to be applicable in complex recurrent networks of cultured neurons.

This paper is also very important to the present work, since I am interested (among other things) in verifying these results (i.e., rank-order based coding) *in silico* (using similar methods), where no experimental restrictions limit the spatial or temporal resolution.

Kermany et al. (2010)

The authors consider experiments where a cultured network is stimulated in various ways. They model different stimuli by electrically stimulating the networks at different sites or sources. They ask which features in the activity, following stimulation, is informative and thus encodes the stimulation site. They use a support vector machine (SVM) to estimate the information content of different features of the neural activity following stimulation. They find that temporal coding schemes, in particular rank-order based coding, outperform coding schemes based on firing rates in spatial classification tasks.

This paper is very important to the present work, since I follow the special technique used to estimate the information content. In fact, I perform the same

experiment with my *in silico* network, where spatial and temporal resolution are not limited. In this way, I can confirm the results obtained by Kermany and colleagues.

Eckmann et al. (2008)

This is an extensive experimental study of the phenomenon of pioneer neurons. Results are based on long-term measurements from rat cortical neuronal cultures. The authors identify 'precursor events' of network bursts, consistent with the idea of 'pioneer neurons' as heralds of network events. More specifically, Eckmann and colleagues observe that the 'pre-bursts' are initiated by a single 'leader neuron' and that the identity of this leader is stable over long time-scales. The authors also observe that the identity of the leaders is informative of the 'identity' of the network bursts, which they assess in terms of the 'signature' of network bursts (the number of spikes per neuron during the network burst). Eckmann and colleagues conclude: "[...] the leaders play a role in the development of the bursts and [we] conjecture that they are part of an underlying sub-network that is excited first and then acts as a nucleation center for the burst."

In summary, the authors give compelling experimental evidence for the existence and stability of leader neurons.

I conclude this short description of relevant experimental results – regarding neuronal cultures – with the remark that, despite of the simplifications in *in vitro* networks compared with *in vivo* networks, neuronal cultures are still *tremendously* complex systems, with effects on many spatial scales and many timescales (see, e.g., Haroush and Marom (2015) and Haroush and Marom (2014)).

1.2.2 Spiking neural networks and mean-field theory

Artificial spiking neural networks (SNNs) and in particular random SNNs have gained much attention in theoretical neuroscience. Two very different and readable reviews of the topic are Ponulak and Kasinski (2011) and Paugam-Moisy and Bohte (2012). Randomly connected SNNs have contributed considerably to the understanding of neural function. They provide generic models for experimentally

observed dynamics in cortical activity associated with phenomena such as short-term memory, attentional biasing, or decision-making (Rolls (2008); Rolls and Deco (2010); Rolls (2016)). In addition, SNNs deepen our understanding of such phenomena because their activity dynamics can often be described analytically in terms of a mean-field theory (Feng (2003); Gerstner et al. (2014); Rolls and Deco (2010); Amit and Tsodyks (1991a,b); Amit et al. (1997); Amit and Brunel (1997a,b); Brunel (2000a); Brunel and Wang (2001); Mattia and Del Giudice (2002)). In this thesis, I do *not* contribute to this endeavour, i.e. to mean-field theory. For a mean-field description of spiking networks of *adapting* neurons, see Guido Gigante's thesis (Gigante (2006)). Aspects of the modelling of *neuronal cultures* with SNNs are the subject of the next keyword.

1.2.3 The modelling of cultured neuronal networks

Papers on modelling of the activity of cultured networks can be divided into what *aspects* precisely they attempt to model, and what kind of *model* is used to reach this goal. The aspects which are modelled reach from the existence of network spikes via the profile of network spikes and inter-network-spike-interval (INSI) statistics to the modelling of the existence of privileged neurons ('pioneer neurons,' 'early-to-fire neurons,' 'leader neurons') which herald NS. Previous attempts to model the *privileged* neurons are discussed in a *separate* keyword section, since the privileged neurons are so central to this thesis.

The first serious modelling attempt which could explain NS (as observed *in vitro*) is the paper Tsodyks et al. (2000). The authors use networks of LIF neurons connected by frequency-dependent synapses. They show that neurons in this network discharge (during NS) in a specific order depending on the average firing rate. This motivates them to sort neurons by their average firing rate in many figures, and I am following this custom in this thesis. In Tsodyks's model, NS terminate due to depression, as is observed experimentally by Eytan and colleagues (see above). Also, Tsodyks and colleagues show that NS are completely abolished if certain neurons with intermediate firing rates are made ineffective. These are essentially the

neurons which I refer to as pioneers, and I will repeat and extend this experiment by Tsodyks and colleagues. They write: "We explain this result by the fact that these neurons not only have low firing rates, and therefore effectively strong excitatory synapses, but are also close enough to threshold to trigger the avalanche of the firing activity leading to the crescendo PB." With 'PB,' they mean 'population burst,' i.e. NS. This single sentence by Tsodyks and colleagues summarises much of the work which I offer here. However, I considerably extend the seminal analysis by Tsodyks and colleagues in that I describe much of the pioneer-phenomenology in greater detail. Tsodyks and colleagues also compute cross-correlations between all individual neurons and the peak of the population burst and plot the maximum of this cross-correlation for each neuron, where the neurons are sorted by discharge rate. In this way, they identify the privileged neurons. In this thesis, I will extend this plot and in this way I give useful additional information. In summary, Tsodyks and colleagues show that 'population bursts' and pioneer neurons arise naturally in heterogeneous networks with LIF neurons and dynamics synapses. This opens the door for an *extensive* analysis of the phenomenon *in silico*. I hope to offer this analysis.

Loebel and Tsodyks (2002) analyse the computational ability of Tsodyks's model by giving a mean-field description of the model. They observe that the depression considerably extends the computational abilities of the network. They are furthermore able to predict the response of the network to various inputs and so gain a deeper understanding of the mechanisms of synchronisation and the computations such a network can perform on its inputs.

Gritsun and colleagues model inter-event-interval statistics (Gritsun et al. (2011)) and burst profiles (Gritsun et al. (2010)). Persi and colleagues attempt to model both (Persi et al. (2004a), Persi et al. (2004b)). Masquelier and Deco (2013) make an attempt to study the laws governing inter-event-interval statistics using *both* fatigue mechanisms (depression and adaptation) and facilitation.

Vladimirski et al. (2008) also model cultures with LIF neurons and depressing synapses. They observe that heterogeneity improves the robustness of the episodic

behaviour. In other words, they show that heterogeneity (of depression) renders the network behaviour robust to variations in excitability.

Wiedemann and Lüthi (2003) observe that "subgroups of easily recruitable neurons serve as amplifiers [...] thereby initiating a cascade-like recruitment of neurons."

1.2.4 Self-organised criticality

Self-organised criticality (SOC) is a phenomenon in which certain complex systems self-organise into a state similar to thermodynamic criticality, however without the need of fine-tuning. This phenomenon is first described by Bak, Tang and Wiesenfeld in their 1988 paper (Bak et al., 1988). Typical examples for critical systems are sand piles, superconductors, the earth (here the 'excitations' are earthquakes), and forests (here the 'excitations' are forest fires). Systems in the state of SOC are characterised among other things by avalanche sizes (the sizes of the excitations of the system) and durations which are distributed by a power-law, so that there is no typical scale in the size and the duration of the excitations. Furthermore, the power spectrum is often of the form $1/f$, so that the small-frequency limit gives rise to long correlations in time. More information can be found in Jensen (1998) and Bak and Chen (1991). I keep the description of this phenomenon rather short, because SOC is only *indirectly* relevant for this thesis.

Cortical slices and the brain are also prominent physical systems which have been linked to SOC. As is visible from the references below, neural systems can operate in a *sub-critical*, a *critical* or a *super-critical* regime (depending heavily on the balance between excitation and inhibition). When making this classification, it emerges that the critical systems are *at the edge* between stochastic and ordered activity. Evidence for critical behaviour vs. sub-critical or super-critical behaviour in such systems (either real neural systems or model systems) is Scarpetta and de Candia (2014), Chialvo (2010), Rubinov et al. (2011), Wang et al. (2011), Massobrio et al. (2015), Pasquale et al. (2008), Beggs and Plenz (2003). Evidence for slightly sub-critical behaviour *in vivo* is also given in Priesemann et al. (2014). Poil et al.

(2012) show how critical dynamics of avalanches and oscillations arises *jointly* in networks with balanced excitation and inhibition. SOC in adaptive neural networks is also studied in Christian Meisel's thesis (Meisel, 2009).

The system which I studied is *super-critical*. My 'avalanches,' the NS, are *far away* from being distributed in a scale-free way. Due to the evidence for either SOC, or slightly sub-critical behaviour, in neural systems, it would be interesting to study privileged neurons and rank-order based coding in such systems. I will come back to this point in the discussion.

1.2.5 Privileged leader neurons: modelling studies

In the literature, there seem to dominate two general approaches to model pioneers. One is based on a simple *percolation* scheme. The second, more realistic approach, is based on networks of integrate-and-fire (IF) neurons, often combined with frequency-dependent synapses. The first modelling attempt which clearly identifies privileged neurons is at the same time one of the first papers which models the type of synchronisation often seen in neuronal cultures (Tsodyks et al. (2000), networks of LIF neurons). The main actors in the development after this impetus are *Cyrille Zbinden* and *Jean-Pierre Eckmann,* (using percolation and IF networks), *Olav Stetter* (also using percolation and IF networks), and researchers around Jordi Soriano (quorum percolation). Note that these three sources are interconnected in that many publications about leaders originate jointly from two or even all of these authors (and collaborating researchers).

Cyrille Zbinden's contribution is mainly distributed in his PhD thesis (Zbinden (2010)) and in Zbinden (2011). He simulates a *structured* network (with a Gaussian distribution of thresholds) of leaky integrators to study what makes a leader neuron a leader neuron. He assigns a leadership score to each neuron, so that neurons with a high score are leaders and finds that leaders are excitatory neurons with a low firing threshold. Furthermore, he detects a tendency that leaders send signals to many excitatory neurons and few inhibitory neurons (thus exhibiting *influence*) and receives only few signals from other excitatory neurons.

Olav Stetter (see Stetter (2012)) models the initiation of NS using discrete time dynamics, quorum percolation and networks of LIF neurons, using different topologies (both random and locally clustered). One of his main results is a formula for the mean of the latency (spike time relative to NS) of a neuron in terms of its degree (number of in-going fibres). Next, he finds evidence that clustering increases the variability of average latencies in the network. Olav Stetter also attempts to reconstruct network topology from activity, using *generalised transfer entropy* (see also Stetter et al. (2012) and Orlandi et al. (2014)).

Important references – which have not been mentioned so far – for work about the properties of leaders in terms of quorum percolation are Breskin et al. (2006); Cohen et al. (2010); Schmeltzer et al. (2014); Soriano et al. (2008); Eckmann et al. (2010). The review article Eckmann et al. (2007) summarises Eckmann's and Soriano's 'philosophy' about neuronal cultures (including quorum percolation) by considering the physics of these *in vitro* preparations.

A recurring theme in these works is the relation between 'leader' on the one hand and 'excitatory in-degree' on the other hand. As we will see, I do not observe the necessity for such a correlation.

1.2.6 The neural code and rank-order based representation

The question of what 'is' the neural code employed by neurons is a strongly debated question in theoretical neuroscience. Readable review articles are Thorpe et al. (2001) and Decharms and Zador (2000). Thorpe et al. (2001) argue that the speed of processing in neural systems is often too fast for rate-coding to be applicable. Also, they argue that rank-order based coding (ROC) is very efficient and easy to implement. I give more evidence for the applicability of ROC in this thesis.

A famous book about the neural code is Rieke (2008). Ju et al. (2015) stimulate a culture optogenetically and are able to make the culture differentiate between different musical styles (which they deliver optically) and thus demonstrate that the ability to process and integrate complex spatio-temporal information is an intrinsic property of generic cortical networks.

1.2.7 Topology and modular organisation

By the *topology* of a neural network, one means the collective information of which neuron is connected to which neuron. This information is a directed graph in a natural way. Sometimes, with *topology*, one means the *labelled, directed* graph which consists of the connectivity pattern *together* with the values of the synaptic weights (which then label the edges of the graph).

Thus, *graphs* and *networks*, which abstractly consist of *nodes* connected by *edges*, are a central topic in neural science. Network science is, in a sense, a meta-theory which unifies and applies to many particular theories inside the field of complex systems. It is important whenever single non-linear units interact within complex networks. Applications of networks and graphs are found in many fields, whenever they deal with networks of coupled oscillators, networks of biological oscillators, Josephson junction arrays, excitable media, neural networks, spatial games, genetic control networks, or many other self-organising systems (see Watts and Strogatz (1998) and references therein). Graphs are also very prominent in foundational research in theoretical physics, as Feynman graphs in quantum field theory (Peskin, 2018), or as graphs labelling states in loop quantum gravity (Rovelli, 2004).

A very popular aspect of network theory is the aspect of the 'six degrees of separation' (Watts and Strogatz, 1998). This says that in many real-world networks, the average distance between two randomly chosen nodes grows logarithmically with the total number of nodes. As a consequence, the average distance between two nodes is often in the order of five even in very large networks, a phenomenon which led to the famous 'Erdös number' with respect to co-authorship networks in the mathematical and physical sciences. Networks with this scaling behaviour of inter-nodal distance are given the suggestive name 'small-world networks'. They arise naturally as networks which are initially completely clustered or regular, and in which then some of the edges are 'rewired' in a random fashion (Watts and Strogatz, 1998). Thus these networks lie somewhere between networks which are completely random and networks which are completely regular.

Núñez-Amaral et al. (2000) study small-world networks, give a clear definition of

when a network is 'small-world,' and give examples. Newman (2000) is an extensive review of the small-world phenomenon. Strogatz (2001) is also a review article about real-world networks. The paper Watts and Strogatz (1998), which was already cited above, explores models of the small-world, the small-world phenomenon, they give examples, and they discuss that the phenomenon is associated with enhanced signal propagation speed, computational power and synchronisability. Thus, e.g., infectious diseases spread more easily. Other works devoted to the spread of information on graphs are Clementi et al. (2015) and Durrett (2010).

A different aspect is the phenomenon that degree-distributions are *scale-free*, typically following a *power-law*. Networks of this type are referred to as 'scale-free networks.' One of the three neural networks which I consider in this thesis is a scale-free network. Scale-free networks are often also small-world networks. They arise as developing networks, where new edges are created which attach preferably to already well-connected nodes[12] ('preferential attachment'), see Barabási and Albert (1999).

In this thesis, I shall focus my attention on a new network type which has not gained much interest so far. It is a random network which is very heterogeneous, in that connection probabilities vary among the units – uniformly inside a certain interval. I am also interested in the generic topic of spread of information, in this case the spread of activity inside neural networks. I will show that certain synchronisation events are heralded by privileged neurons, and that this effect is most pronounced in *broadly heterogeneous* networks. A reference for random graphs with arbitrary degree distributions is Newman et al. (2001).

Neural networks *in vivo* are typically highly modular on larger scales, which means that if neurons B and C are connected both to neuron A, there is a higher probability (compared to baseline) that B is also connected to C. Modular networks arise naturally as follows: One starts with a certain number of sub-networks $N_1, \ldots, N_r$, such that neurons within a certain sub-network are highly connected, and then weakly couples the different sub-networks with each other. A summary of

[12] Well-connected nodes in a network are called *hubs*, see Wills and Meyer (2019).

different network topology measures useful in neuroscience is Rubinov and Sporns (2010).

The networks which I consider in this thesis are *unstructured random* networks *without* imprinted modularity. However, as mentioned already above, one of the networks I consider has a feature which has not gained much attention in the modelling literature: it is *highly heterogeneous*. The reasons why I did not include modularity in these networks in spite of the experimental evidence *in vivo* for this network property are multi-fold. First, I wanted to find a *minimal* model for the effects which I shall describe. Second, there is some evidence (Marom and Shahaf (2002)) that neuronal cultures *in vitro* can evolve into unstructured *random* networks without considerable modularity, and one of my aims is the minimal modelling of certain aspects of the activity of these cultures. Finally, one may adopt a philosophy that something which is achievable in random networks *without* modularity is also be possible in networks *with* precise anatomical connections. Thus, if even unstructured random networks express privileged neurons, as long as synchronisation occurs, then the possibility that these neurons encode information and the precise neuronal code used might be *invariants*, so that the coding principles may apply also to the precisely structured networks of neuronal assemblies *in vivo*.

I finish this subsection with two remarks. The first is that there exists a very simple family of random graphs where connections are established *independently* of each other and randomly. These graphs are referred to as Erdös-Rényi graphs, and the degree distribution in such a network is binomial (Erdős and Rényi (1960)).

The second remark is almost needless: *real* neural networks *in vivo* are tremendously complex and specific, far away from random networks (see, e.g., Shepherd (2003)).

1.3 Aims

How spiking activity reverberates through neuronal networks, how evoked and spontaneous activity are interrelated, and how the combined activities of neurons

represent external stimulation are pivotal questions in neuroscience. In the first part of the this thesis, after this introductory chapter, I aim to describe minimal models of unstructured spiking networks *in silico* which can show a representational behaviour and information-processing capabilities which have been observed so far (if at all) only in networks where heterogeneity of neuronal background currents and hence an unnecessary artificiality in the choice of parameters enters in a crucial way. I first want to show that my new network shows the same advantageous and appealing behaviour of its tailor-made ancestors: privileged 'pioneer neurons,' which discharge reliably early and in reproducible 'paths' in the early phase of recruitment during synchronisation events, in particular if the latter are caused by gentle stimulation. After that, I will come to the central results of this thesis regarding the question how the network's dynamics enables representational capabilities. My aim is to show how gentle external stimulation can be subsequently read out from diverging activity-fluctuations by means of spike-order codes – when the privileged pioneer neurons are used for read-out. Here, I apply two methods which have not – to the best knowledge of the author – been applied *in computo* to this particular set of questions. In particular, I aim to show conclusively that the 'source' of stimulation is faithfully represented by the discharge order of pioneers and not by any rate-based features, thus demonstrating the viability and importance of this type of 'temporal' code. I aim to show in the same part that the distinctive role of pioneer neurons is owed to a combination of exceptional sensitivity to – and considerable influence on – network activity, an effect which makes pioneers act effectively as the strongest amplifiers in the network. All these effects, which have been observed only in considerably more bespoke models in the past (if at all), are described in greater detail and are extended considerably. Furthermore, I attempt to go far beyond mentioning phenomenology by explaining in detail the mechanisms which are responsible for the pioneer-effect and many phenomena which are derived from it. Along these lines, I show that broadly heterogeneous but random connection-topology not only increases the number of pioneer neurons in entirely unstructured networks, but also renders the emergence of pioneer neurons more

robust to changes in the excitatory-inhibitory balance. Concluding the first part of this thesis, I offer a new minimal model for the emergence and representational role of pioneer neurons, as observed experimentally *in vitro* (Eytan and Marom, 2006; Shahaf et al., 2008; Kermany et al., 2010), putting emphasis on a type of unstructured random connectivity which has not been studied much in the literature before – in contrast to homogeneous random networks of Erdös-Rényi type and scale-free networks with hubs. The latter two networks are also considered in this thesis, but only in so far as they serve to provide a counter-example to the generality in which the effects I am interested in are expressed.

In the last chapter of the results' part, I aim to point the reader to an intriguing but subtle way to detect relatively strong synapses in my *in silico* network, a method which might conceivably also succeed in *in vitro* neuronal networks, since all quantities needed can be measured *ex vivo*. I do this by computing (in my *in silico* network) a bespoke ratio of two different expressions which compute similarly but differently the same quantity: the so-called precedence probability, the probability that neuron i in the network will deposit its first spike within a synchronisation event before neuron j does.

Chapter 2

Phenomenology in silico

2.1 A concise description of my networks

My recurrent networks consist of 400 excitatory and 100 (Sahara et al., 2012) inhibitory LIF neurons[1] each, connected by frequency-dependent (dynamical) conductance synapses. Background currents which effectively shift the leak reversal potentials are constant in time and do not vary from neuron to neuron within the class of excitatory neurons (or inhibitory neurons, respectively). Spontaneous activity occurs because the effective leak reversal potentials are above threshold. There is no noise in the definition of my model. Heterogeneity of neurons is solely due to heterogeneity in the topology, be it tiny or considerable. This is a variant – and I would argue an improvement – of Tsodyks's model where heterogeneity is due to heterogeneous background currents.

I consider three different types of connection topologies: One homogeneous topology of Erdös-Rényi (ER) type, one network with scale-free connectivity and hubs, and one network with pronounced topological heterogeneity. In the last of the three networks, each neuron essentially has its own connection probability. Mean connection density is 20 % in all cases, consistent with figures obtained *in*

[1] This means that the network is rather small. One of the reasons for the small size is that I need very long simulations ($\mathcal{O}(10^3$ s)) in order to verify the effects I am interested in. But the time a simulation takes scales as N^2, where N is the number of neurons (due to the synaptic dynamics).

vitro and *in vivo* (Perin et al. (2011)). Note also that the weights Δg_{ij} of synapses are chosen randomly from a clipped Gaussian distribution with certain mean values. These mean values of the weights are different for the different type of connections (excitatory-to-excitatory, excitatory-to-inhibitory, etc.) and are *different* in the three networks considered. The reason is that *comparable behaviour* in the three networks is only possible with slight variations in the excitation-inhibition-balance, due to the different degree-distributions. See appendix for more details about the networks and the parameters. I want to emphasise at this point that all three networks are *unstructured*, i.e. they are random networks *without* tailor-made circuitry or *particular* built-in 'anatomical' connections.

My network comprises a minimal model for the effects I am interested in. Depression of synapses (one example of a fatigue mechanism) is needed in order to obtain all-or-nothing synchrony (interrupted by phases of near-silence), and – as we shall see – heterogeneity is needed in order to robustly obtain the so-called pioneer neurons (see below), which are at the centre of this thesis.

The network is programmed in C once using discrete time steps (Euler steps in conjunction with the first-order exponential integrator method) and once using a programming technique where the simulator jumps from spike to spike (event-based simulator). I tested my C code by making sure that both types of simulators give equivalent results, and by translating Tsodyks's model (Tsodyks et al. (2000)) into a model with conductance synapses and simulating this model. I obtained the same network behaviour as Tsodyks did. Note that many of the choices of parameters of my models are inspired by Tsodyks et al. (2000). Yet, the model is quite different, because I do not employ heterogeneous background currents.

The three different topologies are visualised in Fig. 2.1 in terms of representative examples of their degree distributions. See the figure caption for details.

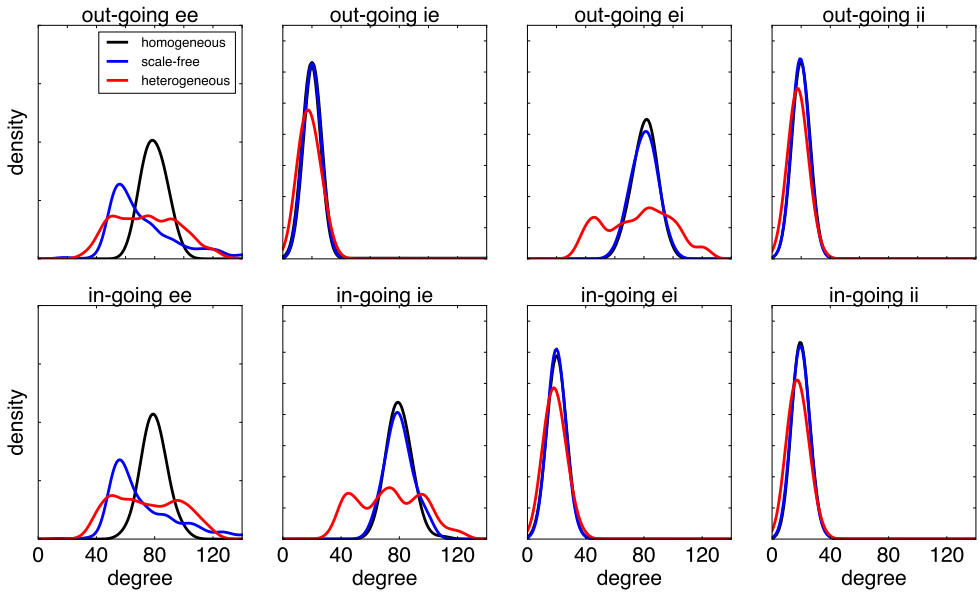

Figure 2.1 | Example degree histograms for the visualisation of the three topologies considered. The label 'ie' means 'excitatory-to-inhibitory' and similarly for the other labels. One observes that out-going and in-going connections of type 'excitatory-to-excitatory' are almost flat in a certain range (up to finite-size fluctuations, see below) in the heterogeneous case. By looking at the same quantity, one sees that the scale-free case serves to interpolate between the homogeneous and heterogeneous case. The statistics is here created from a single realisation of each network in order to illustrate finite-size fluctuations. Note that, in the heterogeneous network, there are no correlations between the number of in-going fibres ('in-degree') of a neuron and the number of its out-going fibres ('out-degree'). This is, of course, not visible from this figure, which does not show correlations between degrees. The colour code employed here for the different topologies will be used throughout the entire thesis and should suggest 'the hotter, the more heterogeneous.'

2.2 Macroscopic behaviour of the networks

In this section, I essentially describe only the *macroscopic* appearance of my three networks, i.e. all elements of the behaviour which manifest themselves from the point of view of the (summed) population activity. The majority of the details of the microscopic behaviour, i.e. the behaviour of individual neurons, will be postponed until later. When I look at the macroscopic behaviour, I first discuss the *spontaneous* activity and then the *evoked* activity. I will show most of the phenomenology only for the heterogeneous network, because *qualitatively,* the effects I am interested in, typically also arise in the same form in the other two networks (if parameters are fine-tuned). However, there is a strong tendency that many of the effects I am

interested in are *quantitatively* much less pronounced in the *non*-heterogeneous case. Here, I adopt the terminology that only my broadly heterogeneous network is considered *heterogeneous*, and both the (truly) homogeneous network and the scale-free network are referred to as *homogeneous* or *non-heterogeneous* networks. The reason for this choice of terminology is that, as we shall see, and as mentioned above, the scale-free network behaves very similar to the (truly) *homogeneous* network (w.r.t. the microscopic activity), in spite of the scale-free degree distribution (including hubs).

2.2.1 Spontaneous activity

Fig. 2.2 shows elements[2] of the spontaneous behaviour of the population of excitatory neurons. The population activity is visualised in the form of raster plots (A) and spike-count[3] histograms (B). Pronounced synchronisation events are readily visible and are marked by red stars in subplot A and by red dashed lines in subplot B. These events are referred to as *network spikes*. The term 'network spike' will be abbreviated as 'NS' in the sequel[4]. Subplot C shows the power spectral densities of the activity of *single* neurons (averaged over all neurons). Although the network is deterministic, sparse connectivity ensures (apart from NS) irregular and asynchronous activity in many neurons (Brunel, 2000b; Mattia and Del Giudice, 2002). The average power spectral density of individual neuron firing (apart from NS) resembles the power spectrum of Poisson spikes with a refractory period (which here corresponds to excised periods of NS) (Spiridon and Gerstner, 1999). As random connectivity entails some non-uniformity in all three network types, the power spectral densities of individual neurons are quite diverse (Pena et al., 2018), with some neurons firing more frequently and regularly (albeit at different rates) and others firing more rarely and irregularly. Finite-size fluctuations of population activity (Brunel, 2000b; Mattia

[2] In a small fraction of realisations of the networks, the activity does *not* express activity as shown in the figure. In these rare cases, synchronisation events are not all-or-none. See appendix for more details.

[3] Only spikes by excitatory neurons are being counted.

[4] In some experiments, I have deliberately *removed* NS from the traces. Specifically, in these cases, I removed the window – beginning 35 ms before and ending 35 ms after the peak of the NS – from the data files. See Fig. 2.5 for the average shape of NS.

and Del Giudice, 2002) induce pairwise correlations between individual neuron spikes (apart from NS), which on average are moderate for homogeneous and scale-free networks ($\rho_{hom} = 0.29 \pm 0.21$, $\rho_{sf} = 0.18 \pm 0.18$) and weak for heterogeneous networks ($\rho_{het} = 0.01 \pm 0.013$). See the appendix for my definition of spectra and cross-correlations.

The parameters of the networks (in particular the strength of excitation) are chosen so that the rate of NS is typically in the order of 1 Hz in all three cases (i.e. topologies). The coefficient of variation of the inter-network-spike interval (INSI, the interval between consecutive NS) is usually close to 0.6, i.e. the point process associated with the synchronisation events is approximately located 'in the middle' between regularity and irregularity.

Why do network spikes occur? The network is designed to be at a point where a larger-than-average positive fluctuation of the excitatory population activity is amplified by the excitatory units: a positive feed-back process occurs which recruits essentially all excitatory neurons. Inhibitory neurons discharge continuously (i.e. all the time) and provide a constant background of inhibition to all excitatory neurons which balances and re-normalises the (effective) leak reversal potentials, so that many of the excitatory neurons are inactive between NS. The approximately exponentially rising activity in the initial phase of a NS is terminated when the excitatory-to-excitatory synapses become depressed due to the high activity of the neurons which are pre-synaptic to these synapses. This terminates the NS and leads to a strong drop in activity. When the activity has its next larger-than-average positive deviation, and when the depressed synapses have recovered, the next NS can be generated. The generation of NS is approximately a threshold-governed phenomenon: NS occur when the excitatory population activity reaches a threshold. This will become important later, although this concept of a population-activity-threshold is only approximately valid, because a given value of the population activity is consistent with a myriad of combinations of individual activities.

The serial correlation coefficients (with positive lag) of the inter-network-spike intervals are close to zero (not shown). This is in accordance with the idea that NS

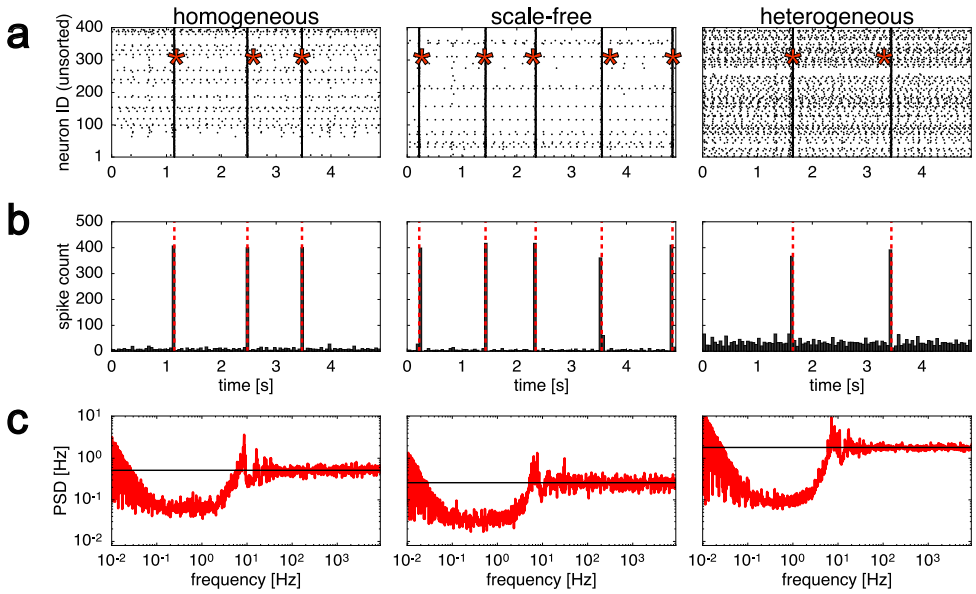

Figure 2.2 | Spontaneous, macroscopic activity. Raster plots of the excitatory activity (**a**), time-dependent spike-count histograms (**b**), and power spectral densities of the individual excitatory activities, averaged over all excitatory neurons (**c**). NS are marked by asterisks in subplot **a** and by dashed red lines in subplot **b**. The horizontal black lines in **c** indicate the mean values of the individual (excitatory) average firing rates.

are *renewal* events, i.e. a NS acts as a reset of the network variables[5] (see Dranias et al. (2015) for evidence for this *in vitro*).

Figure 2.3 shows the firing rates of all excitatory units, which are sorted by their discharge rate. Most neurons do not fire between NS and discharge precisely once during NS. This is the reason why most neurons in Figure 2.3 have the same firing rate (which is identical to the NS rate, red dashed line). Inhibitory neurons fire continuously at a rate of about 30 Hz and do not increase their firing during NS.

My networks are tuned so that NS are all-or-nothing events. In order to verify the all-or-none character, I used the following procedure. I introduced an activity-threshold θ with a value $\theta = 1.1 \cdot \langle A \rangle$, where $\langle A \rangle$ is the average value of the time-dependent excitatory population activity. Whenever the excitatory population activity crosses θ from below, I determine the corresponding instance of time as t_1. Next, the earliest instance of time after t_1 where the excitatory population activity

[5] In the LIF neuron driven by un-correlated noise, a reset of the membrane voltage also acts as a reset, i.e. the resulting spike train is a renewal point process.

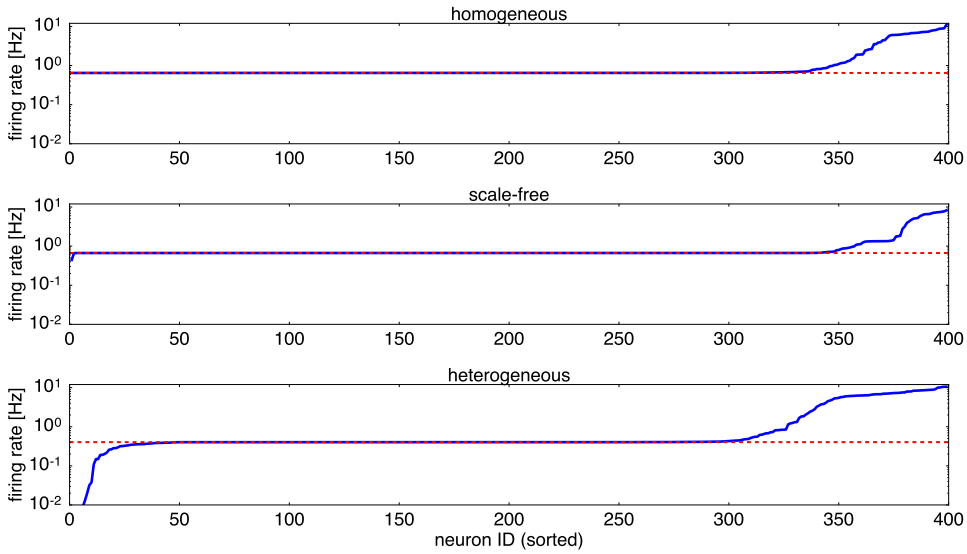

Figure 2.3 | Distribution of individual firing rates. Typical distribution of firing rates of excitatory neurons for the three networks considered. Neurons are sorted by their discharge rate. The dashed red lines denote the NS rates.

crosses θ from *above* is determined as t_2. Then I determine the global maximum A_{max} of the activity inside the interval $[t_1, t_2]$. I then construct a histogram of all the values A_{max}. In short: I establish a histogram of activities associated with activity peaks ('histogram of peak-activities'). If NS are all-or-none, I expect two well-separated peaks in this histogram: One peak which consists of activity values corresponding to local maximums of small sub-threshold fluctuations; and one peak which consists of the maximums of activity which occur during NS. Figure 2.4 A shows that these two types of peaks together with their separation are indeed observed, confirming the all-or-none character of NS. In Figure 2.4 A, activity is measured in units of the average activity. One observes that the relative size of NS is slightly smaller in the heterogeneous case. This is due to a higher level of activity *between* NS in the heterogeneous case.

I note that the same procedure (as described above) for detection of *activity peaks* was used for the detection of (proper) *network spikes* (in particular their instances of time). The only difference is that in the latter case the threshold was

chosen to be $0.5 \cdot \max(A)$ instead of $1.1 \cdot \langle A \rangle$, where $\max(A)$ is the global maximum of the population activity, so that *sub-threshold* peaks are *excluded*.

Figure 2.4 B shows a black dot for each individual spike emitted by an excitatory neuron, with an x-value corresponding to the logarithm of the time from the spike to the next NS, and with a y-value representing the so-called *sorted neuron ID*, i.e. the least active neuron has ID 1, the next least active neuron has ID 2 and so on. This way of sorting will be used frequently in the sequel and is indicated by the label 'neuron ID (sorted)'. (I already used it in Figure 2.3.) The distribution of black dots in the three plots corresponding to the three networks in Figure 2.4 B is reminiscent of the letter 'π'. The right 'leg' corresponds to individual spikes with a large time to the next NS. These spikes occur *shortly after* a NS, which is why the time to the next NS is large. The left 'leg' corresponds to spikes which occur *shortly before* an NS and this is why the time to the next NS is very small. The separation between the two 'legs,' indicated by the red line, shows that for the approximately 350 least active excitatory neurons, there is almost no activity between NS, so that for these neurons a *first spike* within a certain NS is well-defined. This will become important later.

Figure 2.5 shows the average shape of the excitatory activity during NS (red) as well as all shapes of the individual NS which occurred in the simulation (black curves). One observes that the shape of NS is rather stereotypical.

2.2.2 Evoked activity

NS can be evoked by stimulating a small set of excitatory neurons. This simply means that the spikes which are directly evoked ('receptive sheath'), i.e. directly enforced by stimulation, raise the population activity above threshold so that the positive feedback process associated with NS is initiated. I note that the set of neurons which are stimulated can be very small in order to evoke NS: $\mathcal{O}(10)$ stimulated neurons are enough (when stimulation is carried out by simultaneously enforcing spikes, see below) in general in order to make sure that many stimulations cause a NS. Note by the way that NS can also be evoked by stimulating *inhibitory* neurons

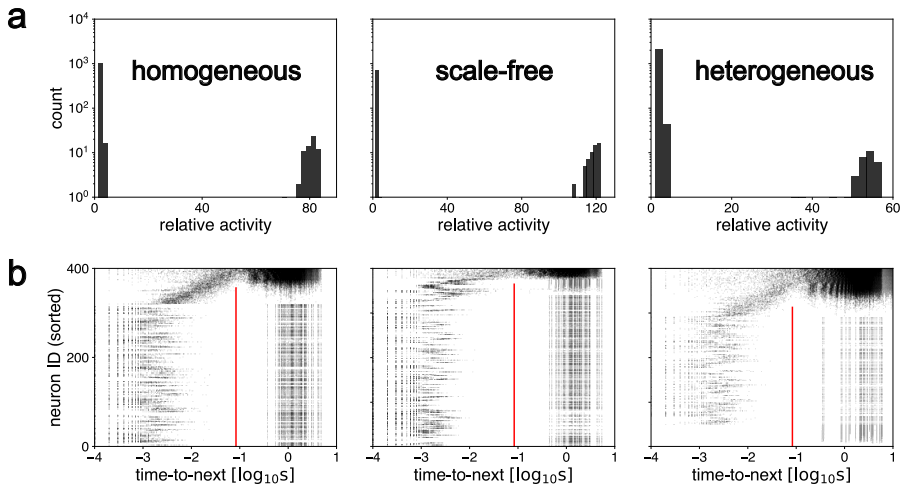

Figure 2.4 | Histogram of peak activities and activity relative to NS. (a) Histogram of peak-activities. The peak at the left of each of the three plots originates from local maximums of sub-threshold activity fluctuations. The peak at the right in each of the three plots comes from activity-peaks which occur during NS. The separation of the two peaks in each of the three figures shows that NS are all-or-none events. Activity is measured in units of the average activity. **(b)** For each individual spike (in a random subset of the collections of all spikes) emitted by one of the excitatory neurons, I plot a dot whose x-value shows the logarithm of the latency from the spike to the next NS, and whose y-value is the sorted neuron ID, i.e. the excitatory neuron N_1 with the lowest average activity has ID 1, the least active neuron more active than N_1 has ID 2, and so on. The plot shows that neurons with sorted ID in the range of the red bar do not discharge between NS, because there is a clear separation between dots on the left which represent spikes which occurred shortly before an NS, and dots on the right which represent spikes which occurred shortly after an NS. Thus, for neurons in the range of the red bar, a *first spike* within an NS is well-defined. The latency-value at which the red bar is plotted (called 'fake latency,' because it is here not associated with a proper physical event like a stimulation) represents a latency which can be used as a criterion in order to determine to which NS a spike emitted by a relatively inactive neuron belongs. Note that, in the heterogeneous network, all neurons from $ID = 265$ on discharge sometimes *between* NS in long simulations. This is not clearly visible here because I have only plotted a random subset of the set of all spikes, in order to keep the figure size restricted.

(not shown).

There are at least two possibilities of how to stimulate neurons. One can enforce a spike in the neurons which are targeted, or one may induce a membrane voltage jump (by some amount ΔV) in the target neurons. The difference is that in the latter case the stimulation may fail to evoke a spike in some of the target neurons. For clarity and simplicity, I decide to use the *first* method. When I stimulate networks, I make sure that the enforced spikes occur *simultaneously*. Also, the (directly) enforced spikes are *not written into the spike-data files* which I analyse. This will become important later.

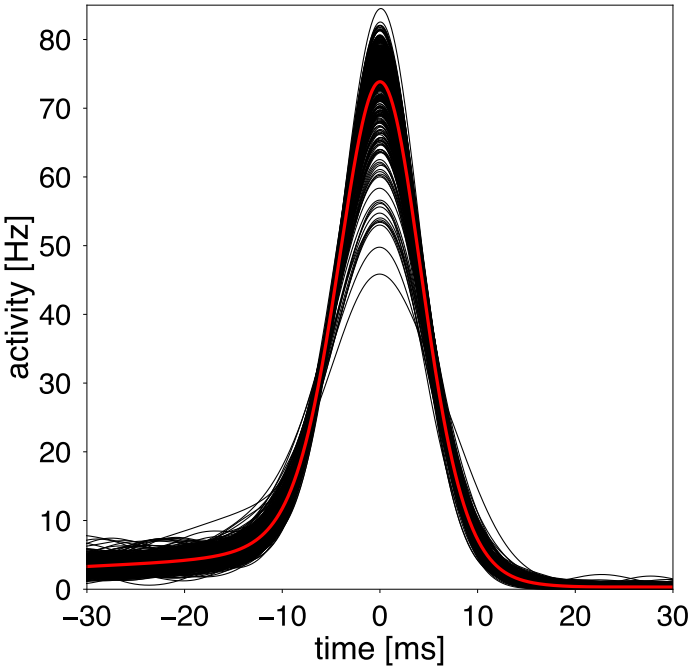

Figure 2.5 | Shape of NS. Average activity-shape (red) of NS in a heterogeneous network and individual activity-shapes of all NS (black) which occurred in this simulation.

Now the question arises of how to detect a successful stimulation, i.e. a stimulation which causes a NS. I expect that a NS which is evoked by stimulation occurs shortly after the stimulation. However, even if the time lag between stimulation and NS is very small, there is *a priori* a small, but positive, probability that the NS would have occurred even without the stimulation, so that it would be inappropriate to say that the stimulation causes the NS. However, this problem does not arise in my case, as I will explain now.

Figure 2.6 *A* shows an epoch of activity (in blue) in which stimulation occurs at irregular intervals[6]. Stimulation instances are shown as vertical red dashed lines. The figure shows some spontaneous (i.e. not evoked by stimulation) NS: they are marked by an asterisk. The stimulation labelled with a minus sign is ineffective, i.e. it does not evoke a NS. The two stimulations labelled with a plus sign cause the

[6]Stimulation instances in this thesis are always modelled as a homogeneous Poisson process with a rate in the order of the spontaneous NS rate.

subsequent NS. This classification of stimulations and NS is based on Figure 2.6 B. There, I plot (this is based on Gigante et al. (2014)) for each stimulation instance a red dot with an x-value which is the latency from the *last* NS to this stimulation, and a y-value which is the latency from the stimulation to the *next* NS. The red dots are, apart from one outlier, separated in two regions: one below and one above the blue rectangle. How does this separation arise? If a stimulation causes a NS, then the time from stimulation to next NS is small (group of red dots below the blue rectangle). If the stimulation is not successful, there is still a reduction of synaptic resources associated with the stimulation, and this drop in synaptic efficacy increases the time to the next NS. This mechanism is responsible for the existence of the 'forbidden region' in the space of times from stimulation to the next NS (blue rectangle). Thus, the blue rectangle can be used to uniquely divide stimulations into successful and unsuccessful. Figure 2.6 A and all classifications of stimulations (into successful or unsuccessful) and NS (into evoked or spontaneous) in the sequel are based on these observations. Finally, I plotted in Figure 2.6 B a distribution of dots ('surrogate events') which would be expected if instances of stimulations and instances of NS were entirely *independent* of each other. The black dots are sampled from this distribution and the number of black dots is equal to the number of red dots. The fact that there are numerous black dots inside the 'forbidden' blue rectangle confirms my strategy described above for classifying stimulations as successful or unsuccessful and for classifying NS as evoked or spontaneous.

Figure 2.7 shows activity shapes of spontaneous (black) NS and evoked NS (in colour), where the different colours serve to distinguish the different stimulation sites. A stimulation site is here (and in the sequel) simply a small collection of randomly chosen neurons (since my networks do not have a 'spatial' structure), such that the groups corresponding to different sites are disjoint. One observes that NS shapes are rather stereotypical, so that the different origins of NS cannot be deduced from the respective NS shape. This observation will be corroborated later by means of a different method (see section 2.4).

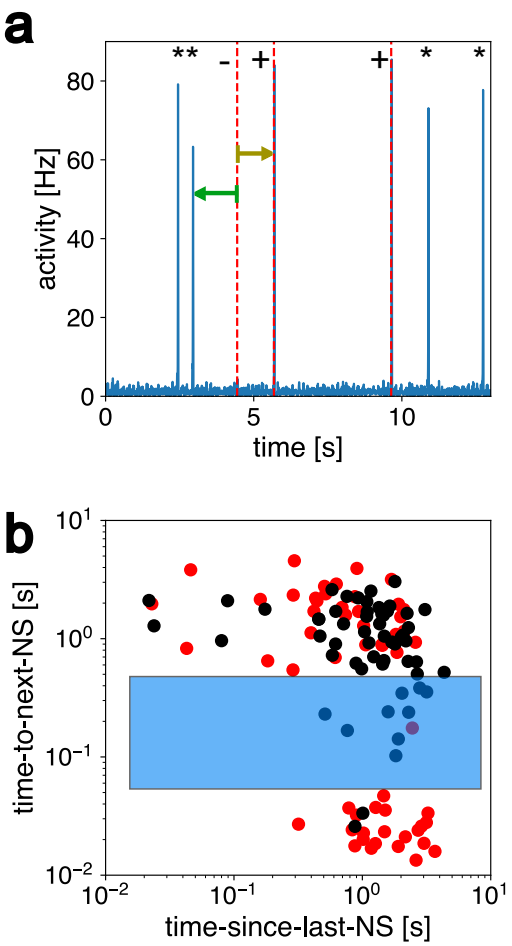

Figure 2.6 | Evoked activity. (**a**) An epoch of activity of the network where stimulation occurs. The activity is drawn in blue, and a few NS are visible. The red dashed lines denote instances of stimulation. The NS labelled with an asterisk are spontaneous, i.e. not evoked by stimulation. The stimulation labelled with a minus is ineffective, i.e. does not evoke a NS. The two stimulations labelled with a plus sign are effective, i.e. evoke NS. Here and in the sequel, stimulation occurs with statistics of a homogeneous Poisson process with a rate in the order of the rate of spontaneous NS. The green arrow illustrates 'time-since-last-NS' and the yellow arrow 'time-to-next-NS' for the ineffective stimulation (see subplot **b**). (**b**) For each instance of stimulation, I plot a red dot at a position whose x-value is the latency from the last NS to this stimulation, and whose y-value is the the latency from the stimulation to the next NS. The distribution of red dots is separated in a region above the blue rectangle and a region below the blue rectangle (apart from one outlier). This clear separation suggests that red dots below the blue rectangle, with small time-to-next NS, correspond to effective stimulations, and that red dots above the blue rectangle, with large time-to-next-NS, correspond to ineffective stimulations. The efficacy of the stimulations is further confirmed by the fact that the 'forbidden' region (blue rectangle) contains many black dots, since the black dots are sampled from the distribution which would be expected if stimulations were entirely ineffective. See also main text. Subplot **b** is based on Gigante et al. (2014).

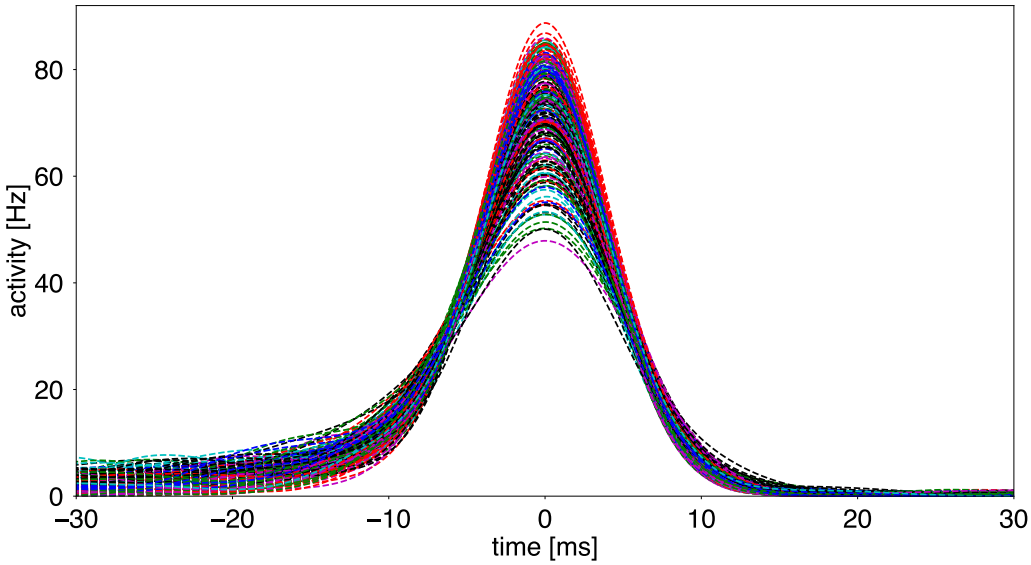

Figure 2.7 | Shapes of evoked NS. This figure shows the activity-shape of NS in a simulation where stimulation occurs at five different sites. Spontaneous NS are shown as black dashed lines, and the NS evoked by stimulation at one of the five sites are shown in colour, where the different colours (red, blue, green, magenta, and cyan) correspond to the different stimulation sites. The figure shows that the shape of NS is rather uninformative about the site where stimulation occurred. This will be confirmed later by means of a different approach.

2.3 Microscopic behaviour of the networks

It turns out that spikes of different neurons within NS have different typical *latencies* relative to the *peak* of the NS. Probability distributions of latencies $t_{spike} - t_{NS}$ of *first* spikes for three different neurons (showing different characteristic behaviour) are shown in Figure 2.8. The curve depicted in black shows a neuron which has its maximal probability of firing *in coincidence* with the peak of the NS. The curve shown in dark grey arises from a neuron which typically fires *after* the peak of NS. Finally, the light grey curve originates from a neuron which fires typically and consistently several milliseconds *before* the peak of the NS. These neurons are by definition the *pioneer neurons* and they are at the centre of my investigations. For the moment, I rely on this rather vague definition of 'pioneer.' In section 2.5, I give a precise definition.

What is the connection between the average firing rate (or sorted neuron ID)

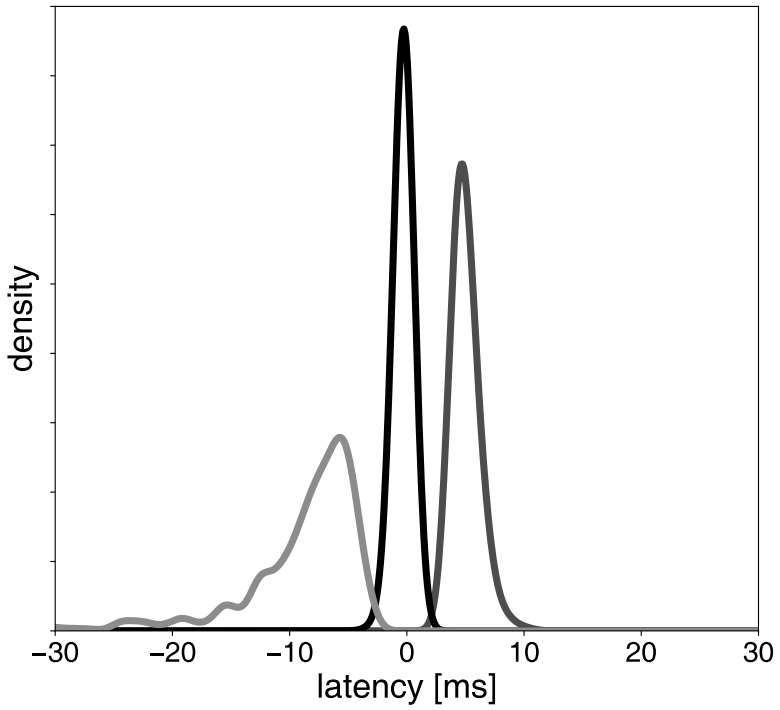

Figure 2.8 | Characteristic latency distributions of three representative neurons. Distributions of latencies of individual first-spikes relative to the peak of NS for three representative neurons. The peak of the NS by definition has latency 0 ms. The neuron whose latency distribution is shown in black fires predominantly with NS (sorted neuron ID 150). The neuron whose latency distribution is shown in dark grey fires typically after NS (sorted neuron ID 40). The neuron whose latency distribution is shown in light grey fires predominantly before NS (sorted neuron ID 300) and hence is a pioneer.

and the sign of the (typical) latency between individual first spike and (peak of) NS? This can be deduced from Figure 2.9 A. This figure shows that essentially the higher the sorted neuron ID, the lower (more negative) is the average latency. Neurons with sorted neuron ID between 0 and 50 fire consistently *after* the peak of the NS. Neurons with sorted ID between 60 and 250 fire most often *during* the peak of the NS. Neurons with ID between 260 and 320 fire consistently *before* the NS and thus are pioneers[7]. Finally, neurons with ID greater than 320 fire continuously, i.e. all the

[7]This rather vague definition of 'pioneer' will be sufficient for the moment. In section 2.5, I give a precise definition which there is shown to be equivalent to this simple definition in terms of the sorted neuron ID.

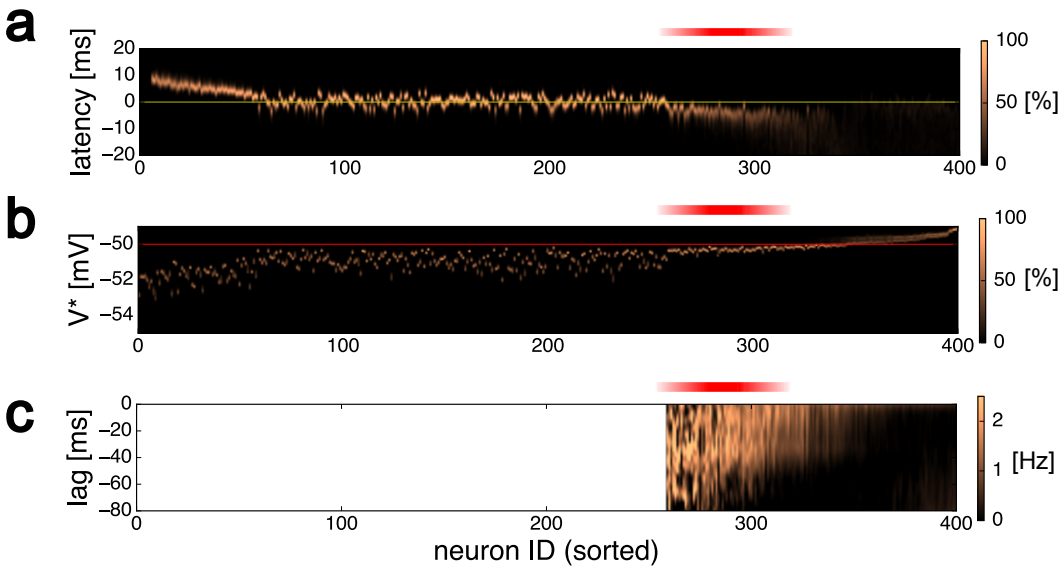

Figure 2.9 | Latency distributions, membrane $*$-voltage distributions, and spike-triggered average activity of individual neurons. Excitatory neurons are sorted horizontally by mean activity (sorted neuron ID). Red shading marks the pioneer range (ID 260 to 320). **(a)** Latency distributions (lightness-coded and arranged vertically) of first spikes of individual neurons, relative to associated peak of NS. Zero latency (yellow line) is defined by the peak (maximum) of the activity during NS. Neurons in the pioneer range fire reliably before the NS (negative latencies). **(b)** Distribution of $*$-voltage in individual neurons (again lightness-coded and arranged vertically), during intervals without NS, relative to firing threshold (horizontal red line). Neurons in the pioneer range have membrane potential just below threshold. **(c)** Average deviation $\Gamma_n(\tau)$ of population activity at lag τ (lightness-coded and arranged vertically), conditioned on individual spikes of neuron n (see main text and appendix). Spikes of neurons in the pioneer range are consistently preceded by positive deviations. Note that deviation $\Gamma_n(\tau)$ is not defined below ID 265 (see main text).

time[8]. This figure is based on a long simulation of a network with heterogeneous connectivity. If parameters are fine-tuned, qualitatively (but not quantitatively) similar behaviour can also be observed with the other two networks.

The reason for this monotonous relation between firing rate and latency is easy to understand. The greater the sensitivity of a neuron to network activity, the greater is its firing rate (and hence its sorted neuron ID), and the earlier it is recruited during

[8]Fig. 2.9 A shows (distributions of) latencies of first-spikes to the peak of NS. For this, a critical latency needs to be chosen from where to measure first-spikes. I refer to this critical latency as *fake latency*, because it is (here) not a latency associated with a true physical event (such as a stimulation). The fake latency can be chosen with the help of Fig. 2.4B. In my case it is about −80 ms. Because neurons with ID bigger than 320 fire continuously, the support of their first-spike-latency distribution is close to the fake latency, which is outside the latency-range of the plot (2.9 A). Therefore the latency-densities of the most active neurons are almost zero (in the range shown) and hence appear very dark in the plot. Later, when I evoke some of the NS by stimulation, I will choose the critical latency of evoked NS (and only of the evoked NS) as the latency of the stimulation, so that first-spikes are measured from the stimulation on.

NS. The sensitivity can be measured in terms of the distribution of the membrane $*$-voltage. The $*$-voltage is a fake voltage which evolves according to the same differential equation as the proper membrane voltage, *but it is never reset*[9]. Therefore, the $*$-voltage can be above threshold (which is impossible for the proper membrane voltage). Also, there is often *approximately* a simple monotonous relation between firing rate and the mean of the $*$-voltage: the higher the latter, the higher the former (this assumes that the standard deviation of the $*$-voltages are comparable for the different neurons). This qualitative and approximate monotonous relation is demonstrated in Figure 2.9 B. From there it is also visible that pioneers (ID 260 to 320, see red shaded bars above the three subplots) reside just below threshold (red line). This is one reason for their distinctive sensitivity and their early recruitment.

Figure 2.9 C shows the average shape $\Gamma_n(\tau) := A_{exc}(t_{spike,n}+\tau) - \langle A_{exc} \rangle$ (i.e., a spike-triggered average of the excitatory population activity) – as a function of τ – for the most active individual neurons n (sorted by average firing rate) and for negative τ, where the average is over all spikes of neuron n which occur *between* NS (see appendix). Here, A_{exc} is the excitatory population activity and $\langle A_{exc} \rangle$ its mean. Put simply: I show the average shape of the excitatory activity (with mean subtracted) preceding spikes of neuron n (for all n). Thus we can see how the summed activity deviates from its mean prior to individual spikes of pioneers. It is not only made sure that the individual spikes which enter this statistics are *between* NS, but also that the window of the activity which is averaged over does *not* contain a NS. The idea here is to see the average shape of activity around individual spikes without *contamination* by the large NS. The plot is only defined from ID 265 upwards, because neurons with lower firing rate never discharge between NS. Also, the plot is only defined for negative τ because the activity *after* the individual spike usually contains a NS, which I want to exclude by definition of the statistics.

One observes that spikes of pioneers are associated the excitatory activity having a positive deviation of ≈ 2 Hz from its mean for a time window T preceding the spike, where the length of T depends on the precise neuron ID. The most early pio-

[9] See Fig. 1.1 and Fig. 1.2 for illustration.

neers have $T \approx 40$ ms. As the neuron ID decreases to late pioneers, the T increases up to 80 ms. Since T is in the order of $\mathcal{O}(10\text{ ms})$, a small and short positive (but significant) deviation of the activity from its mean (with about 30 to 60 more spikes than usual in this time window) suffices for recruiting pioneers. This is another aspect of their distinctive sensitivity.

The average distribution (between NS) of recovered resources of excitatory synapses projecting from the excitatory neuron n to other neurons is shown in Figure 2.10 A (for all n, see Tsodyks et al. (2000)). The location of the red bar, representing (here and elsewhere) the range of pioneers, shows that the privileged pioneer neurons have out-going synapses which are almost completely recovered between NS (in the asynchronous state). In this sense, pioneers are also very influential on population activity. However, I want to emphasise that pioneers are *not* hubs in general (see Fig. 3.1 A below). In my model, the *only* connection between 'being a pioneer' on the one hand and the 'degree' (out or in) on the other is that pioneers have relatively few inhibitory ancestors (see Fig. 3.1 B below), which is the main reason why they are located 'just below threshold.'

Figure 2.10 B shows the average amplitude of the post-synaptic potential (see appendix for details about its computation) caused by a spike in neuron n in neurons which are post-synaptic to neuron n. As expected, plotting this for all n, where the neurons are sorted by their firing rate, resembles the shape of the distribution of recovered resources (Figure 2.10 A). In particular, the higher the average firing rate of a neuron, the smaller is its average PSP caused in its post-synaptic partners, because the synapses involved are more depleted.

However, when the stationary post-synaptic deviation from the reference level (see appendix) of the membrane voltage caused by *all* spikes in neuron n is considered (for all n sorted by firing rate), the resulting plot *increases* with sorted neuron ID, because the drop in the synaptic efficacy is more than counter-balanced by the fact that there are more pre-synaptic spikes (since the firing rate increases with sorted neuron ID). Thus, if I *de-efferentiate* or *silence* (i.e. cut all out-going connections of neurons) selective groups of neurons with similar sorted neuron ID, silencing

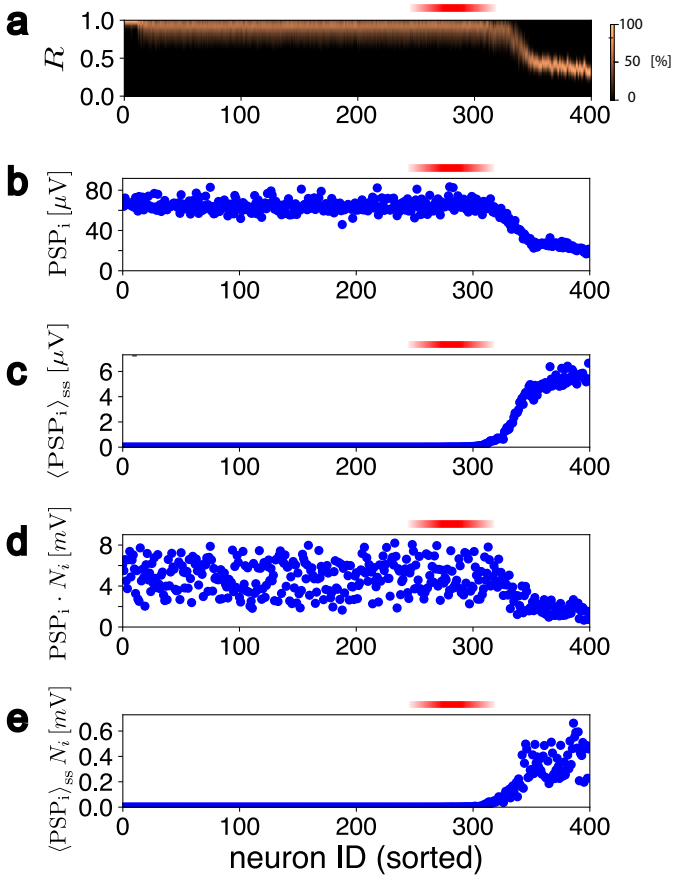

Figure 2.10 | Influence of excitatory neurons. Red shading marks the pioneer range (ID 260 to 320). **(a)**: Probability density of synaptic resources R, average over all efferent synapses of a given neuron, during intervals between NS. Resources decrease monotonically with mean activity. The most active neurons to retain substantial resouces are neurons in the pioneer range. **(b)** Amplitude of post-synaptic potentials $\mathrm{PSP_i}$ elicited by single spikes, averaged over all efferent synapses. **(c)**: Steady-state post-synaptic potential $\langle \mathrm{PSP_i} \rangle_{\mathrm{ss}}$ elicited by Poisson spiking at individual mean rate of neuron, averaged over all efferent synapses. Note that increasing firing rate overcompensates diminishing resources. **(d)**: Same as (b), but *summed* over all efferent synapses. **(e)**: Same as (c), but *summed* over all efferent synapses.

the most active excitatory neurons should cause the strongest drop in the average activity level. This will become important in a moment. All this shows that w.r.t. influence – as measured by PSPs – pioneers are *not* extremal: they lie 'in the middle' as does their firing rate.

This observation is challenged by actually carrying out the experiment described above (see Tsodyks et al. (2000)), see Figure 2.11 A: I de-efferentiate neurons with

sorted neuron ID in the interval [0,30], then in the interval [10,40], and so on, and plot the resulting average NS rate as a function of the sorted neuron ID of the most inactive neuron which is de-efferentiated. One observes that de-efferentiating pioneers has a very strong effect: NS are completely abolished if the earliest pioneers are made ineffective (Tsodyks et al. (2000)). Figure 2.11 A shows two curves (in green and blue) which correspond to two different realisations of the network. The blue curve has the property that the NS rate becomes positive again when neurons which are more active than pioneers are being silenced. I want to emphasise at this point that this behaviour (which does not occur in the case of the curve shown in green) is the *typical* behaviour, i.e. it occurs in most realisations of the network (not shown). But how is that possible? If silencing neurons which are more active than pioneers causes the strongest drop in the average level of activity (as suggested by Figure 2.10 C), why does the NS rate become positive again in this case?

The NS rate depends on two factors. It depends on the average level of the excitatory activity, and on the size of the activity-threshold for NS generation (and on the variability of the activity). The fact that (in most realisations) only silencing pioneers abolishes NS could be explained if making pioneers ineffective significantly *increases the threshold for NS generation.* Silencing neurons which are more active than pioneers would lead to a positive NS rate again if the lower average level of excitation (as compared to silencing pioneers) was paired with a *zero increase* of the threshold for NS generation (as compared to silencing pioneers), so that the difference between threshold and average activity-level is smaller when making neurons impotent which are more active than pioneers. This could explain why NS are again observed if these neurons are made ineffective.

This is precisely the case, as shown in Figure 2.11 B. There I estimate the threshold for NS generation as a function of the neurons which are silenced. I do this by measuring the highest value of the excitatory population activity which occurred well between NS. The logic here is that due to the stochastic behaviour of the activity, it will 'just miss' the threshold at least once in long simulations. Therefore the highest value of activity between NS should set a lower limit for the threshold

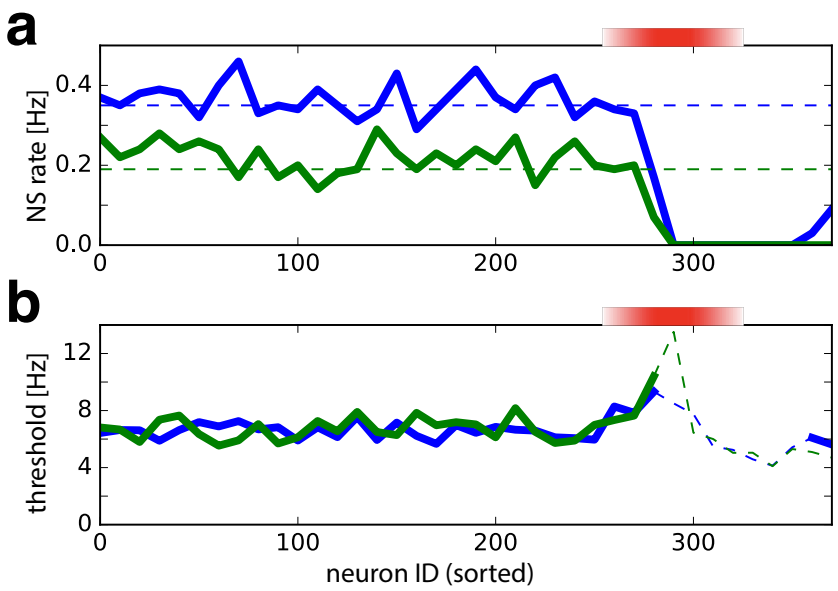

Figure 2.11 | Effect of silencing groups of neurons. Red shading marks the pioneer range (ID 260 to 320). **(a)** Rate of NS as a function of N, for modified networks with neuron cohort $N \leq \text{ID} \leq N+30$ silenced by de-efferentiation. Results are shown for two realisations (blue and green) with different spontaneous NS rates (dashed lines). NS cease when neurons in the pioneer range are silenced. Typically, NS are recovered when neurons above this range are silenced (*e.g.*, blue trace). **(b)** Silencing pioneer neurons elevates threshold for triggering NS. Threshold population activity (in Hz), after silencing neurons $N \leq \text{ID} \leq N+30$. Two realisations are shown (blue and green traces). Over much of the pioneer range, only a lower bound for the threshold could be established (dashed traces), because even the largest observed fluctuations failed to trigger a NS.

for NS generation. Figure 2.11 B shows that the threshold does in fact increase if pioneers are made ineffective. This fact can be explained by the assumption that pioneers are the first link in the positive feedback-chain which supports the NS. If they are silenced, the feedback-chain breaks and thus the threshold is increased. All this shows again that pioneers are not only sensitive, but also rather influential.

2.4 Information and dynamics: rank-order based coding

How information is encoded in the spiking activity of neurons is a popular topic in theoretical neuroscience. The final aim must be to understand neural coding *in*

vivo. However neuronal cultures provide a useful model system for studying coding, and probing the neurons is much easier in this case. As already mentioned earlier, Kermany and colleagues (Kermany et al. (2010)) studied decoding by modelling different stimuli as stimulation of geometrically distant parts of the culture (i.e. different 'loci' or 'sites'). Furthermore, they investigate decoding by means of different coding schemes by considering the following four different coding schemes: the vector of individual average-firing rates ('neuronal rates'), the time-dependent population activity ('temporal rates'), the time-to-first-spike (TFS) vector (equivalent to the vector of latencies from first-spike to the peak of the NS) and the order ('rank order') of first-spikes of individual neurons. The first two schemes may be considered as rate codes, the latter two as temporal codes. See Fig. 2.12 for illustration.

Kermany and colleagues then study decoding by enquiring how much informa-

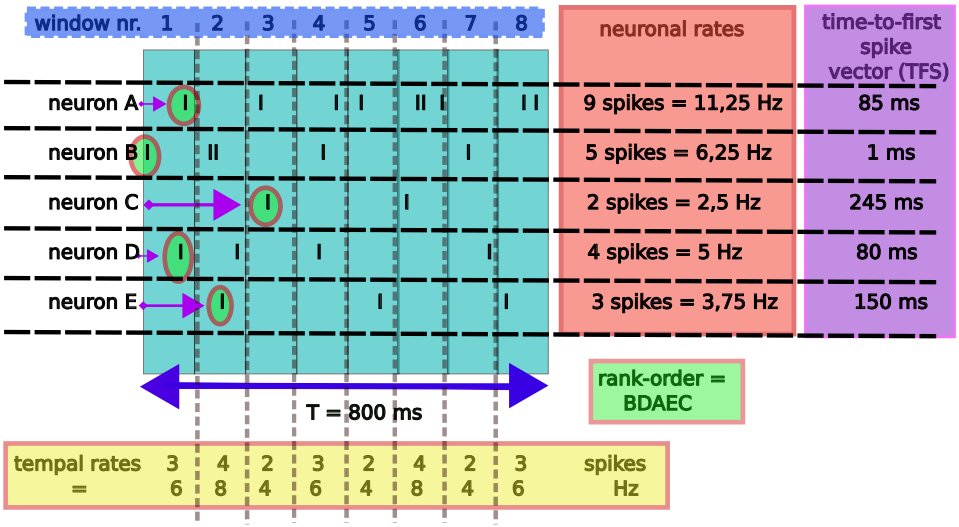

Figure 2.12 | Coding schemes considered. Illustration of neuronal rates (red), TFS vector (purple), rank-order (green) and temporal rates (yellow). The vertical black lines in the region with light-blue background are the spike times of the five neurons which are presented.

tion about the site of stimulation is contained in the activity-features corresponding to the different coding schemes. They do this by employing a support vector machine[10] (SVM). First they train the SVM by providing examples of feature-label pairs

[10]The SVM which I use here is based on the Python package 'sklearn' which implements the LIBSVM algorithm, see Chang and Lin (2011).

(where the labels are simply the IDs of the stimulation sites) based on the first half of the data. Then the SVM is fed with the feature-data of the second half of the data and the percentage of successful predictions of the label is recorded ('classification accuracy'). Kermany and colleagues observe that temporal coding schemes outperform rate codes in spatial classification tasks. Note that obviously rank-order is contained in the TFS scheme, so the former is a 'compacted' version of the latter. Kermany and colleagues also demonstrate that in many cases the TFS scheme does not add additional information to the rank scheme. In other words, there is virtually no loss of information when TFS vectors are compacted to rank-order strings.

Kermany and colleagues remove the directly stimulated neurons ('receptive

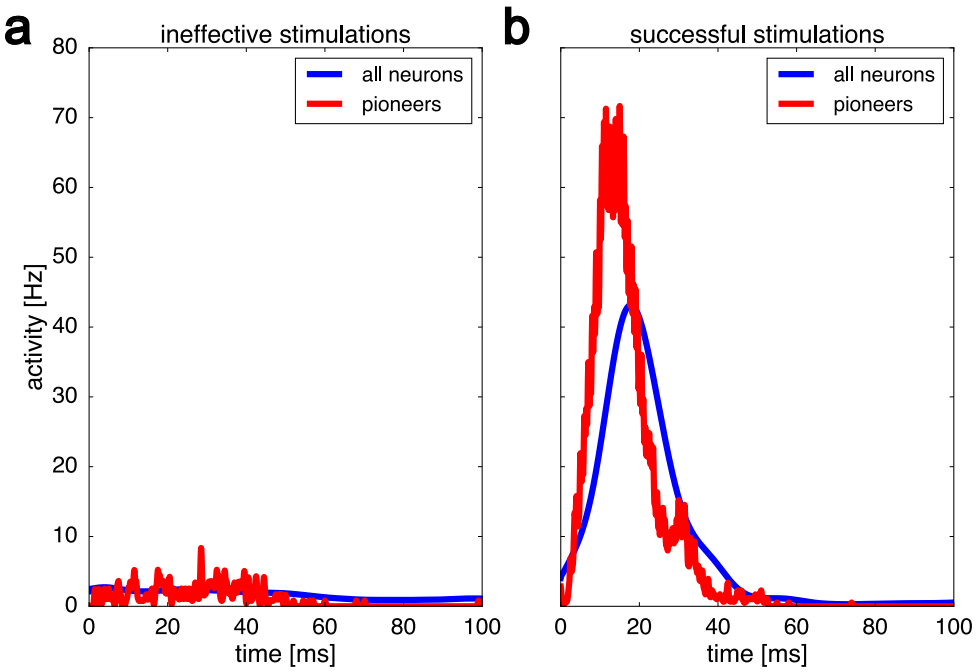

Figure 2.13 | Activity following successful and unsuccessful stimulation. Average activity of all excitatory neurons (ID 1 to 400) and of pioneer neurons (ID 260 to 320), over 100 ms following stimulation. **(a)** Unsuccessful stimulations. **(b)** Successful stimulations.

sheath') from the data so that only activity propagating deeper into the network is used to assess TFS and rank-order. This makes the classification task highly non-trivial.

I perform the same experiments with my *in-silico* network. I only consider

successful stimulations. Fig. 2.13 shows the time-dependent firing rate of the population of pioneers and of the entire population of neurons for both successful and unsuccessful stimulations. I consider information coding in two situations. The first situation has five different stimulation sites and readout is carried out by groups of ten neurons. The second situation has twelve stimulation sites and readout is done by groups consisting of 30 neurons.

I also remove the receptive sheath by not writing (directly) evoked individual spikes into the data files which enter the analysis. In addition, first spikes of NS which are evoked by stimulation are counted from the instance of stimulation on instead from the instance which corresponds to the proper fake latency (which is employed for spontaneous NS, see footnote 8).

The result is shown in Fig. 2.14 (see caption for details). One observes that rate-based schemes are impotent: the classification task is solved at chance level in this case. In contrast, the two temporal schemes are quite informative about the site where stimulation occurred, at least when pioneers are used for read-out. And again the TFS scheme does not provide more information than the spike-order scheme.

It is possible to interpolate between the neuronal-rates scheme and the TFS scheme. For this, I divide the entire time window about NS in bins of a fixed length T_{bin} and compute the average firing rate of each neuron in each bin. This gives rise to a matrix (r_{ij}) of rates where one index labels the bins and one index labels the neurons. For large bin size, this corresponds to the neuronal-rates scheme, and for small bin size this corresponds effectively to the TFS scheme (because each neuron typically deposits only *one* spike per NS), apart from the fact that the information of TFS is then presented less compactly. Performing the same experiment as before as a function of bin size yields the plot shown in Fig. 2.15. The information content is small for large bin size, consistent with what has been shown earlier. If the bin size decreases, the information content of the population of pioneers increases until a bin size of 5 ms is reached. From that point on, the information content formally decreases again, but this is presumably due to incomplete convergence of

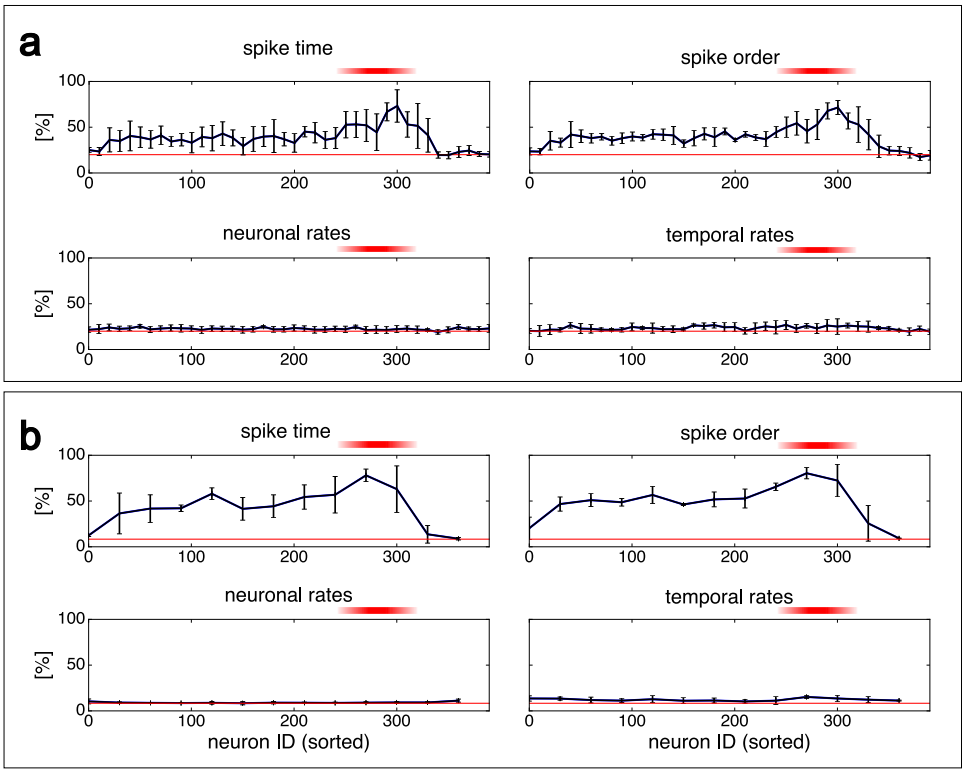

Figure 2.14 | Classification performance of different decoding schemes, based on different groups of neurons. (a) 5 stimulation sites and 10 readout neurons. **(b)** 12 stimulation sites and 30 readout neurons. Red shading marks the pioneer range (ID 260 to 320). Results for 'spike time,' 'spike order,' 'neuronal rates,' 'temporal rates' are shown separately. Percentage of correct classification $\alpha(N)$ of one of five (twelve) stimulated locations is shown, based on the activity of neurons with sorted ID $[N, N+10]$ (**a**) or $[N, N+30]$ (**b**). Chance performance is 20% in subplot **a** and 8.3% in subplot **b**.

the algorithm: the matrix (r_{ij}) is very large in this case and learning is more difficult.

A different point of view on this behaviour can be obtained by the following procedure (compare Shahaf et al. (2008)). I first measure the order O_A in which the neurons in a certain group G discharge during a certain NS NS_A. Then I measure, in which order O_B these neurons (i.e. again the neurons in group G) discharge in a different NS NS_B. Since the pioneers' discharge order is very informative of the site where stimulation occurs (as we have seen), I expect that the orders O_A and O_B are 'more similar' when G is the set (or a subset) of pioneers and when NS_A and NS_B are both caused by stimulation at the same site, than if NS_A and NS_B were evoked by stimulation at different sites. Fig. 2.16 C and D shows that this is precisely the

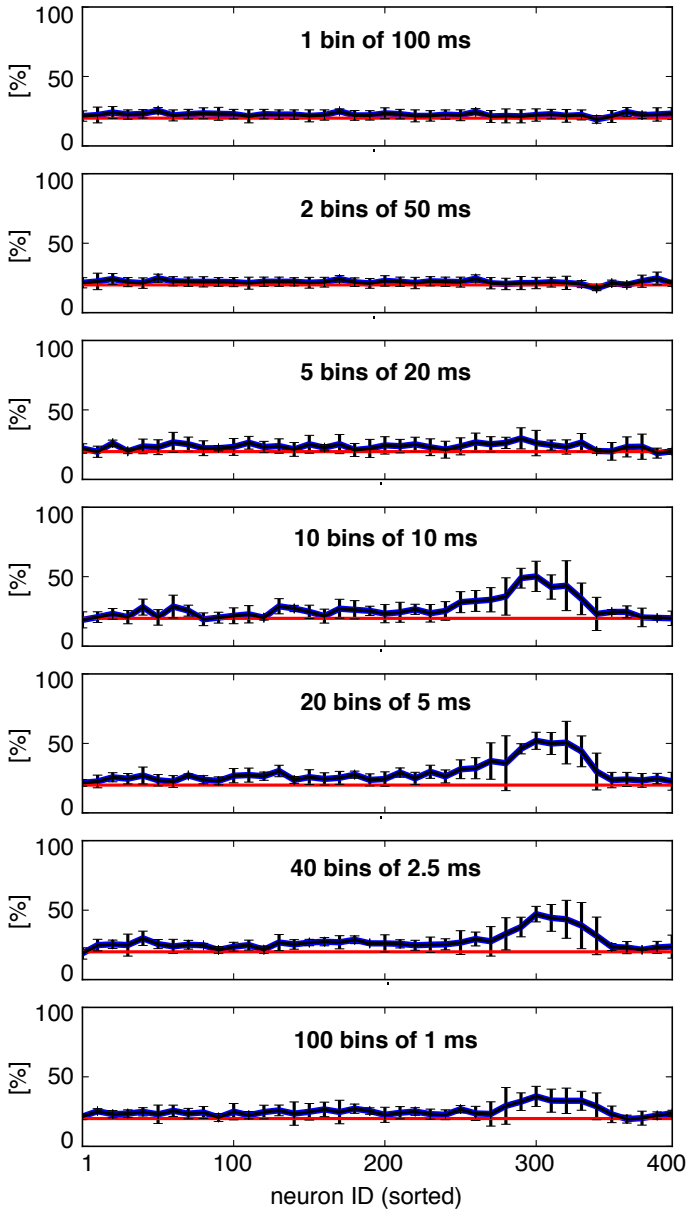

Figure 2.15 | Interpolation between rate-based and spike-based decoding. Here I only show the case of 5 stimulation sites with 10 readout neurons. The activity of groups of $n = 10$ neurons over 100 ms) was analysed in k time-bins, forming a rate vector of length $k \cdot n$. Decoding performance is shown for different groups of neurons and different bin sizes.

case. Here, I adopt a very simple measure for similarity of discharge orders based on (essentially) counting the minimal number of transpositions (pair-exchanging permutations) needed for transforming one spike order into the other (see appendix for

details). There is also a tendency that spike orders of pioneers prior to *spontaneous* NS are less similar than spike-orders of NS evoked by stimulation at one given site and that spike orders obtained from NS evoked at different sites are less similar than spike orders obtained from spontaneous NS. Figure 2.16 also shows (see subplots A and B) that spike orders are extremely stereotypical when the analysis is done using groups of neurons which contain only non-pioneers. Therefore, the pioneers' ability to encode information provided by differential spatial stimulation is due to differential sensitivity to stimulation at different sites.

To confirm these observations and to comprehensively compare all groups of excitatory neurons, I establish the distribution of spike-order similarity (SOS), both within and between classes of NS. As before, NS classes are understood to be 'spontaneous,' 'evoked at site 1,' 'evoked at site 2,' and so on. Given two distributions of SOS values, with means $\mu_{1,2}$ and standard deviations $\sigma_{1,2}$, I expressed their 'distance' in terms of the z-score, $z = |\mu_2 - \mu_1|/(\sigma_1 + \sigma_2)$. The results are shown in Figs. 2.16 E and F, for the two experiments with $k = 5$ and $k = 12$ stimulation sites, respectively. In both experiments, the difference between 'within-class' and 'between-class' distributions was most pronounced when SOS was computed for pioneers (ID 260 to 320). This demonstrates conclusively that the 'spike order' of pioneers is more informative about stimulation site than that of any other group of excitatory neurons.

2.5 Automatic detection of pioneers

For several reasons, it is desirable to have an 'automatic procedure' for detecting pioneer neurons. One of these reasons is that one may obtain a 'landscape' which shows how many pioneers exist in different networks as a function of the average weights[11]. Another more fundamental reason is that I defined pioneers rather vaguely. I looked at latency distributions and decided quite provisionally that neurons with a sorted neuron ID between 260 and 320 are pioneers. This will be rectified

[11] This is the topic of the next section.

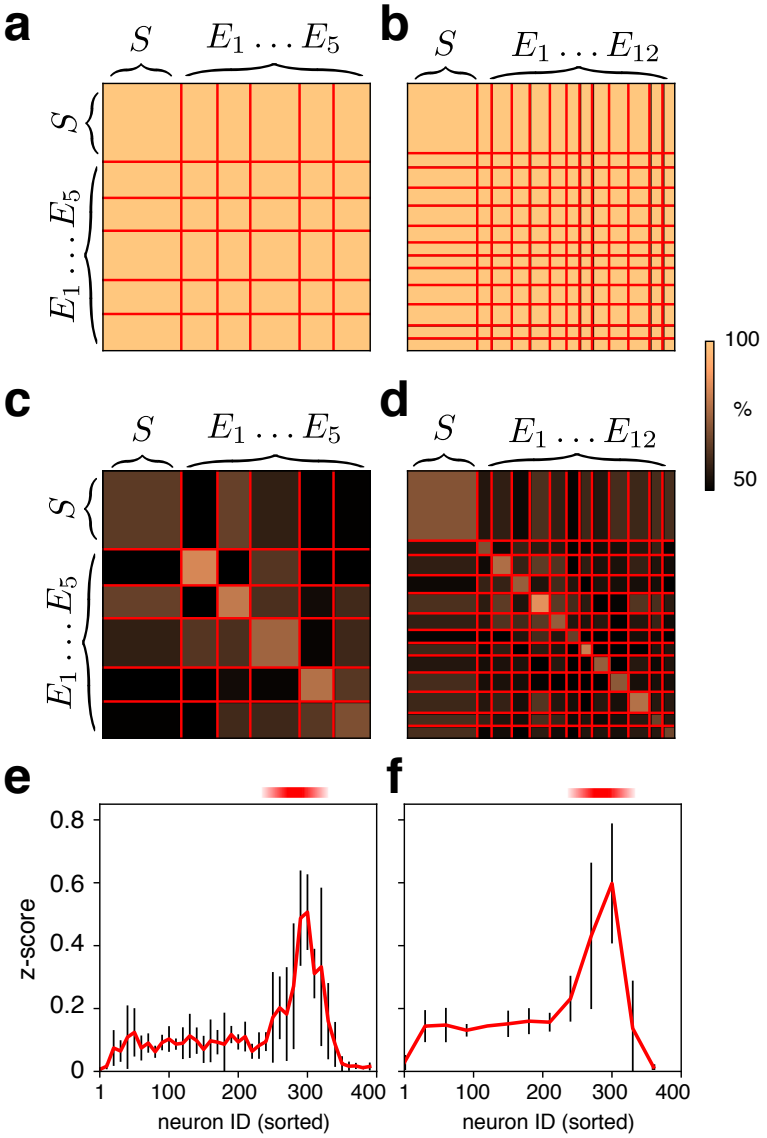

Figure 2.16 | Matrices of average 'spike order similarity' (SOS) during different NS. Observed NS were classified as 'spontaneous' (S), 'evoked at site 1' (E_1), 'evoked at site 2' (E_2), and so on, and class boundaries between sorted NS are marked by red lines. **(a-d)** Representative sets of non-pioneer or pioneer neurons, average SOS of pairs of NS (fraction of maximal similarity, colour scale) **(a)**: Non-pioneer SOS, five stimulation sites (k = 5 and n = 10). **(b)**: Non-pioneer SOS, twelve stimulation sites (k = 12 and n = 30). **(c)**: Pioneer SOS, five stimulation sites (k = 5 and n = 10). **(d)**: Pioneer SOS, twelve stimulation sites (k = 12 and n = 30).**(ef)** Distance between SOS distributions (mean and std. dev. of z-score), within-class and between-class, for sets of neurons with contiguous ID. **(e)**: Five stimulation sites and sets of ten neurons (k = 5 and n = 10, contiguous ID in range $[N, N + n - 1]$). **(f)**: Twelve stimulation sites and sets of thirty neurons (k = 5 and n = 10, contiguous ID in range $[N, N + n]$).

now.

Pioneers should be pioneers because (i) their average latency is negative, because (ii) the absolute value of the average latency is rather large and because (iii) the standard deviation of the latency is (relatively) small. In order to quantify this, I consider the *latency consistency*

$$1/CV(\tau) := -\frac{\langle\tau\rangle}{\Delta\tau}, \tag{2.1}$$

where $\langle\tau\rangle$ is the average latency and where $\Delta\tau$ is the standard deviation of the latency. Due to the minus sign, one would expect that pioneers have the highest latency consistency (i.e. large and positive).

The latency consistency as a function of the sorted neuron ID is plotted in Fig. 2.17A. A suitable threshold ($\theta = 1.5625$) for the latency consistency is plotted in red. One observes that the neurons between ID 260 and 320 are in fact above threshold. However, there are also many other neurons which are above threshold, both neurons which are less active and neurons which are more active than pioneers (as defined so far). In order to deal with this, I now also look at the (absolute value of the) coefficient of variation $CV(V^*) := \frac{\Delta V^*}{V_{thr} - \langle V^*\rangle}$ of the $*$-voltage. Since pioneers are distinguished by their characteristic sensitivity, I would expect that pioneers 'are also situated close to threshold,' so that $|CV(V^*)|$ is rather large. This is shown in the same figure (subplot B). The threshold for the 'voltage variability' $|CV(V^*)|$ plotted in red is $\theta = 0.64$. In order to show that my pioneers, as previously defined, are essentially identical with the neurons where both latency-consistency and voltage variability are above the respective thresholds, I plot the 'false alarms' and the 'misses,' counting (for simplicity) the criterion based on neuron ID as the 'true' measure. If this terminology (from signal detection theory) is accepted, I can declare that in Fig. 2.17 the misses are plotted in red and the false alarms in magenta. Obviously, my criterion (the thresholds) is chosen to be rather 'liberal' in that there are more false alarms than misses. In any case, the number of wrong classifications is in the order of 10, so one can see that the new 'operational' criterion for

(automatically) detecting pioneers essentially agrees with my provisional definition based on neuron ID. From now on, I will consider pioneers as those neurons which are pioneers according to the *new* criterion. In particular, I am now able to detect pioneers in an automated way, which will be important in the next section. My previously obtained results will remain essentially unchanged under this substitution of 'definition,' as shown in Fig. 2.17. Note that this Figure also shows that in principle considering the $*$-voltage would be sufficient to detect pioneers, at least in the example shown. However I do not treat this as being conclusive: in 'arbitrary' networks a neuron might be close to threshold without being a pioneer, so in making landscapes (as I want to do now) I am on the safer side if I also include a condition based on latency-consistency. If this were not needed in certain cases, it would do no harm.

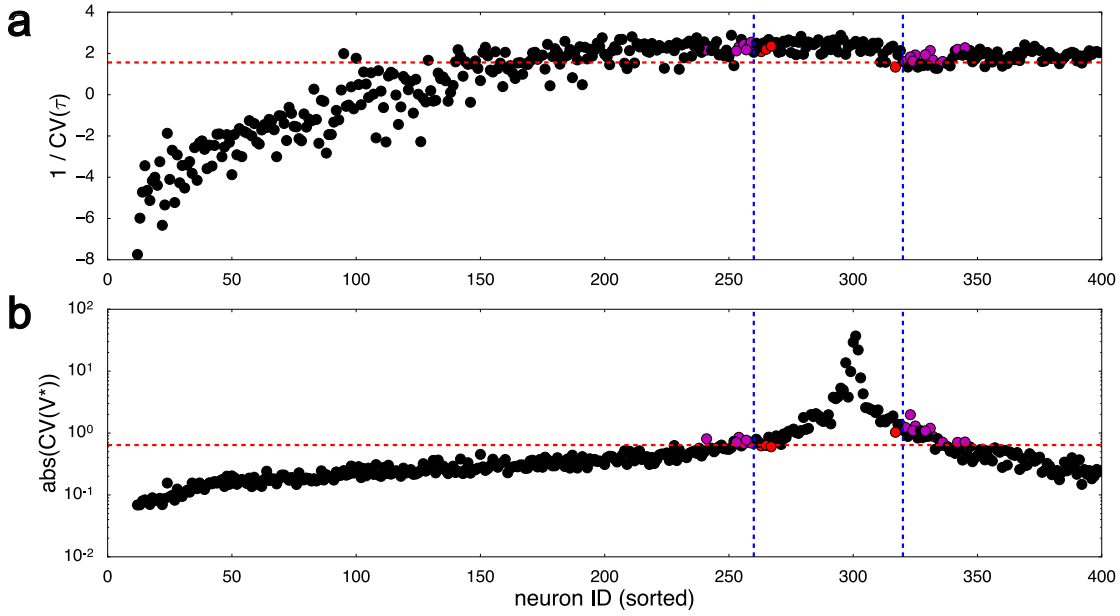

Figure 2.17 | Latency consistency and voltage variability for all excitatory neurons. (a) Latency-consistency (LC) as a function of sorted neuron ID. The range of pioneers, as previously defined, is marked by vertical blue lines. The LC-threshold for the new criterion (for detecting pioneers) is the horizontal red line. Misses are plotted in red and false alarms in magenta. For all black dots the two definitions (of being a pioneer) agree. **(b)** Same as **a**, but the absolute value of the voltage variability is considered. Note that in classifying dots as 'correct,' 'false alarms,' and 'misses,' I consider (in both subplots) the *conjunction* of the two threshold-based criteria.

2.6 The dependence on recurrent excitation and inhibition

All what I have described so far is phenomenology which applies in *particular* networks with a certain choice of parameters. It would thus be interesting to obtain a 'landscape' which shows how the phenomenology depends on the parameters, in particular the choices for the average weights. In principle, there are four average weights: excitatory-to-excitatory, excitatory-to-inhibitory, inhibitory-to-excitatory and inhibitory-to-inhibitory. Varying them all would however be overkill, in a sense, and not in practice realisable. Following Gigante et al. (2015), I varied only excitation and inhibition jointly, so that the average weights[12] are parametrised as $\omega_{0,ee} = r_E \overline{\omega}_{0,ee}$, $\omega_{0,ie} = r_E \overline{\omega}_{0,ie}$, $\omega_{0,ei} = r_I \overline{\omega}_{0,ei}$ and $\omega_{0,ii} = r_I \overline{\omega}_{0,ii}$, where the over-lined quantities are certain convenient reference values. The scaling factors r_E and r_I are varied between 0 and 2. The entire procedure is carried out for all three network topologies, and the (over-lined) reference values of the average weights are identical in all three cases. In order to get a wide 'image' of the macroscopic activity, I measure the average interval between NS (T_{INSI}, INSI = 'inter-network-spike interval'), the coefficient of variation of the INSI (CV_{INSI}) and the ratio A_{max}/A_{mean} between the maximum A_{max} of the time-dependent excitatory population activity and its temporal mean A_{mean}. This ratio is a simple dimensionless measure of how all-or-none the synchronisation events are. Note that for many choices of the parameters, NS will not be all-or-none, due to a different balance between excitation and inhibition. In these cases, the statistics of the INSI simply measure the statistics of 'relatively large' events, such as avalanches.

The result is shown in Fig. 2.18 (see also Gigante et al. (2015)). Note that in creating this figure, I create several ($\mathcal{O}(10)$) realisations at each point in the landscape and plot the mean in each ensemble of individual realisations. This procedure enables me to assess also the stability of the phenomena at each point in the landscape.

[12] In my networks, the weights Δg_{ij} associated with connections are drawn from a clipped Gaussian distribution with certain mean values (see appendix). These mean values are denoted as $\omega_{0,xy}$, where $x, y \in \{e, i\}$.

Qualitatively, the activity is divided in two regions: one above the blue dashed lines and one below them. In the region *around* the blue lines, the avalanches are separated by large intervals, they occur rather irregularly and the activity ratio A_{max}/A_{mean} is large. This is the (super-critical) regime where NS occur (in an all-or-none fashion). From what is known for single neurons, I expect that increasing excitation or decreasing inhibition at this point leads to tonic activity, i.e. the synchronisation events are large, all-or-none and very frequent. This is confirmed by looking at the bottom-right corner of the subplots in Fig. 2.18. There, the average INSI interval is small, the CV_{INSI} is very small. Also, due to the large frequency of the synchronisation events, A_{mean} increases so that the activity ratio becomes small.

Decreasing excitation or increasing inhibition should lead to a regime where avalanches are not all-or-none, occur more frequently than in the default regime (defined by the blue dashed curves), occur irregularly, and where the activity ratio is small due to the smallness of A_{max}. As one can see, this is precisely the case. Also, there is no qualitative difference between the different topologies at this macroscopic level.

So far, I have only reproduced results by other researchers and one has only seen what seems rather obvious. The next step is to count how many pioneers exist (using the method described in the previous section) at the points in the landscape which are denoted by red dots.

The result is shown in Fig. 2.19. As one can observe (subplot A), there are consistently much more pioneers in the heterogeneous network. Subplot B shows that the fraction of realisations where NS occur are also generally higher in general. Thus the entire phenomenon (NS and existence of pioneers) is much more stable and consistent in the heterogeneous case. I give an explanation for this in the next chapter. Note that having more pioneers potentially also leads to a more pronounced information encoding by the rank-order of pioneers. This gives a belated justification for the heterogeneous topology which I construct.

Fig. 2.19 also shows that different topologies express highly disparate dynamics for identical (r_E, r_I) values (see subplot C). For example, NS rates differ up to fivefold

between topologies.

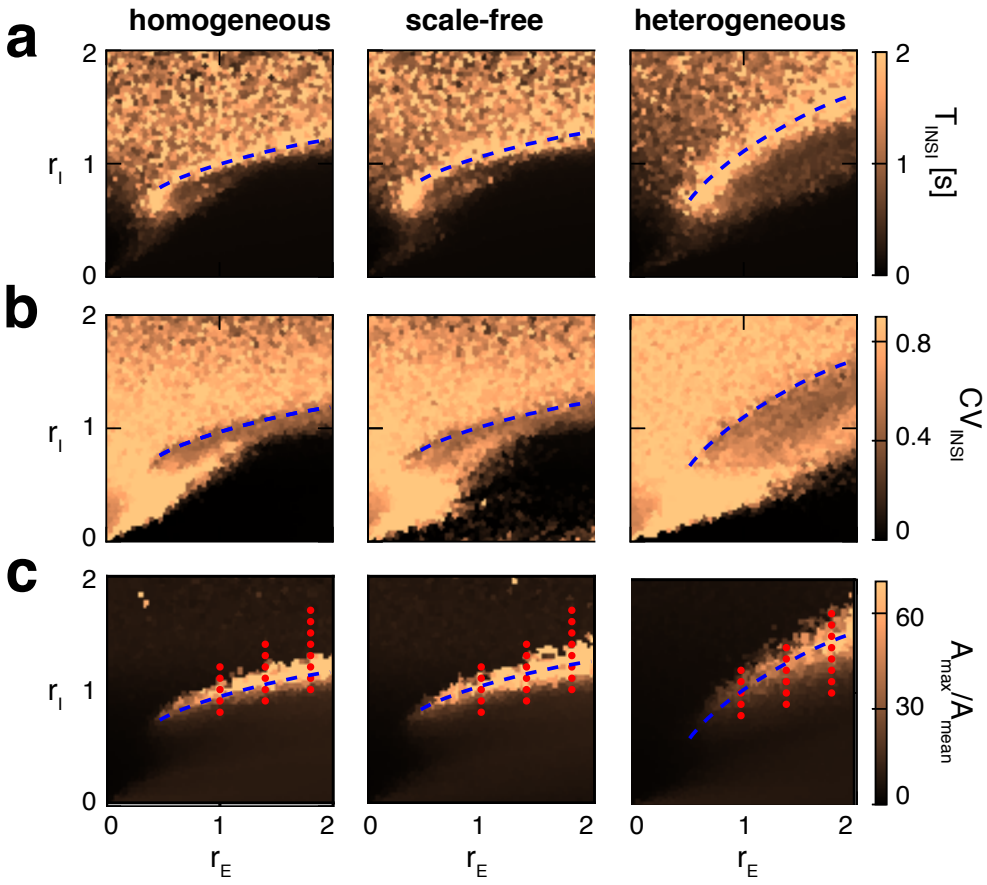

Figure 2.18 | Macroscopic dynamics, excitation/inhibition strength, and type of connectivity. Dynamical characteristics of synchronisation events, for different types of connectivity and for different absolute and relative strengths of excitation, r_E, and inhibition, r_I. Blue, dashed curves indicate the transitional region which, for each type of connectivity, separates the excitation-dominated regime of 'tonic' dynamics from the inhibition-dominated regime with 'asynchronous' and 'avalanche-like' dynamics (see text). Red dots mark the (r_E, r_I) value pairs further investigated in Fig. 2.19. **(a)** Average interval between NS, T_{INSI}; **(b)** Coefficient of variation of interval, CV_{INSI}; **(c)** Activity ratio, A_{max}/A_{mean}.

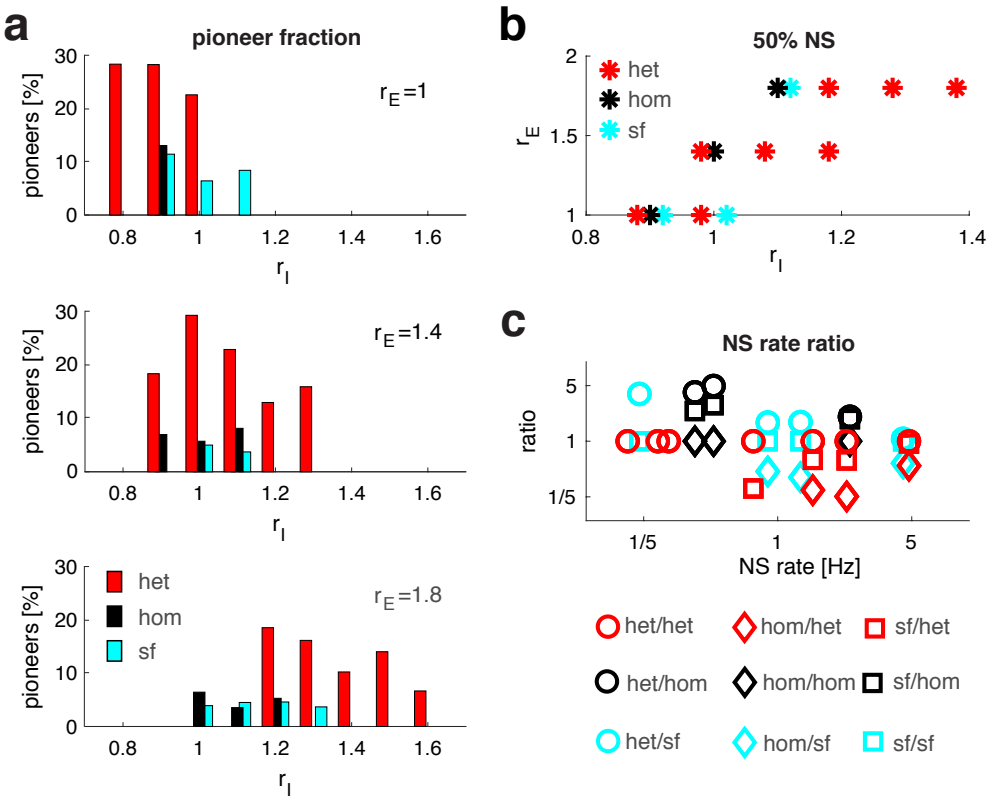

Figure 2.19 | Microscopic and macroscopic dynamics at selected connection strengths r_E, r_I. Fraction of pioneer neurons and NS rates in networks with $r_E \in \{1.0, 1.4, 1.8\}$ and $r_I \in \{0.8, 1.0, \ldots, 1.6, 1.8\}$ (red dots in Fig. 2.18c). **(a)** Percentage of pioneer neurons in networks of different topology, for selected values (r_E, r_I). (see text). **(b)** Consistent NS in different network realisations. Value pairs (r_E, r_I) that produced NS in $\geq$ 50% of realisations. **(c)** Disparity of NS rates in networks with different topologies, but identical values of (r_E, r_I). Ratio of NS rates, sorted by reference rate, for all ordered topology pairs. For example, at three identical value pairs (r_E, r_I), NS rates f_{hom} and f_{het} could be established for homogeneous and heterogeneous networks. Ratios f_{het}/f_{hom} are shown against f_{hom} (black circles) and ratios f_{hom}/f_{het} against f_{het} (red diamonds).

Chapter 3

Mechanisms in silico

In this chapter, I shall analyse the pioneer-effect in more detail. In particular, it is my aim to link the existence of pioneers to the heterogeneous topology. In order to do so, I have developed a two-step-procedure. My argument is that the heterogeneity, in particular the degree-distribution, shapes the appearance of the $*$-voltage-statistics of individual neurons in the network (see section 3.1), and that, in turn, the $*$-voltage-statistics shapes the first-spike-histograms (i.e. latency-distributions), see section 3.2.

3.1 Voltage-statistics from degree-statistics

My first observation is that in all three of my networks, the mean and the standard deviation of the $*$-voltage of each neuron are determined approximately by the degrees (and the average *ee*-weight) as follows:

$$\left\langle V^* \right\rangle \approx -0.12\ \mathrm{mV} \cdot D_{inh} - 48.3\ \mathrm{mV}, \tag{3.1}$$

$$\mathrm{std}(V^*) \approx 2.1 \cdot 10^7 \frac{V}{S} \cdot \sqrt{D_{exc}} \cdot \omega_{0,ee}. \tag{3.2}$$

Here, D_{exc} is the number of excitatory neurons which are pre-synaptic to the given neuron (*excitatory in-degree*), and D_{inh} is the number of inhibitory neurons which are pre-synaptic to the given neuron (*inhibitory in-degree*). These approximate equations are verified in Figs. 3.1 D and E, respectively. Of course, these equations are *approximate* and *phenomenological* equations. I do not expect these equations to hold exactly in general networks, where the mean and the standard deviation of the $*$-voltage depend on many factors, such as *both* degrees, the ISI-statistics of the pre-synaptic spike trains, etc.

In order to verify that, qualitatively, the *statistics* of the $*$-voltage is determined

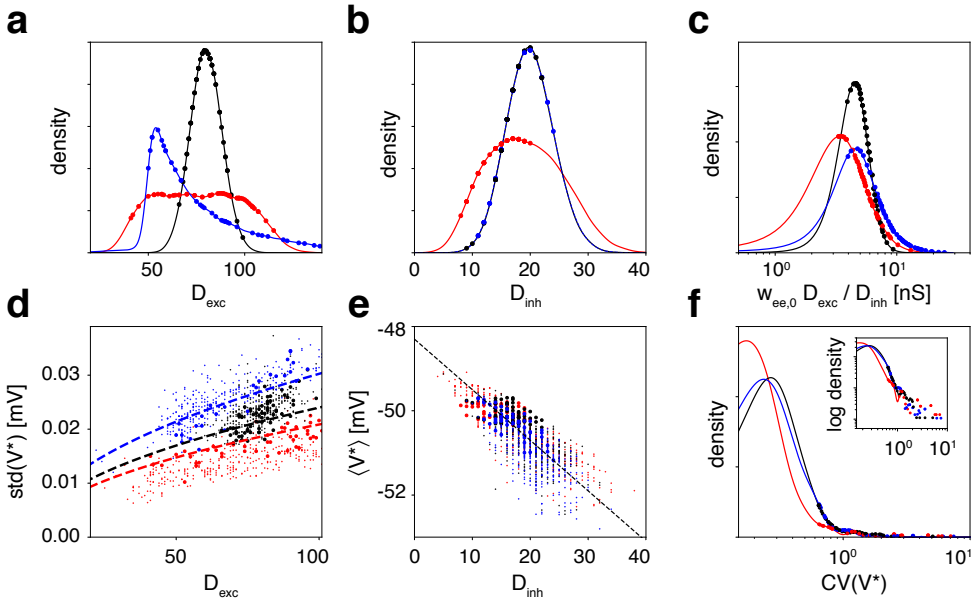

Figure 3.1 | Afferent connectivity and $*$-voltage $V^\star$ during periods without NS, for different network topologies. Colours distinguish heterogeneous (red), homogeneous (blue), and scale-free topology (black). Symbols mark pioneer neurons in a representative realisation. **(a)** Distribution of afferent excitation, D_{exc}, averaged over multiple network-realisations. **(b)** Distribution of afferent inhibition, D_{inh}, averaged over multiple network-realisations. **(c)** Distribution of 'effective afference ratio', $\omega_{ee,0} D_{exc} / D_{inh}$, averaged over multiple network-realisations. **(d)** Dependence of the standard deviation of $*$-voltage, std $\left(V^\star\right)$ on D_{exc}. Regression curves indicate proportionality std$(V^\star) \approx 2.1 \cdot 10^7 \frac{\mathrm{V}}{\mathrm{S}} \cdot \sqrt{D_{exc}} \cdot \omega_{ee,0}$. **(e)** Dependence of mean $*$-voltage, $\langle V^\star \rangle$, on D_{inh}. Regression line indicates proportionality $\langle V^\star \rangle \approx -0.12\ \mathrm{mV} \cdot D_{inh} - 48.3\ \mathrm{mV}$. **(f)** Distribution of sensitivity, CV$(V^\star)$, with right tail on logarithmic scale (inset).

by the *statistics* of the degree distributions in the above way, I plotted the mean and standard deviation of the $*$-voltage for each neuron in an ensemble of 400 neurons in two ways. The red dots in Fig. 3.2 show for all 400 excitatory neurons

in one network of each type the mean and the standard deviation of the $*$-voltage as *measured*. The black dots are as numerous as the red dots, and for each black dot a value for D_{exc} and for D_{inh} has been obtained by sampling (independently[1]) the measured degree-distributions as shown in Fig. 3.1 A and B, and computing the mean and standard deviation of the $*$-voltage according to equations 3.1 and 3.2. Fig. 3.2 shows that the measured distributions and the 'theoretical' distributions based on degree-sampling agree qualitatively. Thus I have effectively deduced the voltage-statistics from the degree-statistics.

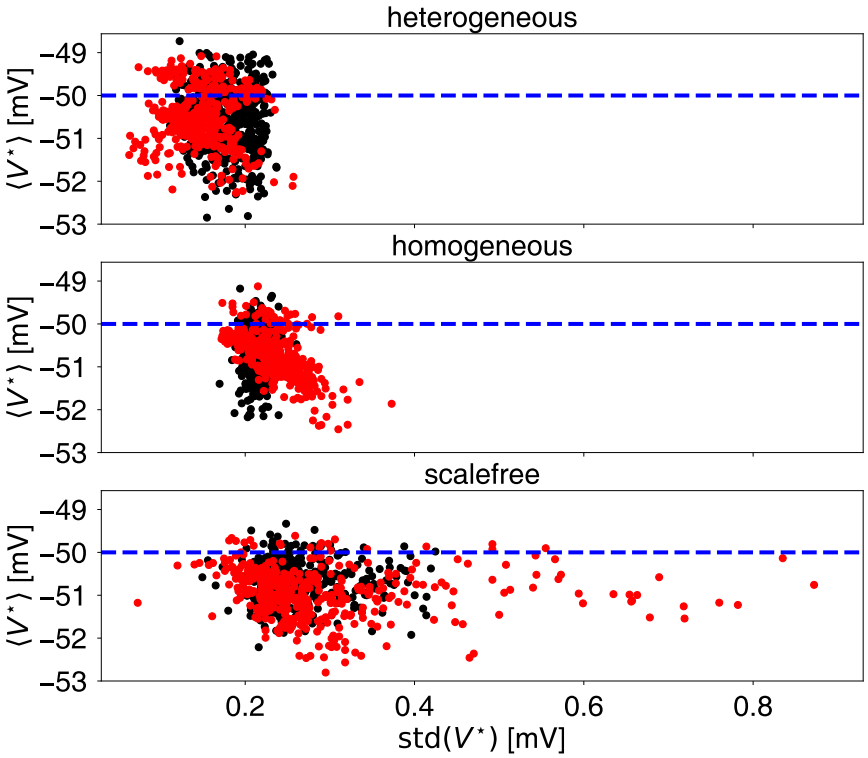

Figure 3.2 | Distribution of mean and standard deviation of $*$-voltage (theory vs. observed). Red dots show individual means and standard deviations of $*$-voltages as observed, black dots show distribution expected theoretically (see main text for details).

[1] In all three topologies, there is no correlation between D_{exc} and D_{inh}.

3.2 Latency-statistics from voltage-statistics

In order to obtain latency-statistics from voltage-statistics, one needs a theory for the latencies. I employed the following 'toy theory.' The $*$-voltage distribution of neuron i is approximately Gaussian:

$$p_i(V) = \frac{1}{\sqrt{2\pi\Delta V_i^{*2}}} \exp\left(-\frac{(V-\langle V\rangle_i)^2}{2\Delta V_i^{*2}}\right), \tag{3.3}$$

I now assume that, during NS, this distribution undergoes a 'rigid motion' which is described by an 'activation function' u:

$$p_i(t,V) = \frac{1}{\sqrt{2\pi\Delta V_i^{*2}}} \exp\left(-\frac{(V-\langle V\rangle_i - u(t))^2}{2\Delta V_i^{*2}}\right), \tag{3.4}$$

where

$$u(t) = \alpha \exp\left(-\frac{t^2}{2\sigma_{NS}^2}\right), \tag{3.5}$$

and where $\alpha = 0.56$ mV and where $\sigma_{NS} = 23$ ms. These parameters are rather arbitrary and have been chosen so that there is a good qualitative agreement between the theory (which I am about to develop) and results obtained from simulations.

The next step is to compute the probability flux $F_i(\tau)$ through threshold for each neuron:

$$F_i(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau}\int_{V_{thr}}^{\infty} p_i(\tau,V)\mathrm{d}V = \dot{\mu}(\tau)\exp\left(-\frac{[V_{thr}-\langle V\rangle_i - u(\tau)]^2}{2\Delta V_i^{*2}}\right), \tag{3.6}$$

which I want to use as an approximation to the latency-distribution. However, this quantity can become negative. For this reason, I half-wave-rectified it and re-normalised it (to unity). This yields the following approximation for the latency distribution:

$$p_{lat,i}(\tau) = N[\dot{\mu}(\tau)]_+ \exp\left(-\frac{[V_{thr}-\langle V\rangle_i - u(\tau)]^2}{2\Delta V_i^{*2}}\right), \tag{3.7}$$

where N is a normalisation constant. This distribution can be computed (for each neuron) as soon as the mean and the standard deviation of the $*$-voltage is known.

The result is shown in Fig. 3.3. Subplots A and C show the mean of the latency vs. the variability of the $*$-voltage, and subplots B and D show the latency-consistency (LC) vs. the variability of the $*$-voltage. Subplots A and B show the quantities as *measured*, and subplots C and D shows the same as subplots A and B, however with the *predicted* latency-statistics. Subplot A shows that the mean latencies are much broader and more negative in the heterogeneous case. The pioneers (fat dots) are the neurons with the most negative latencies. There are fewer pioneers in the homogeneous networks, and the few pioneers in the homogeneous networks have smaller latencies. Subplot B shows that LC is also generally higher in the heterogeneous case. Note that I consider here the same three networks for which I presented the phenomenology in the respective earlier chapter. I.e., I consider here the three networks with *different* parameters, but at a point where the macroscopic dynamics is similar. In Fig. 2.19 on the other hand, I compared networks with *identical parameters*. Thus the demonstration that there are also more pioneers in the heterogeneous network when the networks are compared at points where the *macroscopic dynamics is similar* (but the average weights different) has been missing so far.

Now the more interesting part of Fig. 3.3 are, of course, subplots C and D showing the theory. However, there is a large quantitative deviation between subplot A and C on the one hand and subplot B and D on the other. In order to explain this, remember that the activation function $u(\tau)$ was *fixed*, i.e. identical in all three topologies. I expect that having more pioneers is reflected partially by a broader shape of the activation function $u(\tau)$. However, if I made $u(\tau)$ broader in the heterogeneous case, I would put in the pioneers *by hand*, which I want to avoid. In principle, the 'toy theory' would have to be iterated: compute latency-distributions from a *fixed* activation function first, then compute a new activation function from the obtained latency-distributions, obtain updated latency-distributions, compute a new activation function based on these, etc. I only show the first iteration in this iterative procedure. It is highly interesting and non-trivial that already in this *first iteration*, qualitative differences appear between the three topologies, as shown

in subplots C and D. As is visible there, the *qualitative appearance* of subplots A and B is reproduced. And presumably, iterating the 'toy theory' would increase the qualitative agreement. Since the activation function was identical in all three topologies, I conclude that the characteristic appearance of the $*$-voltage statistics *is the cause for the broader latencies and the higher LC in the heterogeneous case*, and thus is *also the cause for the broader shape of the 'actual' activation function.* Since I could reproduce the $*$-voltage statistics from the degree-distributions (section 3.1), this means that there exists a direct link between the heterogeneity and the stability of the pioneer-effect.

3.3 Chains of sensitivity

In the last two sections, I have shown how the characteristic heterogeneous degree-distribution leads to a characteristic distribution of the $*$-voltage statistics, and how the latter leads to the enhanced pioneer-effect. However, the presentation was somewhat formal. I have not said *precisely* what *property* of the $*$-voltage statistics is responsible for the more pronounced pioneers. I will now attempt to trace the effect back.

This is based on *chains of sensitivity*. The recruitment process during NS presumably works as follows: First, neurons (pioneers) with a high sensitivity S_1 are recruited. The spikes of these neurons raise the population activity even further, so that neurons with a sensitivity $S_2 = S_1 - \Delta S$ are recruited. Their spikes make neurons with sensitivity $S_3 = S_2 - \Delta S$ discharge, and so on. So in order to obtain a *broad* recruitment process, the number of neurons with sensitivities in each interval $[S_2, S_1]$, $[S_3, S_2]$ has to be relatively small. If I measure the sensitivity by means of the variability $CV(V^*)$ of the $*$-voltage, I see that one needs a relatively broad distribution of $CV(V^*)$. Or equivalently, since a broader shape of this distribution means that the curve assumes generally *smaller* values, I expect that the distribution of voltage variabilities assumes smaller values in the heterogeneous case, so that in

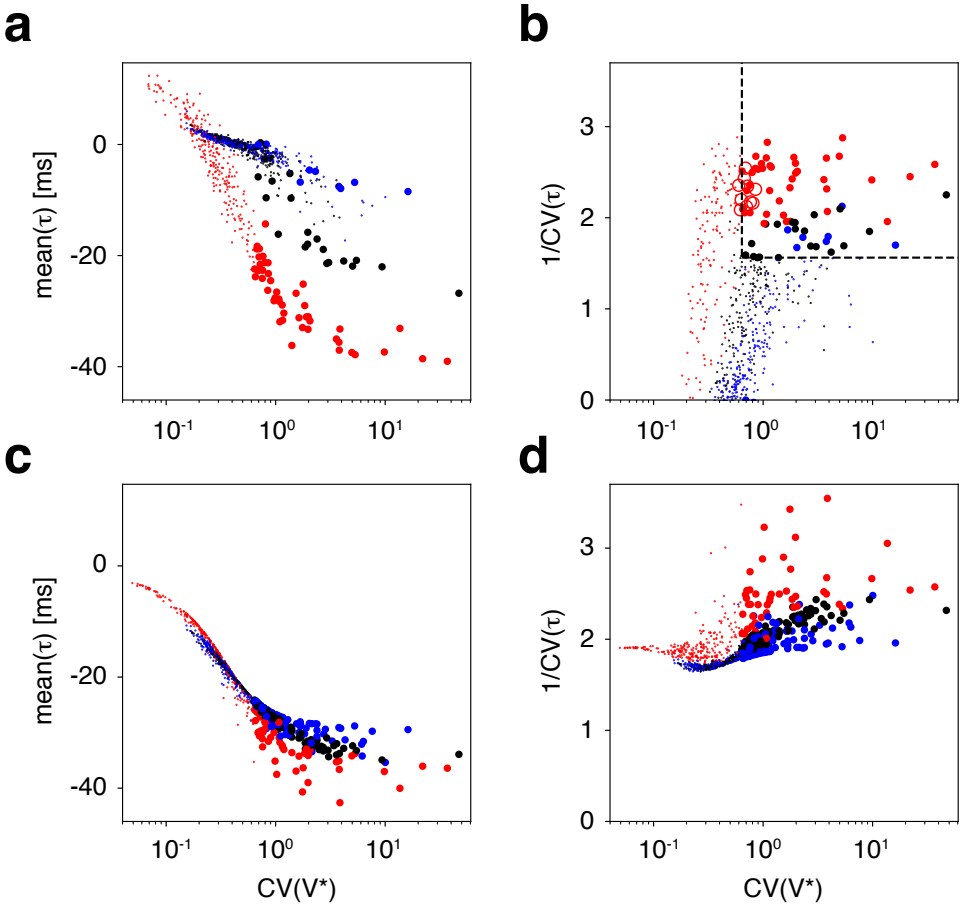

Figure 3.3 | Spiking behaviour and sensitivity compared for different network topologies. Colours distinguish neurons of heterogeneous (red), scale-free (black), and homogeneous networks (blue). Pioneers are marked by large dots, other neurons by small dots. **(ab)** Full effects of network topology (including different time-courses of population activity). **(a)** Comparison of mean latency $\langle\tau\rangle$ during NS initiation and sensitivity CV($V^\star$) during periods without NS. **(b)** Comparison of consistency of latency $1/\mathrm{CV}(\tau)$ and sensitivity CV($V^\star$). The sensitivity-consistency criterion (dashed lines) for pioneers and the sorted-ID criterion largely overlap (discrepancies are marked by red circles). **(cd)** Isolated effects of network topology (for identical time-course of population activity). Corresponding comparisons of latency, consistency, and sensitivity when an identical, rigid time-course of population activity is prescribed for all network topologies.

fact in each of the above 'sensitivity intervals' there are fewer neurons. Fig. 3.1 F shows that this is indeed the case. From there it is seen that the distribution of voltage variabilities peaks at smaller values in the heterogeneous case, hence assumes generally smaller values than the 'homogeneous distributions' for moderate sensitivities, and has a tail at high sensitivities which is as pronounced as in the other two cases (see inset).

Thus I can summarise the influence of the heterogeneous topology in one sentence: *Broad degree-distributions lead to broad distributions of voltage variabilities, which leads to broad chains of sensitivity, where in each sensitivity interval* $[S_2, S_1]$, $[S_3, S_2]$, ..., *there are fewer neurons. Thus, at each step in the recruitment process, fewer neurons are recruited in the heterogeneous case, and hence the entire recruitment process is broader and slower.*

Chapter 4

Inferring strong synapses from activity measurements

A frequent problem in neuroscience is to obtain knowledge about a biological neural network from activity measurements (see, e.g., Stetter et al. (2012), Mishchencko et al. (2011)). This is a difficult 'reverse-engineering' problem. Here, I show how strong synapses between units in an *in vitro* network can be detected from the measurement of neural activity. I assume here that the network shows a super-critical activity consisting of large bursts interrupted by periods of near-silent activity, as is often observed. My strategy is to describe and verify a procedure *in silico*, which then may be also applied *in vitro*, since all quantities needed can be measured in the latter case. As we will see, the method can be applied whenever time is 'absolute,' so that first-spikes can be defined. More generally, the procedure applies whenever a distinguished spike (e.g. first-spike) exists for each neuron in the population which is recorded. For simplicity, however, I assume in the sequel that there exists a network spike, so that the 'distinguished' spike of each neuron is the first spike within a NS (as earlier).

There are several ways to 'compute' the probability that neuron i in the network will deposit its first spike *before* unit j during NS. The most appropriate way to do

so is to measure this quantity directly. Thus I look at all NS and consider during how many of them neuron i fires before neuron j. I call the resulting probability $P_{i\to j}$ and I record this for all neuronal pairs (i, j) of which both neurons rarely discharge between NS, so that it is meaningful to talk about first spikes.

There is a second way to 'compute' these precedence probabilities. I record for each neuron i when it discharges relative to NS and estimate from this a distribution $p_{lat,i}$ of latencies. This is done for all neurons. If neurons i and j were *entirely independent*, I would expect that the precedence probability is equal to the 'convolution' $Q_{i\to j}$

$$Q_{i\to j} = \int_{-\infty}^{\infty} p_{lat,i}(\tau)\left(\int_{\tau}^{\infty} p_{lat,j}(\sigma)\mathrm{d}\sigma\right)\mathrm{d}\tau. \tag{4.1}$$

The ratio $R_{i\to j} := P_{i\to j}/Q_{i\to j}$ (more precisely: its deviation from unity) is therefore a measure for the mutual dependence of neuron i and j. If there is a strong synapse from i to j, and if neuron j fires usually after i, I expect that the difference $t_j - t_i$ of the latencies of the first-spikes (t_j and t_i, respectively) from j and i is smaller (than if the synapse were weak). The reason is that discharge of neuron i strongly excites neuron j and thus it is likely that the latency of the first spike of neuron j is decreased, so that the difference between the latencies also decreases. This increases the overlap between the latency-distributions of neuron i and j and thus decreases the above convolution $Q_{i\to j}$. On the other hand, a strong synapse between neuron i and j can*not* influence the true precedence probability $P_{i\to j}$ (assuming that i discharges first in most cases), because at the point where neuron i discharges, the order of the spikes is pre-determined, and the strong synapse from i to j on the other hand can only show its effect when i already discharged, so that the subsequent excitation of neuron j cannot alter the order. All this shows that the ratio of both precedence probabilities should be indicative of strong synapses. I show below that this is actually the case.

I now determine the joint distribution $p_{forward}(R, \Delta g)$ of $R_{i\to j}$ and $\Delta g_{i\to j} (= \Delta g_{ji})$, which is the joint distribution of my causality-measure and the weight from neuron i

to j. The ensemble of this distribution is the ensemble of all excitatory-to-excitatory synapses in the network where the weight is non-zero. This distribution is labelled with a 'forward,' because it measures the 'forward' causality where the ratio $R_{i\to j}$ is the one from i to j and the weight is also the one from i to j, i.e. $\Delta g_{i\to j}$. There is a 'control case,' the 'backward direction,' where I determine the joint distribution $p_{backward}(R,\Delta g)$, which is the distribution of the joint probability of $R_{i\to j}$ and $\Delta g_{j\to i}(=\Delta g_{ij})$. I expect that the forward distribution is significantly elevated when R and Δg are relatively large, and I do not expect such an effect in the backward direction.

Now, if Δg and R are correlated, I expect the 'correlation' $\frac{p(R,\Delta g)}{p(R)p(\Delta g)}$ to be different from unity. Here, the two distributions in the denominator are the respective marginal distributions. In order to make 'distance to unity' and thus correlations visible, I now plot the logarithm of this correlation as a function of R and Δg. This is shown in Fig. 4.1 once in the forward direction and once in the backward direction. In the forward direction, a strong positive correlation between R and Δg is visible, which is not there in the backward direction. This establishes the above claims.

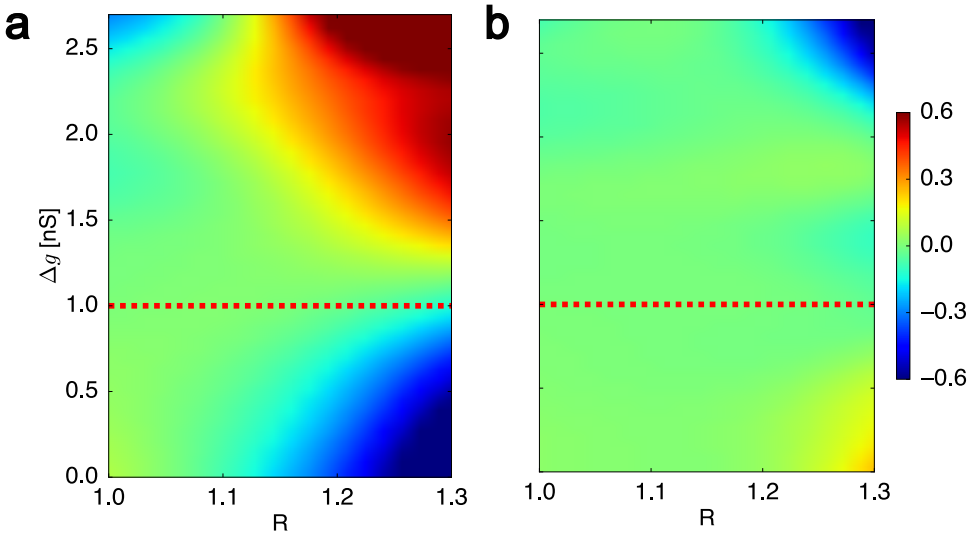

Figure 4.1 | Correlation between R and Δg. Natural logarithm of the ratio $\frac{p(R,\Delta g)}{p(R)p(\Delta g)}$, as a function of R and Δg, whose deviation from zero shows correlations between R and Δg. **(a)**: forward direction; **(b)**: backward direction. In the 'forward-case,' large weights are correlated with strong causality. The dashed red line is the value of the average weight.

Chapter 5

Summary, discussion and outlook

In this thesis, I have described heterogeneous, unstructured, random networks of LIF neurons connected with frequency-dependent and conductance-based synapses. Such networks give rise to a 'super-critical' macroscopic activity with network-wide bursts ('network spikes') interrupted by phases of near-silence, in agreement with findings *in vitro* (Marom and Shahaf (2002)), and resembling activity described by Tsodyks and colleagues (Tsodyks et al. (2000)). The *microscopic* dynamics manifests itself by privileged 'pioneer neurons' which herald network spikes and whose spike order is indicative of the source of stimulation, as is also observed experimentally (Eytan and Marom (2006); Shahaf et al. (2008); Kermany et al. (2010)). I believe that I analysed the phenomenology and mechanisms involved in the 'pioneer-effect' in great detail, describing a plethora of effects *in silico*, where temporal and spatial resolution are not limited.

Specifically, I described an unstructured network with all-or-nothing synchronisation events (NS), where the NS are terminated by synaptic depression (as observed experimentally, see Eytan and Marom (2006)), and with a macroscopic activity as shown in Fig. 2.2. Three network topologies were considered: a homogeneous

network of Erdös-Rényi type, a heterogeneous network where each neuron has its own connection probability, and a 'scale-free' network which serves to interpolate between the former two networks (Fig. 2.1). The three networks were designed with different average weights, so that the activity was comparable in terms of existence of NS and NS rate. Additionally, I verified that NS are all-or-none events (Fig. 2.4). The next step was to stimulate the network in order to evoke NS. A procedure for classifying NS as spontaneous or evoked was introduced (based on Gigante et al. (2015)). The microscopic dynamics manifests itself by the existence of pioneer neurons whose latency distribution is such that these neurons discharge consistently early (Fig. 2.9 A). The reason for this behaviour is their characteristic sensitivity, as measured by the $*$-voltage distribution (Fig. 2.9 B). In addition, I reproduced an experiment by Tsodyks and colleagues, where groups of neurons are silenced based on their average firing rate. I observed that making pioneers ineffective abolished NS (Tsodyks et al. (2000)). In order to explain the paradoxical result that NS are observed again when neurons *more active* than pioneers are silenced, I measured the threshold for NS generation and showed in this way that this threshold is increased when pioneers are silenced (Fig. 2.11). Along these lines, I also showed that pioneers are *not* distinguished by a particularly large excitatory in-degree (Fig. 3.1A). In our work, the only visible correlation between 'degree,' on the one hand, and 'being a pioneer,' on the other, is that the inhibitory in-degree does not assume large values for pioneers (Fig. 3.1B).

Then, I studied how the dynamics influences representational capabilities by studying how information (provided spatially) is encoded in my networks. I reproduced earlier experiments *in vitro* (Kermany et al. (2010)) by showing that the discharge order of pioneers is indicative of where a stimulation occurred. Rate-based schemes were not able to to so, in contrast (Fig. 2.14). I confirmed this result by computing matrices of rank-order similarities (Fig. 2.16).

Next, I introduced a technique for automatically detecting pioneers. The reason is that I needed an operational criterion in order to assess how the existence of pioneers depends on the network parameters. The latter was than determined in

the form of a landscape (Fig. 2.18). The operational criterion for being a pioneer was based on the conjunction of two criteria: pioneers have a high latency consistency and a high voltage variability.

In chapter 3, I attempted to understand the mechanism behind the connection between heterogeneity on the one hand and the abundance of pioneers on the other. I derived the characteristic distribution of $*$-voltage statistics in the three networks from the respective degree distributions, and, in turn, derived – using a simple 'toy' theory – the latency distributions from the distribution of $*$-voltage statistics. I summarised this effect by showing that heterogeneous degrees make the distribution of voltage variabilities flatter, which, in turn, leads to sparsely populated sensitivity chains, which finally lead to a more pronounced pioneer effect.

Finally, I described a 'Bayesian-like' approach to detecting strong synapses (see chapter 4). This makes it possible to infer elements of the structure (synapses with high synaptic weights) of real neural networks solely from activity measurements.

On a very general scale, my work contributes to the topic of "spread of rumour" on random graphs. More specifically, it adds insight to the ubiquitous discussion of possible neural codes: even unstructured random networks without any type of cellular inhomogeneity express pioneer neurons whose discharge order is indicative of where a (spatial) stimulation occurred. Pioneer or 'leader' neurons have already been studied by several research groups. Much of this research activity detected connections between 'being a pioneer' and the 'excitatory in-degree.' In my networks, such a connection was not present (as already mentioned above). Pioneers were distinguished by having particular average discharge rates, but these were not simply determined by the excitatory in-degree. Also, pioneers were influential by comprising the first link in the positive feedback-loop supporting NS, but did not show exceptionally large out-degree (because in my networks in-degrees and out-degrees are not correlated). The main reason for the early discharge of pioneers, however, was their characteristic sensitivity to small fluctuations of the population activity: pioneers are situated just below threshold. But why are recruitment paths and rank-order strings of pioneers very dissimilar if the network is stimulated at

different sites? I have shown that this is the main reason why pioneers encode the stimulation site by means of the rank. The answer to the previous question is presumably that the pioneers' sensitivity is by itself somewhat heterogeneous: some pioneers have a very high sensitivity, some a moderately high sensitivity. In the very fist phase of the initiation of NS, after a stimulation at a specific site, activity percolates along paths of least resistance. If a moderately sensitive pioneer happens to have many fathers which have already been active at time t due to their direct activation, it will discharge very early, otherwise it may discharge later. Thus, the *chain of sensitivity* which the pioneers comprise interferes with the information about paths of least resistance which a particular stimulation site gives rise to, so that effectively different rank-orders emerge.

During my studies, I have also attempted to couple my *in silico* networks unidirectionally in order to demonstrate information *transmission* (Bauermeister et al. (2015)) from an 'upstream' network towards a 'downstream' network (not shown in this thesis). This information transmission has been demonstrated *in vitro* by Levy and colleagues (Levy et al. (2012)). However, my first attempts to do so were based on coupling an upstream *in vitro* network to an *in silico* downstream partner. I succeeded in transmitting order-based information to the downstream network (not shown). Upstream NS can be classified as 'spontaneous', 'evoked at site 1,' etc., so that the NS which occur downstream can be classified as 'spontaneous,' 'evoked by a spontaneous NS upstream,' 'evoked by a NS upstream which was there evoked at site 1,' etc. I was able to show that the information where upstream stimulation occurred could be read out *downstream* – under *the assumption that the evoked downstream NS precedes* the evoking upstream NS. This 'reversed' mode of transmission is consistent with the 'fast' mode where the transmission is not severely delayed. However, I did not present these results in this thesis, because I was unable to reproduce this in a more consistent scheme, where two *in silico* networks (of my type) are being coupled. The main obstruction was that I was not able to determine conditions under which many of the peaks of evoked downstream NS occur *before* the peaks of the evoking upstream NS, so that the recruitment phase downstream

is generally faster. By the way, it is easy to see why NS have to occur in this *reverse* order for information transmission to work: the upstream NS acts as a reset of the upstream network, and the uni-directional coupling from upstream to downstream destroys *any* information contained in the recruitment phase of the downstream NS, if the order of the two NS is 'sequential' instead of 'reversed.' Thus, future work has to establish conditions under which two neuronal modules can be coupled coherently and such that the order of NS 'upstream' and 'downstream' is *reversed* very frequently.

My results do not directly have relevance *in vivo*, since super-critical behaviour *in vivo* is mainly associated with sleep (sleep spindles) or epileptiform, i.e. pathological, activity. Therefore, the most important proposal for future work would be to analyse rank-order based coding in the critical or slightly sub-critical regime. It is to be expected that this change in the character of the synchronisation events severely interferes with coding schemes based on the rank of neurons which herald the events (which then would be avalanches). Evidence for sequential recruitment *in vivo* abounds, on the other hand: Contreras et al. (2013) (anaesthetised rats somatosensory and auditory cortex), Luczak et al. (2007) (somatosensory) Luczak and Barthó (2012) and Luczak and MacLean (2012) (somatosensory), Stark et al. (2015) (hippocampus), Rolston et al. (2007) (cultures), Matsumoto et al. (2013) (hippocampus ex vivo) Peyrache et al. (2010) (prefrontal cortex sleeping rats) Carrillo-Reid et al. (2015) (primary visual cortex).

Understanding the encoding of information in neural population activity is important for grasping the fundamental computations underlying brain function, and for interpreting signals that may be useful for the control of prosthetic devices. My work might contribute to first steps in this direction, which asks for 'gentle steering' of neuronal device.

Appendices

Appendix A

Network design and parameters

I simulated the collective activity of assemblies of 400 excitatory and 100 inhibitory neurons (leaky integrate-and-fire, LIF) (Tuckwell, 2005), connected randomly by means of conductance synapses with short-term dynamics (Tsodyks et al., 1998). Spontaneous activity was evoked by a uniform and constant background current injected into all neurons. As neuron models were identical, the only source of heterogeneity was the connectivity (20 % mean density). Three types of connectivity were investigated: 'homogeneous random' (Erdös-Rényi), 'scale-free' (Barabási and Albert, 1999), and 'heterogeneous random'.

A neural simulator was programmed in C and verified against existing simulators, as well as by reproducing the results of (Tsodyks et al., 2000). Time was discretised in steps of 0.5 ms and numerical integration was performed with the first-order exponential integration method. To compute power spectra, smaller integration steps of 0.1 ms were used. To ensure representative results, I investigated multiple realisations of every network architecture (typically more than 10). Each type of connectivity expressed generally consistent behaviour, although event rates and average activity levels varied between realisations.

A.1 Neurons

The time-dependent membrane voltage $V(t)$ was governed by the differential equation

$$\frac{\mathrm{d}V}{\mathrm{d}t}(t) = \frac{E_L - V(t)}{\tau_m} + \frac{R_m I_b}{\tau_m} + \frac{R_m I_{syn}(t)}{\tau_m}, \tag{A.1}$$

where E_L is the leak reversal potential, τ_m is the membrane time constant, R_m is the membrane resistance, I_b is the background current, and I_{syn} is the synaptic current, see below. Whenever the voltage reached the threshold V_{thr}, it was reset immediately to V_{res}, where it remained for a refractory period τ_{ref}. The parameters of the neuron model were as follows: $E_L = -70$ mV; $\tau_m = 30$ ms for excitatory neurons and $\tau_m = 10$ ms for inhibitory neurons; $R_m = 40$ MΩ for excitatory neurons and $R_m = 50$ MΩ for inhibitory neurons; $I_b = 525$ pA for excitatory neurons and $I_b = 420$ pA for inhibitory neurons; $V_{thr} = -50$ mV; $V_{res} = -65$ mV; $\tau_{ref} = 3$ ms for excitatory neurons and $\tau_{ref} = 2$ ms for inhibitory neurons. Note that the background currents raise the equilibrium potential over the threshold level, ensuring spontaneous activity (with hypothetical rates $\nu_{exc} = 12$ Hz and $\nu_{inh} = 36$ Hz in the absence of connectivity). Note further that the model is defined without noise. Initial membrane voltages were assigned randomly from the interval $[V_{res}, V_{thr}]$. To avoid onset artefacts, the initial two seconds of activity were ignored.

A.2 Synapses

The synaptic state is described by four time-dependent variables (Tsodyks et al., 1998): the instantaneous fractions of recovered, active, and inactive (synaptic) resources ($R(t)$, $E(t)$, and $I(t)$, respectively) and the fraction of *recovered* resources $u(t)$ recruited by pre-synaptic spikes. These non-dimensional variables satisfy equations 1.10, 1.11, 1.12, 1.13 and 1.14 (see chapter 1). The axonal conduction delay was uniform and 0.5 ms. τ_{rec} is the recovery time constant; τ_I is the inactivation time constant; τ_{facil} is the facilitation time constant; and U is a parameter associated with resource utilisation. Parameter values are as follows (subscript 'ee' stands

for 'excitatory-to-excitatory,' 'ie' stands for 'excitatory-to-inhibitory,' 'ei' stands for 'inhibitory-to-excitatory,' 'ii' stands for 'inhibitory-to-inhibitory'): $\tau_{I,ee} = \tau_{I,ie} = 3$ ms and $\tau_{I,ei} = \tau_{I,ii} = 10$ ms; the values for U, τ_{rec} and τ_{facil} were randomly chosen and hence varied from synapse to synapse. Values were chosen from Gaussian distributions with mean $U_{ee} = U_{ei} = 0.3$; $U_{ie} = U_{ii} = 0.04$; $\tau_{rec,ee} = \tau_{rec,ei} = 0.8$ s; $\tau_{rec,ie} = \tau_{rec,ii} = 0.1$ s; $\tau_{facil,ie} = \tau_{facil,ii} = 1$ s. The standard deviation of each distribution was half the respective mean. However, Gaussian distributions were clipped and restricted to a physically possible range (i.e., positive values for time constants and values between zero and unity for U). For *ee*- and *ei*-synapses, τ_{facil} was zero (no facilitation).

The synaptic current $I_{syn,i}$ of the i-th neurons was

$$I_{syn,i}(t) = g_{exc,i}(t)(E_{exc} - V(t)) + g_{inh,i}(t)(E_{inh} - V(t)), \tag{A.2}$$

where the reversal potentials were chosen as $E_{exc} = 0$ and $E_{inh} = -70$ mV. The conductances $g_{exc,i}$ and $g_{inh,i}$ are given by

$$\begin{aligned} g_{exc,i}(t) &= \sum_{j \text{ exc}} \Delta g_{ij} E_{ij}(t) \\ g_{inh,i}(t) &= \sum_{j \text{ inh}} \Delta g_{ij} E_{ij}(t), \end{aligned}$$

where the sum is over all excitatory (inhibitory respectively) neurons. Δg_{ij} is the matrix of synaptic weights, and E_{ij} is the (time-dependent) matrix of resources in the active state.

The assignment of neuron and synapse parameters was modelled on (Tsodyks et al., 2000).

A.3 Connectivity matrix

In homogeneous random (Erdös-Rényi) networks, each ordered neuron pair (i, j) formed a synaptic connection $i \to j$ with 20 % probability. Over all neurons, the

degree of connectivity thus followed a Gaussian distribution. Scale-free networks were obtained with the 'preferential attachment' procedure (Barabási and Albert, 1999), such that connectivity followed a power-law distribution with a mean connectivity of 20 %. Heterogeneous random networks were generated as follows. Every neuron i was individually assigned four random numbers, $\lambda_{pre,exc}$, $\lambda_{post,exc}$, $\lambda_{pre,inh}$, and $\lambda_{post,inh}$, each drawn independently from the interval $[0,\delta]$, where $\delta = 0.2$ is the mean connection density. In a second step, every ordered neuron pair i,j was individually assigned two random numbers, ξ and η, drawn independently from $[0,1]$. An excitatory projection $j \to i$ was established, if neuron j was excitatory and $\xi < \lambda_{pre,exc}$. Similarly, an inhibitory projection $j \to i$ was established, if j was inhibitory and $\xi < \lambda_{pre,inh}$. Projections $i \to j$ were established if j was excitatory and $\eta < \lambda_{post,exc}$, or if j was inhibitory and $\eta < \lambda_{post,inh}$. This procedure resulted in a random graph with mean connection density of 20 %. Heterogeneity arises because each neuron exhibits an individual connection density, with independent out-degree and in-degree.

Established projections were assigned a synaptic weight Δg_{ij}, each chosen randomly and independently from a (clipped) Gaussian distribution with mean ω and standard deviation $\omega/2$ (clipping ensured $\Delta g_{ij} > 0$). Not established projections were assigned $\Delta g_{ij} = 0$. Mean values were chosen such as to obtain spontaneous activity with pronounced synchronisation events ('network spikes', see below) at rates of $O(10^0$ Hz). Specifically, I chose $\omega_{ee,0} = 1150$ pS, $\omega_{ei,0} = 8500$ pS, $\omega_{ie,0} = 5$ pS, $\omega_{ii,0} = 200$ pS for homogeneous random networks; $\omega_{ee,0} = 1450$ pS, $\omega_{ei,0} = 9500$ pS, $\omega_{ie,0} = 5$ pS, $\omega_{ii,0} = 200$ pS for scale-free networks; $\omega_{ee,0} = 1000$ pS, $\omega_{ei,0} = 8500$ pS, $\omega_{ie,0} = 5$ pS, $\omega_{ii,0} = 200$ pS for heterogeneous random networks.

Almost all realisations of random connectivity resulted in spontaneous network activity including large synchronisation events. This was the case for ≈ 90 % of the homogeneous random networks, ≈ 80 % of the scale-free networks, and ≈ 100 % of the heterogeneous random networks. In the remaining realisations, spontaneous activity failed to ignite network spikes (in an all-or-nothing fashion).

Appendix B

Power spectra and cross-correlations

For the computation of power spectra and cross-correlations, I divided a long simulation ($\approx$ 1000 s, resolution 0.1 ms, with NS removed) into bins of length $T = 20\ s$, thus creating an ensemble of ≈ 50 time-traces. For computation of power-spectra of individual neurons, I computed the Fourier-transform

$$\tilde{\rho}_i(\omega) = \int \mathrm{d}t\, \rho_i(t) \exp(\mathrm{i}\omega t) \tag{B.1}$$

of the spike-trains of each neuron i, integrated over single bins. The power spectrum $S_i(f)$ of the activity of the ith excitatory neuron was then determined as

$$S_i(f) = \frac{\left\langle |\tilde{\rho}_i(2\pi f)|^2 \right\rangle}{T}, \tag{B.2}$$

where the average is taken over the ensemble of bins. This was averaged over excitatory neurons to yield Fig. 2.2 C.

For the computation of cross-correlations, I determined the normalised 'scalar

product'

$$C_{ij}(\tau) = \frac{\left\langle \int \mathrm{d}t (\rho_i(t) - \nu_i)(\rho_j(t+\tau) - \nu_j) \right\rangle}{\sqrt{\left\langle \int \mathrm{d}t (\rho_i - \nu_i)^2(t) \right\rangle} \sqrt{\left\langle \int \mathrm{d}t (\rho_j - \nu_j)^2(t) \right\rangle}} \tag{B.3}$$

for each pair of distinct neurons i and j which discharge at least once between NS (where ν_i and ν_j are the average firing rates between NS). The integral is again over single bins and the average is over the ensemble of all bins. Spike trains were regularised by square-shaped kernels with a width of 0.5 ms. Then I considered the ensemble of cross-correlations $C_{ij}(0)$ at lag $\tau = 0$ for each pair of distinct excitatory neurons which both discharge between NS. For this ensemble, the mean and standard deviation were stated.

Appendix C

Histograms and densities

Histograms and densities were computed as follows. In Fig. 2.2 B, rectangular bins of size 50 ms were used. In Fig. 2.4 A, rectangular bins of size 2 (for the relative activity) were used. In all other cases, densities were estimated with Gaussian kernels. Kernel width and sampling resolution for population activity (spike density) were 3 ms and 0.1 ms, respectively. For latency distributions, the corresponding values were 0.8 ms and 0.1 ms, with approximately 400 samples per kernel. For voltage distributions they were 0.1 mV and 10 μV, with 20,000 samples. For the fraction of recovered resources R, kernel width and voltage distribution were 0.05 and 0.01, with $O(10^4)$ samples.

Recovered resources $R_i(t)$ of neuron i were averaged over synapses $i \to j$ to N_i post-synaptic neurons j

$$R_i(t) = \frac{\sum_j R_{ji}(t)}{N_i}, \tag{C.1}$$

where $R_{ji}(t)$ is recovered resources of synapse $i \to j$ at time t. Densities $p_i(R)$ were established over all time points *excluding* NS (*i.e.*, time-points more than 35 ms before or after a NS).

Population activity is understood to be activity of the excitatory subpopulation and is sometimes given as *absolute* activity (in Hz) and sometimes as *relative* activity (*i.e.*, in units of the average activity level).

Appendix D

Encoding of external stimulation

To assess the extent to which network activity encodes external stimulation, I perturbed spontaneous network activity in some simulations. External stimulation targeted particular subsets of excitatory neurons (10 or 30 randomly selected neurons) and forced a single spike in each target neuron. Each subset of targets was considered a 'stimulation site' and up to 12 non-overlapping sites were used. Subsequent network activity (*i.e.*, 100 ms of activity, exclusive of the forced spikes) was characterised in terms of four features, following (Kermany et al., 2010). Two features were based on firing rates $a_i(t)$ of neurons i: temporal profile of population activity $A(t) = \sum_i a_i(t)$, and spatial profile of population activity $A_i = \int a_i(t)dt$. Two further features were based on spiking activity of neurons i in the interval between stimulation and the subsequent NS: timing of first spikes t_i and rank order of first spikes o_i. Rank order was obtained by sorting negative spike latencies with respect to the subsequent NS (for example, the negative latency vector $(-20\text{ ms}, -10\text{ ms}, -15\text{ ms}, -17\text{ ms})$ would yield the rank order vector $(1,4,3,2)$).

To analyse the information encoded by different activity features, simulations were divided into a training and a test set. Following (Kermany et al., 2010), the

training set was used to train a support vector machine (SVM, from Python library LIBSVM, module 'sklearn', (Chang and Lin, 2011)) to classify the stimulated location on the basis a particular feature. The test set was used to determine classification performance (fraction of correct classifications of the stimulated site), providing a lower bound for the 'true' information about stimulation site encoded by a particular activity feature.

Appendix E

Silencing of neurons

To assess the relative importance of different subsets of excitatory neurons, I wished to render ineffective the members of any particular subset. To do so, I retained the spikes of such neurons, but suppressed all post-synaptic effects. A reduced frequency of NS in partially de-afferentiated networks revealed the relative importance of the manipulated subset of neurons.

Appendix F

Spike-triggered average population activity

To assess the relation between population activity and individual neuron spikes, I computed the 'spike-triggered deviation' $\Gamma_i(\tau)$ as follows

$$\Gamma_i(\tau) = \frac{\int \big(A(t) - \langle A \rangle\big) \rho_i(t-\tau) \mathrm{d}t}{\int \rho_i(t) \mathrm{d}t} \tag{F.1}$$

where $A(t)$ is time-dependent population activity, $\langle A \rangle$ is its temporal mean, $\rho_i(t)$ is the spike sequence (Dirac comb) of neuron i, and τ is the latency between activity time t and spike time $t-\tau$. The computation was restricted to periods between NS, and the normalisation term $\int \rho_i(t)$dt is the number of spikes fired by neuron i between NS. In principle, $\Gamma_i(\tau)$ measures influence and sensitivity (for $\tau > 0$ and $\tau < 0$, respectively) to population activity of neuron i. However, as many neurons spike only shortly before NS, I mostly obtain information about negative latencies, that is, about sensitivity to population activity. For this reason, the spike-triggered deviation in Fig. 2.9 C is restricted to negative latencies. Moreover, it is defined only for (sorted) neuron ID > 260, as less active neurons never spike between NS.

Appendix G

Estimation of post-synaptic effects

To estimate the differential effect of neuron i on post-synaptic neurons j throughout the network, I proceeded as follows. For every synaptic target j, I formed the difference $W_{ji} \equiv V_j' - V_j$ between the hypothetical $*$-voltage $V_j'(t)$ that would have resulted from a single additional spike of neuron i at time t_{sp} and the actual the $*$-voltage $V_j(t)$, which may be approximated as

$$W_{ji}(t) \approx \frac{\tau_I U \Delta g R_m (E_{\text{exc}} - \langle V_j \rangle)}{(\tau_m - \tau_I)(1 + U \tau_{\text{rec}} \nu_i)} \Theta(t - t_{\text{sp}}) \Big[e^{-(t-t_{\text{sp}})/\tau_m} - e^{-(t-t_{\text{sp}})/\tau_I} \Big]. \tag{G.1}$$

where ν_i is the asynchronous firing rate of neuron i (between NS), $\langle V_j \rangle$ is the expected $*$-voltage of neuron j, and τ_I, U, Δg, τ_{rec} are parameters of the synapse in question. Note that I neglect conduction delays and assume the driving force to be constant. The expression for $W_{ji}(t)$ peaks at time

$$t_{\max} = \frac{\tau_m \tau_I}{(\tau_m - \tau_I)} \log\left(\frac{\tau_m}{\tau_I}\right), \tag{G.2}$$

so that the post-synaptic potential in neuron j that is triggered by the additional spike in neuron i at time t_{sp} is $W_{ji}(t_{\max})$. The cumulative post-synaptic effect of all

spikes in neuron i is given by the stationary limit

$$\left\langle W_{ji} \right\rangle_{\mathrm{ss}} = \frac{R_m \Delta g (E_{\mathrm{exc}} - \left\langle V_j \right\rangle) \tau_I U \nu_i}{1 + U \tau_{\mathrm{rec}} \nu_i}. \tag{G.3}$$

which is approximately equal to $\tau_m \cdot \nu_i \cdot W_{ji}(t_{\max})$.

In Fig. 2.10 B and C, the differential effects of neuron i are averaged over all N_i post-synpatic neurons j

$$\mathrm{PSP}_i \equiv \frac{1}{N_i} \sum_j W_{ji}(t_{\max}), \qquad \left\langle \mathrm{PSP}_i \right\rangle_{\mathrm{ss}} \equiv \frac{1}{N_i} \sum_j \left\langle W_{ji} \right\rangle_{\mathrm{ss}}. \tag{G.4}$$

In Fig. 2.10 D and E, the products $\mathrm{PSP}_i \cdot N_i$ and $\left\langle \mathrm{PSP}_i \right\rangle_{\mathrm{ss}} N_i$ are plotted (summed effect).

Appendix H

Modification of the Levenshtein edit distance

To quantify dissimilarity in the rank order or 'first spikes' observed in different contexts, I modified the Levenshtein edit distance (Levenshtein, 1966) used in previous studies (Shahaf et al., 2008). Whereas the Levenshtein metric is useful for strings with same and/or different 'letters,' in the present situation all rank order strings contain the same 'letters' (because all neurons fire at least one spike and rare missing spikes can be 'filled in' at the highest rank). Now consider two strings $s_1 s_2 \dots s_n$ and $s_{\pi(1)} s_{\pi(2)} \dots s_{\pi(n)}$, where π is an appropriately chosen permutation. The number of inversions L, which is the number pairs (i, j) such that $i < j$ but $\pi(i) > \pi(j)$, ranges from 0 (if strings are identical) to $L = \frac{n(n-1)}{2}$ (if strings are inverted). Accordingly, I adopted

$$L_n = \left(1 - \frac{2L}{n(n-1)}\right) 100\% \tag{H.1}$$

as normalised measure of similarity.

Bibliography

Amit DJ, Brunel N (1997a) Dynamics of a recurrent network of spiking neurons before and following learning. Network: Computation in Neural Systems 8(4):373–404

Amit DJ, Brunel N (1997b) Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. Cerebral cortex (New York, NY: 1991) 7(3):237–252

Amit DJ, Tsodyks M (1991a) Quantitative study of attractor neural network retrieving at low spike rates: I. substrate—spikes, rates and neuronal gain. Network: Computation in neural systems 2(3):259–273

Amit DJ, Tsodyks M (1991b) Quantitative study of attractor neural networks retrieving at low spike rates: Ii. low-rate retrieval in symmetric networks. Network: Computation in Neural Systems 2(3):275–294

Amit DJ, Brunel N, et al. (1997) Global spontaneous activity and local structured (learned) delay period activity in cortex. Cerebral Cortex 7:237–252

Bak P, Chen K (1991) Self-organized criticality. Scientific American 264(1):46–53

Bak P, Tang C, Wiesenfeld K (1988) Self-organized criticality. Physical Review A 38(1):364

Barabási AL, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509–512

Bauermeister C, Keren H, Braun J (2015) Coherent coupling of in vitro neuronal slices onto in silico networks. In: 11th Bernstein Conference, Heidelberg, Germany.

Bear MF, Connors BW, Paradiso MA (2007) Neuroscience, vol 2. Lippincott Williams & Wilkins

Beggs JM, Plenz D (2003) Neuronal avalanches in neocortical circuits. Journal of Neuroscience 23(35):11167–11177

Breskin I, Soriano J, Moses E, Tlusty T (2006) Percolation in living neural networks. Physical review letters 97(18):188102

Brunel N (2000a) Dynamics of networks of randomly connected excitatory and inhibitory spiking neurons. Journal of Physiology-Paris 94(5-6):445–463

Brunel N (2000b) Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. Journal of Computational Neuroscience 8:183–208

Brunel N, Wang XJ (2001) Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. Journal of computational neuroscience 11(1):63–85

Burkitt AN (2006a) A review of the integrate-and-fire neuron model: I. homogeneous synaptic input. Biological cybernetics 95(1):1–19

Burkitt AN (2006b) A review of the integrate-and-fire neuron model: Ii. inhomogeneous synaptic input and network properties. Biological cybernetics 95(2):97–112

Carrillo-Reid L, Miller JeK, Hamm JP, Jackson J, Yuste R (2015) Endogenous sequential cortical activity evoked by visual stimuli. Journal of Neuroscience 35(23):8813–8828

Chang CC, Lin CJ (2011) Libsvm: A library for support vector machines. ACM transactions on intelligent systems and technology (TIST) 2(3):27

Chialvo DR (2010) Emergent complex neural dynamics. Nature physics 6(10):744

Clementi A, Silvestri R, Trevisan L (2015) Information spreading in dynamic graphs. Distributed Computing 28(1):55–73

Cohen O, Keselman A, Moses E, Martínez MR, Soriano J, Tlusty T (2010) Quorum percolation in living neural networks. EPL (Europhysics Letters) 89(1):18008

Contreras EJB, Schjetnan AGP, Muhammad A, Barthó P, McNaughton BL, Kolb B, Gruber AJ, Luczak A (2013) Formation and reverberation of sequential neural activity patterns evoked by sensory stimulation are enhanced during cortical desynchronization. Neuron 79(3):555–566

Dayan P, Abbott LF, et al. (2001) Theoretical neuroscience, vol 806. Cambridge, MA: MIT Press

Decharms RC, Zador A (2000) Neural representation and the cortical code. Annual Review of Neuroscience 23(1):613–647

Dranias MR, Westover MB, Cash S, VanDongen AM (2015) Stimulus information stored in lasting active and hidden network states is destroyed by network bursts. Frontiers in integrative neuroscience 9:14

Durrett R (2010) Some features of the spread of epidemics and information on a random graph. Proceedings of the National Academy of Sciences 107(10):4491–4498

Eckmann JP, Feinerman O, Gruendlinger L, Moses E, Soriano J, Tlusty T (2007) The physics of living neural networks. Physics Reports 449(1-3):54–76

Eckmann JP, Jacobi S, Marom S, Moses E, Zbinden C (2008) Leader neurons in population bursts of 2d living neural networks. New Journal of Physics 10(1):015011

Eckmann JP, Moses E, Stetter O, Tlusty T, Zbinden C (2010) Leaders of neuronal cultures in a quorum percolation model. Frontiers in computational neuroscience 4:132

Erdős P, Rényi A (1960) On the evolution of random graphs. Publ Math Inst Hung Acad Sci 5(1):17–60

Eytan D, Marom S (2006) Dynamics and effective topology underlying synchronization in networks of cortical neurons. Journal of Neuroscience 26(33):8465–8476

Feng J (2003) Computational neuroscience: a comprehensive approach. Chapman and Hall/CRC

Gardiner C (2009) Stochastic methods, vol 4. Springer Berlin

Gerstner W, Kistler WM, Naud R, Paninski L (2014) Neuronal dynamics: From single neurons to networks and models of cognition. Cambridge University Press

Gigante G (2006) Stochastic networks of adapting spiking neurons: oscillations, bursts and resonant properties. PhD thesis

Gigante G, Deco G, Del Giudice P (2014) Spontaneous and evoked population bursts: history-dependent response, differential role of neuronal adaptation, synaptic short-term depression, and time scales inference. In: 9th FENS Forum of Neuroscience, Milan, Italy

Gigante G, Deco G, Marom S, Del Giudice P (2015) Network events on multiple space and time scales in cultured neural networks and in a stochastic rate model. PLoS Computational Biology 11(11):e1004547

Gritsun T, le Feber J, Stegenga J, Rutten WL (2011) Experimental analysis and computational modeling of interburst intervals in spontaneous activity of cortical neuronal culture. Biological Cybernetics 105(3-4):197–210

Gritsun TA, Le Feber J, Stegenga J, Rutten WL (2010) Network bursts in cortical cultures are best simulated using pacemaker neurons and adaptive synapses. Biological Cybernetics 102(4):293–310

Haroush N, Marom S (2014) Evidence for dynamic excitation-inhibition ratio in networks of cortical neurons. arXiv preprint arXiv:150100081

Haroush N, Marom S (2015) Slow dynamics in features of synchronized neural network responses. Frontiers in computational neuroscience 9:40

Honerkamp J (1993) Stochastic dynamical systems: concepts, numerical methods, data analysis. John Wiley & Sons

Izhikevich EM (2007) Dynamical systems in neuroscience. MIT press

Jensen HJ (1998) Self-organized criticality: emergent complex behavior in physical and biological systems. Cambridge University Press, New York.

Ju H, Dranias MR, Banumurthy G, VanDongen AM (2015) Spatiotemporal memory is an intrinsic property of networks of dissociated cortical neurons. Journal of Neuroscience 35(9):4040–4051

Kandel ER, Schwartz JH, Jessell TM, of Biochemistry D, Jessell MBT, Siegelbaum S, Hudspeth A (2000) Principles of neural science, vol 4. McGraw-hill New York

Kermany E, Gal A, Lyakhov V, Meir R, Marom S, Eytan D (2010) Tradeoffs and constraints on neural representation in networks of cortical neurons. Journal of Neuroscience 30(28):9588–9596

Koch C (1999) Biophysics of computation: processing in single neurons. Oxford University Press.

Levenshtein VI (1966) Binary codes capable of correcting deletions, insertions, and reversals. In: Soviet Physics Doklady, vol 10, pp 707–710

Levy O, Ziv NE, Marom S (2012) Enhancement of neural representation capacity by modular architecture in networks of cortical neurons. European Journal of Neuroscience 35(11):1753–1760

Lindner B (2002) Coherence and stochastic resonance in nonlinear dynamical systems. Logos-Verlag

Loebel A, Tsodyks M (2002) Computation by ensemble synchronization in recurrent networks with synaptic depression. Journal of Computational Neuroscience 13(2):111–124

Luczak A, Barthó P (2012) Consistent sequential activity across diverse forms of up states under ketamine anesthesia. European Journal of Neuroscience 36(6):2830–2838

Luczak A, MacLean JN (2012) Default activity patterns at the neocortical microcircuit level. Frontiers in Integrative Neuroscience 6:30

Luczak A, Barthó P, Marguet SL, Buzsáki G, Harris KD (2007) Sequential structure of neocortical spontaneous activity in vivo. Proceedings of the National Academy of Sciences 104(1):347–352

Markram H, Wang Y, Tsodyks M (1998) Differential signaling via the same axon of neocortical pyramidal neurons. Proceedings of the National Academy of Sciences 95(9):5323–5328

Marom S, Shahaf G (2002) Development, learning and memory in large random networks of cortical neurons: lessons beyond anatomy. Quarterly Reviews of Biophysics 35(1):63–87

Masquelier T, Deco G (2013) Network bursting dynamics in excitatory cortical neuron cultures results from the combination of different adaptive mechanism. PLoS One 8(10):e75824

Massobrio P, Pasquale V, Martinoia S (2015) Self-organized criticality in cortical assemblies occurs in concurrent scale-free and small-world networks. Scientific reports 5:10578

Matsumoto K, Ishikawa T, Matsuki N, Ikegaya Y (2013) Multineuronal spike sequences repeat with millisecond precision. Frontiers in Neural Circuits 7:112

Mattia M, Del Giudice P (2002) Population dynamics of interacting spiking neurons. Physical Review E 66(5):051917

Meisel C (2009) Self-organized criticality in adaptive neural networks. PhD thesis, Diploma thesis, Albert-Ludwigs-Universitat Freiburg, Germany.

Mishchencko Y, Vogelstein JT, Paninski L (2011) A bayesian approach for inferring neuronal connectivity from calcium fluorescent imaging data. The Annals of Applied Statistics pp 1229–1261

Morin FO, Takamura Y, Tamiya E (2005) Investigating neuronal activity with planar microelectrode arrays: achievements and new perspectives. Journal of Bioscience and Bioengineering 100(2):131–143

Newman ME (2000) Models of the small world. Journal of Statistical Physics 101(3-4):819–841

Newman ME, Strogatz SH, Watts DJ (2001) Random graphs with arbitrary degree distributions and their applications. Physical review E 64(2):026118

Núñez-Amaral L, Barthelemy M, Scala A, Stanley H (2000) Classes of behavior of small-world networks. Tech. rep.

Orlandi JG, Stetter O, Soriano J, Geisel T, Battaglia D (2014) Transfer entropy reconstruction and labeling of neuronal connections from simulated calcium imaging. PloS one 9(6):e98842

Papoulis A, Pillai SU (2002) Probability, random variables, and stochastic processes. Tata McGraw-Hill Education

Pasquale V, Massobrio P, Bologna L, Chiappalone M, Martinoia S (2008) Self-organization and neuronal avalanches in networks of dissociated cortical neurons. Neuroscience 153(4):1354–1369

Paugam-Moisy H, Bohte S (2012) Computing with spiking neuron networks. Handbook of natural computing pp 335–376

Pena RFO, Vellmer S, Bernardi D, Roque AC, Lindner B (2018) Self-consistent scheme for spike-train power spectra in heterogeneous sparse networks. Front Comput Neurosci 12(9):1–19

Perin R, Berger TK, Markram H (2011) A synaptic organizing principle for cortical neuronal groups. Proceedings of the National Academy of Sciences 108(13):5419–5424

Persi E, Horn D, Segev R, Ben-Jacob E, Volman V (2004a) Neural modeling of synchronized bursting events. Neurocomputing 58:179–184

Persi E, Horn D, Volman V, Segev R, Ben-Jacob E (2004b) Modeling of synchronized bursting events: the importance of inhomogeneity. Neural Computation 16(12):2577–2595

Peskin ME (2018) An introduction to quantum field theory. CRC press

Peyrache A, Benchenane K, Khamassi M, Wiener SI, Battaglia FP (2010) Sequential reinstatement of neocortical activity during slow oscillations depends on cells' global activity. Frontiers in Systems Neuroscience 3:18

Poil SS, Hardstone R, Mansvelder HD, Linkenkaer-Hansen K (2012) Critical-state dynamics of avalanches and oscillations jointly emerge from balanced excitation/inhibition in neuronal networks. Journal of Neuroscience 32(29):9817–9823

Ponulak F, Kasinski A (2011) Introduction to spiking neural networks: Information processing, learning and applications. Acta neurobiologiae experimentalis 71(4):409–433

Priesemann V, Wibral M, Valderrama M, Pröpper R, Le Van Quyen M, Geisel T, Triesch J, Nikolic D, Munk MH (2014) Spike avalanches in vivo suggest a driven, slightly subcritical brain state. Frontiers in Systems Neuroscience 8:108

Rieke F (2008) Spikes: Exploring the Neural Code. MIT Press.

Risken H (1996) Fokker-planck equation. In: The Fokker-Planck Equation, Springer, pp 63–95

Rolls E (2008) Memory, Attention and Decision-making. Oxford University Press.

Rolls E, Deco G (2010) The Noisy Brain: Stochastic Dynamics as a Principle of Brain Function. Oxford University Press.

Rolls ET (2016) Cerebral cortex: principles of operation. Oxford University Press

Rolston JD, Wagenaar DA, Potter SM (2007) Precisely timed spatiotemporal patterns of neural activity in dissociated cortical cultures. Neuroscience 148(1):294–303

Rovelli C (2004) Quantum gravity. Cambridge university press

Rubinov M, Sporns O (2010) Complex network measures of brain connectivity: uses and interpretations. Neuroimage 52(3):1059–1069

Rubinov M, Sporns O, Thivierge JP, Breakspear M (2011) Neurobiologically realistic determinants of self-organized criticality in networks of spiking neurons. PLoS computational biology 7(6):e1002038

Sahara S, Yanagawa Y, O'Leary DD, Stevens CF (2012) The fraction of cortical gabaergic neurons is constant from near the start of cortical neurogenesis to adulthood. Journal of Neuroscience 32(14):4755–4761

Scarpetta S, de Candia A (2014) Alternation of up and down states at a dynamical phase-transition of a neural network with spatiotemporal attractors. Frontiers in systems neuroscience 8:88

Schmeltzer C, Soriano J, Sokolov IM, Rüdiger S (2014) Percolation of spatially constrained erdős-rényi networks with degree correlations. Physical Review E 89(1):012116

Shahaf G, Eytan D, Gal A, Kermany E, Lyakhov V, Zrenner C, Marom S (2008) Order-based representation in random networks of cortical neurons. PLoS Computational Biology 4(11):e1000228

Shepherd GM (2003) The synaptic organization of the brain. Oxford University Press.

Soriano J, Martínez MR, Tlusty T, Moses E (2008) Development of input connections in neural cultures. Proceedings of the National Academy of Sciences 105(37):13758–13763

Spiridon M, Gerstner W (1999) Noise spectrum and signal transmission through a population of spiking neurons. Network: Computation in Neural Systems 10(3):257–272

Stark E, Roux L, Eichler R, Buzsáki G (2015) Local generation of multineuronal spike sequences in the hippocampal ca1 region. Proceedings of the National Academy of Sciences 112(33):10521–10526

Stetter FO (2012) Bursting dynamics and topological structure of in vitro neuronal networks. PhD thesis, Niedersächsische Staats-und Universitätsbibliothek Göttingen

Stetter O, Battaglia D, Soriano J, Geisel T (2012) Model-free reconstruction of excitatory neuronal connectivity from calcium imaging signals. PLoS computational biology 8(8):e1002653

Stratonovich RL (1967) Topics in the theory of random noise, vol 1. CRC Press

Strogatz SH (2001) Exploring complex networks. nature 410(6825):268

Thompson RF (2000) The brain: A neuroscience primer. Macmillan

Thorpe S, Delorme A, Van Rullen R (2001) Spike-based strategies for rapid processing. Neural Networks 14(6-7):715–725

Tsodyks M, Pawelzik K, Markram H (1998) Neural networks with dynamic synapses. Neural Computation 10(4):821–835

Tsodyks M, Uziel A, Markram H, et al. (2000) Synchrony generation in recurrent networks with frequency-dependent synapses. Journal of Neuroscience 20(1):825–835

Tsodyks MV, Markram H (1997) The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. Proceedings of the national academy of sciences 94(2):719–723

Tuckwell HC (2005) Introduction to theoretical neurobiology., vol 1. Cambridge University Press.

Van Kampen NG (1992) Stochastic processes in physics and chemistry, vol 1. Elsevier

Vladimirski BB, Tabak J, O'Donovan MJ, Rinzel J (2008) Episodic activity in a heterogeneous excitatory network, from spiking neurons to mean field. Journal of Computational Neuroscience 25(1):39–63

Wang SJ, Hilgetag C, Zhou C (2011) Sustained activity in hierarchical modular neural networks: self-organized criticality and oscillations. Frontiers in computational neuroscience 5:30

Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world'networks. nature 393(6684):440

Wiedemann UA, Lüthi A (2003) Timing of network synchronization by refractory mechanisms. Journal of Neurophysiology 90(6):3902–3911

Wills P, Meyer FG (2019) Metrics for graph comparison: A practitioner's guide. arXiv preprint arXiv:190407414

Zbinden C (2010) Leader neurons in living neural networks and in leaky integrate and fire neuron models. PhD thesis, University of Geneva

Zbinden C (2011) Leader neurons in leaky integrate and fire neural network simulations. Journal of Computational Neuroscience 31(2):285–304