VRIJE
UNIVERSITEIT
BRUSSEL

Proef ingediend met het oog op het behalen van de graad van
Master in de Fysica en Sterrenkunde

Connecting Neurons

A Stochastic Model of Information Flow in Cultured Networks of Human Neurons

Hannah Pinson

August 17th, 2018

Supervisor: Prof. Dr. Vincent Ginis
Co-supervisor: Prof. Dr. Max Tegmark

Faculty of Sciences and Bioengineering Sciences
Department of Physics and Astronomy

Summary

How is the structure of a neural network related to its function?

Up to the present day this simple question remains a largely unanswered question. Slow but steady scientific progress has been made in studying a related, but more limited problem: how do specific dynamical patterns emerge from the complex network structure, and how does this dynamical activity shape the network in its turn?

Until recently, it was technically impossible to record the activity of a substantial number of neurons at once. This explains why the former question is still situated at the very frontiers of neuroscience, physics and computer science. In this master thesis project, we build on the recent spectacular advancements in neuronal recording technology, induced pluripotent stem cell techniques and (big) neuronal data analysis to shed light on the relation between activity and connectivity at a very high level of detail.

The outcome of this thesis is a set of numerical procedures that, starting from a recording of a neural network, result in precise information about the nature of the dynamical activity patterns, coupled to detailed insights in the underlying network structure. As a proof-of-concept, we apply these numerical procedures to recordings of an in vitro network of stem cell derived human cortical neurons. Our findings on the spontaneously arising neuronal dynamics are in agreement with previous studies: a self-organized increase in connectivity leads to global synchronization patterns, while sparsely connected neurons exhibit a typical slow rate spontaneous activity. We add to these findings methods to derive detailed information about the underlying connectivity, such as transmission delays and the direction of information flow. We hope our work will induce at least a small acceleration towards insights in the relation between form an function of neural networks.

Samenvatting

Hoe is de structuur van een neuraal netwerk gerelateerd aan zijn functie?

Tot op de dag van vandaag blijft deze simpele vraag een grotendeels onbeantwoorde vraag. Langzame maar gestage wetenschappelijke vooruitgang wordt geboekt bij het bestuderen van een gerelateerd, maar beperkter probleem: hoe komen specifieke dynamische patronen tevoorschijn uit de complexe netwerkstructuur en hoe vormt deze dynamische activiteit het netwerk op zijn beurt?

Tot voor kort was het technisch onmogelijk om de activiteit van een aanzienlijk aantal neuronen tegelijk te registreren. Dit verklaart waarom het nog steeds zo moeilijk is om de eerste vraag te beantwoorden en waarom deze vraag de grenzen van de neurowetenschap, de fysica en de computerwetenschappen aftast. In deze masterthesis bouwen we voort op de recente spectaculaire doorbraken in de neuronale meettechnologie, geÃńnduceerde pluripotente stamceltchnieken en (grote) neuronale gegevensanalyse om een beter inzicht te krijgen op de relatie tussen activiteit en connectiviteit, en dit op een zeer fijn niveau van detail.

De uitkomst van dit werk is een verzameling numerieke procedures die, uitgaande van een opname van een neuraal netwerk, resulteren in nauwkeurige informatie over de aard van de dynamische activiteitspatronen, gekoppeld aan gedetailleerde inzichten in de onderliggende netwerkstructuur. Als een proof-of-concept passen we deze numerieke procedures toe op opnamen van een in vitro netwerk van menselijke corticale neuronen, afgeleid uit stamcellen. Onze bevindingen over de spontaan ontstane neuronale dynamiek zijn in overeenstemming met eerdere studies: een zelfgeorganiseerde toename van de connectiviteit leidt tot globale synchronisatiepatronen, terwijl schaars geconnecteerde neuronen een spontane activiteit vertonen met een lagere karakteristieke tijdsschaal. We voegen aan deze vindingen methoden toe om gedetailleerde informatie te verkrijgen over de onderliggende connectiviteit, zoals transmissievertragingen en de richting van de informatiestroom. Op die manier hopen we dat ons werk een kleine versnelling zal veroorzaken in de verkenning van de relatie tussen vorm en functie in het menselijk brein.

Acknowledgements

I am foremost very grateful to my advisors, Prof. Max Tegmark and Prof. Vincent Ginis. Both of them helped me reach the academic dream of conducting research at MIT, Vincent by gently pushing me in the right direction, Max by finally pulling me in. Both of them exhibit an almost unnatural form of optimism. I can only hope this thesis will prove it to be somewhat justified.

I would furthermore like to thank my fellow researchers at MIT, especially Nima Dehghani, a relentless and amicable source of neuroscientific knowledge, and Rionna Flynn, a diligent student who has done a wonderful job helping us analyze the data during the summer. Not directly involved in this project, but very appreciated nevertheless, are Cris, John and Kevin, along with all the members of the Poggio lab, the Center for Brains, Minds and Machines at MIT, and everyone who was patient enough to discuss my questions near Gadi's fine-tuned coffee machine.

Special thanks, above all, to Lucas and my dad; Lucas for his love and for joining me in this adventure, my dad for his love and for hiding his wish for me to return home (I will!). I remember that my mother once said neuroscience would be a good field for me; as always, I didn't listen, and I studied physics and computer science instead. This thesis is dedicated to her. Even though I am not able to tell her this anymore, as always, she turned out to be right.

-Boston, August 17th 2018.

Contents

1

Introduction

At every instant in time, an endless collection of cognitive processes take place in our human brains, or perhaps in any brain; all of them are interesting at least, and absolutely fascinating at best. That we use these very cognitive processes to examine cognition itself only adds to the fascination. A persistent part of all human thinking is focused on understanding the working of our own minds, and by extension the working of our own brains. But, however imaginative we might be, the practical exploration of how the mind emerges from the human brain has always been constrained both in ethical and in technical ways [1] [2]: even if we would be willing to damage the brain of a living person along the way, we would by no means be able make a detailed measurement of the activity all 86 billion or so neurons [3] at once.

[1] *How advances in neural recording affect data analysis.* Stevenson and Kording,2011.

[2] *Emerging ethical issues in neuroscience.* Farah,2002.

[3] *Equal numbers of neuronal and non-neuronal cells make the human brain an isometrically scaled-up primate brain.* Azevedo et al.,2009.

It's a major achievement of human thinking, then, that we realized we can study large and complex systems by connecting our insights across different scales. Trying to measure the activity of all neurons in the brain at once might be as useless as tracking every single molecule in a gas to obtain the temperature. Problems of this sort can be tackled by approaching the problem at different levels, and by devising plausible mechanisms that relate the properties at the smallest, local scales to the behavior of the system at the largest, global scale. We thus study the brain at all levels, from the dynamics of single neurons to how brain regions are communicating during different states of consciousness; and we hope to connect more and more of these levels in an effort to understand the emergence of the utterly fascinating high-level cognitive processes. This thesis, then, is situated on one of the smallest levels: its focus is the neuronal activity of a small network of human cortical neurons cultured in a dish.

At this level, a lot of fundamental questions remain to be answered. The general question is of course how *network structure* affects *neuronal dynamics*—and also the other way round: how neuronal dynamics shapes the network structure. The neurons in a network *spike*, i.e., each neuron sends out short electrical signals to the other neurons; it is these spikes

and the overall spike firing patterns that form the basis of communication, computation and eventually cognition in the brain. (Of course, there is far more to this picture than just spikes, ranging from specific proteins to which languages you learned; but these are outside the scope of both this thesis and the level in the brain structure hierarchy it is focused on). These firing patterns is what we will mean by neuronal dynamics. The general question thus amounts to asking how the firing patterns are related to how the neurons are connected.

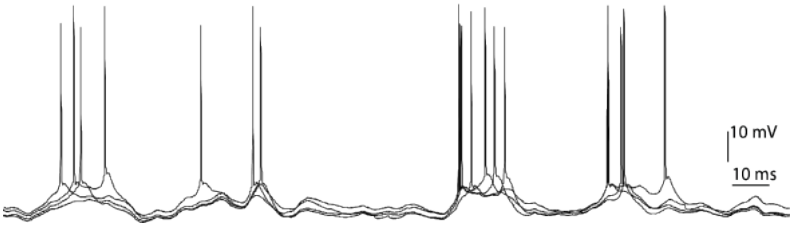


Figure 1.1: Neuronal variability: the exact same constant stimulus is applied to a single in vitro neuron in four different trials; the resulting spike trains are shown on top of each other. source: gerstner2014neuronal.

Within this general framework we will focus on the more specific problem of *neuronal variability*. It has long been known that the response of single neurons as well as the response of neuronal networks to the exact same electrical stimulus is usually different from trial to trial [4] [5] [6]. The spikes sometimes appear to happen in an almost completely random fashion. In terms of neural encoding—the idea that the outside world, entering in the form of external stimuli, should somehow be represented in the activity of the neurons in order for it to be processed—this seems highly counterintuitive: if a structured stimulus results in random activity, how is the neural system then encoding this stimulus, and how could it possibly compute something about the outside world? The common answer is that in this case most of the information seems to be captured not in the exact spike times, but in the rates at which the neurons fire those spikes. In other words, if someone would like to reconstruct the external stimulus from the neuronal dynamics, he or she would succeed better when using the spike rate following the stimulus, than when considering the exact timing of the spikes [7]. However, not all spiking happens completely randomly. Depending on the type, location and function of the neurons, different studies have shown that the timing of the spikes sometimes contains information as well [8,9]. The information about the stimulus is thus contained in a so-called time code, a rate code, or in a combination of both [10] [11].

In those cases where only the spike rate carries the information, the spiking processes can be described as Poisson processes: the spikes are produced randomly and independently, but with a fixed average firing rate (often denoted with the symbol λ). Poisson models (and slightly more complicated extensions thereof) are thus ubiquitous in spike train analysis. Often, one wishes to model the variability in this spike *rate* as well—after all, if the information is contained in the spike rate, this spike rate should be able to change depending on the stimulus received or the neural computation taking place [12,13]. This can be done by using a doubly stochastic process, i.e., a Poisson or related process where the

[4] *Reading a neural code*. Bialek et al.,1991.

[5] *A relationship between behavioral choice and the visual responses of neurons in macaque MT*. Britten et al.,1996.

[6] *Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey*. Bair and Koch,1996.

[7] *Spikes: exploring the neural code*. Rieke and Warland,1999.

[8] *Deciphering the spike train of a sensory neuron: counts and temporal patterns in the rat whisker pathway*. Arabzadeh et al.,2006.

[9] *Bursting neurons signal input slope*. Kepecs et al.,2002.

[10] *Spike timing and spike rate make complementary contributions to perceptual decisions in rat S1 and S2 cortex*. Zuo et al.,2015.

[11] *Neural coding: rate and time codes work together*. Seth,2015.

[12] *Stimulus dependence of neuronal correlation in primary visual cortex of the macaque*. Kohn and Smith,2005.

[13] *Stimulus onset quenches neural variability: a widespread cortical phenomenon*. Churchland et al.,2010.

spike rate is in itself a random variable [14,15].

In short, the neuronal variability is often discarded as noise and replaced by statistical properties such as the spike rate, but the question remains how it actually arises: are neuronal systems just intrinsically noisy, or are we missing a part of the picture? In other words: is all the variability noise, or are we missing a part of the signal? At this point it is instructive to consider the variability resulting from three different phenomena [16]. The first one is an intrinsic source of noise that arises even in single, unconnected neurons. This noise is related to fluctuations in the permeability of the cell membrane [17]; since changes in this permeability lie at the basis of spike train generation, the spikes can potentially be generated at random, and under certain conditions *even in the absence of an external stimulus*. This is then called *spontaneous activity*.

The two other sources of variability arise in networks of connected neurons. One of them concerns the transmission of signals across synapses: experiments with double electrode recordings revealed that only 10 to 30 % of spikes are reliably transferred across chemical synapses from one neuron to the next [18] [19]. The reliability of signal transmission is thus a stochastic element that definitely influences the dynamics of connected neurons.

The last source of variability we consider leads us back to the original, general question of how network structure relates to neuronal dynamics. We will focus this discussion on cortical neurons. Numerous experiments have indicated that spiking neurons in cortex of animals exhibit a strong degree of temporal irregularity, and that their inter spike interval distributions are usually in strong agreement with those generated by processes with Poisson like statistics [20] [21] [22]. Since active cortex are processing information, this cannot be pure, intrinsic noise alone. The first argument for the variability in the output of each neuron could be that cortical neurons receive synaptic input with large variability: this variability might be due to fluctuations in the sensory stimuli the cortical neurons receive, or it might result from the stochastic action of their synapses. But since cortical cells in vivo have thousands of synaptic contacts (and since the structure of a neural cell leads up to a integration or summation of its inputs over a typical time window), this variability should more or less average out, yielding an output with very small residual variability. Unless of course the fluctuating synaptic inputs are substantially correlated. Experiments show that spike trains of pairs of neurons in cortex can be correlated within a relatively narrow time scale (order 10ms) [23] [24], but only a small fraction of the total activity is tightly correlated in this way. Another explanation is to be found in the network structure.

Neurons can be *excitatory* or *inhibitory*: a signal or spike stemming from an excitatory neuron will increase the probability of spiking in a connected neuron, a signal stemming from an inhibitory neuron will

[14] *Point processes*. Cox and Isham,1980.

[15] *Variance as a signature of neural computations during decision making*. Churchland et al.,2011.

[16] *Neuronal dynamics: From single neurons to networks and models of cognition*. Gerstner et al.,2014.

[17] *Detecting and estimating signals in noisy cable structures, I: Neuronal noise sources*. Manwani and Koch,1999.

[18] *The probability of transmitter release at a mammalian central synapse*. Hessler et al.,1993.

[19] *Redistribution of synaptic efficacy between neocortical pyramidal neurons*. Markram and Tsodyks,1996.

[20] *Power spectrum analysis of bursting cells in area MT in the behaving monkey*. Bair et al.,1994.

[21] *An intracellular analysis of the visual responses of neurones in cat visual cortex.*. Douglas et al.,1991.

[22] *The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs*. Softky and Koch,1993.

[23] *Dynamics of neuronal interactions in monkey cortex in relation to behavioural events*. Vaadia et al.,1995.

[24] *Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties*. Gray et al.,1989.

decrease that probability. It has been shown that theoretical models of randomly, uniformly connected networks where these two types of connections are *balanced* can result in spiking activity with Poisson like statistics [25] [26] [27] [28], even modeling the presence of intrinsic noise or other stochastic elements. This because the behavior is chaotic, in the sense that slight changes in the initial conditions lead to drastically different firing activities. The activity thus merely looks random, but is actually produced in a deterministic system. Anatomical studies have furthermore shown that the connections in cortex are not uniformly distributed but instead can cluster into local groups of highly connected neurons; theoretical models incorporating these types of network structure give rise to doubly stochastic dynamics [29] [30]. Further structural considerations, such as distance-dependent connectivity, improved the extent to which this balanced network theory can explain some aspects of neural variability [31].

[25] *Chaos in neuronal networks with balanced excitatory and inhibitory activity*. Van Vreeswijk and Sompolinsky,1996.

[26] *Chaotic balanced state in a model of cortical circuits*. Vreeswijk and Sompolinsky,1998.

[27] *Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons*. Brunel,2000.

[28] *Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons*. Ostojic,2014.

[29] *Slow dynamics and high variability in balanced cortical networks with clustered connections*. Litwin-Kumar and Doiron,2012.

[30] *Two layers of neural variability*. Churchland and Abbott,2012.

[31] *The spatial structure of correlated neuronal variability*. Rosenbaum et al.,2017.

Despite this recent progress, we are still a long way from fully understanding the link between network structure and activity, and by extension, between network structure and function. And even though theoretical network models are shedding light on various aspects of this interesting problem, unambiguous experimental verification of these models is hard to obtain. In fact, to verify the exact relation between network structure and activity, one needs to know the exact network structure and its exact activity. In fact, one needs *a directed graph of information flow* [32], where the nodes represent single neurons and the edges represent the directed transmission of neural signals. Such a graph can only be constructed if one can measure the activity of multiple connected neurons, and if one can subsequently attribute the resulting spiking patters to the neurons that produced them. Given the very small scales involved, both in space (typical neural structures have sizes of the order of a couple of μm), time (a typical spikes lasts around $1ms$) as activity (spikes correspond to voltages in the order of ten to hundreds of mV), as well as the intricate three dimensional structure neural networks can form, this has been proven technically very challenging. Furthermore, inserting electrodes in the brain of living animals, and certainly humans, damages the underlying cells and network structure.

[32] *Cliques of neurons bound into cavities provide a missing link between structure and function*. Reimann et al.,2017.

In this master thesis project, we constructed a directed graph of the information flow between active human cortical neurons growing on top of a high-resolution multi-electrode array. The project brings together—for the first time—some recent spectacular advancements in science: induced pluripotent stem cells (iPSC) that can be driven to become human cortical neurons and their supporting glia cells [33]; the development of multi electrode arrays for the extracellular recordings of neuronal ensembles with an unprecedented resolution [34]; the development of algorithms to, from these extracellular recordings, attribute the activity to the correct neurons in a fully automated way [35], and finally, the needed computational resources and procedures to handle large amounts of data [36]. The aspects specific to this project are shown in Fig. 1.2.

[33] *The human brain in a dish: the promise of iPSC-derived neurons*. Dolmetsch and Geschwind,2011.

[34] *High-resolution CMOS MEA platform to study neurons at subcellular, cellular, and network levels*. Müller et al.,2015.

[35] *A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo*. Yger et al.,2018.

[36] *Big data: A review*. Sagiroglu and Sinanc,2013.

The use of stem cell derived cortical neurons in an in vitro setup has its advantages, but it also has obvious drawbacks: how the network formed by these neurons relates to the structures that develop in vivo is not exactly known [37] [38]. A substantial advantage of the in vitro approach, however, is the fact we are able to exactly control the external input the network receives, by either stimulating it at certain sites or by not stimulating it at all. In this thesis a network that was not stimulated at all is studied; it furthermore lacked the inhibitory connections to bring it a balanced network state. Even such a network forms spontaneous connections and shows spontaneous activity and variability, even when there is no meaningful computational task at hand. Eventually, such studies could hopefully tell us more about what should be considered meaningful signal and what is merely noise, and their results can act as a baseline to compare with networks that do receive stimulation.

[37] *Modeling human cortical development in vitro using induced pluripotent stem cells.* Mariani et al.,2012.
[38] *Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome.* Paşca et al.,2011.

The main body of work for this thesis then consisted in constructing a pipeline of numerical procedures that take a recording of the neuronal activity, and output a directed graph of information flow, together with detailed information about the nature of these connections in terms of in transmission delays. This pipeline consists of the following steps:

1. identifying the spiking patterns belonging to separate neurons,
2. calculating the auto- and cross-correlations between pairs of neurons,
3. calculating the so-called network autoregressive kernels, which reflect the flow of information between the different neurons and capture the information about their transmission delays,
4. and finally the construction and visualization of a directed graph of information flow.

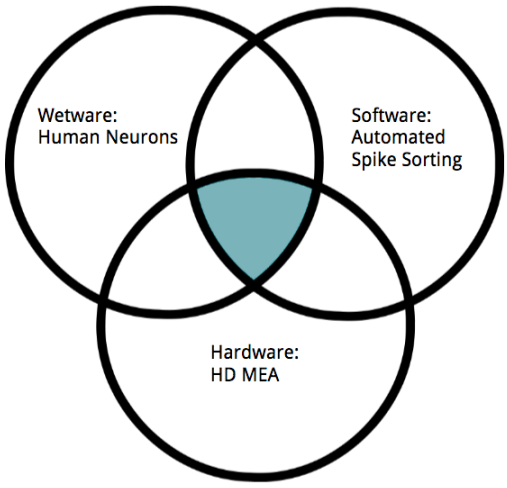


Figure 1.2: The recent scientific advancements that lie at the basis of this project.

We furthermore propose a straightforward and easy to implement classification scheme that relates the activity patterns of single neurons

to their connectivity. Along with these numerical procedures we present some analytical results, mainly from the field of stochastic process analysis, that helped us gain insight in our results. Finally, we will apply these procedures to the data recorded from our in vitro networks of cortical neurons. This setup, although a highly simplified version of real in vivo cortical networks, proved to be an excellent testbed to tune and verify our proposed numerical procedures.

Reading Guide

Directly following this first introductory chapter, we continue with a chapter in which we discuss the basic properties of neurons in general, and cultured cortical neurons in particular.

In chapter 3, we elaborate more on the physical models underlying deterministic and spontaneous action potential generation. We devote a large section to a detailed analysis of stochastic spike spike train processes, as these insights will be of great value to interpret our final results.

In chapter 4, we discuss our proposed method of detailed connectivity analysis. We again relate this to a stochastic process, in order to further enhance our understanding of the methods we use.

In chapter 5, we comment on the practical aspects of the project: we present some details about the experimental setup, along with descriptions of the needed numerical computations. We also introduce our straightforward method to classify spiking patterns.

Finally, in chapter 6, we report in a proof-of-concept way on our obtained results, and we elaborate on how they relate to previous studies and more realistic setups.

2

Introduction to Cultured Cortical Neural Networks

2.1 Single Neurons

2.1.1 Neuron Morphology

There exist a great many type of neurons, and they could be classified in a great many ways: based, e.g., on their shape, function, location, activity or neurotransmittor production [1]. Their actual shape depends on their function and the environment they develop in; it is for example the case that stem cell derived neurons kept in vitro will in general develop different from in vivo neurons of the same type [2] [3]. But it is of course useful to discuss the common and most significant elements of all these different neuron morphologies. A schematic representation of these different structures is given in Fig. 2.1.

[1] *Fundamental neuroscience*. Squire et al.,2012.

[2] *Modeling human cortical development in vitro using induced pluripotent stem cells*. Mariani et al.,2012.

[3] *Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome*. Paşca et al.,2011.

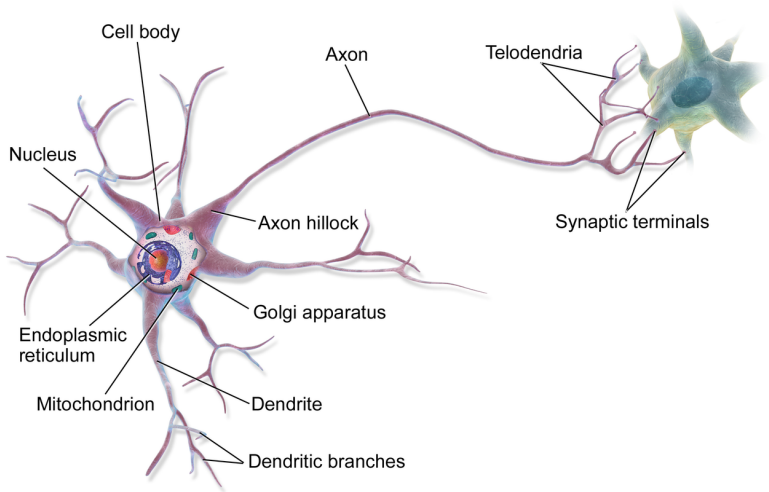


Figure 2.1: Schematic representation of a multipolar neuron, and how its axon is leading up to the next neuron.

The structures that we will most often use to interpret our results are the following:

- *The dendrites*, often forming a structure that is metaphorically called a 'dendritic tree'. These structures receive the extracellular signals and transmit them to
- *the soma*, or neuron cell body, where all these extracellular signals are integrated and passed on to
- *the axon initiation segment (AIS)*, where the action potentials (see later) originate. This structure is followed by
- *the axon*; which transfers the action potentials from the AIS to the synaptic terminals, and finally this structure branches to end up in
- *the synapses*. Here the signals are transferred from the axon terminals to the dendrites of the next neuron.

2.1.2 *Spikes and Action Potentials*

Although, as we discussed in the introduction, it is not always the neuron spikes *in themselves* that carry the information, they do form the building blocks of the processes we eventually want to study. Often they are taken to be binary events: either the neuron emits a spike (1), or it is silent (0). This abstraction removes both the neuron morphology as well as the underlying mechanism of action potential generation from the picture; but, as will become clear later, both these concepts—and their combination in terms of action potentials traveling in and between different neurons—will turn out to be important for our analysis. Here we will first discuss the action potential in itself; in the next section, we will explain how the action potentials are generated and transported within a single neuron.

In essence, action potentials are the result of a rapid de- and repolarization of the neuron cell membrane: the potential difference across the membrane abruptly 'spikes' from its resting value, and then it rapidly falls down again. After this voltage spike the membrane voltage is for a short period in time even lower than its baseline value before the spike; this is called the hyperpolarization phase or refractory period. The typical shape of an action potential is schematically shown in Fig. 2.2. We will discuss the general model underlying action potential generation, called the Hodgkin-Huxley model [4], in depth in section 3.1. This model uses a set of coupled differential equations to described how the interplay between concentrations of sodium and potassium ions on the one hand, and the cell membrane on the other hand, leads up to the generation of an action potential. Here we will start with a general introduction to the process, an introduction we will subsequently connect to the neuron morphology.

[4] *A quantitative description of membrane current and its application to conduction and excitation in nerve.* Hodgkin and Huxley,1952.

The rapid increase in the voltage is possible due to a feedback mechanism between the permeability of the membrane and the potential differ-

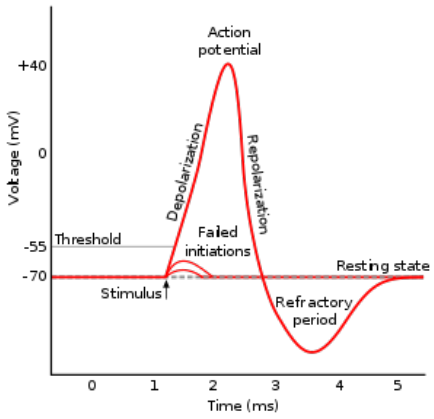


Figure 2.2: Schematic representation of a typical action potential.

ence across the membrane. At rest, the membrane has a negative resting potential difference denoted V_r. A slight increase in this voltage, due to the arrival of a stimulus or sometimes due to stochastic effects, opens up certain so-called *ion channels* in the membrane, which allow ions to diffuse from the outside of the cell to the inside. This depolarizes the membrane further. The feedback consists of the ion-channels being voltage dependent, i.e., more channels open as the voltage increases, yielding an ever faster voltage increase or depolarization. In the same process, however, different types of ions and of ion channel states are involved: the channels that were initially opening up become *inactivated*, and other voltage-dependent channels that can transport ions from inside the cell to the outside start to open, such that the membrane potential difference rapidly decreases again. It finally reaches a value that is below the original membrane voltage V_r. After that, the systems needs a certain period of time to restore itself to its original state.

The feedback mechanism only really 'takes off' once a certain voltage threshold is crossed. If an incoming stimulus is not strong enough to depolarize the membrane above the threshold, no action potential will be initiated (see Fig. 2.2). The needed strength of the stimulus to lead up to an action potential initiation depends naturally on the voltage across the membrane: during the hyperpolarized phase or refractory period, the stimulus needs to be even stronger for the threshold to be crossed. This is called the relative refractory period; a period during which the same stimulus is less likely to induce a action potential, compared to when it would arrive when the cell membrane patch was in its resting state. During the first part of the refractory period, corresponding to the time period in which the neuron is generating an action potential, it is not physically possible for the membrane to generate an second action potential. This period is called the absolute refractory period.

2.1.3 *Initiation and Transmission of Action Potentials*

Action potentials make up the communication units between different neurons, and they can be seen as the result of a local computation:

- the input consists of action potentials from other neurons the neuron at hand receives through its synapses. The chemical synapses are located at the end of its dendrites;
- the computation is the local integration of the signals through the dendritic tree and the soma. If the integrated signal reaches the axon initial segment (AIS) and is above the action potential initiation threshold, a new axon potential will be initiated at this site.
- This initiated action potential is the output of the system and can be transferred to the next neuron by traveling down the axon and its branches to the axon terminals and synapses.

The action potential initiated at the AIS can travel down the axon because it is regenerated: the currents resulting from the change in voltage spread across the axon and depolarize adjacent patches of the membrane, often (but not always) inducing a new action potential and thus continuing the signal propagation [5] [6]. Once the traveling action potential arrives at the end of the axon, it is potentially transmitted in a chemical synapse: a junction between the axon of the original neuron and the dendrites of the next neuron (often called the postsynaptic neuron), where, induced by the action potential, neurotransmitter molecules are released to bind on the receptors of the postsynaptic cell. This process is illustrated in Fig. 2.3.

[5] *Evidence for electrical transmission in nerve*. Hodgkin,1937.
[6] *Axon physiology*. Debanne et al.,2011.

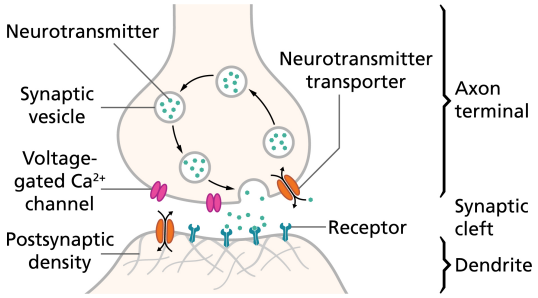


Figure 2.3: Schematic representation of a chemical synapse.

The binding of the neurotransmitters to the receptors of the postsynaptic cell again induces a change in the voltage of the postsynaptic cell membrane due to activation of certain voltage dependent ion channels. However, the binding might induce two opposite changes: if it increases the voltage (and thus depolarizes the cell membrane), the synapse is called excitatory; if the binding affects the voltage in the opposite way (polarizing the cell membrane further), the synapse is called inhibitory. Either way these changes propagate passively to nearby regions of the postsynaptic dendritic membrane. Unlike in the axon this process is not regenerative: typically the signals decay exponentially with the distance

from the synapse. But an excitatory signal might eventually reach the axon hillock, and if this voltage stimulus is still sufficiently large, a new action potential might be initiated at this site. It is far more likely, however, that different excitatory signals from presynaptic neurons have to arrive together to ignite an action potential at the axon hillock. As we mentioned earlier, this can be seen as a kind of local computation; and the complexity of this computation is enhanced through the existence of inhibitory signals that can modulate the input-output relation even further.

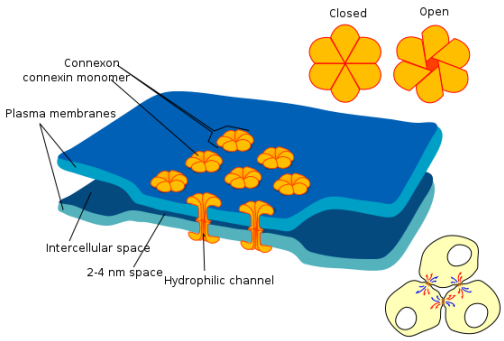


Figure 2.4: Electrical synapse.

Chemical synapses are not the only way in which neurons could transfer their action potentials. A faster connection is achieved when the neurons have both their membranes connected in an electrical synapse: so-called connexons then allow the ionic current of the presynaptic action potentials to be directly transfered to the postsynaptic cell. A schematic representation of an electrical synapse is shown in Fig. 2.4. This process is much faster than the transferal in a chemical synapse, because there is no need for the (relatively) slow diffusion of neurotransmitters. It can only be formed, however, in cells that are close to each other: usually, the cells approach within 4-5 nm of each other, whereas 20 to 40nm separates the cells connected by a chemical synapse [7].

[7] *Principles of neural science*. Kandel et al.,2000.

2.1.4 *Overview of Typical Timescales and Dimensions*

The following overview of typical timescales and dimensions that will turn out to be important for interpreting our results are compiled from Kandel et al. [2000] [8] and Debanne et al. [2011] [9].

[8] *Principles of neural science*. Kandel et al.,2000.

[9] *Axon physiology*. Debanne et al.,2011.

Spatial Dimensions. Typically

- The soma has a diameter of **50 μm** or more.
- The axon has a diameter between **0.2 μm and 20 μm**.
- For a chemical synapse, the distance between cell pre- and postsynaptic cell membranes is between **20 nm and 40 nm**. These cell mem-

branes are the membranes of the synaptic terminals of the axon and the dendrites (or soma) of the next cell.

- For an electrical synapse, the distance between cell pre- and postsynaptic cell membranes is between **4 nm and 5 nm**. These cell membranes are the membranes of the somas.

Transmission Speeds and Delays. Typically

- the electrical synaptic delay is **virtually absent**. The transmission is usually bidirectional.
- the chemical synaptic delay is **at least 0.3ms, and usually 1-5ms or longer**. The transmission is unidirectional.
- axonal conduction velocity depends on the axon diameter and on axon myelination (a structural property that greatly enhances conductance speed). For unmyelinated axons, the conduction velocity is typically **0.4 to 2.0 m/s**.

Amplitudes and Signal to Noise. Typically

- action potential amplitudes depend on the location in the neuron and the received external stimulus. Extra-cellular voltage recordings are typically **between 5 and 500 mV**.
- the action potential amplitude is highest in the axon initial segment (AIS), lower in the soma and in the rest of the axon, and very small in the dendritic structures.
- The signal to noise ratio naturally depends on the measuring device; in our setup the largest amplitudes are up to **180** σ_{noise} while the axonal signals (not including the AIS) are usually **1-2** σ_{noise}. Dendritic signals are unlikely to be detected or are buried in the noise.

Spike Rate Properties. Typically

- the absolute refractory period equals the duration of the action potential (**1ms**)
- relative refractory period takes around **4-5ms**
- the spike rate of spontaneous activity is **a couple of spikes per second**.

2.2 *Connected Cultured Cortical Neurons*

As we mentioned before, measuring the activity of cortical neural networks in vivo is a technically challenging task, and the complicated and dense structure of those networks makes it even more daunting (or even sheer impossible) to analyze their detailed activity patterns and the related fundamental network properties. Studying such networks in vitro alleviates this task greatly—but has the drawback of creating a perhaps unrealistic situation. However, it has been shown that cortical networks cultured in vitro are comparable to immature, developing neocortex networks in vivo [10] [11]. Studying developing networks in vitro can thus at least shed light on the interplay between early network *formation* and activity.

[10] *Spontaneous neuronal activity in developing neocortical networks: from single cells to large-scale interactions*. Luhmann et al.,2016.

[11] *Spontaneous neuronal discharge patterns in developing organotypic mega-co-cultures of neonatal rat cerebral cortex*. Baker et al.,2006.

There are in general three different types of in vitro networks one can obtain: animal *brain slices* will have an intricate, fully developed network structure that is more difficult to analyze, but is of course very close to the in vivo structure; *dissociated cultures* are obtained from (chemically) dissociating a premature or mature animal brain region of interest and extracting its neural cells [12]; and finally, as in our case, one can use a *stem cell culture*, where neural cells and their supporting glia cells are derived from induced pluripotent stem cells. We will denote the latter two categories as *cultured* neurons.

[12] *Dissociated cortical networks show spontaneously correlated activity patterns during in vitro development*. Chiappalone et al.,2006.

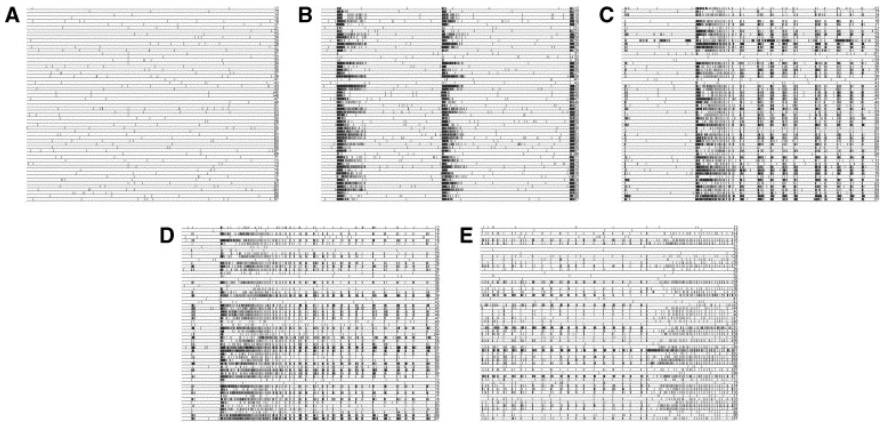


Figure 2.5: Raster plots showing the recorded spike trains of a developing dissociated cortical network at different days in vitro (DIV). Time runs horizontally, while the vertical axis corresponds to the different recorded neurons. (A) 7 DIV; (B) 14 DIV; (C) 21 DIV; (D) 28 DIV; (E) 35 DIV. Adapted from Chiapalone et al., 2016.

After a few days in culture, neurons start to connect to each other with functionally active synapses, forming a random network and displaying spontaneous activity. Over time the networks shift from electrical to chemical transmission, while at the same time the axonal connectivity increases. At this stage the (dissociated) networks typically show repetitive *burst* discharges [13] [14] [15] [16], i.e., they show spike trains where a couple of spikes occur in very rapid succession, followed by relatively longer periods of 'silence'. These burst can be synchronized over large fractions of the culture. An example of these kinds of patterns developing over time in a network of cultured neurons is shown in Fig. 2.5.

[13] *The mechanisms of generation and propagation of synchronized bursting in developing networks of cortical neurons*. Maeda et al.,1995.

[14] *Dissociated cortical networks show spontaneously correlated activity patterns during in vitro development*. Chiappalone et al.,2006.

[15] *Network dynamics and synchronous activity in cultured cortical neurons*. Chiappalone et al.,2007.

[16] *Spontaneous neuronal activity in developing neocortical networks: from single cells to large-scale interactions*.

The precise structure of these spatio-temporal patterns depends on a number of factors [17], including:

- the properties of the neurons, i.e., their structure, membrane properties, and their timescales of recovery and refraction,
- the flow of activity through the network, determined by the effective connectivity,
- the timescales of signal propagation, i.e., the axonal and synaptic transmission speeds and the timescale of dendritic integration,
- the balance between inhibitory and excitatory connections.

Several models and simulations based on the above factors have provided insight in how this synchrony could arise [18] [19] [20] [21] [22] [23]. We want to emphasize two important facts: first, it is worth noting that the spatio-temporal patterns depend on the ratio between inhibition and excitation in the network; but, unlike in the case of chaotic behavior in a balanced network, the presence of inhibitory neurons is not a necessary condition for synchronization to emerge. In our developing network of purely excitatory neurons, there should thus be some kind of synchronization present, presumably depending on the connectivity pattern. Furthermore, it has been shown experimentally [24] that very young cultures exhibit a random topology, which over time evolves to a so-called small-world topology (i.e. the neurons start to form local clusters) [25]. This kind of network balances integration of network areas with segregation of specialized processing units, which increases the network efficiency [26]. The fact that this topology can arise without external stimulation points to the presence of intrinsic biological mechanisms.

[17] *Dynamics and plasticity in developing neuronal networks in vitro*. van Pelt et al.,2005.

[18] *Oscillations, complex spatiotemporal behavior, and information transport in networks of excitatory and inhibitory neurons*. Destexhe,1994.

[19] *The mechanisms of generation and propagation of synchronized bursting in developing networks of cortical neurons*. Maeda et al.,1995.

[20] *Inhibition can disrupt hypersynchrony in model neuronal networks.*. Deyo and Lytton,1997.

[21] *Emergent oscillations in a realistic network: the role of inhibition and the effect of the spatiotemporal distribution of the input*. Pauluis et al.,1999.

[22] *Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons*. Brunel,2000.

[23] *Changing excitation and inhibition in simulated neural networks: effects on induced bursting behavior*. Kudela et al.,2003.

[24] *Emergence of a small-world functional network in cultured neurons*. Downes et al.,2012.

[25] *Emergence of a small-world functional network in cultured neurons*. Downes et al.,2012.

[26] *Complex networks: Structure and dynamics*. Boccaletti et al.,2006.

3

Neural Dynamics

3.1 Deterministic Action Potential Generation

In this section, we will discuss action potential generation and their underlying models. Action potentials are the results of currents flowing through the ion channels of the neuronal cell membrane, and Hodgkin and Huxley [1] were the first to derive differential equations for these dynamics through a extensive series of experiments on the giant axon of the squid.

[1] *A quantitative description of membrane current and its application to conduction and excitation in nerve.* Hodgkin and Huxley,1952.

3.1.1 The Hodgkin-Huxley model

The cell membrane of a neuron separates the interior of the cell from the extracellular space; a difference in ion concentration in those two regions naturally results in an electric potential across the membrane. The cell membrane itself consists of a lipid bilayer that acts as an electrical isolator. However, there exist specific proteins throughout the layers that act as ion channels: some of them are capable of actively pumping ions across the membrane (and are therefore adequately called *ion pumps*), other merely form channels through which the ions can migrate. More than 200 different ion channels are known to date [2], but the original Hodgkin-Huxley model is based on the dynamics of three specific ion channels: a potassium (K) channel, a sodium (Na) channel, and a general channel through which all sorts of ions can 'leak'. The general dynamics of action potential generation does not depend on the specificity of the ions, however; as we will soon see, the Hodgkin-Huxley model reveals that it is the role they take and the interplay between them that matters.

[2] *Channelpedia: an integrative and interactive database for ion channels.* Ranjan et al.,2011.

Passive State

Let us first consider the state of the system when no action potential is being generated. This *passive state* is a steady state governed by two different "forces": an electrical force and a thermodynamic force. The thermodynamic force arises due to the difference in concentrations in- and outside the cell for each type of particle; the electrical force of course arises due to the difference in the concentration of charges.

A typical mammalian cell in the passive state has a concentration of around 18 millimoles Na^+ inside the cell, compared to 150 millimoles outside. In contrast, the concentration of K^+ is much higher *inside* the cell: typically around 135 millimoles inside, compared to 3 millimoles outside (for more details, see [3]). Let us focus on the K^+ molecules for a moment: due to their concentration gradient, the K^+ molecules tend to move *outwards*. This movement causes the intracellular potential to become more negative, offsetting the outward flow, and at some membrane potential difference E_K the system reaches a dynamic equilibrium (if only potassium ions were present and conducted). The value of E_K can be calculated from thermodynamic principles (specifically, using the Nernst equation [4]), and for the above typical mammalian cell and temperature $T = 37^\circ C$, the value is $E_K = -102mV$. Similarly, the value for the equilibrium potential of the sodium ions would be $E_{Na} = +56mV$.

[3] Larry Squire, Darwin Berg, Floyd E Bloom, Sascha Du Lac, Anirvan Ghosh, and Nicholas C Spitzer. *Fundamental neuroscience*. Academic Press, 2012

[4] *Zur kinetik der in lösung befindlichen körper*. Nernst,1888.

Due to the presence of multiple ion concentrations—all types of kations, and all types of anions—none of these dymical equilibria is actually reached, and the system is instead in a steady state with a potential difference across the membrane that lies in between the equilibrium values. This potential difference is called the resting potential, and is denoted V_r. The exact value can be calculated through the Goldman-Hodgkin-Katz equation [5] and lies, for a typical mammalian cell, usually between $-60mV$ and $-75mV$. If one knows the Nernst equilibrium potential for each of the ions separately as well as the conductance of the membrane due to the ion channels, V_r can be calculated straightforwardly from noting that, in a steady state, the total current density should be zero:

[5] *Potential, impedance, and rectification in membranes*. Goldman,1943.

$$\begin{aligned} i_{tot} &= 0 && (3.1)\\ &= i_{Na} + i_K + i_L && (3.2)\\ &= G_{Na}(V_r - E_{Na}) + G_K(V_r - E_K) + G_L(V_r - E_L) && (3.3) \end{aligned}$$

where i denotes a current density, and G is the conductance per unit area for each type of ion. Therefore:

$$V_r = \frac{G_{Na}E_{Na} + G_K E_K + G_L E_L}{G_{Na} + G_K + G_L}. \quad (3.4)$$

The difference $(V_r - E_X)$ thus acts as a driving force that keeps the system of ions of type X from reaching its dynamical equilibrium. Hodgkin and Huxley also derived that the conductances G_{Na} and G_K are actually voltage dependent, which will be discussed in the next paragraph.

Gating Models

Before we elaborate on the deterministic equations that describe the gating mechanisms in the Hodgkin and Huxley model, we would like to emphasize that the ion channels actually open and close stochastically. The Hodgkin and Huxley model thus corresponds to an average taken over a sufficiently large amount of experiments. This is illustrated in Fig. 3.1.

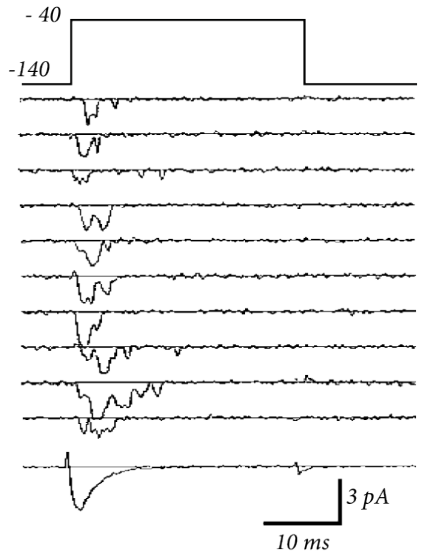


Figure 3.1: Stochastic channel activation. The current flowing to the membrane after application of a voltage step (top row) shows step-like changes and is different in each trial (subsequent traces). The bottom trace is the average over many traces, which corresponds to the description in the Hodgkin-Huxley model. Source: Gerstner et al, 2014.

The deterministic gating model, then, is based on the assumption that populations of channels of the same type undergo changes collectively; this means that we can describe the population as transitioning between the states of all *open* and all *closed*, depending on the voltage across the membrane. This transition is governed by voltage dependent rate coefficients $\alpha(V)$ (closed $\rightarrow$ open) and $\beta(V)$ (open $\rightarrow$ closed). Thus, if we denote the fraction of an ion channel population that is open with n, we can write:

$$\frac{dn}{dt} = -\beta(V)n + \alpha(V)(1-n) \tag{3.5}$$

$$\Leftrightarrow \tau_n(V)\frac{dn}{dt} = -n + n_\infty(V) \tag{3.6}$$

where we defined

$$\tau_n(V) \equiv \frac{1}{\alpha(V) + \beta(V)} \tag{3.7}$$

with dimension of time, and

$$n_\infty(V) \equiv \frac{\alpha(V)}{\alpha(V) + \beta(V)} \tag{3.8}$$

a dimensionless quantity. For time-dependent voltages, i.e., $V = V(t)$, the solution to Eq. 3.6 is in general given by:

$$n(t) = n(t_0)h(t) + \int_{-\infty}^{+\infty} h(t-t')n_\infty(t')dt' \tag{3.9}$$

with

$$h(t) = e^{-\frac{t-t_0}{\tau_n(t)}}\Theta(t-t_0). \tag{3.10}$$

The solution to Eq. 3.6 for a given constant voltage (thus τ_n constant and n_∞ constant) can be derived easily from solving the differential equation, or one could use Eq. 3.9:

$$n(t) = n(t_0)e^{-\frac{t-t_0}{\tau_n}} + n_\infty(1 - e^{-\frac{t-t_0}{\tau_n}}). \tag{3.11}$$

Thus, if the voltage across the membrane remains fixed, the fraction of open channels n eventually reaches the equilibrium value n_∞, and it does so with time constant τ_n. Hodgkin and Huxley used 3 such variables in their description of the gating mechanisms:

- m describes the fraction of open Na^+ channels
- h describes the fraction of inactivated (see below) Na^+ channels
- n describes the fraction of open K^+ channels

The equilibrium values and time constants as a function of the membrane for these gating variables are shown in Fig. 3.2. These values were experimentally determined by Hodgkin and Huxley. They also found, from fitting different functions to their data, that the conductances of the membrane were related to these variables in the following way:

$$\frac{1}{R_{Na}} = g_{Na}m^3h, \tag{3.12}$$

and

$$\frac{1}{R_K} = g_K n^4, \tag{3.13}$$

where g_{Na} and g_K are constants denoting the maximum conductance per unit area of the membrane.

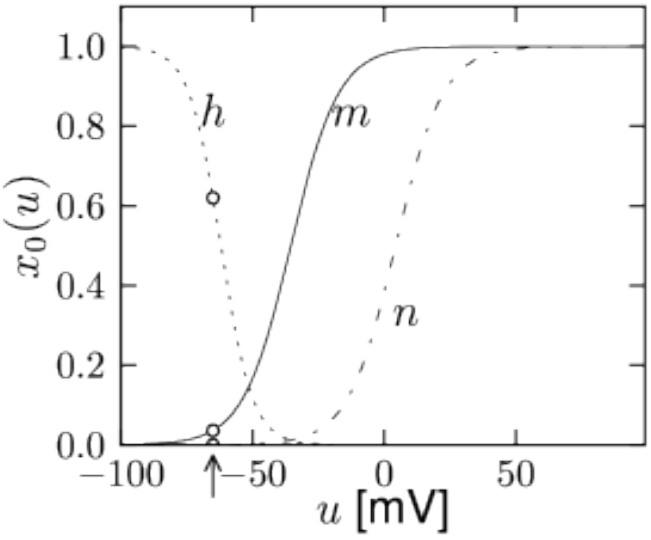


Figure 3.2: The curves h_∞, m_∞ and n_∞ in function of the membrane potential difference u.

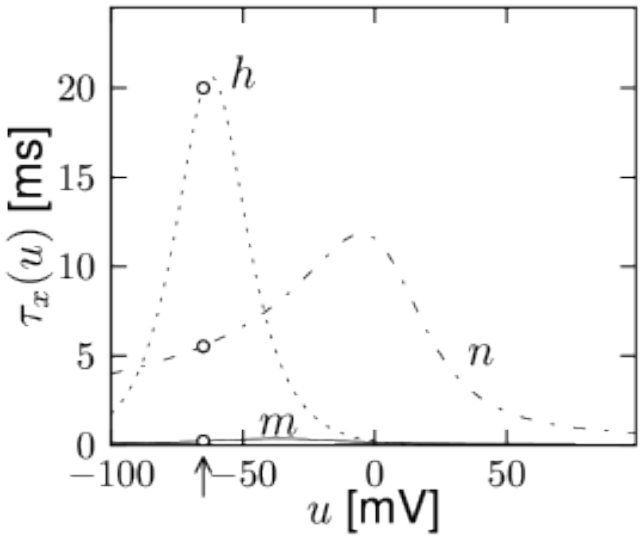


Figure 3.3: The curves τ_h, τ_m and τ_n in function of the membrane potential difference u.

Action Potentials

Finally, to understand how an action potential can be generated, it is instructive to translate the above descriptions into an electrical circuit. Such a circuit is shown in Fig. 3.4. The membrane itself can be modeled by a capacitor with capacitance C; each channel type is represented by a resistor: the sodium channel resistance R_{Na}, the potassium channel resistance R_K and the resistance of the general 'leak' channel R_L; and, knowing that the sodium and potassium channels are actually voltage dependent, they are pictured with a diagonal arrow. The corresponding Nernst potential differences E_{Na}, E_K and E_L make up the batteries in the circuit. The potential difference across the membrane (thus the capacitor) is denoted u.

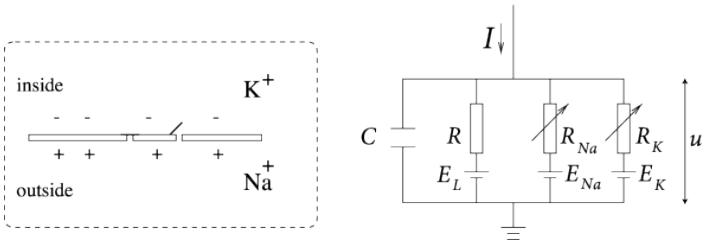


Figure 3.4: Visualization of the resting potential difference across the cell membrane V_r due to different concentrations of ions in- and outside the cell; and this biophysical description represented as an electrical circuit. Source: Gerstner et. al., 2014.

We can now derive what happens when an external current I_{ext} arrives in the circuit. This current will be split in a current I_C that charges the capacitor, and in different components I_k which pass through the ion channels. Thus:

$$I_{ext}(t) = I_C(t) + \sum^{k} I_k(t) \tag{3.14}$$

$$= I_C(t) + \frac{1}{R_{Na}}(u - E_{Na}) + \frac{1}{R_K}(u - E_K) + \frac{1}{R_L}(u - E_L). \tag{3.15}$$

Since $C = q/u$, with q the charge on the capacitor, we can write $I_C = C\frac{du}{dt}$. Rearranging Eq. 3.15 and using the expressions for the conduc-

tances (Eq. 3.12 and Eq. 3.13), we arrive at:

$$C\frac{du}{dt} = -g_{Na}m^3h(u - E_{Na}) - g_K n^4(u - E_K) - g_L(u - EL) + I_{ext}. \quad (3.16)$$

Together with the differential equations for the gating variables:

$$\tau_m \frac{dm}{dt} = -m + m_\infty, \quad (3.17)$$

$$\tau_n \frac{dn}{dt} = -n + n_\infty, \quad (3.18)$$

$$\tau_h \frac{dh}{dt} = -h + h_\infty, \quad (3.19)$$

this equation (Eq. 3.16) makes up the Hodgkin Huxley model.

3.2 *Spontaneous Activity*

In a deterministic model of action potential generation such as the Hodgkin-Huxley model, the ion channels in the considered patch of cell membrane are treated as a population of channels. It is assumed that the population properties, namely the fraction of channels that are in a certain state, continuously and deterministically evolve depending on the membrane potential difference. In reality, however, thermal noise drives conductance fluctuations in ion channels. This leads to the channels opening and closing randomly, and instead of a continuous and deterministic process, the process of ion permeation through the membrane is thus foremost a discrete and stochastic process. However, one might naturally expect that the deterministic model is recovered in the limit of large channel populations.

We will summarize two different analytical approaches to these stochastic dynamics, but they essentially lead to the same conclusion: channel fluctuations can lead to spontaneous action potentials, i.e., *an action potential can be generated without an external stimulus*. The first analytical approach [6] models the ion channels through Markov chains: each channel is assumed to randomly fluctuate between a discrete number of possible states. The transition probabilities between these states are (assumed to be) only dependent on the current state and the current membrane voltage. The exact transition matrix to be used can be derived from experimental data (for the ubiquitously used giant axon of the squid, one could, e.g., use the results of [7]). Once the transition matrix is determined, the Markov chain can be simulated in order to compare the results with the outcome of the deterministic Hodgkin-Huxley model (using experimentally determined parameters derived from a similar setup, e.g., measurements on the giant axon of the squid). As one would intuitively expect, the models only converge for large numbers of channels. For small numbers of ion channels, or, equivalently, for a smaller area of membrane patch at the same ion channel density, the models diverge significantly: the mean firing rate is non-zero even in the absence

[6] *Limitations of the Hodgkin-Huxley formalism: effects of single channel kinetics on transmembrane voltage dynamics*. Strassberg and DeFelice,1993.

[7] Isabel Llano, Christina K Webb, and Francisco Bezanilla. Potassium conductance of the squid giant axon. single-channel studies. *The Journal of general physiology*, 92(2):179–196, 1988

of an external stimulus in the stochastic case, and it is higher with respect to the deterministic model in the presence of an external stimulus, as can be seen from Fig. 3.5.

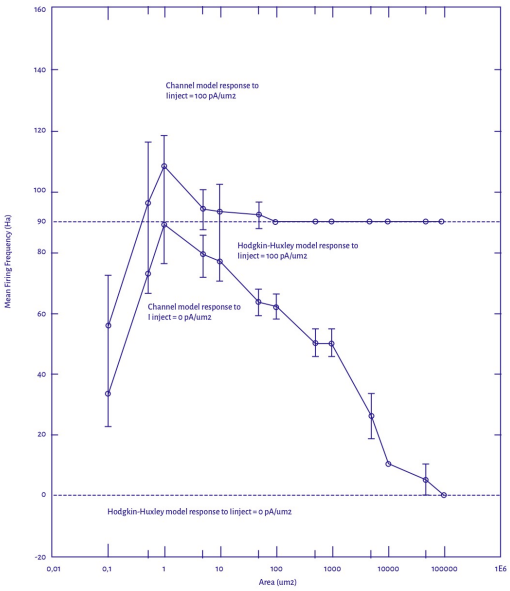


Figure 3.5: The mean firing rate is non-zero even in the absence of an external stimulus. Source: adapted from Strassberg and DeFelice

The second analytical approach [Chow and White, 1996] reformulates the problem by drawing an analogy between action potential generation and a particle escaping a potential well by jumping over a barrier. The analysis starts from a separation of the time scales: the sodium current has a faster time scale than the membrane potential difference, which in turn is faster that the time scales involving the potassium current (for details, see section 3.1). This allows to approximate the system by a bistable Langevin equation where the location of the particle corresponds to the membrane potential, and the inverse number of sodium channels plays the role of temperature. The channel noise, in this description playing the role of thermal noise, can occasionally "kick" the membrane potential over the barrier. The mean firing rate, which we will denote λ, can then be calculated using Kramers' classic result for barrier escape [Kramers, 1940]. It turns out that λ is an exponential function of the absolute number of sodium channels N_{Na}:

$$\lambda \sim e^{-N_{Na}}. \tag{3.20}$$

This corresponds approximately to the analysis of Strassberg and DeFelice (for a quick comparison, one can look at Fig. 3.5 where a log-linear scale is used; a detailed comparison is given in [8]). The inter spike interval (ISI) distribution, i.e., the probability distribution of the times between subsequent action potentials, can also be derived from this model [9]:

$$p_{ISI}(\tau) = \lambda(N_{Na})e^{-\lambda(N_{Na})\tau} \tag{3.21}$$

This indicates that the action potentials are fired randomly and independently: the ISI distribution corresponds to a Poisson process with mean

[8] *Spontaneous action potentials due to channel fluctuations.* Chow and White,1996.

[9] *Fokker-planck equation.* Risken,1996.

firing rate $\lambda(N_{Na})$. A more realistic model would incorporate the refractory period, which would result in a so-called a *renewal process*. Poisson processes and renewal processes as models for neuronal activity will be discussed in depth in section 3.3.

3.3 Spike Train Analysis

3.3.1 Spike Train Analysis of Single Neurons

Autocorrelation Functions and the Power Spectral Density. The autocorrelation function of a continuous stationary random process $x(t)$ observed between $t = -T$ and $t = +T$ is defined by

$$R_{xx}(t) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t+\tau)x(t)dt. \tag{3.22}$$

Let

$$x_T(t) \equiv \begin{cases} x(t) & \text{if } -T \leq t \leq T \\ 0 & \text{otherwise.} \end{cases}$$

We can write down the Fourier transform of this function:

$$\widetilde{x}_T(\omega) = \int_{-\infty}^{+\infty} x_T(t)e^{-i\omega t}dt = \int_{-T}^{T} x(t)e^{-i\omega t}dt, \tag{3.23}$$

and define the power spectral density $S(\omega)$ of $x(t)$ as

$$S_{xx}(\omega) = \lim_{T\to\infty} \frac{1}{2T}\widetilde{x}_T(\omega)\widetilde{x}_T^*(\omega). \tag{3.24}$$

The power spectral density has a useful physical interpretation: if $x_T(t)$ is a current flowing through a unit resistor during the interval $(-T, T)$, then $S(\omega)$ is the average power dissipated with frequencies between ω and $\omega + d\omega$. It is thus a useful characterization of a random process in the frequency domain. The power spectral density is furthermore related to the autocorrelation of the function through the *WienerâĂŞKhinchin* theorem:

$$R_{xx}(\tau) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} S_{xx}(\omega)e^{i\omega t}d\omega \tag{3.25}$$

and

$$S_{xx}(\omega) = \int_{-\infty}^{+\infty} R_{xx}(\tau)e^{-i\omega t}d\tau, \tag{3.26}$$

stating that the power spectral density and the autocorrelation function are related through (inverse) Fourier transforms. Given that these functions contain the same information, we can use the function that best suits our needs to analyze a certain process, and obtain its equivalent in the time or frequency domain when desired.

Correlation and Covariance. The tacit assumption when discussing correlation functions is (usually) that the values of these functions have been standardized such that they lie between -1 and 1. The fact that the

term correlation function is used where actually the covariance function is given might be confusing. We will try to clarify this where needed.

Poisson Processes with Finite Event Duration. We start with the case where a single neuron produces spikes which occur independently and in a random manner, such that the underlying stochastic process can be characterized as a Poisson process with a constant spiking rate λ. Let us recall some basic properties of such a Poisson process. The probability that exactly k spikes occur in a time interval (o,T) can be written as:

$$p_k(t) = \frac{(\lambda T)^k}{k!} e^{-\lambda T}. \tag{3.27}$$

The time intervals Δ between subsequent events, in the field of neuroscience conventionally called the *inter spike intervals* (ISI), are distributed according to:

$$p_{ISI}(\tau) = \lambda e^{-\lambda \tau} \tag{3.28}$$

A neuron typically produces a series of spikes with a particular shape, a shape which we will call the spike waveform or simply the *waveform*, and which we will denote with $f_w(t)$. The waveform in itself reflects the underlying process of de- and hyperpolarization as well as the method of measurement. But, once the spike times are known, one could construct a signal with any possible waveform—this could be done, as we will do, to smoothen the point process to be used in a numerical calculation. We can describe such a spike train $x(t)$ as $f_w(t)$ convolved with a series of Dirac delta distributions centered at the N spike times t_i:

$$x(t) = \sum_{i=1}^{N} f_w(t - t_i) = \sum_{i=1}^{N} f_w(t) * \delta(t - t_i), \tag{3.29}$$

and, to gain insight in the correlation functions calculated from the data, we can examine how different temporal patterns of the Dirac delta train will influence the theoretical correlation functions.

We will first derive an expression for the average of the signal $x(t)$. To this end we can start by considering exactly k spikes, occurring in an interval T. Since the spikes arrived independently, the probability of a spike occurring in a subinterval dt follows a uniform distribution with probability dt/T. The time average of the signal over this interval T is then given by:

$$E[x_k(t)] = \frac{k}{T} \int_0^T f_w(t - z) dz = \frac{k}{T} \int_{-\infty}^{+\infty} f_w(t - z) dz \tag{3.30}$$

where the last equation holds due to the compact support of $f_w(t)$. Thus, if we use the well-known result $E[k] = \sum_{k=0}^{\infty} k p(k) = \lambda T$, the (ensemble) average of the signal $x(t)$ is given by:

$$E[x(t)] = \sum_{k=0}^{\infty} E[x_k(t)] p_k(t) = \lambda \int_{-\infty}^{+\infty} f_w(t - z) dz = \lambda \int_{-\infty}^{+\infty} f_w(u) du. \tag{3.31}$$

We can now compute the power spectral density, and subsequently the autocorrelation function. Starting from the Fourier transform of Eq. 3.29:

$$\widetilde{x}_T(\omega) = \widetilde{f}_w(\omega) \sum_{i=1}^{N} e^{-i\omega t_i}, \tag{3.32}$$

we can write:

$$\widetilde{x}_T(\omega)\widetilde{x}_T^*(\omega) = \widetilde{f}_w(\omega)\widetilde{f}_w^*(\omega) \sum_{i=1}^{N}\sum_{j=1}^{N} e^{-i\omega(t_i - t_j)} \tag{3.33}$$

$$= \widetilde{f}_w(\omega)\widetilde{f}_w^*(\omega)\big(N + \sum\sum_{i\neq j} e^{-i\omega(t_i - t_j)}\big). \tag{3.34}$$

It can be shown that for large N and $\mid \omega \mid \gg \frac{1}{T}$, and in case the t_i and t_j are distributed independently, the double summation represents a number with amplitude of order $\sqrt{N}$. On letting N and T tend to infinity we can thus retain only the first term:

$$S_{xx}^{\backslash 0}(\omega) = \lim_{T\to\infty} \frac{1}{2T}\widetilde{x}_T(\omega)\widetilde{x}_T^*(\omega) = \frac{1}{2}\lambda\widetilde{f}_w(\omega)\widetilde{f}_w^*(\omega). \tag{3.35}$$

This equation is only valid for $\mid \omega \mid > 0$, however, which can be seen from the divergence of Eq. 3.34 when $\omega = 0$ and N going to infinity. This is why we added the $\backslash 0$ in superscript. The DC term can be obtained from the Fourier transform of the average of $x_T(t)$, and the complete power spectral density is given by:

$$S_{xx}(\omega) = \frac{1}{2}\lambda\widetilde{f}_w(\omega)\widetilde{f}_T^*(\omega) + 2\pi\delta(\omega)E[x(t)]^2. \tag{3.36}$$

Finally, on taking the inverse Fourier transform, we get (see Eq. 3.25 and Eq. 3.31):

$$R_{xx}(\tau) = \lambda\int_{-\infty}^{+\infty} f_w(t)f_w(t-\tau)dt + (\lambda\int_{-\infty}^{+\infty} f_w(t)dt)^2. \tag{3.37}$$

We thus see that, if we let $E[x(t)] = 0$, the autocorrelation function of a spike train with random and independent spikes simply equals the waveform autocorrelation function, up to multiplication with a scalar factor λ, the average spike rate that characterizes the underlying Poisson process. It's easy to see that, if the signal $x(t)$ would have been a spike train without convolved waveforms:

$$x(t) = \sum_{i=1}^{N} \delta(t - t_i) \tag{3.38}$$

the autocorrelation function [10] is given by, for large N,

$$R_{xx}(\tau) = \lambda\delta(\tau) + \lambda^2. \tag{3.39}$$

[10] this is the covariance function, the actual correlation amounts to (after subtracting the mean term and normalizing by the variance λ) $R_{xx}(\tau) = \delta(\tau)$.

Renewal Processes. If we want to incorporate refractory periods and transmission delays, we have to generalize our expression for the (auto)correlation function to include cases where the occurrence of a spike is dependent on the time passed since a previous spike. In this section we will derive

the expression for the autocorrelation, which can be split in the following way:

$$R_{xx}(\tau) = R_{xx}^{future}(\tau) + R_{xx}^{past}(\tau) + R_{xx}(0) \tag{3.40}$$

$$= R_{xx}^{future}(\tau) + R_{xx}^{future}(-\tau) + R_{xx}(0) \tag{3.41}$$

where $R_{xx}^{future}(\tau) = 0$ for $\tau \leq 0$. It will turn out useful to disentangle this expression even further:

$$R_{xx}(\tau) = R_{xx}^{zm,future}(\tau) + R_{xx}^{zm,future}(-\tau) + R_{xx}^{zm}(0)\delta(\tau) + \langle x \rangle^2, \tag{3.42}$$

where *zm* denotes *zero mean*, indicating the relation with the zero mean signal $x^{zm}(t) = x(t) - \langle x \rangle$. Furthermore, to derive the expression for $R_{xx}^{zm,future}(\tau)$, we will make use of functions that are naturally only defined for input values $\tau > 0$; it is, for example, not sensible to talk about the inter spike interval (ISI) distribution for intervals that are negative or zero. These functions are extended across the whole domain by making use of a Heaviside step function.

$R_{xx}^{zm,future}(\tau)$ is naturally related to the conditional probability that a spike occurs at a time $t + \tau$ when there was a spike at time t. We will denote this conditional probability distribution $C(\tau)$—with the implicit understanding that this function is extended across the whole domain by using a Heaviside step function. It is important to note that $C(\tau)$ is a conditional probability for *any* spike to occur at time $t + \tau$, not only the first one. $C(\tau)$ is thus given by the sum of the probabilities for:

- the first inter spike interval (ISI) to equal τ
- the first inter spike interval to equal $s < \tau$, and the second inter spike interval to equal $\tau - s$
- the first spike interval to equal s, the second inter spike interval to equal s', and the third inter spike interval to equal $\tau - s - s'$

We can thus write:

$$C(\tau) = p_{ISI}(\tau) + \int_0^\tau p_{ISI}(s)p_{ISI}(s-\tau)ds + \int_0^\tau \int_0^\tau p_{ISI}(s)p_{ISI}(s')p_{ISI}(\tau - s - s')dsds' + \dots \tag{3.43}$$

We can extend the inter spike interval distribution p_{ISI} across the whole domain by using a Heaviside step function, which allows us to extend the integral boundaries from $-\infty$ to $+\infty$. We can now see that Eq. 3.43 is in fact a recurrent equation:

$$C(\tau) = p_{ISI}(\tau) + \int_0^\infty p_{ISI}(s)C(s-\tau)ds, \tag{3.44}$$

and this recurrent equation can be solved by Fourier transforming to the frequency domain, where the convolution becomes a multiplication:

$$\widetilde{C}(\omega) = \frac{\widetilde{p}_{ISI}(\omega)}{1 - \widetilde{p}_{ISI}(\omega)}. \tag{3.45}$$

If we know the Fourier transform of the inter spike interval probability distribution $\widetilde{p}_{ISI}(\omega)$, we can thus obtain the conditional probability $\widetilde{C}(\omega)$. Subsequently we can relate this to the autocorrelation function: the conditional probability should be multiplied by the average rate at which spikes are produced, i.e., the function $R_{xx}^{zm,future}(\tau)$ is given by:

$$R_{xx}^{zm,future}(\tau) = \lambda C(\tau). \tag{3.46}$$

Then, given the symmetry of $R_{xx}(\tau)$ (Eq. 3.42), its Fourier transform, the power spectral density, is for $\mid \omega \mid > 0$ given by:

$$S_{xx}^{\backslash 0}(\omega) = 2\pi\Big(R_{xx}^{zm}(0) + 2\mathcal{R}e\big(\widetilde{R}_{xx}^{zm,future}(\omega)\big)\Big), \tag{3.47}$$

and by adding the DC term and using Eq 3.46, we arrive at:

$$S_{xx}(\omega) = 2\pi\Big(R_{xx}^{zm}(0) + 2\lambda\mathcal{R}e\big(\widetilde{C}(\omega)\big) + \delta(\omega)\langle x\rangle^2\Big). \tag{3.48}$$

We can perform a sanity check at this point. If we compute this power spectral density (Eq. 3.48) for a Poisson process, we should arrive, after taking the inverse Fourier transform, at Eq. 3.39. The probability density distribution for the inter spike intervals in a Poisson process is given by Eq. 3.28, and after extending it to the whole domain, we get:

$$p_{ISI}(\tau) = u(\tau)\lambda e^{-\lambda\tau} \tag{3.49}$$

with $u(\tau)$ the Heaviside step function. Its Fourier transform is given by:

$$\widetilde{p}_{ISI}(\omega) = \frac{\lambda}{\lambda + i\omega}, \tag{3.50}$$

and it is straightforward to arrive at

$$\frac{\widetilde{p}_{ISI}(\omega)}{1 - \widetilde{p}_{ISI}(\omega)} = \frac{\lambda}{i\omega}. \tag{3.51}$$

Taking the real part of this purely imaginary quantity yields zero, and we finally arrive at

$$S_{xx}(\omega) = 2\pi\Big(R_{xx}^{zm}(0) + \delta(\omega)\langle x\rangle^2\Big). \tag{3.52}$$

which, upon taking the inverse Fourier transform, indeed corresponds to the earlier obtained result Eq. 3.37 where $R_{xx}^{zm}(0) = \lambda$.

The difference in the autocorrelation function between a Poisson and a renewal process is thus captured by the conditional probability—entirely as one would expect. Because we can relate this conditional probability to the inter spike interval distribution, we will now discuss a family of ISI distributions that give rise to stochastic spiking with a (relative) refractory period.

Relative refractory period. To model such a relative refractory period, intuitively, the probability for short inter spike intervals should be smaller than in the case of a Poisson process. The family of gamma probability distributions of different orders κ captures this behavior:

$$p_{ISI}^{(\kappa)}(\tau) = (\kappa\lambda)^{\kappa}\tau^{\kappa-1}\frac{e^{-\kappa\lambda\tau}}{(\kappa-1)!}. \tag{3.53}$$

$\kappa = 1$ yields the exponential distribution of a Poisson process (Eq. 3.28), while order 2 also introduces a relative refractory period through the factor τ; higher orders not only yield longer refractory periods, but also correspond to increasingly narrow ISI distributions, resulting in highly regular spiking processes. This is illustrated in Fig. 3.6.

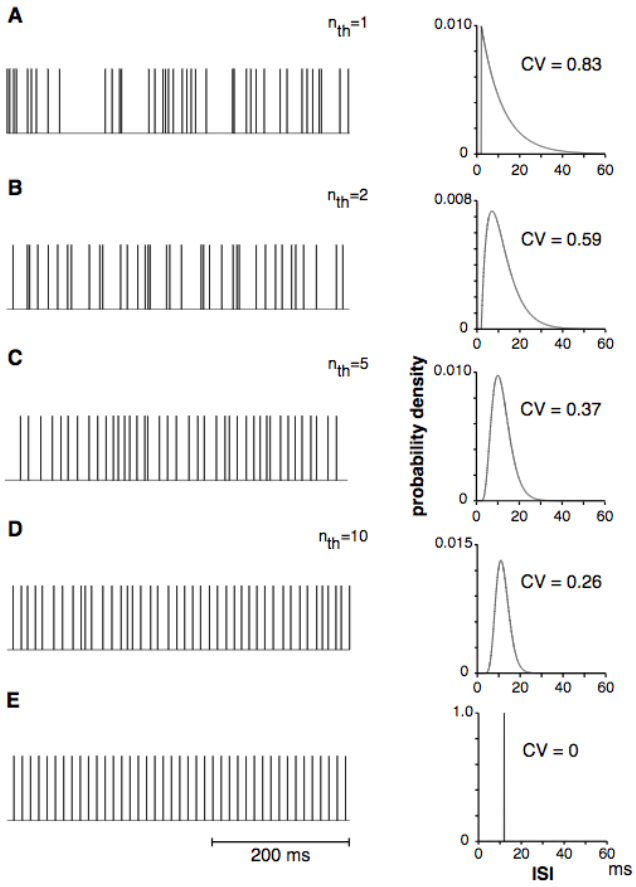


Figure 3.6: Sample spike trains from renewal processes with different gamma process orders. Source: Gabiani and Koch.

We can now derive expressions for the PSD and autocorrelation function of a neuron spiking stochastically with an average rate λ, and a relative refractory period captured by:

$$p_{ISI}^{(2)}(\tau) = 4\lambda^2\tau e^{-2\lambda\tau}u(\tau), \tag{3.54}$$

i.e., a gamma probability distribution of order $\kappa = 2$. The Fourier transform of this distribution yields:

$$\widetilde{p}_{ISI}(\omega) = \frac{\lambda^2}{(2\lambda + i\omega)^2}, \tag{3.55}$$

and, by Eq. 3.45, we find:

$$\widetilde{C}(\omega) = \frac{1}{\frac{i\omega}{\lambda} - \frac{\omega^2}{4\lambda^2}}. \tag{3.56}$$

After some straightforward calculations, we can obtain the real part of this expression:

$$\mathcal{R}e\big(\widetilde{C}(\omega)\big) = \frac{-4\lambda^2}{\omega^2 + 16\lambda^2} \tag{3.57}$$

$$= -\frac{\lambda}{2}\frac{8\lambda}{\omega^2 + (4\lambda)^2} \tag{3.58}$$

where we rearranged Eq. 3.57 into Eq. 3.58 to reveal that we can use the Fourier transform pair

$$e^{-\alpha|t|} \leftrightarrow \frac{2\alpha}{\omega^2 + \alpha^2} \tag{3.59}$$

in inverse Fourier transforming the PSD:

$$S_{xx}(\omega) = 2\pi\big(R_{xx}^{zm}(0) - \lambda^2\frac{8\lambda}{\omega^2 + (4\lambda)^2} + \delta(\omega)\langle x\rangle^2\big) \tag{3.60}$$

to finally obtain:

$$R_{xx}(\tau) = \lambda\delta(\tau) - \lambda^2 e^{-4\lambda|\tau|} + \lambda^2 \tag{3.61}$$

where we used the previously obtained results $R_{xx}^{zm}(0) = \lambda$ and $\langle x\rangle^2 = \lambda^2$. The average spiking thus changes with respect to the pure Poisson process. Eq. 3.61 corresponds to what one would intuitively expect from the autocorrelation of a process with a relative refractory period: the correlation relatively decreases for small $\mid \tau \mid$ with respect to the pure Poisson process; but, as $\tau \to \infty$, the autocorrelation converges to the autocorrelation of a pure Poisson process. This is visualized in Fig. 3.7.

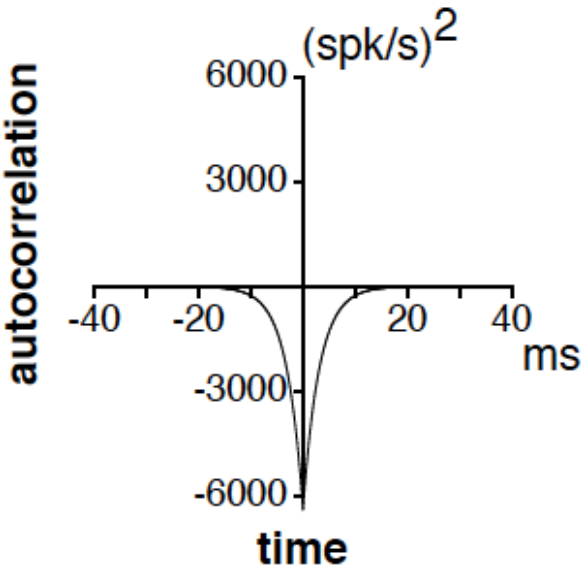


Figure 3.7: Zero mean autocorrelation function corresponding to an ISI distribution following a gamma distribution of order 2. In this example $\lambda = 80Hz$. Source: Gabiani and Koch.

3.3.2 *Spike Train Analysis of Connected Neurons*

Cross correlation functions and transmission delays. We will now consider the correlation function between the spike trains of two different neurons, which we will denote x_1 and x_2. This cross correlation function can be obtained in a way very similar to that of the autocorrelation function, although there are two major differences in its actual shape:

1. there is no trivial zero lag term, and
2. the function might not be symmetric.

By analogy to Eq. 3.42, we can write:

$$R_{x_1x_2}(\tau) = R^{zm,future}_{x_1x_2}(\tau) + R^{zm,past}_{x_1x_2}(\tau) + R^{zm}_{x_1x_2}(0)\delta(\tau) + \langle x_1\rangle\langle x_2\rangle \quad (3.62)$$

where the zero lag term $R^{zm}_{x_1x_2}(0)$ is proportional to the probability that the two neurons fire at the exact same instant. In fact, we can interpret $R^{zm}_{x_1x_2}(0)$ similar to the way we interpreted $R^{zm,future}$ (see Eq. 3.46): as the conditional probability that the second neuron will spike at the same time, given that the first neuron spikes; and this multiplied by the probability that the first neuron spikes. For two pure Poisson with spike rate λ_1 and λ_2 processes we thus get:

$$R^{zm}_{x_1x_2}(0) = 0.\lambda_1 = 0, \quad (3.63)$$

compared to the trivial zero lag term in the autocorrelation of x_1:

$$R^{zm}_{x_1x_1}(0) = 1.\lambda_1 = \lambda_1. \quad (3.64)$$

Suppose now we don't have two independently spiking Poisson neurons, but two neurons that are coupled through a synaptic connection. Neuron 1 fires with an average spike rate of λ_1, and neuron 2 fires with an average spike rate of $\lambda_2 = \frac{1}{2}\lambda_1$. In this (not very realistic) example we assume the spikes of neuron 2 are all the direct consequence of spikes in neuron 1, and that the transmission takes a certain time t_{delay}. In other words, half of the spikes of neuron 1 are reliably transmitted to neuron 2 and result in an action potential there. The conditional probability that a spike is detected in x_2 a time t_{delay} *after* there was a spike detected in x_1, is thus 0.5. On the other hand, the probability that a spike is detected in x_1 a time t_{delay} *before* a spike is detected in x_2 is equal to 1. Therefore:

$$\begin{aligned} R^{zm,future}_{x_1x_2}(t_{delay}) &= 0.5\lambda_1 && (3.65)\\ &= 1\lambda_2 && (3.66)\\ &= R^{zm,past}_{x_2x_1}(-t_{delay}). && (3.67) \end{aligned}$$

It is in general true that $R^{future}_{x_1x_2}(\tau) = R^{past}_{x_2x_1}(-\tau)$ and $R^{future}_{x_1x_2}(\tau) = R^{past}_{x_2x_1}(-\tau)$.

In the above case, we would thus find a sharp peak at $R^{future}_{x_1x_2}(t_{delay})$ with amplitude $0.5\lambda_1$ plus the mean term [11], provided of course $t_{delay} > t_{refrac}$.

[11] or, after standardization, $R^{future}_{x_1x_2}(t_{delay}) = \frac{0.5\lambda_1}{\sqrt{\lambda_1}\sqrt{\lambda_2}}$

In a more realistic network of neurons, the cross correlation functions between neurons might be influenced by the following factors:

- the synaptic transmission reliability
- the postsynaptic reliability of transmission to the soma and axon hillock
- the integration of signals (stemming from different neurons) in the dendritic tree and soma
- the relation between the resting potential difference and the action potential generation threshold
- the average spike rate of the neurons, which might be related to intrinsic noise and all the factors above.

Within one neuron, i.e., between the different structures of a neuron, the cross correlation functions might be much closer to those of a pure transferral of signals like in the simplified example given above. This will turn out to be an important fact that we can use interpreting our data.

Doubly stochastic processes. Finally, although we will not elaborate on or use this further, we note that the cross correlation function between two neurons that are solely correlated through their spike *rates* can be calculated [12] from using:

[12] *Measuring and interpreting neuronal correlations*. Cohen and Kohn,2011.

$$P(x_1, x_2) = \int d\lambda_1 d\lambda_2 P(x_1 \mid \lambda_1) P(x_2 \mid \lambda_2) P(\lambda_1, \lambda_2) \tag{3.68}$$

where $P(\lambda_1, \lambda_2)$ is the joint distribution of the respective spike rates. And, to arrive at the covariance:

$$\langle x_1 x_2 \rangle = \sum_{x_1} \sum_{x_2} P(x_1, x_2). \tag{3.69}$$

4

A Directed Graph of Information Flow

Eventually, we wish to obtain insight in the effective connectivity of the network: which connections are used, their transmission delays, and finally the direction of information flow between the active neuronal structures in the network. So far, we explained how to identify these neuronal structures from the data through a spike sorting process; and we elaborated on the likely form of (and the needed calculation for) the auto- and cross-correlations of the obtained spike trains. In this chapter we discuss how we can obtain the effective connectivity from this set of auto- and cross-correlation results.

A straightforward (but somewhat naive) way to obtain the directed signal flow in the network could be based solely on the cross correlation functions between the spike trains of two neuronal structures. After all, these functions reflect the transmission of spikes (together with some information about the delays at which those happen). The problem is that those functions often exhibit spurious correlations, i.e., they might indicate a relationship that in reality does not exist [1] [2]. For two (coupled or independent) point Poisson processes, the spurious relations are less common than in case of, e.g., low frequency continuous processes, but they nevertheless exist. Consider for example the situation where two independent Poisson processes are recorded for a finite time and with a finite sampling rate; one of them has a very high average spike rate and the other one spikes much slower. Depending on the parameters, the computed cross correlation might look as if spikes of the slower neuron are caused by the faster neuron, just because in a lot of cases there happens to be a spike of the first neuron preceding it. Thus, the information about the spike rates should be included; and this information is to be found in the autocorrelation functions (or, strictly speaking, in the unstandardized autocorrelation functions). Furthermore, in networks where neurons receive signals from multiple neurons (as is almost always the case), the information contained in the correlation functions of *all* connected neurons should be included.

[1] *Why do we sometimes get nonsense-correlations between Time-Series?–a study in sampling and the nature of time-series.* Yule,1926.

[2] *Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models.* Dean and Dunsmuir,2016.

In section 4.1, we will first discuss a general framework to model the

causal relationships between a set of coupled time series. We will then show how, based on a general assumption of gaussianity, we can obtain the parameters of this framework from the computed auto- and cross-correlation functions. Finally, we will explain how these parameters can be converted into a directed graph; and we will also discuss how this graph might deviate from the actual flow of information in the network.

4.1 Autoregressive Kernels

4.1.1 Autoregressive Processes

A multivariate, higher order autoregressive process is a process that can be written in the form [3]:

[3] *An introduction to multivariate statistical analysis*. Anderson,1958.

$$\boldsymbol{x}(t) = \sum_{k=1}^{p} \boldsymbol{A}^{(k)} \boldsymbol{x}(t-k) + \boldsymbol{E}(t), \tag{4.1}$$

where $\boldsymbol{x}(t)$ is a vector of size n that represents the values of the n time-series at time t; p is the order of the model; $\boldsymbol{A}^(k)$ is the matrix of size nxn with entries that relate $\boldsymbol{x}(t)$ to the values of $\boldsymbol{x}$ at distance k in the past, and $E(t)$ is a diagonal matrix adding Gaussian noise at step t. Note that $A^{(k)}$ is a constant matrix that does not depend explicitly on time.

The elements $A_{ij}^{(k)}$ tell us how the values $x_j(t-k)$ influence the value $x_i(t)$. In words, Eq. 4.1 tells us that the current value of a timeseries in the network can be obtained as a linear combination of the past values of all the timeseries, including its own past. The coefficients $A_{ij}^{(k)}$ capture a causal relationship in the sense that they reveal which events in the past are relevant for predicting the current state, given all relevant information (i.e., given the past values of *all* the timeseries up to p timesteps in the past).

If we consider $A_{ij}^{(k)}$ for each time k, we obtain a so-called autoregressive kernel which we will denote $A_{ij}^{(t)}$. Convolving this function with the past of $x_j(t)$ gives the contribution of this j^{th} timeseries to the present value of the i^{th} timeseries, $x_i(t)$—hence the name kernel. The concept slightly differs from the more conventional Volterra or Wiener kernels, however. First of all it is an extension to the multivariate case; and moreover, the considered input consists not only of the signals received from the other elements in the network, but also the timeseries' own past. How these kernels can be estimated, then, is the topic of the next section.

4.1.2 *Gaussian Processes*

The assumption of Gaussianity often yields an effective and practical approach to modeling, and this is not different in the problem we are trying to solve. In this section, we will discuss this assumption of Gaussianity using a one-dimensional time series. More precisely, we will model this timeseries as a Gaussian Process (GP):

A Gaussian Process is a collection of random variables, any finite number of which has a joint Gaussian distribution.

Loosely speaking, we will consider an (in our case ordered) collection of random variables $\{X_1, \ldots, X_n\}$, one variable for each time t; moreover, we will take the joint distributions between any number of variables to be Gaussian. The Gaussian Process in this way defines a distribution over sample functions, where a function is considered the (possibly infinite) vector of particular instances of these ordered random variables. (The latter definition is not entirely rigorous, but sufficient for the current discussion.) This distribution is completely characterized by a mean function $m(t)$ and a covariance function $k(t, t')$:

$$f(t) \sim \mathcal{GP}\big(m(t), k(t, t')\big). \tag{4.2}$$

Before we continue to see how this works, we will review some properties of the classical Gaussian probability distribution. First some notational conventions: $\boldsymbol{x}$ and $\boldsymbol{y}$ symbolize vectors of any number of random variables that are jointly Gaussian distributed, $\boldsymbol{\mu_x}$ and $\boldsymbol{\mu_y}$ the vectors of their respective means. Thus, e.g.,

$$p(\boldsymbol{x} \mid \boldsymbol{\mu_x}, \Sigma_x) \sim (2\pi)^{-D/2} \mid \Sigma_x \mid^{-1/2} \exp\big(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu_x})^T \Sigma_x^{-1} (\boldsymbol{x} - \boldsymbol{\mu_x})\big) \tag{4.3}$$

where D is the length of the vector $\boldsymbol{x}$, and Σ_x is the DxD, symmetric, positive definite covariance matrix. Using shorthand notation:

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu_x}, \Sigma_x). \tag{4.4}$$

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are also jointly Gaussian with covariance matrix K,

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \sim \mathcal{N}\Big(\begin{bmatrix} \boldsymbol{\mu_x} \\ \boldsymbol{\mu_y} \end{bmatrix}, K\Big) \equiv \mathcal{N}\Big(\begin{bmatrix} \boldsymbol{\mu_x} \\ \boldsymbol{\mu_y} \end{bmatrix}, \begin{bmatrix} C_x & B \\ B^T & C_y \end{bmatrix}\Big), \tag{4.5}$$

then the marginal distribution of $\boldsymbol{x}$ is

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu_x}, C_x), \tag{4.6}$$

and the conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{x}$ is:

$$\boldsymbol{y} \mid \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu_y} + BC_x^{-1}(\boldsymbol{x} - \boldsymbol{\mu_x}), C_y - BC_x^{-1}B^T). \tag{4.7}$$

There's an important point to emphasize here: these equations are entirely linear. Let

$$A \equiv BC_x^{-1}, \tag{4.8}$$

and conditional covariance matrix Σ

$$\Sigma \equiv C_y - BC_x^{-1}B^T. \tag{4.9}$$

Then we can see that Σ is independent of $\boldsymbol{x}$, which allows us to interpret $\boldsymbol{y}$ as an affine function of $\boldsymbol{x}$ plus a random noise vector $\boldsymbol{n}$ that is independent of $\boldsymbol{x}$. In other words,

$$\boldsymbol{y} = \boldsymbol{\mu_y} + A(\boldsymbol{x} - \boldsymbol{\mu_x}) + \boldsymbol{n}, \tag{4.10}$$

with the noise vector $\boldsymbol{n}$ satisfying:

$$\langle \boldsymbol{n} \rangle = 0, \quad \langle \boldsymbol{nx} \rangle = 0, \quad \langle \boldsymbol{nn} \rangle = \Sigma. \tag{4.11}$$

If $\boldsymbol{y}$ is the system output and $\boldsymbol{x}$ is the system input, then (Eq. 4.10) describes the linearity of the system. However, within the framework of Gaussian Processes, $\boldsymbol{x}$ is not the input of the system but rather *the output in the past*. Predictions of the future output are thus obtained as a linear combination of outputs in the past (Eq. 4.10)—but the described system itself may not be linear in the sense of the output being a linear combination of the input. The latter would only be the case when the covariance function k is a linear function.

Then, to gain insight in Gaussian Processes, consider the case where all the random variables X_t are taken to be independent; we can then "draw" a sample function $f(t)$ by obtaining at each time t an instance of the random variable X_t that follows a marginal Gaussian distribution characterized by $m(t)$ and $k(t,t)$. If we shift to the vector notation as above, we can let K be the covariance matrix of the joint Gaussian distribution over all variables X_t. Since in the independent case, all off-diagonal elements of K are zero, we can thus just sample from the individual marginal distributions (Eq. 4.6) $X_t \sim \mathcal{N}(\mu_{X_t}, K_{tt}) = \mathcal{N}(m(t), k(t,t))$.

We can obtain more interesting functions by introducing correlations between the variables, however. A common example is the *squared exponential* covariance function with characteristic length-scale l:

$$k(t,t') = \exp(-\frac{1}{2l^2} \mid t - t' \mid^2), \tag{4.12}$$

which tells us both that the process is stationary, and that the covariance is almost unity between nearby variables, but decreases exponentially as the distance in time increases. In other words, the sample function are smooth, in this case over a timescale of length l. Now, in constructing such a sample function $f(t)$ we cannot restrict ourselves to sampling from the marginal distributions of individual variables X_t. We need to take into account the structure captured by the joint distributions. The computational efficiency of Gaussian Processes, then, lies in the fact that this can be achieved from using the conditional distributions (Eq. 4.7).

Suppose we are given a starting value, $f(t_0)$. We wish to obtain a value $f(t_*)$ at another time t_*. The covariance matrix K' of the joint distribution between X_{t_0} and X_{t_*} is given by:

$$K' = \begin{bmatrix} k(t_0, t_0) & k(t_0, t_*) \\ k(t_*, t_0) & k(t_*, t_*) \end{bmatrix}, \tag{4.13}$$

and thus, from (Eq. 4.7),

$$\big(X_{t_*} \mid X_{t_0} = f(t_0)\big) \sim \mathcal{N}\Big(m(t_*) + \frac{k(t_0, t_*)}{k(t_*, t_*)}(f(t_0) - m(t_0)), \quad k(t_0, t_0) - \frac{k^2(t_0, t_*)}{k(t_*, t_*)}\Big). \tag{4.14}$$

4.1.3 *Gaussian Processes and Bayesian Modeling*

Now the groundworks are laid to switch viewpoints: from generating sample functions from a known Gaussian process, to estimating the properties of a Gaussian process given a sample function. We will assume the process is stationary, in which case it is convenient to define zero-mean variables: $\boldsymbol{x} - \boldsymbol{\mu_x} \mapsto \boldsymbol{x}$. We can then rewrite Eq. 4.10:

$$\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{n}. \tag{4.15}$$

The weight matrix A can be estimated from basic linear regression: if the noise is white or can be whitened, the solution is simply given by Eq. 4.8. We will discuss this further below. Here we will first briefly explain how we can incorporate prior knowledge in this estimation by using the framework of bayesian modeling; this will turn out to be important when we wish to regularize our final results. We start by recalling Bayes' rule:

$$p(A \mid \boldsymbol{x}, \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{x}, A)p(A)}{p(\boldsymbol{y} \mid \boldsymbol{x})}, \tag{4.16}$$

or, in words,

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}}. \tag{4.17}$$

For ease of notation, we will rewrite Eq. 4.15 as

$$\boldsymbol{y} = \boldsymbol{x}^T A' + \boldsymbol{n} \tag{4.18}$$

Suppose now we have a prior believe that the weights A' are distributed according to a Gaussian function:

$$A' \sim \mathcal{N}(0, \Sigma_p). \tag{4.19}$$

Then it follows

$$p(A' \mid \boldsymbol{x}, \boldsymbol{y}) \propto p(\boldsymbol{y} \mid \boldsymbol{x}, A')p(A') \tag{4.20}$$

$$\propto \exp(-\frac{1}{2}(\boldsymbol{y} - \boldsymbol{x}^T A')^T \Sigma^{-1}(\boldsymbol{y} - \boldsymbol{x}^T A')) \exp(-\frac{1}{2} A'^T \Sigma_p^{-1} A') \tag{4.21}$$

$$\propto \Big(A' - \frac{\boldsymbol{x}\Sigma^{-1}\boldsymbol{y}}{\boldsymbol{x}\Sigma^{-1}\boldsymbol{x}^T + \Sigma_p^{-1}}\Big)^T \Big(\boldsymbol{x}\Sigma^{-1}\boldsymbol{x}^T + \Sigma_p^{-1}\Big)\Big(A' - \frac{\boldsymbol{x}\Sigma^{-1}\boldsymbol{y}}{\boldsymbol{x}\Sigma^{-1}\boldsymbol{x}^T + \Sigma_p^{-1}}\Big) \tag{4.22}$$

where we used only the terms dependent on the weight matrix A'. From this, we can recognize that the posterior distribution is Gaussian with mean μ_A

$$\mu_A = \frac{\boldsymbol{x}\Sigma^{-1}\boldsymbol{y}}{\boldsymbol{x}\Sigma^{-1}\boldsymbol{x}^T + \Sigma_p^{-1}} \tag{4.23}$$

and covariance matrix Σ_A

$$\Sigma_A = \left(\boldsymbol{x}\Sigma^{-1}\boldsymbol{x}^T + \Sigma_p^{-1}\right)^{-1}. \tag{4.24}$$

Having no prior beliefs corresponds to infinite covariances Σ_p, in which case Eq. 4.22 reduces to the solution of linear regression with correlated noise with covariance Σ. If the noise is uncorrelated, the solution reduces further to the classical linear regression solution Eq. 4.8.

4.1.4 Estimating the Autoregressive Kernels

Saying the timeseries we are trying to model is a sample drawn from a Gaussian process thus entails the following assumptions:

- the function is stationary,
- the function can be non-linear, but
- the function can be completely defined by a mean and covariance function plus the addition of Gaussian noise.

These assumptions yield a straightforward solution to estimating the coefficients of the corresponding autoregressive process based on the conditional distribution: the problem can be solved by linear regression, and the solution is given by Eq. 4.8 (in case the function has zero mean). Note that the only properties of the data used in this estimation are correlation functions (again, in case of zero mean). These might thus as well be the correlation functions associated to a Poisson process, observed with some additional Gaussian noise (the double usage of the term 'process' in both 'Gaussian process' and 'Poisson process' might be somewhat confusing in this case). However, processes such as doubly stochastic processes will not be correctly modeled by this framework.

We can easily extend this linear regression problem (Eq. 4.8) to the multi-variate, higher order case [4], and we finally arrive at the Yule-Walker equations [5] :

$$\begin{bmatrix} C_0 & C_1 & C_2 & C_3 \dots & C_{p-1} \\ C_1 & C_0 & C_1 & C_2 \dots & C_{p-2} \\ C_2 & C_1 & C_0 & C_1 \dots & C_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{p-1} & C_{p-2} & C_{p-3} & \dots & C_0 \end{bmatrix} \begin{bmatrix} A^{(1)} \\ A^{(2)} \\ A^{(3)} \\ \vdots \\ A^{(p-1)} \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_p \end{bmatrix}$$

where C_i is the covariance matrix between all variables at lag i.

[4] *Estimation of parameters and eigenmodes of multivariate autoregressive models.* Neumaier and Schneider,2001.

[5] *On periodicity in series of related terms.* Walker,1931.

4.2 *A Directed Graph of Information Flow*

Once the full matrix A—and thus each autoregressive kernel—has been estimated, obtaining a directed graph boils down to computing a value from the kernel that captures the connectivity. In general words one could say there is a directed connection between two neuronal structures in the network if the kernel for that direction is not uniformly zero, that is, non-zero in a statistically significant way. A significance test is not easy to obtain, however: the estimation of the kernels is a highly nonlinear function of the original data, and the distributions of their estimators are not well established [6]. We will discuss our (pragmatic) approach to this problem in chapter 5.

[6] *Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance.* Kamiński et al.,2001.

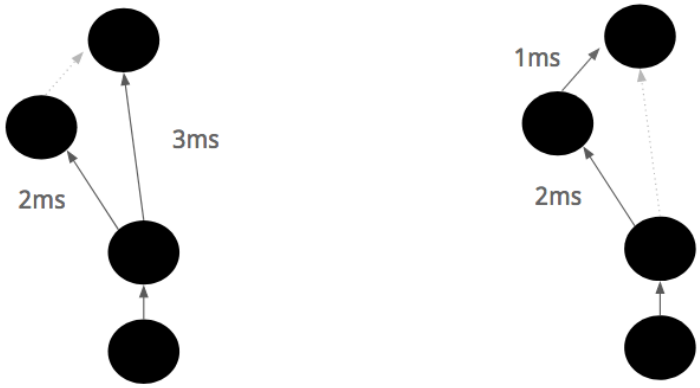


Figure 4.1: Example situations that are 'equivalent' in terms of correlation functions and resulting kernels. Only structural information can discern the true from the spurious connectivity.

Finally, we should mention that this method of determining the directed flow of signals is not always able to capture the real directed flow of information. Consider, e.g., the situation as depicted in Fig. 4.1. On the right hand side, the information flows from the bottom and splits after the second node, such that it arrives with a certain time delay in each of the upper nodes; the signals arrive at the same times, however, in the situation on the right. The correlation functions will thus be exactly the same, and since the kernels are derived from nothing else but the correlation functions, there is no way to discern the true path from the spuriously derived path. This could be resolved by using structural information however; to do this, knowledge about the actual underlying structure could be used as a Bayesian prior. This is again outside the scope of this project however (but might be added to the analysis in future work).

5

Experimental Setup and Methods

5.1 Multi Electrode Array Recordings at High Resolution

The complete sensor array covers a region of $3.85x2.10sq - mm$ with 26400 electrodes, each covering a region of $9.35.45sq - \mu m$. Due to constraints in the electronics implementation, we can only read out from 1024 electrodes at once. However, we can determine which electrodes to record from, and the configuration is changed in a matter of seconds. We sample the data at a rate of 20K Hz; our recordings of several minutes thus have a considerable size in terms of data (a couple of Gb per file).

5.2 Overview of the Data Processing Pipeline

- **Identifying Neuronal Structures**. The output at this stage is a list of detected neuronal structures in terms of their spike times, spatio-temporal footprint and location. The results are only really reliable for structures producing high amplitude spikes; the lower the amplitude, the higher the probability that the detected spike was actually noise. The low amplitude detections will provide us with additional information about the network structure, however. The discussion of the underlying algorithms makes up the first section of this chapter.
- **Spike Train Analysis of Single Neurons.** Based on the spike interval distributions of the high-amplitude structures (which we assume to represent a single neuron), we will try to classify the neurons as producing slow-rate spontaneous activity, as producing activity at both smaller and larger timescales, or as exhibiting an inter spike interval close to the corresponding Poisson distribution.
- **Computing the Correlations.** Based on the output from the first step, we will elaborate on the methods and data transforms used to compute the auto- and cross-correlation (in fact, auto- and cross-covariance) matrices of the spike trains of all neuronal structures.

- **Constructing the Effective Connectivity Graph.** In the last section we provide some additional information on the methods used to determine the kernels and on the methods used to construct the effective connectivity graph.

5.3 *Identifying Neuronal Structures*

In order to model the effective connectivity between different neurons, we have to disentangle the data: each sensor might receive signals from different neurons, and each neuron might be inducing signals in different sensors. Consider a single neuron. The axon hillock and soma usually produce spikes with a high amplitude, with resulting signals being recorded in the sensors directly covering these structures, as well as in other sensors nearby. But an action potential could travel through the other parts of the axon and all its branches as well, and the activity of a single neuron might thus potentially leave traces across a large part of the sensor array. We can also switch perspectives, and look at the problem from the viewpoint of a single sensor: the same sensor might receive signals from different nearby neurons. This is of course especially the case when the sensor covers an area where neuronal structures are dense.

Although some general conclusions can be drawn from the sensor time series alone, detailed insights can only be obtained once the data is disentangled (correctly) into the respective neuron time series. This process, which goes by the name of *spike sorting*, has been the focus of many research efforts in computational neuroscience [1] [2] [3]. Feasible due to the low dimensionality of the recorded data, this process was for a long time mainly performed manually. But recently the advent of recording devices with up to hundreds of electrodes has urged the development of semi-automated to fully automated algorithms to tackle this problem. Below, we will discuss the general approach of those automated spike sorting algorithms.

[1] *A review of methods for spike sorting: the detection and classification of neural action potentials*. Lewicki,1998.

[2] *Past, present and future of spike sorting techniques*. Rey et al.,2015.

[3] *Spike sorting for large, dense electrode arrays*. Rossant et al.,2016.

Spike sorting is usually performed in a setting with relatively low recording resolution, i.e., each sensor or electrode records signals from multiple neurons, and it is very likely that only the high amplitude signals are recorded (i.e., those originating from the action initial segment). In case the spacing between the electrodes is large, such that the probability of one neuron being detected by more than one electrode is very small, the spike sorting problem simplifies further: the detected spikes on each electrode can be assigned to the neuron that produced them based on the spike shape. This is illustrated in Fig. 5.1.

Once all the spikes in the signal have been detected, the (usually correct [4]) assumption is that the different neurons produce spikes with a shape and amplitude typical to that neuron (the latter being only partly due to their difference in distance to the electrode). They can thus either

[4] *Axon physiology*. Debanne et al.,2011.

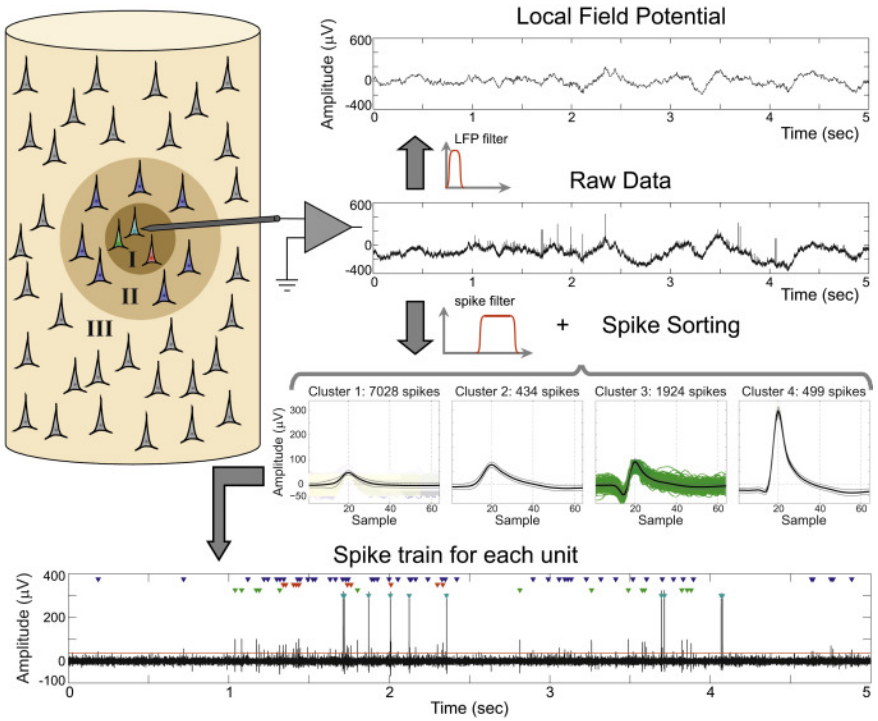


Figure 5.1: Illustration of the spike sorting process used for an in-vivo experiment, where the ratio of electrodes to neurons is usually relatively small. A schematic representation of the spikes in the neighborhood of the considered electrode is given on the left. In this setup, only the spikes from neurons in zone I are close enough to the electrode to be reliably discerned, the other neural activity contributes to the local field potential (an effect not present in our in vitro setup). Once the spikes have detected in the signal, a classification based on the shape reveals which spike times are associated to each neuron (shown at the bottom trace of the figure.) Source: Rey et al, 2915.

be manually assigned to a neuron; or the shape of the spike or action potential can be used as input to a classification algorithm.

In a next step, we can consider the fact that the high amplitude signals originating from the AIS could possibly be recorded in multiple electrodes. This is due to the electrical signal spreading from where it originated; it is thus not always the case that the signal will be recorded at exactly the same instant in each of the neighboring electrodes. The goal can be, somewhat vaguely, formulated in this way: when a neuron spikes, we want to be able to say what the resulting measurement typically 'looks like' in space and time, i.e., we want to determine a spatio-temportal footprint for each neuron. These spatio-temporal footprints can then be matched to the original data in order to determine the spike times for each neuron.

Due to the very high resolution of our recordings, we are, unlike in other spike sorting settings, able to pick up signals from different structures of the same neuron. We deliberately lowered the spike detection threshold to also detect these signals. This, unfortunately, also yields a multitude of false detections (i.e., detection of noise). We will show in the results in which ways this information can still be valuable, however.

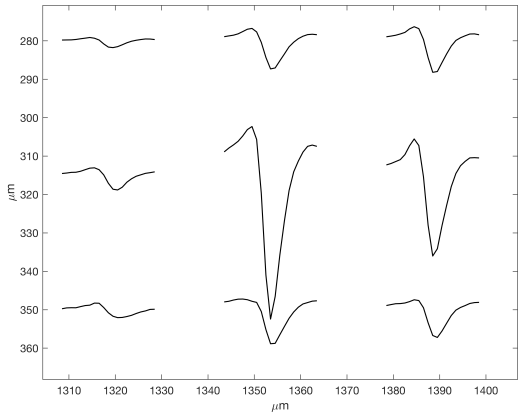


Figure 5.2: Example of a spatio-temporal neuron footprint: the same action potential is detected in multiple neighboring electrodes.

5.4 *Spike Train Analysis of Single Neurons*

Once the spike sorting process is completed, we have a list of the activity of the different neuronal structures in hand, i.e., a list of locations, spike times, and average amplitudes. We will discard the data from the low amplitude structures for the time being—this will be used later on. For now, we can concentrate on the structures emitting spikes with a high amplitude, which presumably correspond to axon initial segments. We will consider this activity to represent a 'single' neuron in this section.

As discussed before, the activity of a single neuron can be classified based on the properties of the spike train; we will consider the categories *Poisson* (i.e., close to Poisson or second order gamma), exhibiting *mixed timescales*, and *slow rate spontaneous* activity. We will try to classify the activity based on the inter spike interval (ISI) distributions, and propose the following easy to implement metrics:

- if the ISI distribution is close to a first- or second order gamma process, the 10% quantile of the distribution will have a relatively small value.
- if the process is not Poisson like, we discriminate by looking at the number of timescales: in a spontaneous activity process, there should be a 'peak' at lower frequencies; in case there are two timescales involved a model fitting two peaks to the distribution should thus yield a better fit.

The discrimination of time scales is more easily discernible using a log transform of the ISI distribution. After this transformation, we fit a two component Gaussian mixture to the log ISI histogram, and use the difference in the means of the obtained distributions as a straightforward indicator of different timescales. This procedure will be illustrated with data exampled in the results section.

5.5 Computing the Correlations

To compute the correlations, we don't bin the data but instead convolve the detected spike times with a waveform which has a substantial width over 0.5ms, corresponding to the natural temporal resolution implied by the spike sorting algorithm. The used function is furthemore Lorentzian, because its longer tails will increase the smoothness.

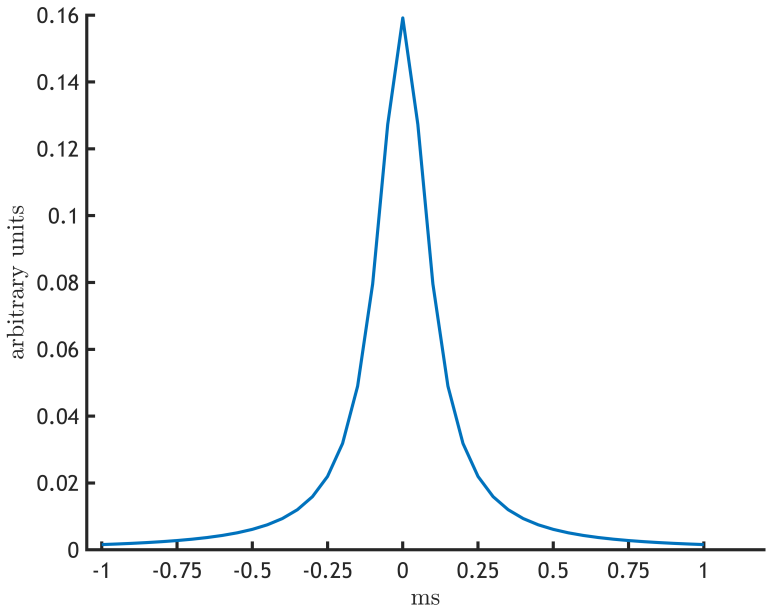


Figure 5.3: Waveform function used to construct the spike trains.

Due to the high computational resources needed to compute the correlation functions up to lags of several hundreds of timesteps, we implemented the computations using a map-reduce procedure.

5.6 Constructing the Effective Connectivity Graph.

The matrix involved in the Yule-Walker equations has a block Toeplitz structure; we can make good use of this structure to efficiently solve the system of equations through applying the Levinson algorithm [5].

[5] *Efficient inversion of Toeplitz-block Toeplitz matrix*. Wax and Kailath,1983.

Then, considering a test of significance: a good approach would be to generate new datasets that retain the underlying stochastic properties but discard the causal connections. An example procedure would be to sample new spike trains using the inter spike interval distributions obtained from the original dataset. Once a surrogate dataset is created, we can again compute the correlations and the kernels; a large collection of surrogate datasets would then yield an estimator of the variance of the kernels. The drawback of this procedure is that it is extremely costly in

terms of computational resources: despite the measures taken to speed up the computations and to handle the large datasets, a reasonable significance test would take days to compute.

For the time being, we thus resorted to a much simpler significance test. We computed the median absolute deviation (the median because this is more robust to outliers) of the absolute values of each kernels, and considered it an effective connection if a values was found above a threshold value $v_{threshold}$ defined by:

$$v_{threshold,ij} = \theta MAD(\mid A_{ij}^{(t)} \mid) \tag{5.1}$$

where θ was taken to be $\theta = 5$.

6

Results and Conclusion

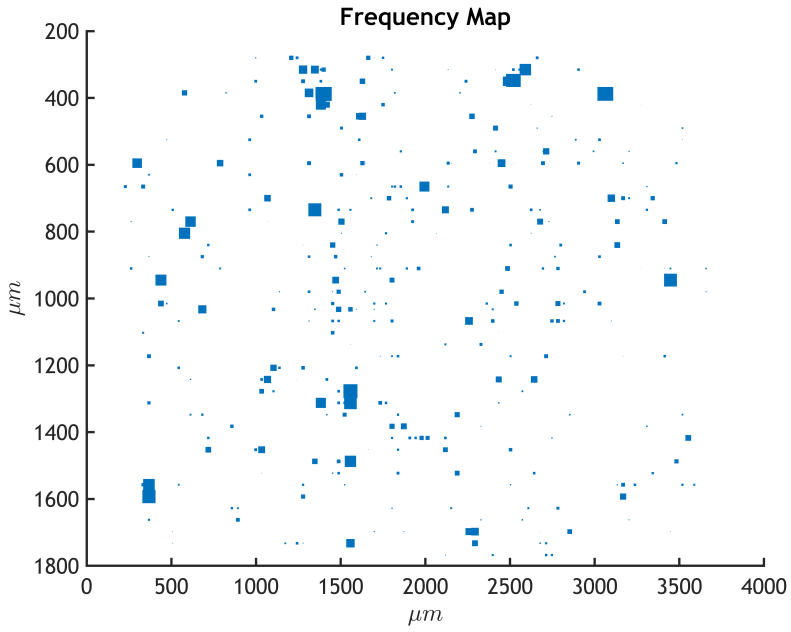
In this analysis, we will restrict ourselves to a single set of recordings from a low density network at day 34 in vitro. The network consisted solely of excitatory neurons and was not stimulated during its development. The whole sensor array is covered by a set of 6 configuration recordings, each of which make up a simultaneous recording of half the electrodes of their respective sensor regions. In other words, at each recording time one sixth of the whole sensor array is recorded simultaneously by a checkerboard pattern of 1024 electrodes, and this for two minutes at a sampling rate of 20 000Hz.

6.1 Spike Sorting Results

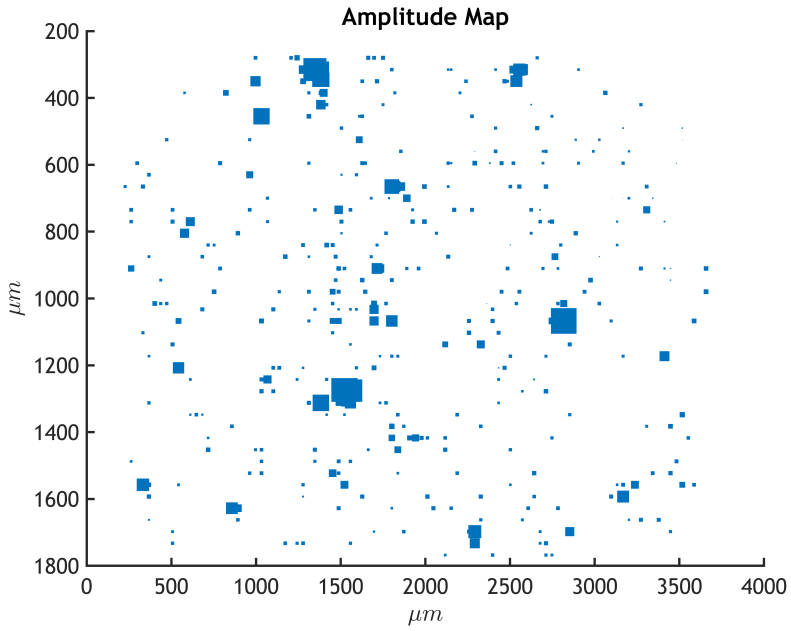
Using the information obtained from the spike sorting procedure, we can create straightforward visualizations that provide us some basic insights in the network and its activity. At this point we can combine the results of different recordings covering the whole sensor array—the spike sorting process is not based on the assumption that the different involved structures were recorded simultaneously (although some information about neuronal structures exactly lying at the boundary of recording sites might be lost). A compiled map of all found average spike rates, together with a similar map of all amplitudes, is shown in Fig. 6.1. We will shortly explain why the average rate map does not provide us with a reliable view of the network; the amplitude map, however, already reveals clusters of high amplitude structures (presumably produced by the axon initial segments (AIS) or other regions close to the soma) and the low amplitude structures (presumably axononal segments) that connect them. In the remainder of this chapter we will denote the former as *somatic* signals, and the latter as *axonal* signals.

Fig 6.2 shows a plot of the average spike amplitude in function of the average spike rate for all detected neuronal structures. If anything, this plot reveals that the axonal structures are confounded with noise: aver-

age spike rates higher than 20Hz are biologically not very plausible, and these rates rather correspond to a multitude of false detections. We can try to separate the axonal from the somatic signals based on amplitude, however, although from activity amplitudes alone we are only able to make a somewhat arbitrary distinction. Structural information, which could be obtained from microscopy pictures, could help to estimate the correct boundary; but for this proof of concept analysis it will turn out to be sufficient to classify the somatic signals as those with an amplitude above $50mV$.



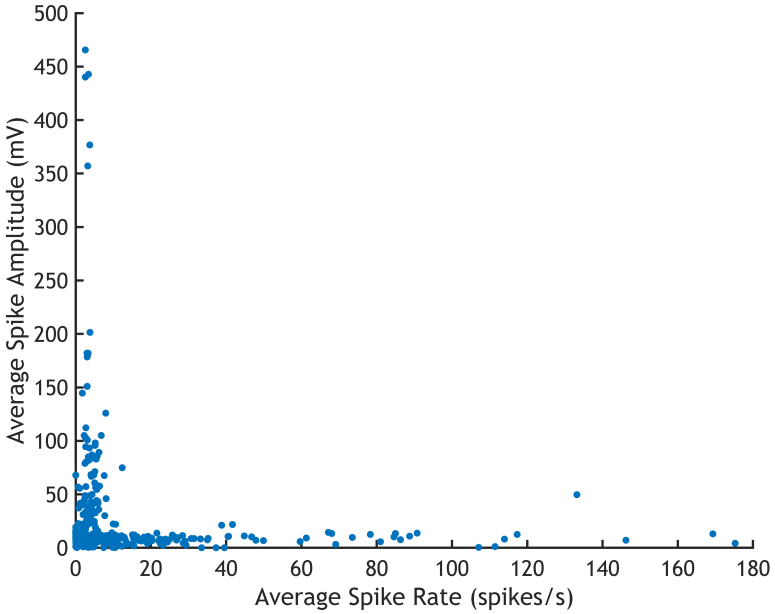
(a)



(b)

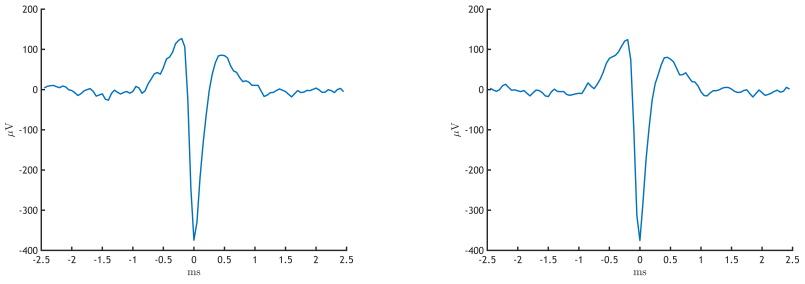
Figure 6.1: Maps of the whole sensor array, showing (a) the average spike rate of the detected neuronal structures, and (b) the average amplitude at the center electrode.

Figure 6.2: amplitude versus frequency

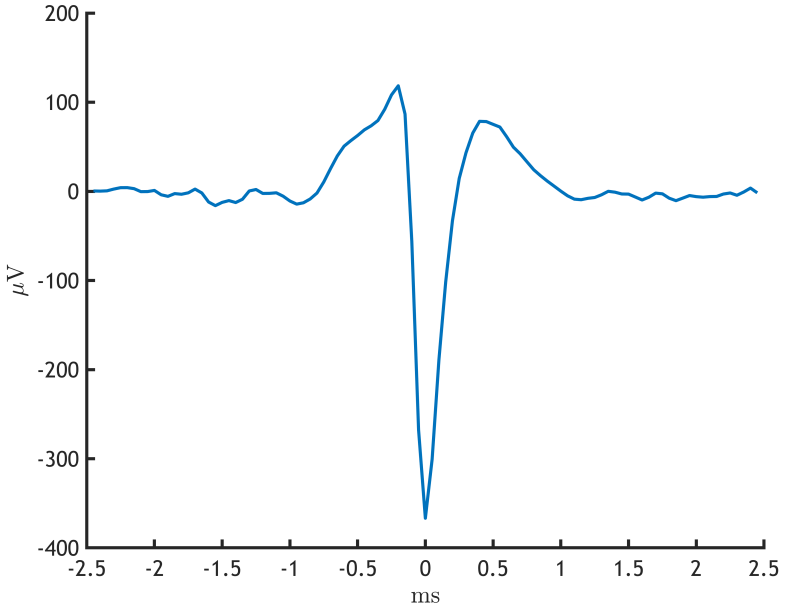


Given the spike times of each detected structure, we can also construct so-called spike triggered averages (STA): centered on each spike time, we extract a 'window' of data, and we subsequently average over the windows obtained for all spike times of the structure. If we do this for the center electrode of a neuronal structure that produces high amplitude signals, the average reveals not much more than what can be seen if we consider a single window (see Fig. 6.3 for an example). This is of course a reflection of the high signal to noise ratio of these spikes, together with the fact that we exactly determined the spike times based on the similarity between waveforms. But because we allowed the spike sorting algorithm to pick up spikes with very low amplitude, confounded with a lot of noise, the spike triggered average of a low amplitude structure reveals much more information. An example is given in Fig. 6.4. Averaging out the noise, a low-amplitude axonal action potential becomes visible. We can verify that this is truly a neuronal structure by constructing an STA over all simultaneously recorded electrodes; this is in fact a movie of what happens (on average, and in space and time) when the neuronal structure spikes. The 'movie stills' also shown in Fig. 6.4 indeed reveal the flow of an axonal action potential [1]

[1] This visible flow of information is in itself a very useful piece of information about the network connectivity, but is not further used within the scope of this project.

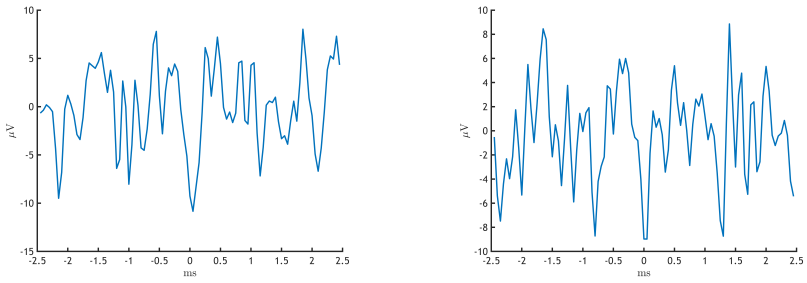


(a) (b)

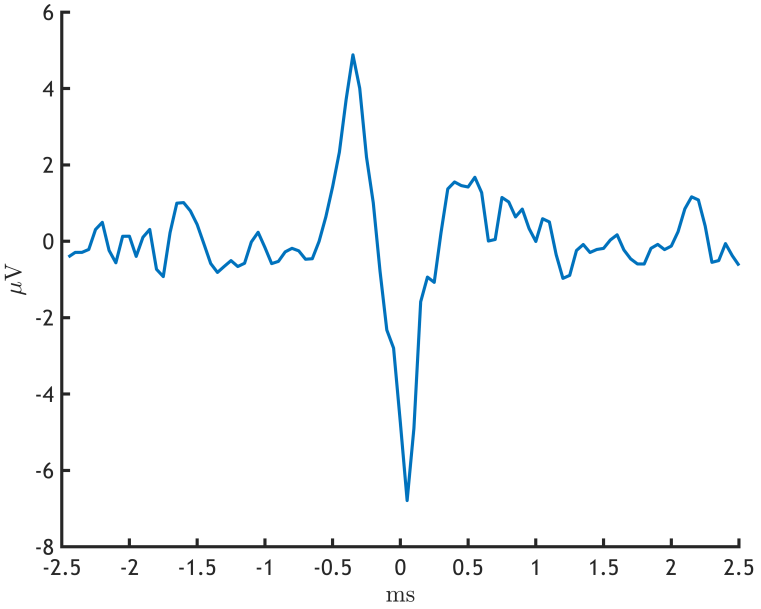


(c)

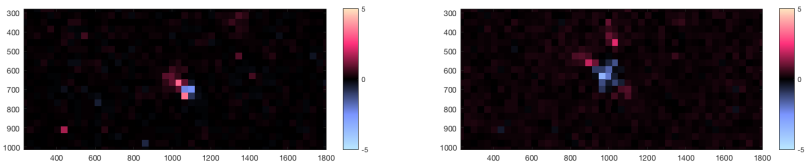
Figure 6.3: Example of high amplitude spikes that were detected by the spike sorting algorithm. In (a) and (b), a window of the original timeseries data centered around the detected peak (the 'spiketime') is shown. (c) shows the average over all such windows of data corresponding to the detected spiketimes.



(a) (b)



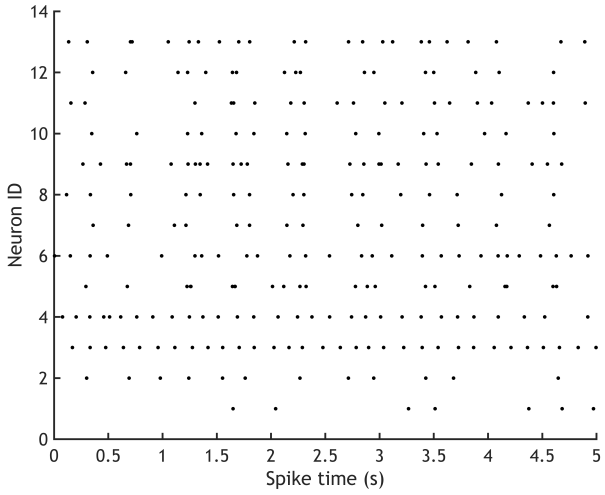
(c)



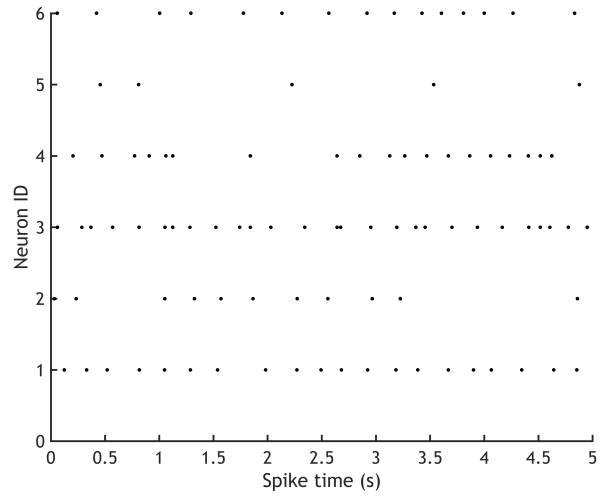
(d) (e)

Figure 6.4: Example of low amplitude axonal signals, buried in noise, that were detected by the spike sorting algorithm. In (a) and (b), a window of the original time series data centered on the detected spike time is shown. (c) shows the average over all such windows (STA). (d) and (c) use the same procedure, but visualize the STA at a single instance in time for the whole recorded part of the sensor array. (d) shows the averaged data exact at the spike time, and (c) reveals the subsequent flow of the signals in two axonal branches at 0.5ms later in time. Red corresponds to positive voltages, blue negative voltages.

Finally, it is very insightful to plot the spiking patterns of all somatic structures in a raster plot, i.e., a plot where each spike for each somatic structure is represented as a dot in function of time. Figure 6.5 shows two such raster plots, obtained from spike sorted recordings covering two qualitatively different regions. The first raster plot corresponds to the region in the upper left corner of the sensor array, where a cluster and some other high-amplitude structures can be seen (see Fig. 6.1 (b)). The second raster plot corresponds to the region in the lower right corner, where high amplitude structures are sparse. The first raster plot (Fig. 6.5 (a)) clearly shows global synchronized activity between the different somatic structures, while Fig. 6.5 (b) shows rather uncorrelated, spontaneous activity at different but relatively slow rates. It is precisely this difference in activity patterns and the relation with the network effective connectivity that we will analyze in the next sections.



(a)



(b)

Figure 6.5: Raster plots corresponding to spike sorted recordings of the upper left corner (a) and the lower left corner (b) of the sensor array.

6.2 *A Study of the Spike Train Statistics*

As described in section 5.4, we will try to classify the somatic spike trains in three different categories: those with statistics close to a Poisson or second order gamma process, those exhibiting mixed time scales, and those that exhibit only the relatively large timescales of pure spontaneous activity. Three distinctive examples for each category are shown in Fig. 6.6.

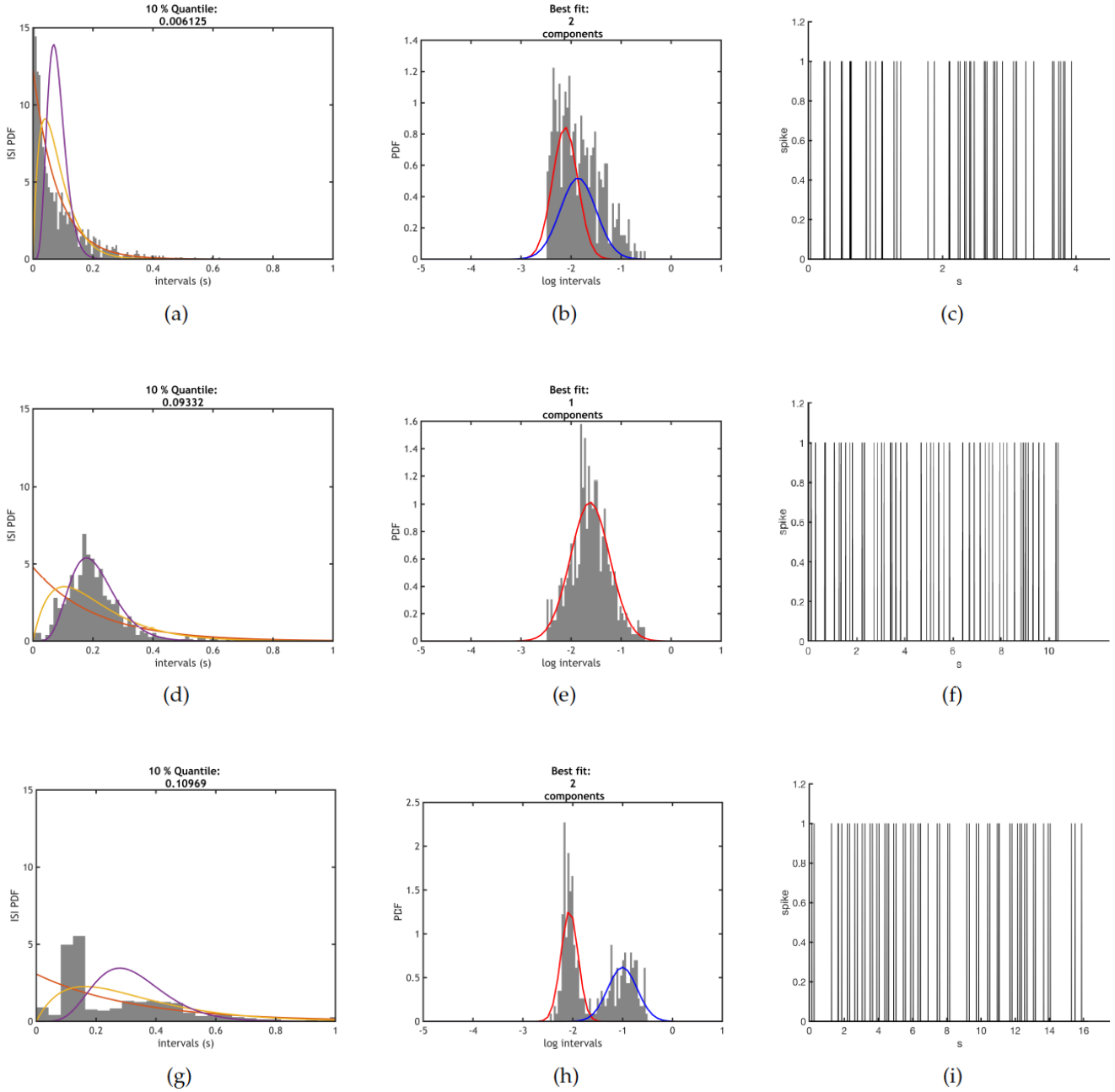


Figure 6.6: Inter spike interval distribution (left), log inter spike interval distribution (middle), and spike trains (right) from three distinctive examples of each considered category. (a), (b) and (c) correspond to a neuron exhibiting Poisson like statistics; (d), (e), (f) to pure spontaneous activity; and (g), (h) and (f) to a neuron emitting spikes at two distinct timescales.

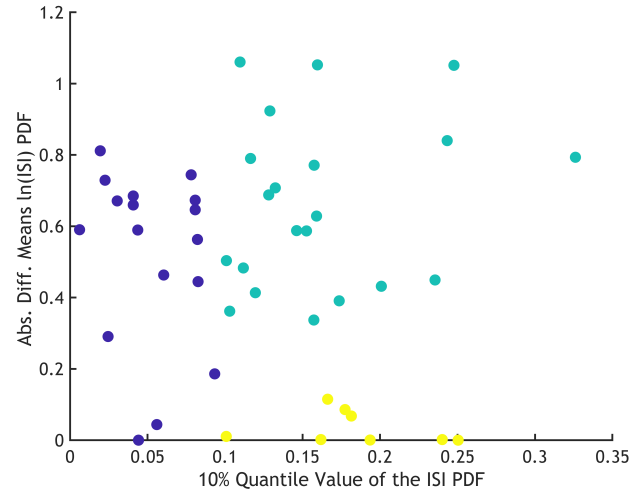
In the first column of Fig. 6.6 (figures (a), (d) and (g)), the inter spike interval distribution is shown. Note the timescale on the x-axis: previously, we discussed spike train statistics of Poisson and second order

gamma processes in view of the refractory period, which is of the order of *ms*; here, the inter spike intervals are spread across multitudes of *0.1s*. The effect of a refractory period, however present, is thus not visible in these plots. We plotted three curves in addition to each histogram of the ISI distributions: the Poisson ISI distribution (orange), the second order gamma process (yellow), and the fifth order gamma process (purple) corresponding to the average spike rate (see Eq. 3.53). Clearly, the histogram in (a) exhibits Poisson like statistics; (d) is most related to a 5th order gamma process, indicating a relatively slow rate activity corresponding to pure spontaneous activity. In fact, we can consider the absence of small inter spike intervals as a generalized refractory period, occurring at a much larger timescale than the original refractory period. (g), then, seems not to be closely matching any gamma process, but instead exhibits two clearly different timescales.

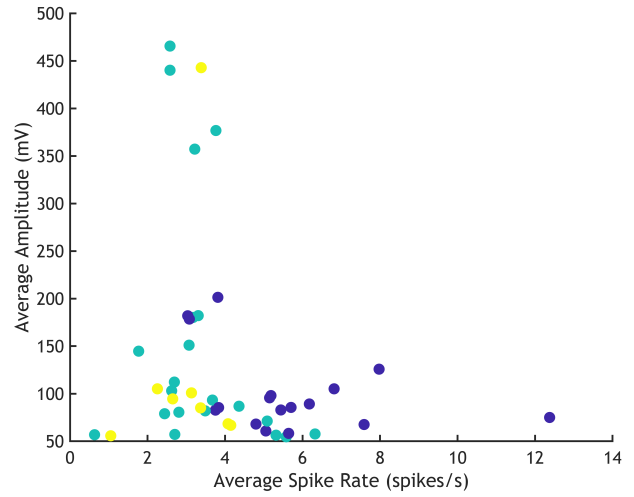
The second column of Fig. 6.6 shows the histograms of the log intervals, which aids in distinguishing peaks at smaller and larger timescales. If the histogram is best fitted by a single Gaussian (which is the case for the spontaneous activity neuron, as can be seen in (e)) that Gaussian distribution is shown in red; if the histogram is better fitted by a two-component Gaussian mixture, the additional Gaussian distribution is shown in blue ((b) and (h)). Note that the log ISI histogram of a neuron with Poisson like statistics is also better fitted by a mixture of two Gaussians, as compared to a fit with a single Gaussian. Finally, the third column of Fig. 6.6 shows snippets of the corresponding spike trains. The limits on the x-axis are rescaled as to show a similar number of spikes for an easy comparison of regularity (or the lack thereof); the scales between the figures are thus substantially different.

We then classified all the somatic activity patterns based on the difference in mean of the fitted Gaussian mixtures (fitted to the log ISI distributions) and the 10% quantile values of the original ISI distributions, as described in section 5.4. The result is shown in Fig. 6.7 (a). The boundary (10%) quantile value was set to $0.1s$, and the boundary value of the difference in means was chosen to be $0.2s$. These values are not derived from first principles, but they seem to reflect the underlying biophysical reality; however, the distinction between the classes is in fact not that clear-cut, and the boundaries values are used in a proof-of-concept manner.

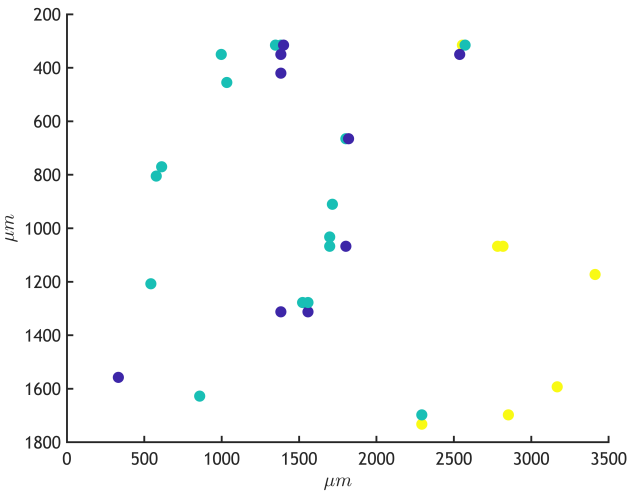
Fig. 6.7 (b) shows the average amplitude in function of the average spike rate for each of the somatic structures, colored according to their class. Members of the Poisson like class naturally show a higher average spike rate than the members of the slow rate spontaneous activity class, while the timescale mixture class members mostly exhibit average rates somewhere in between. Fig. 6.7 (c) then shows the location of the members of each class in the sensor array. This is entirely in accordance with our findings in the previous section (compare with Fig. 6.6).



(a)



(b)



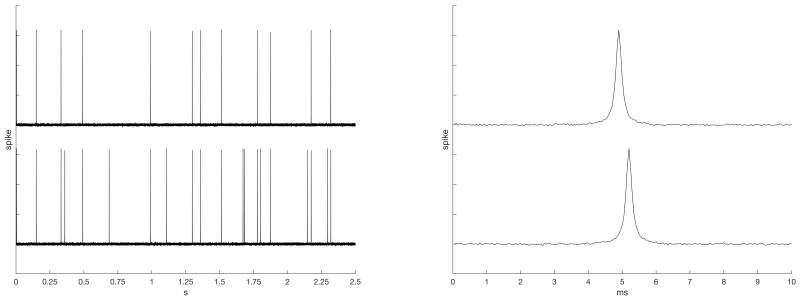
(c)

Figure 6.7: Classification of the single neuron activity (represented by the activity patterns of the somatic structures). (a) shows the absolute value of the difference between the means of the two-component Gaussian mixtures in function of the 10% quantile values of the corresponding ISI distributions. The dark blue dots correspond to Poisson like processes, the yellow dots to slow rate spontaneous activity, and the green dots to double timescale mixtures. The same coloring scheme is used in (b) a scatter plot of the average amplitude in function of the average spike rate, and (c) a scatter plot revealing the location of somatic structures.

6.3 *A Study of the Effective Connectivity*

We can now relate our previous results to the effective connectivity properties we derived from computing the network autoregressive kernels. First of all, it is important to note that the directed flow and transmission of signals in a neural network take place at much smaller timescales than the timescales involved in the spontaneous activity (see chapter 2). We thus computed the correlations and kernels up to lags of 25*ms*, which roughly corresponds to the largest transmission delay a signal between two neurons could have (given our recording area).

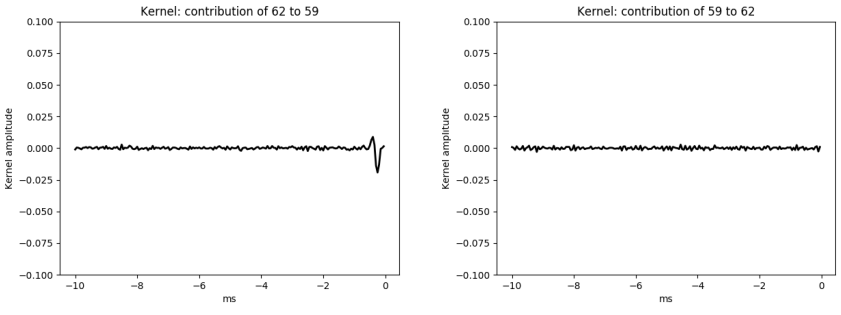
We will discuss a single, distinctive example we derived from our recordings: two closely connected neurons, both classified as being of the time scale mixture type. Fig. 6.8 shows a part of their spike trains. A close inspection of (a) shows that the spike train of the second neuron (bottom trace) correspond to the sum of its own spontaneous activity plus the spikes of the first neuron. When we zoom in on the spike trains (Fig. 6.8 (b)), it seems to be the case that the spikes of the first neuron always precede the spikes of the second neuron.



(a) (b)

Figure 6.8: Spike train snippets of two closely connected neurons. (b) is a zoomed in version of (a).

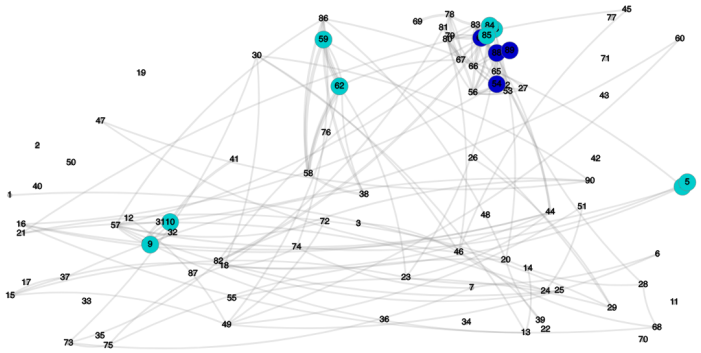
Upon inspecting the computed kernels (shown in Fig. 6.9), we can see that this information is exactly captured by the kernels: the first neuron sends spikes to the second neuron (direction of the transfer) with a transmission delay of around 0.5*ms*, which might indicate a transfer via a chemical synapse.



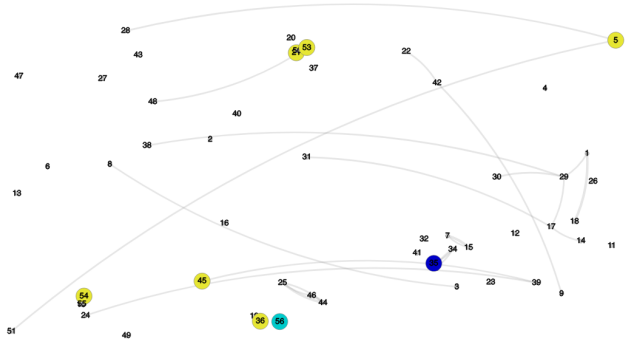
(a) (b)

Figure 6.9: Resulting kernels. (a) a significant contribution from neuron 1 to neuron 2 (which have general neuron IDs 59 and 62); (b) the kernel corresponding to the other direction is not significant.

Finally, we show the complete graphs of information flow as derived from the significant kernels in Fig. 6.10, for the two earlier considered regions in the upper left and lower right corner of the sensor array. It is immediately clear that (a) corresponds to a region of much larger effective connectivity, with even a local cluster of super high connectivity visible; (b) on the other hand is a region where connectivity has not developed beyond some sparse connections. The colors of the nodes again correspond to their classes, with gray indicating axonal structures.



(a)



(b)

Figure 6.10: Graphs of the significant information flows. (a) shows the significant connection in the upper left corner of the sensor array. (b) shows the significant connections in the lower right corner of the sensor array. The small, gray nodes correspond to axonal structures, the gray lines to significant connections. The somatic structures are again colored according to their class.

6.4 *Discussion and Conclusion*

Combining all these results and their visualizations, we can arrive at a straightforward explanation of the measured activity patterns in terms of the connectivity. All neurons seem to show a spontaneous, slow rate activity when not significantly connected. A significant effective connection between two neurons will result in the activity of the receiving neuron to be the (partial) sum of its own spontaneous activity and the activity of the source neuron. This sum of stochastic processes introduces smaller timescales in the spiking pattern of the receiving neuron: if we think of the spontaneous activity as showing a generalized, large refractory period, then this refractory period might be violated by the arriving spikes (since, in a unidirectional connection, the source neuron emits spikes *independent* of the receiving neuron). Connected neurons are thus generally

spiking at multiple time scales. We can even go further, and explain the Poisson like spiking processes in terms of very high connectivity: the more incoming spikes associated to independently spiking source neurons, the more the inter spike interval will lose its dependence on the previous spike. The generalized refractory period associated to the slow rate higher order gamma process thus gradually disappears, and the process starts to look more and more like a second order gamma process with its natural, short refractory period, depending on the number of effective, significant connections.

The proposed method to classify the neuronal activity patterns can in this case thus be directly linked to the connectivity: the difference in means of the two component Gaussian mixture is a first indication of connectivity, because it indicates the existence of at least two different time scales. The 10% quantile value then indicates the degree of connectivity: the higher the connectivity, the more the ISI distribution will shift from the slow rate spontaneous activity distribution to a Poisson or second order gamma distribution, both having a large probability mass in the smaller timescales.

Finally, the relatively fast timescales of axonal and synaptic transmission (as compared to the timescale of spontaneous activity) will result in a highly connected network to appear synchronized (of course depending on the binning of timesteps). Or results are thus in close agreement with what was found in previous experimental studies of in vitro cortical networks. However, the simple underlying mechanism we can discern is most likely not the only one of importance in a real in vivo cortical networks. Apart from the fact that these networks have both inhibitory and excitatory connections, and the fact that they receive natural external stimulation from the earliest stages of development, these types of networks also exhibit a much higher degree of connectivity. This particular experimental setup has proven to be an excellent testbed for our data analysis procedures, however, and we will continue to test and improve our methods in more realistic settings in the future.

Bibliography

Theodore Wilbur Anderson. *An introduction to multivariate statistical analysis*, volume 2. Wiley New York, 1958.

Ehsan Arabzadeh, Stefano Panzeri, and Mathew E Diamond. Deciphering the spike train of a sensory neuron: counts and temporal patterns in the rat whisker pathway. *Journal of Neuroscience*, 26(36):9216–9226, 2006.

Frederico AC Azevedo, Ludmila RB Carvalho, Lea T Grinberg, José Marcelo Farfel, Renata EL Ferretti, Renata EP Leite, Wilson Jacob Filho, Roberto Lent, and Suzana Herculano-Houzel. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5): 532–541, 2009.

Wyeth Bair and Christof Koch. Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural computation*, 8(6):1185–1202, 1996.

Wyeth Bair, Christof Koch, William Newsome, and Kenneth Britten. Power spectrum analysis of bursting cells in area mt in the behaving monkey. *Journal of Neuroscience*, 14(5):2870–2892, 1994.

Robert E Baker, Michael A Corner, and Jaap van Pelt. Spontaneous neuronal discharge patterns in developing organotypic mega-co-cultures of neonatal rat cerebral cortex. *Brain research*, 1101(1):29–35, 2006.

William Bialek, Fred Rieke, RR De Ruyter Van Steveninck, and David Warland. Reading a neural code. *Science*, 252(5014):1854–1857, 1991.

Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4-5):175–308, 2006.

Kenneth H Britten, William T Newsome, Michael N Shadlen, Simona Celebrini, and J Anthony Movshon. A relationship between behavioral choice and the visual responses of neurons in macaque mt. *Visual neuroscience*, 13(1):87–100, 1996.

Nicolas Brunel. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of computational neuroscience*, 8 (3):183–208, 2000.

Michela Chiappalone, Marco Bove, Alessandro Vato, Mariateresa Tedesco, and Sergio Martinoia. Dissociated cortical networks show spontaneously correlated activity patterns during in vitro development. *Brain research*, 1093(1):41–53, 2006.

Michela Chiappalone, Alessandro Vato, Luca Berdondini, Milena Koudelka-Hep, and Sergio Martinoia. Network dynamics and synchronous activity in cultured cortical neurons. *International journal of neural systems*, 17(02):87–103, 2007.

Carson C Chow and John A White. Spontaneous action potentials due to channel fluctuations. *Biophysical journal*, 71(6):3013–3021, 1996.

Anne K Churchland, Roozbeh Kiani, Rishidev Chaudhuri, Xiao-Jing Wang, Alexandre Pouget, and Michael N Shadlen. Variance as a signature of neural computations during decision making. *Neuron*, 69(4): 818–831, 2011.

Mark M Churchland and LF Abbott. Two layers of neural variability. *Nature neuroscience*, 15(11):1472, 2012.

Mark M Churchland, M Yu Byron, John P Cunningham, Leo P Sugrue, Marlene R Cohen, Greg S Corrado, William T Newsome, Andrew M Clark, Paymon Hosseini, Benjamin B Scott, et al. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature neuroscience*, 13(3):369, 2010.

Marlene R Cohen and Adam Kohn. Measuring and interpreting neuronal correlations. *Nature neuroscience*, 14(7):811, 2011.

David Roxbee Cox and Valerie Isham. *Point processes*, volume 12. CRC Press, 1980.

Roger T Dean and William TM Dunsmuir. Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: The importance of constructing transfer function autoregressive models. *Behavior research methods*, 48(2):783–802, 2016.

Dominique Debanne, Emilie Campanac, Andrzej Bialowas, Edmond Carlier, and Gisèle Alcaraz. Axon physiology. *Physiological reviews*, 91 (2):555–602, 2011.

Alain Destexhe. Oscillations, complex spatiotemporal behavior, and information transport in networks of excitatory and inhibitory neurons. *Physical Review E*, 50(2):1594, 1994.

Scott N Deyo and William W Lytton. Inhibition can disrupt hypersynchrony in model neuronal networks. *Progress in neuropsychopharmacology & biological psychiatry*, 21(5):735–750, 1997.

Ricardo Dolmetsch and Daniel H Geschwind. The human brain in a dish: the promise of ipsc-derived neurons. *Cell*, 145(6):831–834, 2011.

Rodney J Douglas, KA Martin, and David Whitteridge. An intracellular analysis of the visual responses of neurones in cat visual cortex. *The Journal of physiology*, 440(1):659–696, 1991.

Julia H Downes, Mark W Hammond, Dimitris Xydas, Matthew C Spencer, Victor M Becerra, Kevin Warwick, Ben J Whalley, and Slawomir J Nasuto. Emergence of a small-world functional network in cultured neurons. *PLoS computational biology*, 8(5):e1002522, 2012.

Martha J Farah. Emerging ethical issues in neuroscience. *Nature neuroscience*, 5(11):1123, 2002.

Wulfram Gerstner, Werner M Kistler, Richard Naud, and Liam Paninski. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.

David E Goldman. Potential, impedance, and rectification in membranes. *The Journal of general physiology*, 27(1):37–60, 1943.

Charles M Gray, Peter König, Andreas K Engel, and Wolf Singer. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338(6213):334, 1989.

Neal A Hessler, Aneil M Shirke, and Roberto Malinow. The probability of transmitter release at a mammalian central synapse. *Nature*, 366(6455): 569, 1993.

AL Hodgkin. Evidence for electrical transmission in nerve. *The Journal of physiology*, 90(2):183–210, 1937.

Alan L Hodgkin and Andrew F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500–544, 1952.

Maciej Kamiński, Mingzhou Ding, Wilson A Truccolo, and Steven L Bressler. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological cybernetics*, 85(2):145–157, 2001.

Eric R Kandel, James H Schwartz, Thomas M Jessell, Department of Biochemistry, Molecular Biophysics Thomas Jessell, Steven Siegelbaum, and AJ Hudspeth. *Principles of neural science*, volume 4. McGraw-hill New York, 2000.

Adam Kepecs, Xiao-Jing Wang, and John Lisman. Bursting neurons signal input slope. *Journal of Neuroscience*, 22(20):9053–9062, 2002.

Adam Kohn and Matthew A Smith. Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *Journal of Neuroscience*, 25(14):3661–3673, 2005.

Hendrik Anthony Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, 7(4):284–304, 1940.

Pawel Kudela, Piotr J Franaszczuk, and Gregory K Bergey. Changing excitation and inhibition in simulated neural networks: effects on induced bursting behavior. *Biological cybernetics*, 88(4):276–285, 2003.

Michael S Lewicki. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 9(4):R53–R78, 1998.

Ashok Litwin-Kumar and Brent Doiron. Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature neuroscience*, 15(11):1498, 2012.

Isabel Llano, Christina K Webb, and Francisco Bezanilla. Potassium conductance of the squid giant axon. single-channel studies. *The Journal of general physiology*, 92(2):179–196, 1988.

Heiko J Luhmann, Anne Sinning, Jenq-Wei Yang, Vicente Reyes-Puerta, Maik C Stüttgen, Sergei Kirischuk, and Werner Kilb. Spontaneous neuronal activity in developing neocortical networks: from single cells to large-scale interactions. *Frontiers in neural circuits*, 10:40, 2016.

Eisaku Maeda, HP Robinson, and Akio Kawana. The mechanisms of generation and propagation of synchronized bursting in developing networks of cortical neurons. *Journal of Neuroscience*, 15(10):6834–6845, 1995.

Amit Manwani and Christof Koch. Detecting and estimating signals in noisy cable structures, i: Neuronal noise sources. *Neural computation*, 11(8):1797–1829, 1999.

Jessica Mariani, Maria Vittoria Simonini, Dean Palejev, Livia Tomasini, Gianfilippo Coppola, Anna M Szekely, Tamas L Horvath, and Flora M Vaccarino. Modeling human cortical development in vitro using induced pluripotent stem cells. *Proceedings of the National Academy of Sciences*, 109(31):12770–12775, 2012.

Henry Markram and Misha Tsodyks. Redistribution of synaptic efficacy between neocortical pyramidal neurons. *Nature*, 382(6594):807, 1996.

Jan Müller, Marco Ballini, Paolo Livi, Yihui Chen, Milos Radivojevic, Amir Shadmani, Vijay Viswam, Ian L Jones, Michele Fiscella, Roland Diggelmann, et al. High-resolution cmos mea platform to study neurons at subcellular, cellular, and network levels. *Lab on a Chip*, 15(13): 2767–2780, 2015.

Walther Nernst. Zur kinetik der in lösung befindlichen körper. *Zeitschrift für physikalische Chemie*, 2(1):613–637, 1888.

Arnold Neumaier and Tapio Schneider. Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software (TOMS)*, 27(1):27–57, 2001.

Srdjan Ostojic. Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nature neuroscience*, 17(4):594, 2014.

Sergiu P Paşca, Thomas Portmann, Irina Voineagu, Masayuki Yazawa, Aleksandr Shcheglovitov, Anca M Paşca, Branden Cord, Theo D Palmer, Sachiko Chikahisa, Seiji Nishino, et al. Using ipsc-derived neurons to uncover cellular phenotypes associated with timothy syndrome. *Nature medicine*, 17(12):1657, 2011.

Quentin Pauluis, Stuart N Baker, and Etienne Olivier. Emergent oscillations in a realistic network: the role of inhibition and the effect of the spatiotemporal distribution of the input. *Journal of computational neuroscience*, 6(1):27–48, 1999.

Rajnish Ranjan, Georges Khazen, Luca Gambazzi, Srikanth Ramaswamy, Sean L Hill, Felix Schürmann, and Henry Markram. Channelpedia: an integrative and interactive database for ion channels. *Frontiers in neuroinformatics*, 5:36, 2011.

Michael W Reimann, Max Nolte, Martina Scolamiero, Katharine Turner, Rodrigo Perin, Giuseppe Chindemi, Paweł Dłotko, Ran Levi, Kathryn Hess, and Henry Markram. Cliques of neurons bound into cavities provide a missing link between structure and function. *Frontiers in computational neuroscience*, 11:48, 2017.

Hernan Gonzalo Rey, Carlos Pedreira, and Rodrigo Quian Quiroga. Past, present and future of spike sorting techniques. *Brain research bulletin*, 119:106–117, 2015.

Fred Rieke and David Warland. *Spikes: exploring the neural code*. MIT press, 1999.

Hannes Risken. Fokker-planck equation. In *The Fokker-Planck Equation*, pages 63–95. Springer, 1996.

Robert Rosenbaum, Matthew A Smith, Adam Kohn, Jonathan E Rubin, and Brent Doiron. The spatial structure of correlated neuronal variability. *Nature neuroscience*, 20(1):107, 2017.

Cyrille Rossant, Shabnam N Kadir, Dan FM Goodman, John Schulman, Maximilian LD Hunter, Aman B Saleem, Andres Grosmark, Mariano Belluscio, George H Denfield, Alexander S Ecker, et al. Spike sorting for large, dense electrode arrays. *Nature neuroscience*, 19(4):634, 2016.

Seref Sagiroglu and Duygu Sinanc. Big data: A review. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on*, pages 42–47. IEEE, 2013.

Anil K Seth. Neural coding: rate and time codes work together. *Current Biology*, 25(3):R110–R113, 2015.

William R Softky and Christof Koch. The highly irregular firing of cortical cells is inconsistent with temporal integration of random epsps. *Journal of Neuroscience*, 13(1):334–350, 1993.

Larry Squire, Darwin Berg, Floyd E Bloom, Sascha Du Lac, Anirvan Ghosh, and Nicholas C Spitzer. *Fundamental neuroscience*. Academic Press, 2012.

Ian H Stevenson and Konrad P Kording. How advances in neural recording affect data analysis. *Nature neuroscience*, 14(2):139, 2011.

A Strassberg and DeFelice. Limitations of the hodgkin-huxley formalism: effects of single channel kinetics on transmembrane voltage dynamics. *Neural computation*, 5(6):843–855, 1993.

E Vaadia, I Haalman, M Abeles, H Bergman, Y Prut, Hi Slovin, and AMHJ Aertsen. Dynamics of neuronal interactions in monkey cortex in relation to behavioural events. *Nature*, 373(6514):515, 1995.

Jaap van Pelt, Ildiko Vajda, Pieter S Wolters, Michael A Corner, and Ger JA Ramakers. Dynamics and plasticity in developing neuronal networks in vitro. *Progress in brain research*, 147:171–188, 2005.

Carl Van Vreeswijk and Haim Sompolinsky. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293): 1724–1726, 1996.

C van Vreeswijk and Haim Sompolinsky. Chaotic balanced state in a model of cortical circuits. *Neural computation*, 10(6):1321–1371, 1998.

Gilbert Thomas Walker. On periodicity in series of related terms. *Proc. R. Soc. Lond. A*, 131(818):518–532, 1931.

Mati Wax and Thomas Kailath. Efficient inversion of toeplitz-block toeplitz matrix. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(5):1218–1221, 1983.

Pierre Yger, Giulia LB Spampinato, Elric Esposito, Baptiste Lefebvre, Stéphane Deny, Christophe Gardella, Marcel Stimberg, Florian Jetter, Guenther Zeck, Serge Picaud, et al. A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo. *ELife*, 7:e34518, 2018.

G Udny Yule. Why do we sometimes get nonsense-correlations between time-series?–a study in sampling and the nature of time-series. *Journal of the royal statistical society*, 89(1):1–63, 1926.

Y Zuo, H Safaai, G Notaro, A Mazzoni, S Panzeri, and M Diamond. Spike timing and spike rate make complementary contributions to perceptual decisions in rat s1 and s2 cortex. *Current Biology*, 25:357–363, 2015.

VRIJE
UNIVERSITEIT
BRUSSEL

In Partial Fulfillment of the Requirements for the Degree of
Master of Science in Applied Sciences and Engineering: Computer Science
Profile Artificial Intelligence

Connecting Neurons

Comparing Computations in Artificial Neural Networks and Cultured Networks of Human Neurons

Hannah Pinson

August 27, 2018

Supervisor: Prof. Dr. Tom Lenaerts
Co-supervisor: Prof. Dr. Max Tegmark

Faculty of Sciences and Bioengineering Sciences
Department of Computer Science

Summary

Artificial neural networks are inspired by their biological counterparts, but in developing those artificial networks we are by no means limited to mechanisms that are plausible in the natural world. Thus far, this biologically inspired design process has yielded artificial neural networks that can be applied in a very wide range of practical situations—from cancer screening to self-driving cars—often with astounding results. To develop these artificial neural networks even further, we can resort to adding purely artificial architectures and enhancements; still it seems there is lot left to learn from the efficient computational processes taking place in biological neural networks.

In previous work [1], we showed how in an in vitro network of cortical human neurons the spiking activity is related to the effective connectivity.

In this thesis, we will elaborate on the methods we used to analyze this large collection of multivariate time series data. We will discuss the algorithm that allowed us to identify neuronal structures, and we will describe the used method to compute effective connectivity from activity data. However, our main goal is to compare the obtained model of the biological network to different related types of artificial neural network models for which the computational power has been studied. To put our analysis in context, we will start by discussing the symbiosis between artificial and biological neural network research in general, and between networks of biological and artificial spiking neurons in particular. The latter artificial models will turn out to be the closest related to our model of the biological network. Given the particular dynamics arising in our culture of cortical neurons, we will also be able to explain why this network can be reasonably approximated by a single layer perceptron network.

Our general results are meant to be a starting point for future research exploring the relation between biological and artificial neural networks. Through developing these methods, we hope to be able to—at some point in the future—increase insight in the computational properties of biological neural networks, as well as in the biologically inspired principles that could further advance the field of artificial neural networks.

[1] *Connecting Neurons: A Stochastic Model of Information Flow in Cultured Networks of Human Neurons*, master thesis physics (unpublished).

Contents

1

Introduction

Biological neural networks and their artificial counterparts have a symbiotic relationship. Understanding of the former sparked the development of the latter, and gained insights in artificial neural networks might in their turn increase our understanding in biological neural networks [1]. However, in further developing artificial neural networks (ANN) we are by no means limited to mechanisms that are plausible in the natural world. To the extent that biological processes or constraints suggest useful approaches, we can adopt them; but we are of course free to dispense of any property or mechanism that seems to impede us from arriving at our goal—and at the same time we can add any purely artificial mechanism that allows us to reach it [2].

[1] *Review of advances in neural networks: Neural design technology stack.* Almási et al.,2016.

[2] *Artificial Intelligence (A Modern Approach).* Russell and Norvig,2010.

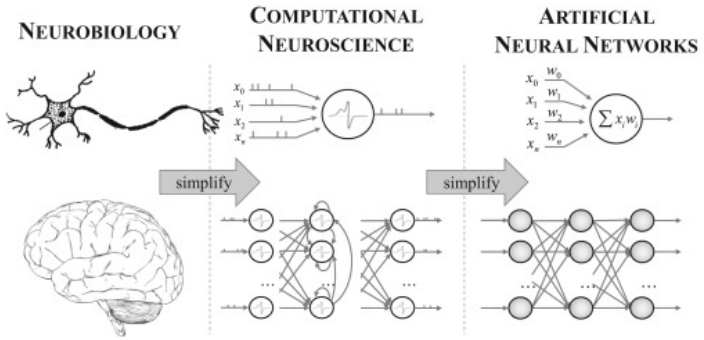


Figure 1.1: Schematic representation of how research in computational neuroscience inspires the design of artificial neural networks. Source: Almasi et al, 2016.

A successful example of this biologically inspired approach is a convolutional neural network [3], a type of artificial neural network that has been very successfully applied in the field of computer vision [4]. Its multi-layered structure is based on the architecture of the visual cortex in the primate brain [5] [6]. But the ubiquitous back propagation algorithm typically used to train ANN like convolutional neural networks was not conceived from an equivalent biological mechanism. This algorithm seems to work so well, however, that it bears a question in the field of neuroscience in its turn: is the back propagation of signals in biological neurons related to learning? [7] [8] The symbiotic relationship between artificial and biological neural networks is thus certainly of the mutual kind.

[3] *Convolutional networks for images, speech, and time series.* LeCun et al.,1995.

[4] *Imagenet classification with deep convolutional neural networks.* Krizhevsky et al.,2012.

[5] *Mathematical description of the responses of simple cortical cells.* Marĉelja,1980.

[6] *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex.* Hubel and Wiesel,1962.

[7] *Experience-dependent changes in extracellular spike amplitude may reflect regulation of dendritic action potential back-propagation in rat hippocampal pyramidal cells.* Quirk et al.,2001.

[8] *Cortical action potential backpropagation explains spike threshold variability and rapid-onset kinetics.* Yu et al.,2008.

The recent successful applications of artificial neural networks in a wide range of domains[9]—from face recognition over cancer screening to self-driving cars—are all involving artificial neural networks that are based on more or less the same abstractions of biological neurons and networks. Within this group of artificial networks, there are of course differences in architecture and design. These are often implemented to facilitate the processing of data with a particular structure: e.g. the convolution operations in a convolutional neural network are well suited for image tasks, while the recurrence operations in recurrent neural networks provide a form of memory that is useful in time series prediction. But neuroscience research has shown that there are in fact a lot more computational mechanisms in the biological brain that could, after some form of abstraction, be used to construct artificial neural networks [10] [11]. It is for example known that computations in the biological brain are based on information encoded in the exact spike times of the neurons, in their average spike *rates*, or sometimes in a combination of both [12]. While the class of artificial neural networks commonly used today are based on spike *rate* encodings, it can be shown that abstractions of biological neural networks that incorporate exact spike timings yield equally expressive and more efficient artificial neural networks [13] [14]. Such artificial neural networks are usually called *spiking neural networks* (SNN).

[9] *Deep learning*. LeCun et al.,2015.

[10] *Review of advances in neural networks: Neural design technology stack*. Almási et al.,2016.

[11] *Theory and simulation in neuroscience*. Gerstner et al.,2012.

[12] *Neuronal dynamics: From single neurons to networks and models of cognition*. Gerstner et al.,2014.

[13] *Networks of spiking neurons: the third generation of neural network models*. Maass,1997b.

[14] *Deep Learning in Spiking Neural Networks*. Tavanaei et al.,2018.

Despite their advantages, the application of SNN is not yet on a par with commonly used ANN. While they are in theory equally expressive but more efficient (in terms of number of artificial neurons needed to compute the same type of function), they usually require more computational resources to simulate: the more realistic the model of the artificial neuron, the more expensive it becomes to simulate large networks to compute the desired outcomes. The Hodgkin-Huxley model, e.g., is a famous and detailed model of deterministic spike generation involving four coupled differential equations [15]: using this model as the building block in large scale simulated networks is almost infeasible. Integrate-and-fire models [16] on the other hand are the simplest models that incorporate exact spike times, but they are not able to capture all parts of realistic neuronal dynamics [17]. A very different approach is developed by Izhikevich [18]: the artificial spiking neuron model is not biologically plausible in itself, but it is very efficient; on the other hand, simulations of networks with these specific (biologically implausible) building blocks yield a large range of network dynamics commonly encountered in biological neural networks. The biological plausibility in function of the implementation cost for different spiking neuron models is shown in Fig. 1.2.

[15] *A quantitative description of membrane current and its application to conduction and excitation in nerve*. Hodgkin and Huxley,1952.

[16] *Neural engineering: Computation, representation, and dynamics in neurobiological systems*. Eliasmith and Anderson,2004.

[17] *How good are neuron models?*. Gerstner and Naud,2009.

[18] *Simple model of spiking neurons*. Izhikevich,2003.

Another aspect of the practical use of artificial neural networks is of course the *learning* of the network parameters. While back propagation of signals can take place in biological networks, the commonly used back propagation algorithm does not seem biologically feasible—amongst other things biological neural networks do not seem capable of computing derivatives [19]. Of course, that does not prohibit us from implement-

[19] *Review of advances in neural networks: Neural design technology stack*. Almási et al.,2016.

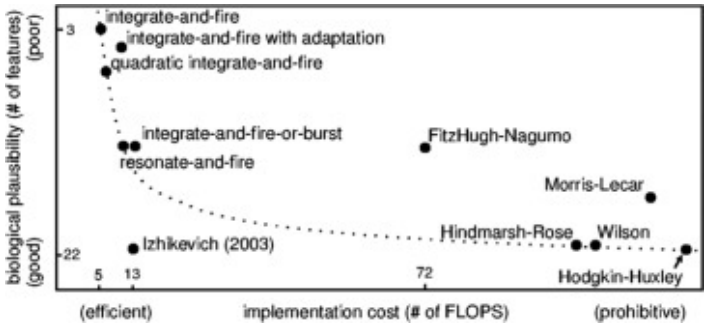


Figure 1.2: Schematic representation of biological plausibility versus implementation cost for different commonly used models of artificial spiking neurons. Source: Almasi et al, 2016.

ing it in simulated SNN [20] [21] [22]. In a biological neural network, however, the relative timing of the spikes in spiking neurons alters the weights, a process called Spike-Timing Dependent Plasticity (STDP) [23] [24]. This thus makes for another learning mechanism that can be used in SNN [25]. In short, spiking neural networks are in theory equally expressive and more efficient; and in practice they can be implemented with different learning algorithms, but they often require more resources than the commonly used artificial neural networks when *simulated*. Their true power will probably only be fully leveraged if we can also use biologically inspired *hardware* instead of the commonly used von Neumann architectures. A successful example of such a very efficient neuromorphic hardware implementation of a spiking neuron network was developed by Merolla et al. [26].

[20] *Error-backpropagation in temporally encoded networks of spiking neurons.* Bohte et al.,2002.

[21] *Training deep spiking neural networks using backpropagation.* Lee et al.,2016.

[22] *Deep Learning in Spiking Neural Networks.* Tavanaei et al.,2018.

[23] *A neuronal learning rule for sub-millisecond temporal coding.* Gerstner et al.,1996.

[24] *Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs.* Markram et al.,1997.

[25] *Simulation of networks of spiking neurons: a review of tools and strategies.* Brette et al.,2007.

[26] *A million spiking-neuron integrated circuit with a scalable communication network and interface.* Merolla et al.,2014.

The symbiotic relationship between artificial and biological neural network research has been very fruitful so far. For its further advancements, however, there seems to be a practical piece missing: until very recently, it was not technically possible to simultaneously record the spiking activity of more than a couple of neurons at once. This severe constraint is of course due to the very small scales involved: in space (typical neural structures have sizes of the order of a couple of μm), time (a typical spikes lasts around $1ms$) and activity (spikes correspond to voltages in the order of ten to hundreds of mV). On top of these scale constraints, the intricate three dimensional structure and the density of neural networks, render it extra challenging to analyse the recorded data. But, for both the understanding of biological neural networks in se, as for insights in their computational properties that could be used to advance the field of artificial intelligence, detailed information on the activity and connectivity of large networks of neurons seems imperative. In fact, one needs *a directed graph of information flow* [27], where the nodes represent single neurons and the edges represent the directed transmission of neural signals. Such a graph can only be constructed if one can measure the activity of multiple connected neurons, and if one can subsequently attribute the resulting spiking patters to the neurons that produced them.

[27] *Cliques of neurons bound into cavities provide a missing link between structure and function.* Reimann et al.,2017.

In this master thesis project, we constructed a directed graph of the information flow between active human cortical neurons growing on top of a high-resolution multi-electrode array. The project brings together—for

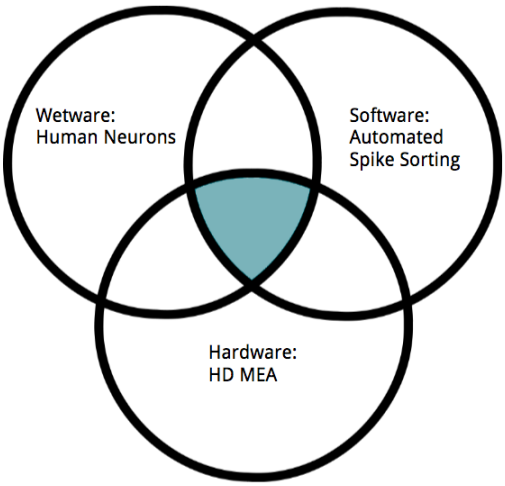


Figure 1.3: The recent scientific advancements that lie at the basis of this project.

the first time—some recent spectacular advancements in science: induced pluripotent stem cells (iPSC) that can be driven to become human cortical neurons and their supporting glia cells [28]; the development of multi electrode arrays for the extracellular recordings of large neuronal ensembles with an unprecedented resolution [29]; the development of algorithms to, from these extracellular recordings, attribute the activity to the correct neurons in a fully automated way [30], and finally, the needed computational resources and procedures to handle large amounts of data [31]. The aspects specific to this project are shown in Fig. 1.3.

The use of stem cell derived cortical neurons in an in vitro setup has its advantages, but it also has obvious drawbacks: how the network formed by these neurons relates to the structures that develop in vivo is not exactly known [32] [33]. A first substantial advantage of the in vitro approach, however, is the fact that the two dimensional network structure developed by these neurons is much easier to analyze. Another substantial advantage is the complete control we have over the input the network receives, by either stimulating it at certain sites or by not stimulating it at all. In this thesis a network that was not stimulated at all is studied. Even in this case—when there is no meaningful computational task at hand—the neurons form spontaneous connections and show spontaneous activity. By studying spontaneous networks we can obtain insights in the fundamental mechanisms of network formation and in the characteristics of spontaneous activity; these results could then also be used as a baseline in future studies where the networks are changing depending on the provided structured input.

The main body of work for this thesis then consisted in constructing a pipeline of numerical procedures that take a recording of the neuronal activity, and output a directed graph of information flow, together with detailed information about the nature of these connections in terms of transmission delays. This pipeline consists of the following steps:

[28] *The human brain in a dish: the promise of iPSC-derived neurons*. Dolmetsch and Geschwind,2011.

[29] *High-resolution CMOS MEA platform to study neurons at subcellular, cellular, and network levels*. Müller et al.,2015.

[30] *A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo*. Yger et al.,2018.

[31] *Big data: A review*. Sagiroglu and Sinanc,2013.

[32] *Modeling human cortical development in vitro using induced pluripotent stem cells*. Mariani et al.,2012.

[33] *Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome*. Paşca et al.,2011.

1. identifying the spiking patterns belonging to separate neurons,
2. calculating the auto- and cross-correlations between pairs of neurons,
3. calculating the so-called network autoregressive kernels, which reflect the flow of information between the different neurons and capture the information about their transmission delays,
4. and finally the construction and visualization of a directed graph of information flow.

We will present a summary of the results obtained from applying this pipeline to a particular recording, and we will finally discuss how this resulting model of a biological network can be mapped to artificial networks of spiking neurons.

1.1 Reading Guide

In the next chapter, we discuss some important properties of biological neural networks in general, and of connected cultured cortical neurons in particular. This chapter thus contains the neuroscientific background needed to put the remainder of this thesis and its results in context.

In chapter 3, we elaborate on the relationship between biological and artificial neural networks, and we summarize some theoretical results on expressibility, efficiency and learnability of neural networks in general. This chapter provides the theoretical framework needed to analyze our results.

Chapter 4 and 5 then contain more detailed discussions of the procedures and algorithms that were used to analyze the neuronal data. Chapter 4 elaborates on the process of spike sorting, a suite of algorithms that allows one to attribute the measured spikes to the correct neurons. Chapter 5 contains the description of the machine learning techniques used to calculate the connectivity and transmission delays from the multivariate time series data.

Finally, in chapter 6, we provide a summary of the obtained connectivity models, and we discuss how these results relate to specific types of artificial neural networks.

2

Cultured Cortical Neural Networks

In this chapter, we briefly discuss some important properties and mechanisms of single biological neurons and of biological neural networks. We will focus the latter discussion on cultured cortical neural networks, since an extended discussion of biological neural networks in general is outside the scope of this thesis. [This chapter is an adapted version of chapter 2 in the master thesis physics.]

2.1 Single Neurons

2.1.1 Neuron Morphology

There exist a great many type of neurons, and they could be classified in a great many ways: based, e.g., on their shape, function, location, activity or neurotransmittor production [1]. Their actual shape depends on their function and the environment they develop in; it is for example the case that stem cell derived neurons kept in vitro will in general develop different from in vivo neurons of the same type [2] [3]. But it is of course useful to discuss the common and most significant elements of all these different neuron morphologies. A schematic representation of these different structures is given in Fig. 2.1.

[1] *Fundamental neuroscience.* Squire et al.,2012.

[2] *Modeling human cortical development in vitro using induced pluripotent stem cells.* Mariani et al.,2012.

[3] *Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome.* Paşca et al.,2011.

The structures that we will most often use to interpret our results are the following:

- *The dendrites*, often forming a structure that is metaphorically called a 'dendritic tree'. These structures receive the extracellular signals and transmit them to
- *the soma*, or neuron cell body, where all these extracellular signals are integrated and passed on to

- *the axon initiation segment (AIS)*, where the action potentials (see later) originate. This structure is followed by
- *the axon*; which transfers the action potentials from the AIS to the synaptic terminals, and finally this structure branches to end up in
- *the synapses*. Here the signals are transferred from the axon terminals to the dendrites of the next neuron.

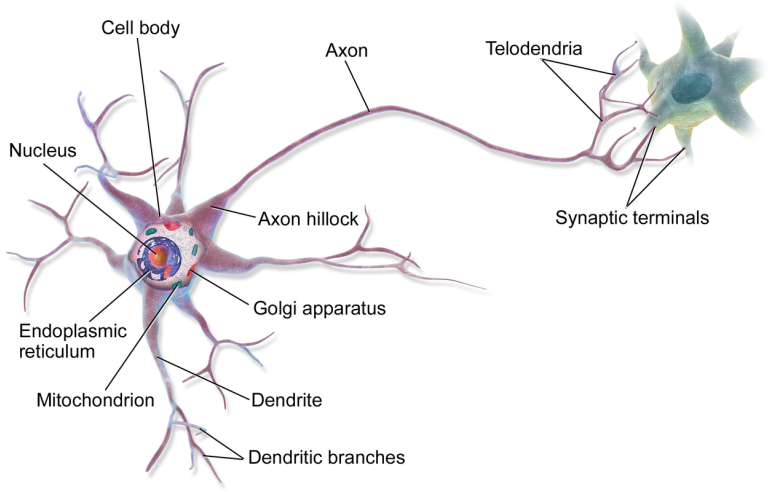


Figure 2.1: Schematic representation of a multipolar neuron, and how its axon is leading up to the next neuron.

2.1.2 *Spikes and Action Potentials*

Spikes—short electrical signals emitted by neurons—form the building blocks of the computational processes we eventually want to study. Often they are taken to be binary events stemming from (point like) neurons: either the neuron emits a spike (1), or it is silent (0). This abstraction removes both the extended neuron morphology as well as the underlying mechanism of action potential generation from the picture; but, as will become clear later, both these concepts—and their combination in terms of action potentials traveling in and between different neurons—will turn out to be important for our analysis. Here we will first discuss the action potential in itself; in the next section, we will explain how the action potentials are generated and transported within a single neuron.

In essence, action potentials are the result of a rapid de- and repolarization of the neuron cell membrane: the potential difference across the membrane abruptly 'spikes' from its resting value, and then it rapidly falls down again. After this voltage spike the membrane voltage is for a short period in time even lower than its baseline value before the spike; this is called the hyperpolarization phase or refractory period. The typical shape of an action potential is schematically shown in Fig. 2.2. We will not discuss the general model underlying action potential generation in great detail here [4], but instead we will continue with a general overview that we can subsequently connect to the neuron morphology.

[4] ref thesis fysica

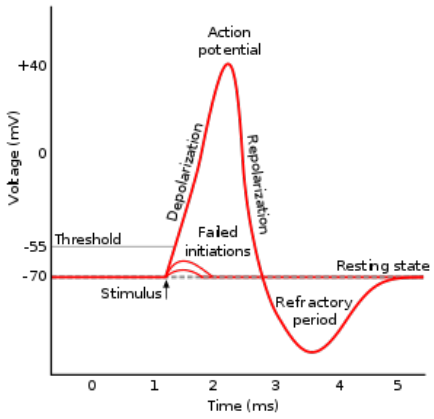


Figure 2.2: Schematic representation of a typical action potential.

The rapid increase in the voltage is possible due to a feedback mechanism between the permeability of the membrane and the potential difference across the membrane. At rest, the membrane has a negative resting potential difference denoted V_r. A slight increase in this voltage, due to the arrival of a stimulus or sometimes due to stochastic effects, opens up certain so-called *ion channels* in the membrane, which allow ions to diffuse from the outside of the cell to the inside. This depolarizes the membrane further. The feedback consists of the ion-channels being voltage dependent, i.e., more channels open as the voltage increases, yielding an ever faster voltage increase or depolarization. In the same process, however, different types of ions and of ion channel states are involved: the channels that were initially opening up become *inactivated*, and other voltage-dependent channels that can transport ions from inside the cell to the outside start to open. This makes the membrane potential difference rapidly decrease again. It finally reaches a value that is below the original membrane voltage V_r. After that, the systems needs a certain period of time to restore itself to its original state.

The feedback mechanism only really 'takes off' once a certain voltage threshold is crossed. If an incoming stimulus is not strong enough to depolarize the membrane above the threshold, no action potential will be initiated (see Fig. 2.2). The needed strength of the stimulus to lead up to an action potential initiation depends naturally on the voltage across the membrane: during the hyperpolarized phase or refractory period, the stimulus needs to be even stronger for the threshold to be crossed. This is called the relative refractory period; a period during which the same stimulus is less likely to induce a action potential, compared to when it would arrive when the cell membrane patch was in its resting state. During the first part of the refractory period, corresponding to the time period in which the neuron is generating an action potential, it is not physically possible for the membrane to generate an second action potential. This period is called the absolute refractory period.

2.1.3 *Initiation and Transmission of Action Potentials*

Action potentials make up the communication units between different neurons, and they can be seen as the result of a local computation:

- the input consists of action potentials from other neurons the neuron at hand receives through its synapses. The chemical synapses are located at the end of its dendrites;
- the computation is the local integration of the signals through the dendritic tree and the soma. If the integrated signal reaches the axon initial segment (AIS) and is above the action potential initiation threshold, a new axon potential will be initiated at this site.
- This initiated action potential is the output of the system and can be transferred to the next neuron by traveling down the axon and its branches to the axon terminals and synapses.

The action potential initiated at the AIS can travel down the axon because it is regenerated: the currents resulting from the change in voltage spread across the axon and depolarize adjacent patches of the membrane, often (but not always) inducing a new action potential and thus continuing the signal propagation [5] [6]. Once the traveling action potential arrives at the end of the axon, it is potentially transmitted in a chemical synapse: a junction between the axon of the original neuron and the dendrites of the next neuron (often called the postsynaptic neuron), where, induced by the action potential, neurotransmitter molecules are released to bind on the receptors of the postsynaptic cell. This process is illustrated in Fig. 2.3.

[5] *Evidence for electrical transmission in nerve*. Hodgkin,1937.
[6] *Axon physiology*. Debanne et al.,2011.

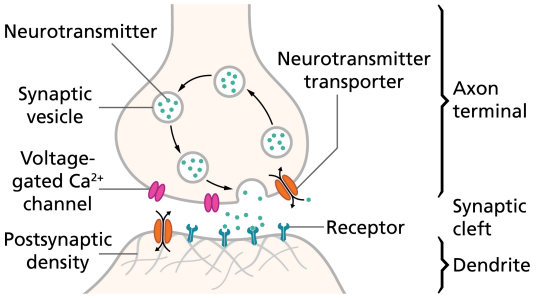


Figure 2.3: Schematic representation of a chemical synapse.

The binding of the neurotransmitters to the receptors of the postsynaptic cell again induces a change in the voltage of the postsynaptic cell membrane due to activation of certain voltage dependent ion channels. However, the binding might induce two opposite changes: if it increases the voltage (and thus depolarizes the cell membrane), the synapse is called excitatory; if the binding affects the voltage in the opposite way (polarizing the cell membrane further), the synapse is called inhibitory. Either way these changes propagate passively to nearby regions of the postsynaptic dendritic membrane as so-called post-synaptic action potentials. Unlike in the axon this process is not regenerative: typically the

signals decay exponentially with the distance from the synapse. But an excitatory signal might eventually reach the soma and the axon initial segment. If the voltage stimulus is still sufficiently large, a new action potential might be initiated at this site. It is far more likely, however, that different excitatory signals from presynaptic neurons have to arrive together to ignite an action potential at the axon initial segment. As we mentioned earlier, this can be seen as a kind of local computation; and the complexity of this computation is enhanced through the existence of inhibitory signals that can modulate the input-output relation even further (see Fig. 2.4).

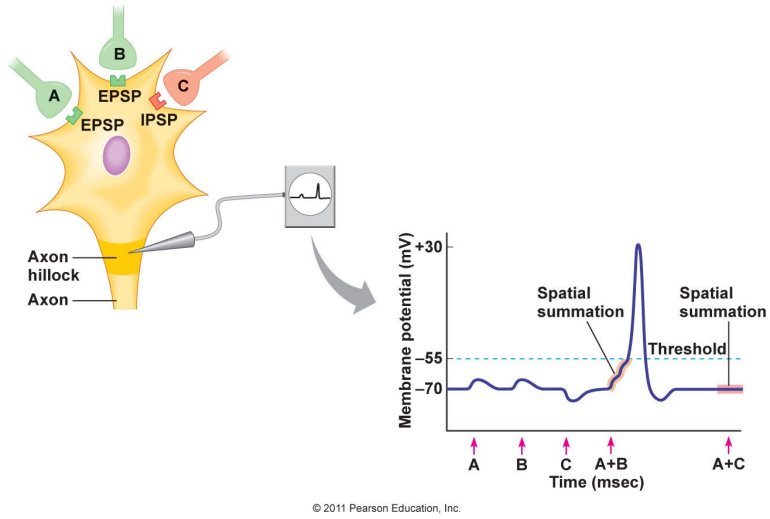


Figure 2.4: Schematic representation of excitatory and inhibitory synaptic connections, together with a representation of the post-synaptic potentials signals they generate. These post-synaptic potentials usually have to overlap significantly for the neuron to initiate a new action potential at the axon initial segment.

Chemical synapses are not the only way in which neurons could transfer their action potentials. A faster connection is achieved when the neurons have both their membranes connected in an electrical synapse: so-called connexons then allow the ionic current of the presynaptic action potentials to be directly transfered to the postsynaptic cell. A schematic representation of an electrical synapse is shown in Fig. 2.5. This process is much faster than the transferal in a chemical synapse, because there is no need for the (relatively) slow diffusion of neurotransmitters. It can only be formed, however, in cells that are close to each other: usually, the cells approach within 4-5 nm of each other, whereas 20 to 40nm separates the cells connected by a chemical synapse [7].

[7] *Principles of neural science.* Kandel et al.,2000.

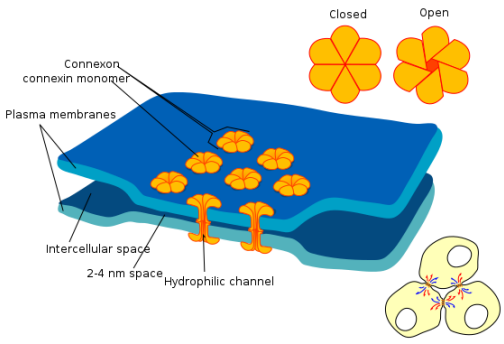


Figure 2.5: Electrical synapse.

2.1.4 *Overview of Typical Timescales and Dimensions*

The following overview of typical timescales and dimensions that will turn out to be important for interpreting our results are compiled from Kandel et al. [2000] [8] and Debanne et al. [2011] [9].

[8] *Principles of neural science*. Kandel et al.,2000.
[9] *Axon physiology*. Debanne et al.,2011.

Spatial Dimensions. Typically

- The soma has a diameter of **50 μm** or more.
- The axon has a diameter between **0.2 μm and 20 μm**.
- For a chemical synapse, the distance between cell pre- and postsynaptic cell membranes is between **20 nm and 40 nm**. These cell membranes are the membranes of the synaptic terminals of the axon and the dendrites (or soma) of the next cell.
- For an electrical synapse, the distance between cell pre- and postsynaptic cell membranes is between **4 nm and 5 nm**. These cell membranes are the membranes of the somas.

Transmission Speeds and Delays. Typically

- the electrical synaptic delay is **virtually absent**. The transmission is usually bidirectional.
- the chemical synaptic delay is **at least 0.3ms, and usually 1-5ms or longer**. The transmission is unidirectional.
- axonal conduction velocity depends on the axon diameter and on axon myelination (a structural property that greatly enhances conductance speed). For unmyelinated axons, the conduction velocity is typically **0.4 to 2.0 m/s**.

Amplitudes and Signal to Noise. Typically

- action potential amplitudes depend on the location in the neuron and the received external stimulus. Extra-cellular voltage recordings are typically **between 5 and 500 mV**.
- the action potential amplitude is highest in the axon initial segment (AIS), lower in the soma and in the rest of the axon, and very small in the dendritic structures.
- The signal to noise ratio naturally depends on the measuring device; in our setup (see chapter 5) the largest amplitudes are up to **180 σ_{noise}** while the axonal signals (not including the AIS) are usually **1-2 σ_{noise}**. Dendritic signals are unlikely to be detected or are buried in the noise.

Spike Rate Properties. Typically

- the absolute refractory period equals the duration of the action potential (**1ms**)
- the relative refractory period takes around **4-5ms**
- the spike rate of spontaneous activity is **a couple of spikes per second** (see below).

2.2 *Neuronal Variability, Rate Codes and Temporal Codes*

It has long been known that the response of single neurons as well as the response of neuronal networks to the exact same electrical stimulus is usually different from trial to trial [10] [11] [12] (see Fig. 2.6). In some cases, the spikes seem to happen in an almost random fashion, in the sense that the exact timing of one spike does not depend on the time of the previous spike [13]. In terms of neural encoding—the idea that the outside world, entering in the form of external stimuli, should somehow be represented in the activity of the neurons in order for it to be processed—this seems highly counterintuitive: if a structured stimulus results in random activity, how is the neural system then encoding this stimulus, and how could it possibly compute something about the outside world?

[10] *Reading a neural code*. Bialek et al.,1991.

[11] *A relationship between behavioral choice and the visual responses of neurons in macaque MT*. Britten et al.,1996.

[12] *Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey*. Bair and Koch,1996.

[13] apart from the fact that it has a reduced probability of occurring during the relative refractory period

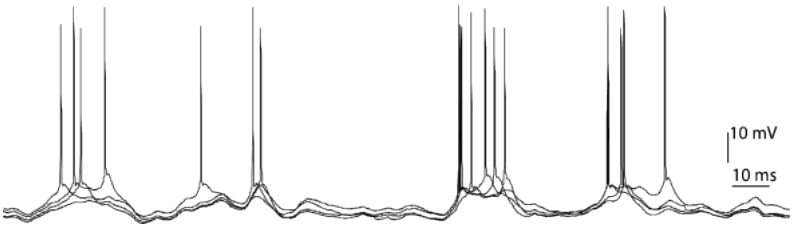


Figure 2.6: Neuronal variability: the exact same constant stimulus is applied to a single in vitro neuron in four different trials; the resulting spike trains are shown on top of each other. Source: Gerstner et al., 2014.

The common answer is that in this case most of the information seems to be captured not in the exact spike times, but in the rates at which the neurons fire those spikes. In other words, if someone would like to reconstruct the external stimulus from the neuronal dynamics, he or she would succeed better when using the spike rate following the stimulus, than when considering the exact timing of the spikes [14]. However, not all spiking happens completely randomly. Depending on the type, location and function of the neurons, different studies have shown that the timing of the spikes sometimes contains information as well [15,16]. The information about the stimulus is thus contained in a so-called temporal code, a rate code, or in a combination of both [17] [18].

If one considers the speed of computations, however, experiments seem to suggest that fast computations by networks of neurons in the cortex must be implemented through temporal codes [19] [20]. The arguments goes as follows: it has been shown that a single cortical area involved in visual processing can complete its computation in just 20-30ms

[14] *Spikes: exploring the neural code*. Rieke and Warland,1999.

[15] *Deciphering the spike train of a sensory neuron: counts and temporal patterns in the rat whisker pathway*. Arabzadeh et al.,2006.

[16] *Bursting neurons signal input slope*. Kepecs et al.,2002.

[17] *Spike timing and spike rate make complementary contributions to perceptual decisions in rat S1 and S2 cortex*. Zuo et al.,2015.

[18] *Neural coding: rate and time codes work together*. Seth,2015.

[19] *Processing speed in the cerebral cortex and the neurophysiology of visual masking*. Rolls and Tovee,1994.

[20] *Brain mechanisms for invariant visual recognition and learning*. Rolls,1994.

[21]. But the firing rates of neurons involved in these computations are usually below 100Hz (thus firing less than 0.1 times per millisecond). To determine the frequency of a completely regular process one needs at least two spike occurrences, meaning that at least 20ms need to pass before the exact spike rate of these neurons can be determined (in case of an irregular process, one evidentially needs more spikes to estimate the average rate). It is thus very unlikely that this fast computation can be achieved based on the information contained in the average spike rates of the neurons involved. However, as we will discuss in the next chapter, almost all of the artificial neural network models currently in use are based on spike rate implementations.

[21] *Biological constraints on connectionist modelling*. Thorpe and Imbert,1989.

2.3 *Connected Cultured Cortical Neurons*

We will now discuss some important results on the nature and activity of connected cultured cortical neurons as they have been described in scientific literature.

As we mentioned before, measuring the activity of cortical neural networks *in vivo* is a technically challenging task, and the complicated and dense structure of those networks makes it even more daunting (or even sheer impossible) to analyze their detailed activity patterns and the related fundamental network properties. Studying such networks in vitro alleviates this task greatly—but has the drawback of creating a perhaps unrealistic situation. However, it has been shown that cortical networks cultured in vitro are comparable to immature, developing neocortex networks in vivo [22] [23]. Studying developing networks in vitro can thus at least shed light on the interplay between early network *formation* and activity.

[22] *Spontaneous neuronal activity in developing neocortical networks: from single cells to large-scale interactions*. Luhmann et al.,2016.

[23] *Spontaneous neuronal discharge patterns in developing organotypic mega-co-cultures of neonatal rat cerebral cortex*. Baker et al.,2006.

There are in general three different types of in vitro networks one can obtain: animal *brain slices* will have an intricate, fully developed network structure that is more difficult to analyze, but is of course very close to the in vivo structure; *dissociated cultures* are obtained from (chemically) dissociating a premature or mature animal brain region of interest and extracting its neural cells [24]; and finally, as in our case, one can use a *stem cell culture*, where neural cells and their supporting glia cells are derived from induced pluripotent stem cells. We will denote the latter two categories as *cultured* neurons.

[24] *Dissociated cortical networks show spontaneously correlated activity patterns during in vitro development*. Chiappalone et al.,2006.

The cultured cells exhibit spontaneous activity, meaning that they produce spikes *even in the absence of external stimuli*. This is related to fluctuations in the membrane potential due to the sometimes random opening and closing of the ion channels [25] [Chow and White, 1996]. What is important is that this activity is typically of a relatively slow rate (a couple of spikes per second) and that it can be modeled as a random process with a 'generalized' relative refraction period: the exact spike time of

[25] *Limitations of the Hodgkin-Huxley formalism: effects of single channel kinetics on transmembrane voltage dynamics*. Strassberg and DeFelice,1993.

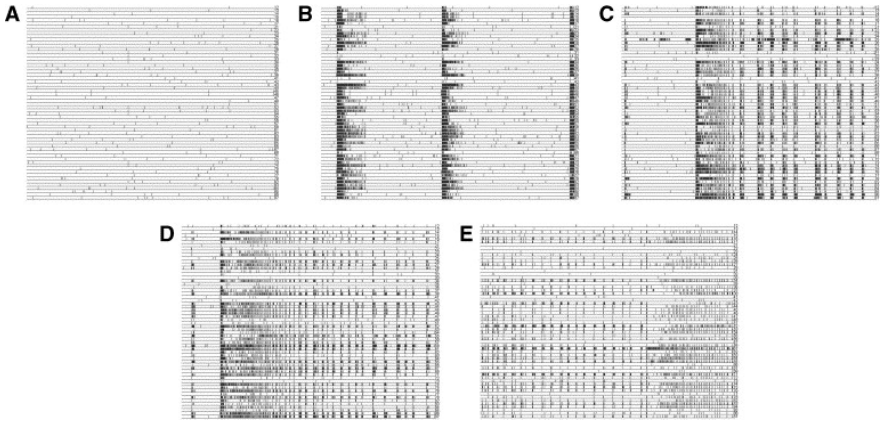


Figure 2.7: Raster plots showing the recorded spike trains of a developing dissociated cortical network at different days in vitro (DIV). Time runs horizontally, while the vertical axis corresponds to the different recorded neurons. (A) 7 DIV; (B) 14 DIV; (C) 21 DIV; (D) 28 DIV; (E) 35 DIV. Adapted from Chiapalone et al., 2016.

the next spike depends on the current spike time, in the sense that the probability of it occurring very quickly (within a range of around 100ms) after the current spikes are lower. The typical timescale of the 'original' refractory period as mentioned above, is around 5ms, hence why we call this property of the spontaneous activity rather a 'generalized' refractory period. Also, unlike for the original refractory period, this generalized refractory period does not imply that the probability of the neuron spiking following an external input signal is decreased (it merely describes the statistical dependence between its 'own' spikes, i.e., the spikes it generates spontaneously).

After a few days in culture, neurons start to connect to each other with functionally active synapses, forming a random network. Over time the networks shift from electrical to chemical transmission, while at the same time the axonal connectivity increases. At this stage the (dissociated) networks typically show repetitive *burst* discharges [26] [27] [28] [29], i.e., they show spike trains where a couple of spikes occur in very rapid succession, followed by relatively longer periods of 'silence'. These burst can be synchronized over large fractions of the culture. An example of these kinds of patterns developing over time in a network of cultured neurons is shown in Fig. 2.7.

The precise structure of these spatio-temporal patterns depends on a number of factors [30], including:

- the properties of the neurons, i.e., their structure, membrane properties, and their timescales of recovery and refraction,
- the flow of activity through the network, determined by the effective connectivity,
- the timescales of signal propagation, i.e., the axonal and synaptic transmission speeds and the timescale of dendritic integration,
- the balance between inhibitory and excitatory connections.

[26] *The mechanisms of generation and propagation of synchronized bursting in developing networks of cortical neurons.* Maeda et al.,1995.

[27] *Dissociated cortical networks show spontaneously correlated activity patterns during in vitro development.* Chiappalone et al.,2006.

[28] *Network dynamics and synchronous activity in cultured cortical neurons.* Chiappalone et al.,2007.

[29] *Spontaneous neuronal activity in developing neocortical networks: from single cells to large-scale interactions.* Luhmann et al.,2016.

[30] *Dynamics and plasticity in developing neuronal networks in vitro.* van Pelt et al.,2005.

Several models and simulations based on the above factors have provided insight in how this synchrony could arise [31] [32] [33] [34] [35] [36]. We want to emphasize two important facts: first, it is worth noting that the spatiotemporal patterns depend on the ratio between inhibition and excitation in the network; but the presence of inhibitory neurons is not a necessary condition for synchronization to emerge. Furthermore, it has been shown experimentally [37] that very young cultures exhibit a random topology, which over time evolves to a so-called small-world topology (i.e. the neurons start to form local clusters) [38]. This kind of network balances integration of network areas with segregation of specialized processing units, which increases the network efficiency [39]. The fact that this topology can arise without external stimulation points to the presence of intrinsic biological mechanisms that drive these developments.

[31] *Oscillations, complex spatiotemporal behavior, and information transport in networks of excitatory and inhibitory neurons.* Destexhe,1994.

[32] *The mechanisms of generation and propagation of synchronized bursting in developing networks of cortical neurons.* Maeda et al.,1995.

[33] *Inhibition can disrupt hypersynchrony in model neuronal networks.*. Deyo and Lytton,1997.

[34] *Emergent oscillations in a realistic network: the role of inhibition and the effect of the spatiotemporal distribution of the input.* Pauluis et al.,1999.

[35] *Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons.* Brunel,2000.

[36] *Changing excitation and inhibition in simulated neural networks: effects on induced bursting behavior.* Kudela et al.,2003.

[37] *Emergence of a small-world functional network in cultured neurons.* Downes et al.,2012.

[38] *Emergence of a small-world functional network in cultured neurons.* Downes et al.,2012.

[39] *Complex networks: Structure and dynamics.* Boccaletti et al.,2006.

2.4 *Overview of Important Properties*

In the next chapter, we will discuss the currently existing mappings between biological neural networks and artificial neural networks. We will summarize here the properties of biological neural networks, discussed above, that will be used in these mappings:

Single Neurons.

- integration and summation of post-synaptic potentials in the dendritic tree,
- the existence of a threshold to initiate action potential generation,
- the information about the input can be contained in the average rate of spikes that make up the output (rate code),
- the information about the input can be contained in the exact spike times (temporal code).

Connected Neurons.

- the output of a single neuron can be transferred to several other neurons through the axonal branches (a one-to-many relationship),
- similarly, the input of a single neuron can consist of an integration and summation of the received signals from several other neurons (a many-to-one relationship),
- significant transmission delays occur when the outputs of the neurons travel down the axons, are transmitted over chemical synapses, and travel as post-synaptic potentials across the dendritic trees.

3

Comparing Biological and Artificial Neural Networks

3.1 Commonly Used Artificial Neural Networks

3.1.1 Summation and Thresholding

In the first artificial neural networks, the artificial neurons were based on two core properties of biological neurons: the combining of incoming signals and the subsequent thresholding [1]. This first artificial neuron was coined the *perceptron* [2]. In modern notation, the perceptron amounts to the following operation:

$$f_{neuron}(\boldsymbol{x}) = \begin{cases} 1 & \boldsymbol{w}^T\boldsymbol{x} + b > 0 \\ 0 & otherwise. \end{cases} \tag{3.1}$$

where $\boldsymbol{x}$ is a vector of scalar inputs, $\boldsymbol{w}$ is a vector of weights, and b is called the bias term. The operation applied to the expression $\boldsymbol{w}^T\boldsymbol{x}$ is usually called the activation function, denoted σ; the activation function of the perceptron is visualized in Fig. 3.1.

[1] *A logical calculus of the ideas immanent in nervous activity*. McCulloch and Pitts,1943.

[2] *The perceptron: a probabilistic model for information storage and organization in the brain.*. Rosenblatt,1958.

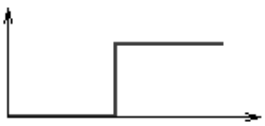


Figure 3.1: Shape of the perceptron activation function, shifted from the origin by the bias term b.

The linear weighted sum of the input, $\boldsymbol{w}^T\boldsymbol{x}$ forms the abstraction of the summation of incoming signals in the dendritic tree. Note that this definition reduces incoming signals to a vector of scalar values that arrive at exactly the same time. The thresholding, then, is captured by the fact that the output only equals 1 if the result of the weighted sum is larger than $|b|$ (if b is taken to be negative). b is thus related to the action potential initiation threshold, but might as well be set to a value that is not physically plausible.

As most artificial neural networks, networks of perceptrons are organized in layers of connected neurons. This is the natural abstraction of the many-to-many connections and the layered organization often found in biological neural networks. An example of a network with an input layer, an output layer, and a single hidden layer in between is shown schematically in Fig. 3.2. The neurons or nodes in each layer are

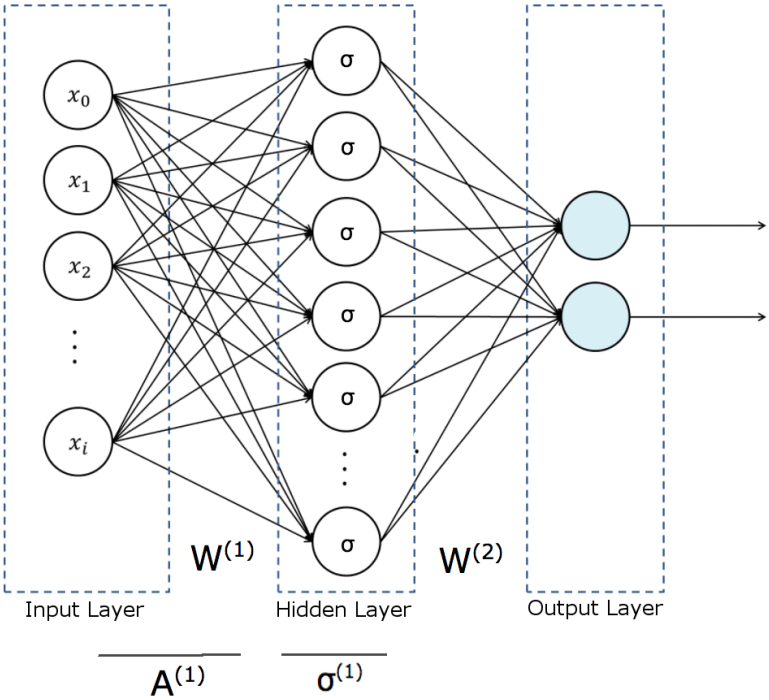


Figure 3.2: Schematic representation of an artificial neural network with a single hidden layer.

connected to the nodes of the next layer by a matrix of weights $\boldsymbol{W}^{(i)}$, consisting of the corresponding weight vectors of each individual neuron (Eq. 3.1) in the next layer. The output $\boldsymbol{f}(\boldsymbol{x})$ (a vector-valued function) of an n-layer network can thus be described as a chain of operations on the input $\boldsymbol{x}$:

$$\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{\sigma}^{(n)}\boldsymbol{A}^{(n)} \cdots \boldsymbol{\sigma}^{(2)}\boldsymbol{A}^{(2)}\boldsymbol{\sigma}^{(1)}\boldsymbol{A}^{(1)}\boldsymbol{x} \tag{3.2}$$

where $\boldsymbol{A}^{(i)}$ is the affine operation applied to the outputs of layer $i-1$ (denoted $\boldsymbol{x}^{(i-1)}$):

$$\boldsymbol{A}^{(i)} = \boldsymbol{W}^{(i)}\boldsymbol{x}^{(i-1)} + \boldsymbol{b}^{(i)}, \tag{3.3}$$

where $\boldsymbol{b}^{(i)}$ is a vector of the bias terms of each neuron in layer i, and $\boldsymbol{\sigma}^{(i)}$ a vector of operators, i.e., $\boldsymbol{\sigma}^{(i)}$ represents the activation function of each neuron in the layer i.

Multilayer perceptron networks and related models (such as Hopfield networks [3] and Boltzmann machines [4]) by definition only produce binary output; and, as we mentioned before, it is implied in the definition of the perceptron that all incoming signals arrive at the same time. However, perceptron networks can still be related to biological networks of spiking neurons in case the biological network exhibits synchronized behavior. For this we have to make an abstraction of action potentials with binary values—a value of 1 corresponds to the neuron spiking within a certain short time window, and a value of 0 corresponds to it being silent. If the firing patterns of all neurons that provide the input for a subsequent neuron are synchronized (up to a few ms), the incoming signals will thus arrive at the same time, and the computations in the biological network can in that case thus be modeled quite well by a

[3] *Neural networks and physical systems with emergent collective computational abilities.* Hopfield,1982.

[4] *A learning algorithm for Boltzmann machines.* Ackley et al.,1985.

perceptron network.

3.1.2 *Rate Coding*

We can now turn our attention to rate coding, i.e., the cases where information is encoded in the spike rate rather than in the actual spike times. To construct an artificial neural network that computes based on spike rates, we have to reinterpret the output of the artificial neuron as the instantaneous spike rate of the spike train it produces. In other words, the activation function σ is now a (usually monotonically increasing) function of the integrated input, outputting the instantaneous spike rate. A common choice for the activation function is the sigmoid function:

$$\sigma(x) = \frac{1}{1 + \exp^{-x}}, \tag{3.4}$$

which is a biologically realistic choice if one considers that that biological neurons, especially in higher cortical areas, have been shown to fire at various frequencies between their minimum and maximum spike rates. But other functions, such as the hyperbolic tangent or the so-called ReLu function, can (and often are) used as well. The latter activation functions are often better suited for the applied learning algorithms [5].

[5] *Deep Learning*. Goodfellow et al.,2016.

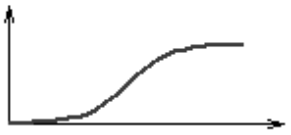


Figure 3.3: Shape of the sigmoid activation function, shifted from the origin by a bias term.

Almost all artificial neural networks commonly used today are based on the rate coding interpretation: e.g., the computed values in feed-forward neural networks, recurrent neural networks and convolutional neural networks should all be interpreted as instantaneous spike rate values when comparing them to biological neural networks. They thus have no intrinsic properties related to the exact timing of spikes, nor do they incorporate any notion of subsequent transmission delays (the simultaneous transmission between layers in an ANN could in some sense also be seen as a transmission delay, but we mean to describe processes where the inputs arrive at different lags.)

3.1.3 *Expressibility, Efficiency and Learnability*

Temporal properties such as transmission delays can give rise to very interesting and complex dynamical processes [6]. Still, the artificial neural networks commonly used today perform extremely well on a wide range of tasks, without incorporating spike timing or even a notion of transmission delay. Before we continue to describe artificial neural networks that incorporate temporal properties, it is perhaps a good moment to ask what we might potentially gain from doing this.

[6] *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Strogatz,2018.

We can discuss the general quality of a neural network using three important factors:

- **expressibility**: what class of functions (in terms of complexity) could the network possibly express?
- **learnability**: can the neural network learn the parameters of the functions it is meant to express, and if so, how fast does it converge?
- **efficiency**: how many resources does the network require, in terms of number of neurons and number of parameters, to express those functions?

The first question (expressibility) yields the same answer for all networks: *any* smooth function. It has been shown that artificial neural networks with nonlinear activation functions are in fact universal approximators [7] [8], which means that they can approximate any smooth function (see Eq. 3.2) to any desired accuracy—even with a single hidden layer. Thus, if the input data (such as a the numerical values corresponding to the pixels in a gray scale picture) can be transformed in a smooth way to a desired output (e.g., to a value that we can interpret as 'this is a picture of a cat'), there exists a network that implements that approximates this function. The accuracy of the approximation is related to the number neurons in the network. This universal approximation property is in itself very useful, but it is by no means the end of the artificial neural network story. The application of ANN can still fail due to inadequate learning, an insufficient number of neurons or simply the lack of a deterministic relationship between input and target. In short, all ANN are equal in theory, but in practice some are definitely more equal than others.

[7] *Approximation by superpositions of a sigmoidal function*. Cybenko,1989.

[8] *Multilayer feedforward networks are universal approximators*. Hornik et al.,1989.

In a practical situation we don't know beforehand the function we wish to approximate—which is of course exactly why we would want to use a technique like an ANN in the first place. Let's assume we at least know that the function embodies a deterministic relationship. The challenge of successfully applying ANN then lies in finding an efficient network architecture, and subsequently in using a suitable learning algorithm, that, starting from a set labeled examples, makes the network approximate the desired function. Based on the properties of the data we can usually determine beforehand which general type of network might be efficient. E.g., a convolutional neural network (CNN) [9] is well suited for tasks and data with what is called 'spatial translational invariance': if we wish to automatically detect a cat in a picture, it should not matter whether the cat is in the left upper corner or in the lower right one. In other words, if the object of interest is translated in space, the network should be able to detect it all the same (hence the term 'invariance'). The architecture of a CNN reflects this invariance and is thus the network of choice for image classification tasks—this is of course all but incidental, since its general architecture is inspired by the biological architecture and cell types in visual cortex of primates [10] [11].

[9] *Convolutional networks for images, speech, and time series*. LeCun et al.,1995.

Translational invariance is just one example of a property that can guide us in the search for efficient networks. In fact, it has been argued

[10] *Mathematical description of the responses of simple cortical cells*. Marĉelja,1980.

[11] *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex*. Hubel and Wiesel,1962.

[12] that, out of all possible functions, the functions that are of actual practical interest often exhibit certain properties such that they can be efficiently approximated by artificial neural networks. Such properties include symmetry (such as translational or rotational invariance), locality, and compositionality (i.e., the function describes a hierarchical process). The latter at least partly explains the success of deep neural networks—networks with multiple layers in between the input and output layer. A network reflecting the hierarchy of the process underlying the data needs less neurons than a network that has to approximate the same process with a single layer of neurons.

[12] *Why does deep and cheap learning work so well?*. Lin et al.,2017.

Apart from the efficiency, there is also the learnability: once a particular network type has been chosen (thus possibly inspired by the nature and structure of the processes that generated the data), a myriad of unknown values regarding actual architecture and network parameters remain to be determined before one arrives at a network that approximates the desired function well. For artificial neural networks it is usually the ubiquitous back propagation algorithm [13] [14] that lies at the heart of this search process. E.g., in the case of the successful CNN developed by Krizhevsky et al. [15], it guided the search towards a network with 8 layers, 500 000 neurons, and suitable values for a staggering 60 million network parameters. The result was a network that could classify the 1.3 million images of the benchmark ImageNet training set in 1000 different categories with a state-of-the-art performance. This example in particular shows once again the interplay between artificial intelligence and neuroscience research: the convolutional neural network is a successful abstraction of the visual processing systems in biological brains, and by studying these abstractions we gain insight in why the biological functions can be performed more efficiently when the visual system is organized in this 'deep', layered structure.

[13] *The roots of backpropagation: from ordered derivatives to neural networks and political forecasting*. Werbos,1994.

[14] *Artificial Intelligence (A Modern Approach)*. Russell and Norvig,2010.

[15] *Imagenet classification with deep convolutional neural networks*. Krizhevsky et al.,2012.

Comparing learnability and training in biological and artificial neural networks is a very interesting topic, but we will not further discuss it within the scope of this thesis. Instead, we will concentrate on the gains in efficiency that might result from using time as a resource for communication and information in artificial neural network models. Such models, thus based on exact spike times, are usually simply called *networks of spiking neurons* [16]. We will describe these models in the next section.

[16] *Networks of spiking neurons: the third generation of neural network models*. Maass,1997b.

3.2 Temporal Coding and Networks of Spiking Neurons

3.2.1 Networks of Spiking Neurons

We discussed before how the timing of computation steps is "trivialized" in the commonly used artificial neural networks: in the first models based on perceptrons, some form of synchrony is assumed; in the more

recent models of ANN, asynchronous and stochastic spike timings are replaced by instantaneous spike rate.

Our goal here is to describe the simplest deterministic model of a spiking neuron that (a) retains the properties of summation and thresholding—thus weighted sums and activation functions—but (b) outputs single spikes and incorporates the notion of transmission delays. We will base our discussion on the results presented by Maass, 1997b [17].

[17] *Networks of spiking neurons: the third generation of neural network models*. Maass,1997b.

(a) can in this case be achieved by keeping track of a value P_i for each neuron i. This value symbolizes the membrane potential at the axon initial segment. We will say the neuron i spikes whenever its potential value P_i reaches a certain threshold value θ_i. (So far, this is entirely equivalent to the assumptions made in the perceptron model.) The potential value P_i is still a weighted sum of incoming signals , but instead of representing these incoming signals by scalar values, we will now represent them by so-called response functions (which achieves (b)). These response functions will be denoted with $k(\tau)$; in essence, they embody the change in the potential value of the considered neuron resulting from a spike generated by another neuron at some time in the past.

Suppose we are at a time t, and we wish to obtain the contribution to P_i of a incoming signal resulting from another neuron j (connected to our currently considered neuron i) spiking at a time s in the past. This contribution can be modeled by the term $w_{ij}k_{ij}(t-s)$ (with $t > s$). Biologically plausible response functions for excitatory and inhibitory connections are shown in Fig.3.4. The shape of these functions can be interpreted in the following way: for times t shortly after the spike time s, the action potential initiated at neuron j is still being transferred across the axon, synapse and post-synaptic structures; until a signal arrives at the axon initial segment of neuron i, the response function is thus zero. The onset of the response function corresponds to the moment the signal arrives at the axon initial segment and starts contributing to a change in the membrane voltage, potentially leading up to the initiation of a new action potential, or spike, at neuron i (see Fig. 2.4). The sharp rise and slower decline of the response functions correspond to the shape of post-synaptic action potentials in real biological neurons.

The weight w_{ij} corresponds as before to the strength of the synaptic connection. In the context of learning, this weight can be replaced by a function $w_{ij} = w_{ij}(t)$, but for this discussion we will use only constant values w_{ij}. When considering both inhibitory and excitatory connections, we can take the values of w_{ij} to be always positive (or, equivalently, we can only consider excitatory connections and use real values for the weights).

To obtain the output of a neuron i, we thus need to consider the past spike times of all its presynaptic neurons. These values need not be saved indefinitely, however; naturally, only the spikes that happened up

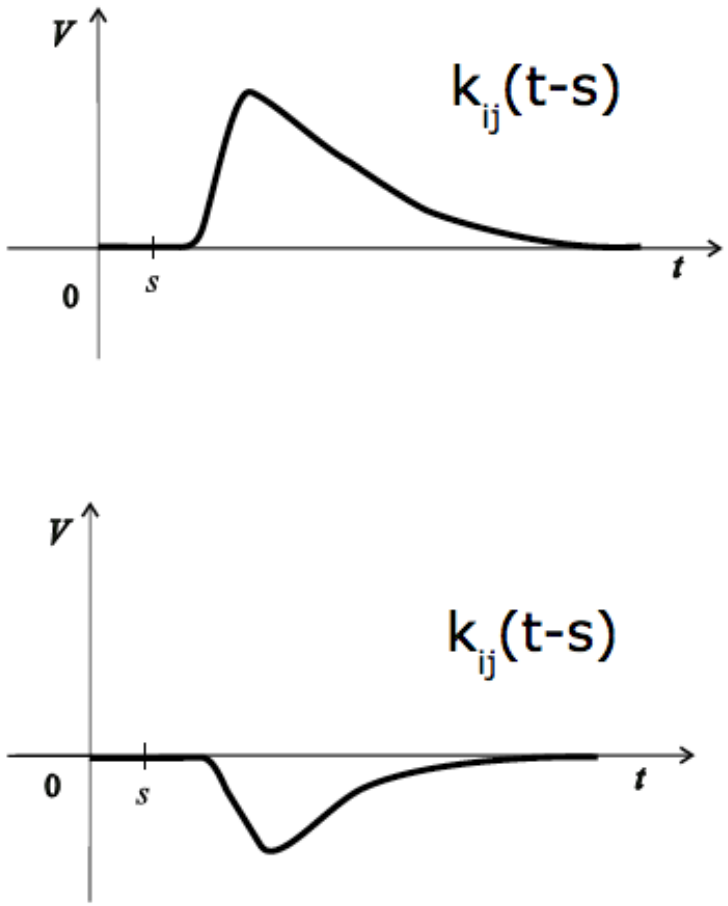


Figure 3.4: Schematic representation of biologically plausible response functions for excitatory connections (upper graph) and inhibitory connections (lower graph). Adapted from Maass, 1997.

to some time τ in the past should be taken into account, and τ roughly corresponds to the interval over which the response function is non-zero. We will denote this collection of past spike times $s_{j,l}$ (where j the number of the neuron and l the number of a spike) simply as ST_{past}. We can now rewrite the equation for the perceptron (Eq. 3.1) to describe the building blocks of networks of spiking neurons with transmission delays:

$$f_i(t, ST_{past}) = \begin{cases} 1 & \sum_j \sum_l w_{ij} k_{ij}(t - s_{j,l}) + \theta_i > 0 \\ 0 & otherwise. \end{cases} \tag{3.5}$$

where the summations run over the number of other neurons in the network ($\sum_j$) and the number of spikes for each of this neuron up to a time τ in the past ($\sum_l$). Note that this equation actually models more than just transmission delays: a transmission delay in itself could be modeled by a response function with a single sharp peak at a time t corresponding to the delay. These response functions would hardly overlap, however, and the threshold would thus never be crossed due to a combination of signals of different presynaptic neurons. The finite extent of the response function thus makes sure that the weighted sum is actually a meaningful

operation. Another aspect is the actual shape of the response functions: to model merely transmission delays, we could use the same response function for each connection, just shifted on the time-axis according to the typical transmission delay for the given connection.

Finally, these abstract spiking neurons can again be connected in layered structures with an input, an output, and a number of hidden layers in between. The input (output) of the network is then encoded in the spike times of the input (output) neurons. E.g., one could use a linear temporal encoding scheme: one can pick a reference time T and a constant c, and the encoding of a real value a is then given by a spike occurring at $T - ac$ [18].

[18] *Pattern recognition computation using action potential timing for stimulus representation*. Hopfield,1995.

3.2.2 *Expressibility and Efficiency*

We will now summarize the theoretical results on the expressibility and efficiency of the networks of spiking neurons as described in the previous section. As an aside, these theoretical results are derived with the assumption that the response functions are continuous and piece-wise linear—but a biologically plausible form can be well approximated, however, by 4 to 5 of such linear segments.

A useful measure of the expressibility of a neural network model with a finite and fixed number of neurons is its corresponding *VC dimension* [19]. The definition of this measure is based on the concept of 'shattering': a model is said to shatter a set of labeled data points, if, for all assignments of labels to those points, there exists a parameter value for the model such that the model correctly evaluates the data. To give a concrete example: a linear function will always be able to solve a problem involving just two data points, whatever the assigned labels; however, depending on the labels, it might not be able to correctly solve a task with three data points. The VC dimension of a model is the maximum number of points that the model can shatter (for a linear function, the VC dimension is thus 2). A first theoretical result compares the expressive power of an artificial spiking neuron with n variable *weights* to the same neuron using n variable *time delays*: variable weights yield a VC dimension of $O(n)$, while variable time delays achieve a VC dimension of $O(n \log(n))$ [20]. In the context of spiking neurons and temporal coding, it can also be shown [21] that the *weights* of the connections are able to play the same role as those of computational units in conventional neural networks.

[19] *An overview of statistical learning theory*. Vapnik,1999.

[20] *On the complexity of learning for spiking neurons with temporal coding*. Maass and Schmitt,1999.

[21] *Networks of spiking neurons: the third generation of neural network models*. Maass,1997b.

An important result is that all networks of spiking neurons are universal approximators [22]. They turn out to be *at least* as computationally powerful as conventional artificial neural networks of the same or similar size, provided the excitatory (inhibitory) response functions have a small linearly increasing (decreasing) segment. The proof of this prop-

[22] *An efficient implementation of sigmoidal neural nets in temporal coding with noisy spiking neurons*. Maass,1995.

erty relies on a correspondence between these linear segments and the approximately linear segment also found in a sigmoid activation function [23]. For some specific example tasks, such as coincidence detection and checking element distinctness, networks of spiking neurons are proven to be far more powerful (or at least efficient) than their conventional rate coding counterparts: where *a single* spiking neuron can compute element distinctness over its n different inputs, a sigmoidal neural net needs at least $\frac{n-4}{2} - 1$ hidden units for this task [24].

[23] *Bounds for the computational power and learning complexity of analog neural nets.* Maass,1997a.

[24] *Networks of spiking neurons: the third generation of neural network models*. Maass,1997b.

4

Spike Sorting Algorithms

4.1 Multi Electrode Array Recordings at High Resolution

Details of the recording device can be found in [Müller et al., 2015]. Here we summarize some if the important properties: the complete sensor array covers a region of $3.85x2.10sq - mm$ with 26400 electrodes, each covering a region of $9.35.45sq - \mu m$. Due to constraints in the electronics implementation, we can only read out from 1024 electrodes at once. However, we can determine which electrodes to record from, and the configuration is changed in a matter of seconds. The data at a rate of 20K Hz; recordings of several minutes thus have a considerable size in terms of data (a couple of Gb per file).

4.2 Identifying Neuronal Structures

In order to compare our biological network with networks of artificial spiking neurons, we first have to reconstruct the biological neural network from its recorded activity. The data we start from is a multivariate time series: it consists of the changing voltages recorded by 1024 electrodes located in the grid of the multi-electrode array (MEA). The neuronal structures naturally do not adhere to the grid structure of the MEA. The first step in thus to 'disentangle' the data: each sensor might receive signals from different neurons, and each neuron might be inducing signals in different sensors. Our final goal is to extract from this time series for each sensor a time series for each neuron, the latter consisting of the series of action potentials the neuron produced at its axon initial segment.

This process, which goes by the name of *spike sorting*, has been the focus of many research efforts in computational neuroscience [1] [2] [3]. Feasible due to the low dimensionality of the recorded data, this process was for a long time mainly performed manually. But recently the advent of

[1] *A review of methods for spike sorting: the detection and classification of neural action potentials*. Lewicki,1998.

[2] *Past, present and future of spike sorting techniques*. Rey et al.,2015.

[3] *Spike sorting for large, dense electrode arrays*. Rossant et al.,2016.

recording devices with up to hundreds of electrodes has urged the development of semi-automated to fully automated algorithms to tackle this problem. Below, we will first discuss the general approach of spike sorting algorithms. We will subsequently provide some more detailed information on the specific algorithm we used.

Let's start by considering a single neuron. The axon initial segment and soma usually produce spikes with a high amplitude, with resulting signals being recorded in the sensors directly covering these structures, as well as in other sensors nearby. But an action potential could travel through the other parts of the axon and all its branches as well, and the activity of a single neuron might thus potentially leave traces across a large part of the sensor array. We can also switch perspectives, and look at the problem from the viewpoint of a single sensor: the same sensor might receive signals from different nearby neurons. This is of course especially the case when the sensor covers an area where neuronal structures are dense.

Spike sorting is usually performed in a setting with relatively low recording resolution, i.e., each sensor or electrode records signals from multiple neurons, and it is very likely that only the high amplitude signals are recorded (i.e., those originating from the action initial segment). In case the spacing between the electrodes is large, such that the probability of one neuron being detected by more than one electrode is very small, the spike sorting problem simplifies further: the detected spikes on each electrode can be assigned to the neuron that produced them based on the spike shape. This is illustrated in Fig. 4.1.

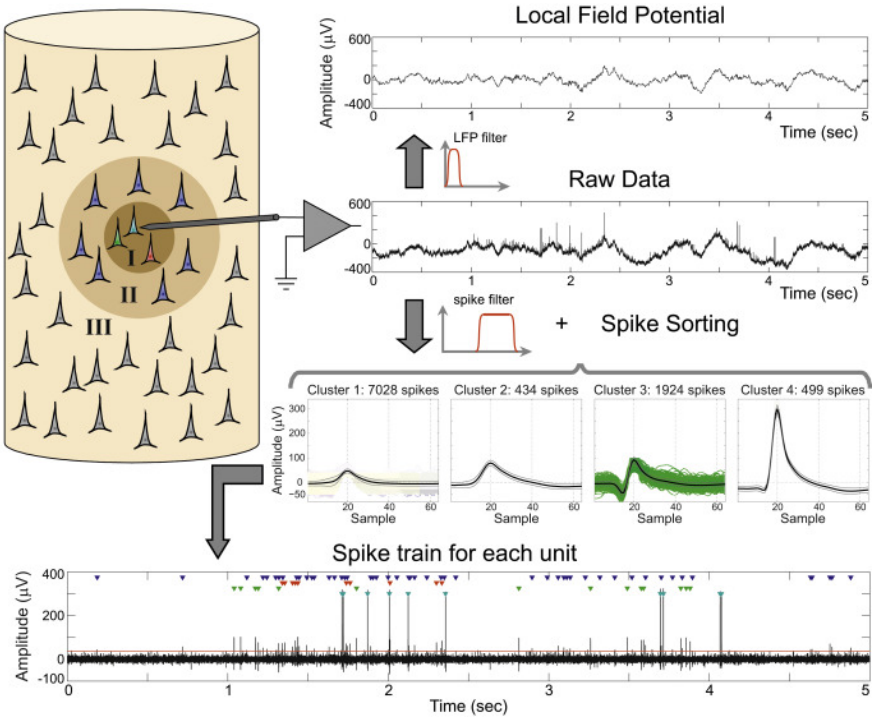


Figure 4.1: Illustration of the spike sorting process used for an in-vivo experiment, where the ratio of electrodes to neurons is usually relatively small. A schematic representation of the spikes in the neighborhood of the considered electrode is given on the left. In this setup, only the spikes from neurons in zone I are close enough to the electrode to be reliably discerned, the other neural activity contributes to the local field potential (an effect not present in our in vitro setup). Once the spikes have been detected in the signal, a classification based on the shape reveals which spike times are associated to each neuron (shown at the bottom trace of the figure.) Source: Rey et al, 2015.

Once all the spikes in the signal have been detected, the (usually correct [4]) assumption is that the different neurons produce spikes with a shape and amplitude typical to that neuron (the latter being only partly due to their difference in distance to the electrode). They can thus either be manually assigned to a neuron; or the shape of the spike or action potential can be used as input to a classification algorithm.

[4] *Axon physiology*. Debanne et al.,2011.

In a next step, we can consider the fact that the high amplitude signals originating from the AIS could possibly be recorded in multiple electrodes. This is due to the electrical signal spreading from where it originated; it is thus not always the case that the signal will be recorded at exactly the same instant in each of the neighboring electrodes. The goal can be, perhaps somewhat vaguely, formulated in this way: when a neuron spikes, we want to be able to say what the resulting measurement typically 'looks like' in space and time, i.e., we want to determine a spatio-temportal footprint for each neuron. These spatio-temporal footprints can then be matched to the original data in order to determine the spike times for each neuron.

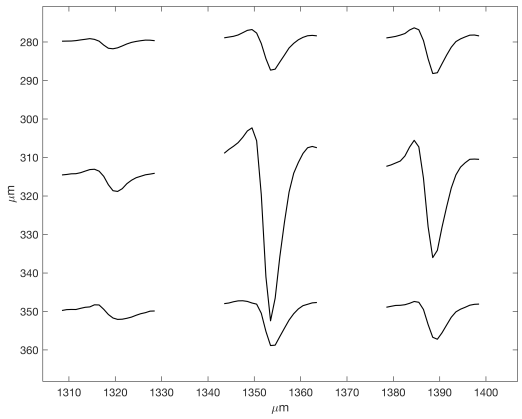


Figure 4.2: Example of a spatio-temporal neuron footprint: the same action potential is detected in multiple neighboring electrodes.

Due to the very high resolution of our recordings, we are, unlike in other spike sorting settings, able to pick up signals from different structures of the same neuron. We deliberately lowered the spike detection threshold to also detect these signals, because the location of the axons provides us valuable information that we could potentially use to validate our results on the effective connectivity of the network. This lower threshold unfortunately also yields a multitude of false detections (i.e., detection of noise).

4.3 *Automated Spike Sorting Algorithms*

We will now discuss the specific, fully automated algorithm we applied to obtain the spike sorted results. The complete spike sorting toolbox is called *Spyking Circus* and is developed by Yger et al.[5]; an open source python implementation is available on GitHub [6]. Given the large-scale

[5] *A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo*. Yger et al.,2018.

[6] https://github.com/spyking-circus

data, it was important to choose an algorithm that could execute the spike sorting in a fully automatic and efficient way; spyking circus is such an algorithm, and its core parts have moreover been designed such that they can be executed in parallel.

Data and Filtering

The input data consist of up to 1024 time series of voltage values sampled at 20 000Hz. Each time series has an associated location, namely the coordinates of the corresponding electrode in the grid. In the first step of the algorithm we want to determine the spike times, where a spike time is defined as the time at which the action potential reaches its maximum (see the typical shape of an action potential, 2.2). The time series data not only consists of action potentials though: there are both low- and high frequency noise components presents, stemming from the electronics and background neuronal processes. Our task of sorting the spikes is made much easier if we apply a bandpass filter to the data first. We choose the filter such that only the frequency components between 500Hz to 7000Hz remain (this is an approximation, in reality the used filter applies a smoother threshold [7]). This filtering removes the mean and the slow wave components in the signal, such that the signal is in essence "flattened out"; in this way, the amplitude of the spikes is no longer influenced by the slow wave components. An example snippet of half a second of bandpass filtered data is shown in Fig.4.3.

[7] *Introduction to digital signal processing and filter design.* Shenoi,2005.

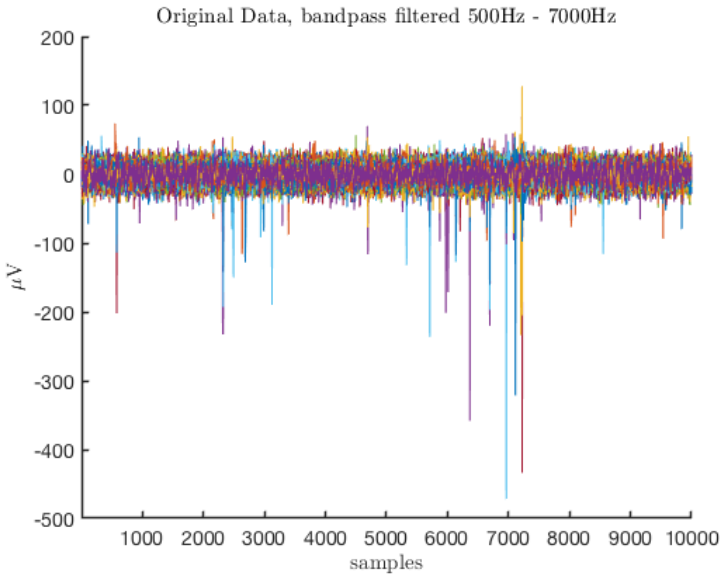


Figure 4.3: Example snippet of the time series data. All channels are displayed on top of each other with different colors.

Zooming in to one of the larger peaks (in terms of minima) for one of the channels in Fig.4.3, we can see the typical shape of an extra-cellular recorded action potential (shown in Fig. 4.4). Since we are recording outside the cells (so-called extracellular recordings), the measured potential differences across the cell membrane during an action potential are in fact of the opposite sign as compared to the standard description of the

potential values (which are defined with respect to voltages *inside* the cell). Spike times in our case thus correspond to local minimal values of the time series data.

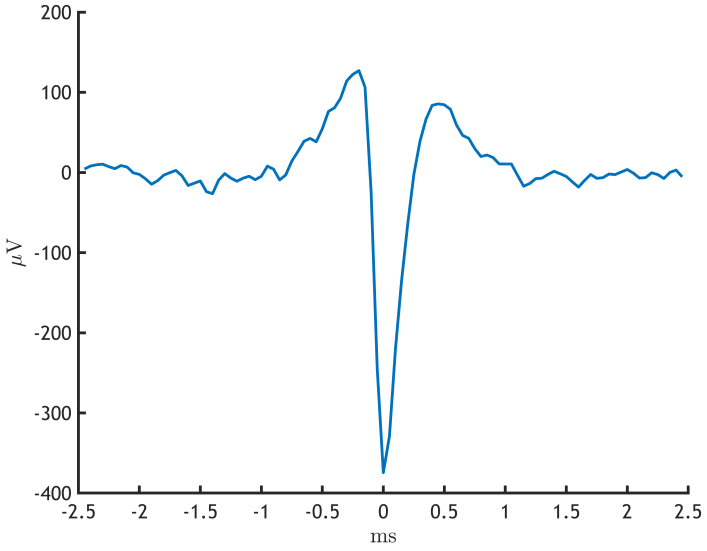


Figure 4.4: Example of an action potential recorded with a very high signal-to-noise ratio.

Spike Detection

Not all local minima are spikes, however; in order to separate the spikes from what remains of the noise after filtering, a threshold needs to be applied. Since the signal-to-noise ratio in each channel is different due to the different qualities of the used electrodes, a different threshold value should be calculated per channel. Usually the median absolute value is used:

$$MAD(\boldsymbol{x}) = median(\mid x_i - median(\boldsymbol{x}) \mid) \tag{4.1}$$

which is a measure similar to the variance. The use of the median makes it more robust, however, for time series data with large outliers—and in our case, we naturally have a lot of outliers in the form of spikes.

The threshold for timeseries i, θ_i, above which local minima are considered spikes is then given by:

$$\theta_i = \lambda MAD_i \tag{4.2}$$

where the scalar λ is a fixed parameter used for all channels. Usually a value between 6 and 7 is chosen. Once the thresholds have been calculated for each channel, it is straightforward to determine the spike times as the local minima that are below the threshold values.

Clustering

Not all detected spikes stem from different spiking 'events'. As we noted before, when a part of a neuron produces an action potential, a spike is usually detected in different electrodes covering the neighborhood of the neuronal structure (see Fig. 4.2). Given a maximal radius in space and a maximal period in time, the algorithm will determine the electrode where the spiking amplitude was maximal (in absolute value). For the time being only the spike time on this electrode will be used.

After this first step, we obtained for each electrode thus a list of spike times for spike events that were found to be maximal on this electrode. The next step is the application of a clustering algorithm based on the shape of the spikes: this will reveal the different groups of spikes, putatively stemming from different neurons.

First, a snippet of data centered on each saved spike time is extracted to obtain the shape of the spike. We set the length of this snippet to 20 samples, or 1ms (the length of a typical action potential). Examples of such waveforms are shown in gray in Fig. **??**. The dotted lines show the obtained threshold values for the electrode at hand.

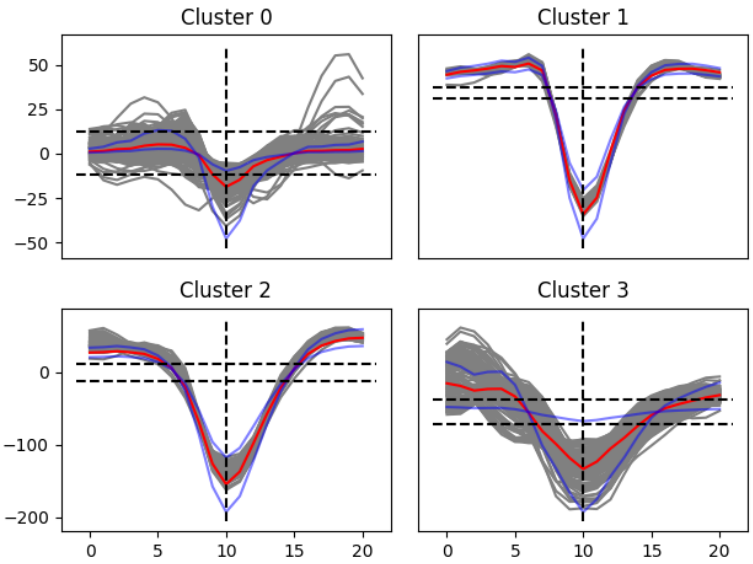


Figure 4.5: Distinctive groups of waveforms collected on a single electrode. In gray the original waveforms, in red the average waveform per group, in blue the standard deviations, and the dotted lines represent the threshold associated to the electrode.

These 20 values for each waveform could be considered as the features for the clustering algorithm; however, it pays of in terms of both computational resources and algorithm performance to apply a feature selection method first. The implemented method is a straightforward principle component analysis (PCA) [8] where the first 5 components are retained. The clustering algorithm is then applied, for each electrode, in the resulting 5 dimensional space where each data point consists of the 5 first principle components of a waveform associated to that electrode.

[8] *Principal component analysis.* Jolliffe,2011.

The clustering algorithm in itself is based on the idea that the cluster

centers are characterized by a higher density than their neighboring points, and that they at the same time are at a relatively large distance from points with higher densities [9]. In the first phase of the clustering algorithm, for each data point x_i, the average distance ρ_i to the S nearest neighboring data points is calculated. $\frac{1}{\rho_i}$ is then used as a proxy for the density. Then, again for each data point x_i, the minimal distance to any other point with a higher density is calculated. This distance is denoted δ_i. The data points can then be ranked based on the ratio $\frac{\delta_i}{\rho_i}$, i.e., the ratio of their minimal distance to a point of higher density (for a center, this should be a large value) to their own density (for a center, this should also be a large value, thus a small value ρ_i). The highest ranked points are putative cluster centers. The number of such putative centers selected is a parameter of the algorithm, naturally related to the maximum number of neurons that could potentially have their maximal amplitude spikes on the same electrode. Given the large electrodes to neurons ratio in our experimental setup, this number of putative cluster centers can be chosen relatively low. Finally, the remaining points can be assigned iteratively to the clusters: starting from points with a lower value ρ_i, each point is assigned to the same cluster as the closest point with a lower value $\rho_{i'}$.

[9] *Clustering by fast search and find of density peaks*. Rodriguez and Laio,2014.

The clustering is thus always performed with a fixed number of clusters. An additional step is added to the process to alleviate this constraint: when the normalized distance [10] between two resulting clusters is smaller than a predefined threshold value σ, the clusters are merged. Example results of this clustering process are shown in Fig.**??**.

[10] *A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo*. Yger et al.,2018.

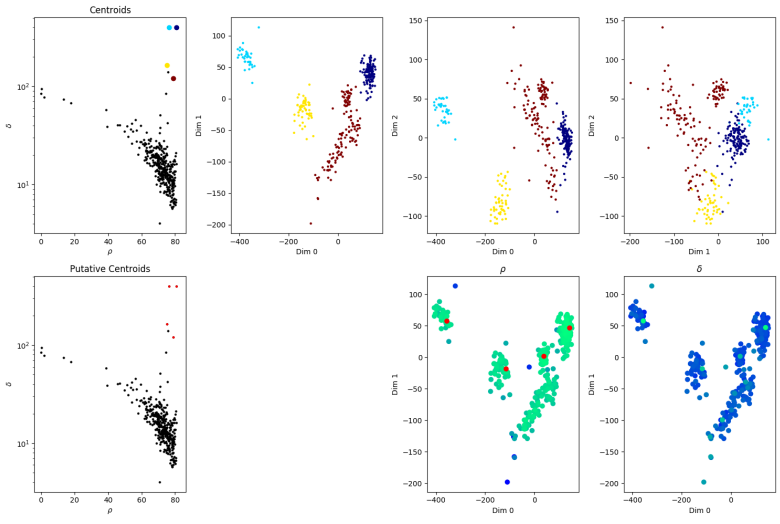


Figure 4.6: Example of the clusters corresponding to the waveforms in Fig. 4.5. The subfigures on the left show the putative centroids (second row) and the finally selected centroids after the merging (upper row). The clusters are shown in PCA for combinations of the first three dimensions in the colored subfigures of the upper row. Finally, the colored subfigures on the second row encode the ρ and δ values for each data point.

Template Extraction and Template Matching

The maximal amplitude spikes have now been assigned to a cluster representing a putative neuron (or, in fact, a neuronal structure). This is a computationally very expensive procedure, depending on the number of

detected spikes. But, one might argue that only a subset of these spikes will be needed to make up a discernible cluster. The spyking circus algorithm takes exactly this approach: it reduces the needed computational power by using only a fraction of all spikes for the clustering phase. Of course, our final goal is still to obtain an list of all spike times for each putative neuron (or neuronal structure). This goal can still be achieved by using the extraction of so-called templates for each cluster/neuron, and by subsequently matching these to the original time series data.

A spatio-temporal template for a neuron can be constructed in the following way: at all the spike times associated to the waveforms belonging to a single cluster, we extract a snippet of the data centered on this spike time for *all* the channels. Averaging these snippets over all spike times associated to the cluster yields a spatio-temporal template of what a spike event of this neuron typically 'looks like' in space and time, i.e., what the typical voltage patterns are in all channels in a short period before, during and after the spike of this neuron. An example is given in Fig. 4.2, where we have shown the template for all electrodes in the sensor array where it is non-zero. Finally, these templates can be matched to the original data by computing the scalar products of these (normalized) templates and windows of the data centered at all the detected spike times. The spike time is then attributed to the best matching template.

Overview of the Used Parameters

```
[detection]
radius         = 50       # Radius [in um] (if auto, read from the prb file)
N_t            = 1          # Width of the templates [in ms]
spike_thresh   = 4          # Threshold for spike detection
peaks          = negative   # Can be negative (default), positive or both
matched-filter = False      # If True, we perform spike detection with matched filters
matched_thresh = 4          # Threshold for detection if matched filter is True
alignment      = True       # Realign the waveforms by oversampling


[filtering]
cut_off        = 500, 7000  # Min and Max (auto=nyquist) cut off frequencies for the band pass butterwo
filter         = True       # If True, then a low-pass filtering is performed
remove_median  = False      # If True, median over all channels is substracted to each channels (moveme


[whitening]
chunk_size     = 10         # Size of the data chunks [in s]
safety_time    = 0.5          # Temporal zone around which templates are isolated [in ms, or auto]
```

```
temporal       = False      # Perform temporal whitening
spatial        = False       # Perform spatial whitening
max_elts       = 1000      # Max number of events per electrode (should be compatible with nb_elts)
nb_elts        = 0.8        # Fraction of max_elts that should be obtained per electrode [0-1]
output_dim     = 5          # Can be in percent of variance explain, or num of dimensions for PCA on wa

[clustering]
extraction     = median-raw # Can be either median-raw (default), median-pca, mean-pca, mean-raw
safety_space   = True       # If True, we exclude spikes in the vicinity of a selected spikes
safety_time    = 0.5       # Temporal zone around which templates are isolated [in ms, or auto]
max_elts       = 5000      # Max number of events per electrode (should be compatible with nb_elts)
nb_elts        = 0.02        # Fraction of max_elts that should be obtained per electrode [0-1]
nclus_min      = 0.01      # Min number of elements in a cluster (given in percentage) [0-1]
max_clusters   = 4         # Maximal number of clusters for every electrodes
nb_repeats     = 3          # Number of passes used for the clustering
smart_search   = True       # Activate the smart search mode
smart_select   = False      # Experimental: activate the smart selection of centroids (max_clusters is
sim_same_elec  = 5          # Distance within clusters under which they are re-merged
cc_merge       = 0.90      # If CC between two templates is higher, they are merged
dispersion     = (5, 5)     # Min and Max dispersion allowed for amplitudes [in MAD]
noise_thr      = 0.8        # Minimal amplitudes are such than amp*min(templates) < noise_thr*threshold
remove_mixture = True       # At the end of the clustering, we remove mixtures of templates
make_plots     = png       # Generate sanity plots of the clustering [Nothing or None if no plots]

[fitting]
chunk_size     = 1          # Size of chunks used during fitting [in second]
gpu_only       = False      # Use GPU for computation of b's AND fitting [not optimized yet]
amp_limits     = (0.3, 5)   # Amplitudes for the templates during spike detection [if not auto]
amp_auto       = True       # True if amplitudes are adjusted automatically for every templates
max_chunk      = inf        # Fit only up to max_chunk
collect_all    = False      # If True, one garbage template per electrode is created, to store unfitted



[extracting]
safety_time    = 1          # Temporal zone around which spikes are isolated [in ms]
max_elts       = 1000       # Max number of collected events per templates
output_dim     = 5          # Percentage of variance explained while performing PCA
cc_merge       = 0.975      # If CC between two templates is higher, they are merged
noise_thr      = 0.8        # Minimal amplitudes are such than amp*min(templates) < noise_thr*threshold
```

5

Effective Connectivity

Eventually, we wish to obtain insight in the effective connectivity of the network: which connections are used, their transmission delays, and finally the direction of information flow between the active neuronal structures in the network. So far, we explained how to identify these neuronal structures from the data through a spike sorting process. In this chapter we discuss how we can use a time series prediction technique that goes by the name of *autoregressive modeling* to obtain the properties of the effective connectivity.

In section 5.1, we will first discuss a general framework to model the causal relationships between a set of coupled time series. We will then show how, based on a general assumption of gaussianity, we can obtain the parameters of this framework from the computed auto- and cross-correlation functions. Finally, we will explain how these parameters can be converted into a directed graph; and we will also discuss how this graph might deviate from the actual flow of information in the network.

5.1 Autoregressive Kernels

5.1.1 Autoregressive Processes

A multivariate, higher order autoregressive process is a process that can be written in the form [1]:

$$\boldsymbol{x}(t) = \sum_{k=1}^{p} \boldsymbol{A}^{(k)} \boldsymbol{x}(t-k) + \boldsymbol{E}(t), \tag{5.1}$$

where $\boldsymbol{x}(t)$ is a vector of size n that represents the values of the n time-series at time t; p is the order of the model; $\boldsymbol{A}^(k)$ is the matrix of size nxn with entries that relate $\boldsymbol{x}(t)$ to the values of $\boldsymbol{x}$ at distance k in the past, and $E(t)$ is a diagonal matrix adding Gaussian noise at step t. Note that $A^{(k)}$ is a constant matrix that does not depend explicitly on time.

[1] *An introduction to multivariate statistical analysis*. Anderson,1958.

The elements $A_{ij}^{(k)}$ tell us how the values $x_j(t-k)$ influence the value $x_i(t)$. In words, Eq. 5.1 tells us that the current value of a timeseries in the network can be obtained as a linear combination of the past values of all the timeseries, including its own past. The coefficients $A_{ij}^{(k)}$ capture a causal relationship in the sense that they reveal which events in the past are relevant for predicting the current state, given all relevant information (i.e., given the past values of *all* the timeseries up to p timesteps in the past).

If we consider $A_{ij}^{(k)}$ for each time k, we obtain a so-called autoregressive kernel which we will denote $A_{ij}^{(t)}$. Convolving this function with the past of $x_j(t)$ gives the contribution of this j^{th} timeseries to the present value of the i^{th} timeseries, $x_i(t)$—hence the name kernel. The concept slightly differs from the more conventional Volterra or Wiener kernels, however. First of all it is an extension to the multivariate case; and moreover, the considered input consists not only of the signals received from the other elements in the network, but also the timeseries' own past. How these kernels can be estimated, then, is the topic of the next section.

5.1.2 Gaussian Processes

The assumption of Gaussianity often yields an effective and practical approach to modeling, and this is not different in the problem we are trying to solve. In this section, we will discuss this assumption of Gaussianity using a one-dimensional time series. More precisely, we will model this timeseries as a Gaussian Process (GP):

A Gaussian Process is a collection of random variables, any finite number of which has a joint Gaussian distribution.

Loosely speaking, we will consider an (in our case ordered) collection of random variables $\{X_1, \ldots, X_n\}$, one variable for each time t; moreover, we will take the joint distributions between any number of variables to be Gaussian. The Gaussian Process in this way defines a distribution over sample functions, where a function is considered the (possibly infinite) vector of particular instances of these ordered random variables. (The latter definition is not entirely rigorous, but sufficient for the current discussion.) This distribution is completely characterized by a mean function $m(t)$ and a covariance function $k(t, t')$:

$$f(t) \sim \mathcal{GP}\big(m(t), k(t, t')\big). \tag{5.2}$$

We will first review some properties of the classical Gaussian probability distribution. Some notational conventions: $\boldsymbol{x}$ and $\boldsymbol{y}$ symbolize vectors of any number of random variables that are jointly Gaussian distributed,

$\boldsymbol{\mu_x}$ and $\boldsymbol{\mu_y}$ the vectors of their respective means. Thus, e.g.,

$$p(\boldsymbol{x} \mid \boldsymbol{\mu_x}, \Sigma_x) \sim (2\pi)^{-D/2} \mid \Sigma_x \mid^{-1/2} \exp\big(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_x})^T \Sigma_x^{-1}(\boldsymbol{x}-\boldsymbol{\mu_x})\big) \tag{5.3}$$

where D is the length of the vector $\boldsymbol{x}$, and Σ_x is the DxD, symmetric, positive definite covariance matrix. Using shorthand notation:

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu_x}, \Sigma_x). \tag{5.4}$$

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are also jointly Gaussian with covariance matrix K,

$$\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix} \sim \mathcal{N}\Big(\begin{bmatrix}\boldsymbol{\mu_x}\\ \boldsymbol{\mu_y}\end{bmatrix}, K\Big) \equiv \mathcal{N}\Big(\begin{bmatrix}\boldsymbol{\mu_x}\\ \boldsymbol{\mu_y}\end{bmatrix}, \begin{bmatrix}C_x & B\\ B^T & C_y\end{bmatrix}\Big), \tag{5.5}$$

then the marginal distribution of $\boldsymbol{x}$ is

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu_x}, C_x), \tag{5.6}$$

and the conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{x}$ is:

$$\boldsymbol{y} \mid \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu_y} + BC_x^{-1}(\boldsymbol{x}-\boldsymbol{\mu_x}), C_y - BC_x^{-1}B^T). \tag{5.7}$$

There's an important point to emphasize here: these equations are entirely linear. Let

$$A \equiv BC_x^{-1}, \tag{5.8}$$

and conditional covariance matrix Σ

$$\Sigma \equiv C_y - BC_x^{-1}B^T. \tag{5.9}$$

Then we can see that Σ is independent of $\boldsymbol{x}$, which allows us to interpret $\boldsymbol{y}$ as an affine function of $\boldsymbol{x}$ plus a random noise vector $\boldsymbol{n}$ that is independent of $\boldsymbol{x}$. In other words,

$$\boldsymbol{y} = \boldsymbol{\mu_y} + A(\boldsymbol{x}-\boldsymbol{\mu_x}) + \boldsymbol{n}, \tag{5.10}$$

with the noise vector $\boldsymbol{n}$ satisfying:

$$\langle \boldsymbol{n} \rangle = 0, \quad \langle \boldsymbol{nx} \rangle = 0, \quad \langle \boldsymbol{nn} \rangle = \Sigma. \tag{5.11}$$

If $\boldsymbol{y}$ is the system output and $\boldsymbol{x}$ is the system input, then (Eq. 5.10) describes the linearity of the system. However, within the framework of Gaussian Processes, $\boldsymbol{x}$ is not the input of the system but rather *the output in the past*. Predictions of the future output are thus obtained as a linear combination of outputs in the past (Eq. 5.10)—but the described system itself may not be linear in the sense of the output being a linear combination of the input. The latter would only be the case when the covariance function k is a linear function.

5.1.3 *Estimating the Autoregressive Kernels*

Now the groundworks are laid to switch viewpoints: from generating sample functions from a known Gaussian process, to estimating the properties of a Gaussian process given a sample function. We will assume the process is stationary, in which case it is convenient to define zero-mean variables: $\boldsymbol{x} - \boldsymbol{\mu_x} \mapsto \boldsymbol{x}$. We can then rewrite Eq. 5.10:

$$\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{n}. \tag{5.12}$$

The weight matrix A can be estimated from basic linear regression: if the noise is white or can be whitened, the solution is simply given by Eq. 5.8. From comparing Eq. 5.12 with the structure of Eq. 5.1, we can see that A corresponds to the coefficients of univariate autoregressive process.

We can easily extend this linear regression problem (Eq. 5.8) to the multi-variate, higher order case [2], and we finally arrive at the Yule-Walker equations [3] :

$$\begin{bmatrix} C_0 & C_1 & C_2 & C_3 \dots & C_{p-1} \\ C_1 & C_0 & C_1 & C_2 \dots & C_{p-2} \\ C_2 & C_1 & C_0 & C_1 \dots & C_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{p-1} & C_{p-2} & C_{p-3} & \dots & C_0 \end{bmatrix} \begin{bmatrix} A^{(1)} \\ A^{(2)} \\ A^{(3)} \\ \vdots \\ A^{(p-1)} \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \vdots \\ C_p \end{bmatrix}$$

where C_i is the covariance matrix between all variables at lag i.

[2] *Estimation of parameters and eigenmodes of multivariate autoregressive models*. Neumaier and Schneider,2001.

[3] *On periodicity in series of related terms*. Walker,1931.

5.2 *Computing the Correlations and the Kernels*

To compute the correlations, we did not bin the data, but instead convolved the detected spike times with a waveform which has a substantial width over 0.5ms (corresponding to the natural temporal resolution implied by the spike sorting algorithm.) The used function is furthermore Lorentzian, because its longer tails will increase the smoothness.

Due to the high computational resources needed to compute the correlation functions up to lags of several hundreds of time steps, we implemented the computations using a map-reduce procedure [4].

Finally, we can note that the matrix involved in the Yule-Walker equations has a block Toeplitz structure; we can make good use of this structure to efficiently solve the system of equations through applying the Levinson algorithm [5].

[4] *MapReduce: simplified data processing on large clusters*. Dean and Ghemawat,2008.

[5] *Efficient inversion of Toeplitz-block Toeplitz matrix*. Wax and Kailath,1983.

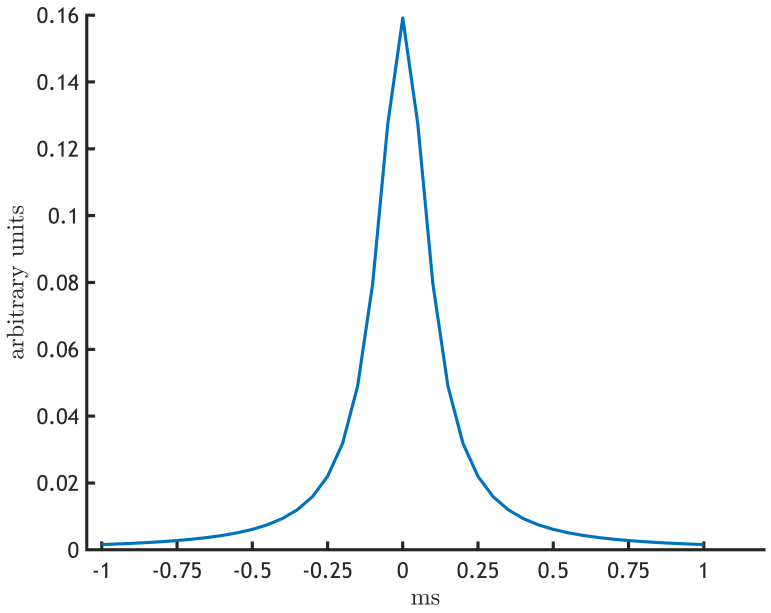


Figure 5.1: Waveform function used to construct the spike trains.

5.3 *A Directed Graph of Information Flow*

Once the full matrix A—and thus each autoregressive kernel—has been estimated, obtaining a directed graph boils down to computing a value from the kernel that captures the connectivity. In general words one could say there is a directed connection between two neuronal structures in the network if the kernel for that direction is not uniformly zero, that is, non-zero in a statistically significant way. A significance test is not easy to obtain, however: the estimation of the kernels is a highly nonlinear function of the original data, and the distributions of their estimators are not well established [6]. A good approach would be to generate new datasets that retain the underlying stochastic properties but discard the causal connections. An example procedure could be to sample new spike trains using the inter spike interval distributions obtained from the original dataset. Once a surrogate dataset is created, we can again compute the correlations and the kernels; a large collection of surrogate datasets would then yield an estimator of the variance of the kernels. The drawback of this procedure is that it is extremely costly in terms of computational resources: despite the measures taken to speed up the computations and to handle the large datasets, a reasonable significance test would take hours to days to compute.

[6] *Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance.* Kamiński et al.,2001.

For the time being, we thus resorted to a much simpler significance test. We computed the median absolute deviation (the median because this is more robust to outliers) of the absolute values of each kernels, and considered it an effective connection if a values was found above a

threshold value $v_{threshold}$ defined by:

$$v_{threshold,ij} = \theta MAD(\mid A_{ij}^{(t)} \mid) \tag{5.13}$$

where θ was taken to be $\theta = 5$.

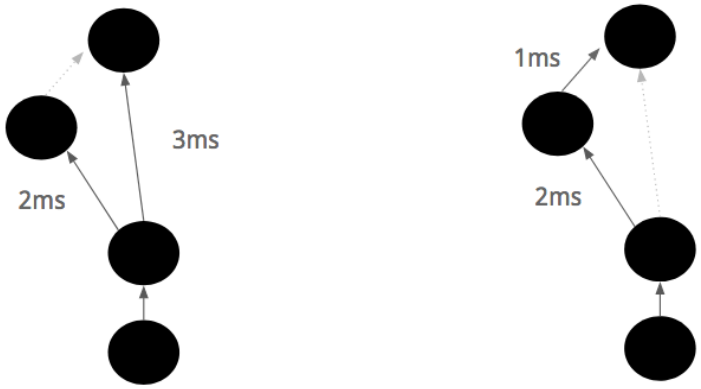


Figure 5.2: Example situations that are 'equivalent' in terms of correlation functions and resulting kernels. Only structural information can discern the true from the spurious connectivity.

Finally, we should mention that this method of determining the directed flow of signals is not always able to capture the real directed flow of information. Consider, e.g., the situation as depicted in Fig. 5.2. On the right hand side, the information flows from the bottom and splits after the second node, such that it arrives with a certain time delay in each of the upper nodes; the signals arrive at the same times, however, in the situation on the right. The correlation functions will thus be exactly the same, and since the kernels are derived from nothing else but the correlation functions, there is no way to discern the true path from the spuriously derived path. This could be resolved by using structural information however; to do this, knowledge about the actual underlying structure could be used as a Bayesian prior. This is outside the scope of this project however (but might be added to the analysis in future work).

6

Results and Conclusion

In this analysis, we will restrict ourselves to a single set of recordings from a low density network at day 34 in vitro. The network consisted solely of excitatory neurons and was not stimulated during its development. The whole sensor array is covered by a set of 6 configuration recordings, each of which make up a simultaneous recording of half the electrodes of their respective sensor regions. In other words, at each recording time one sixth of the whole sensor array is recorded simultaneously by a checkerboard pattern of 1024 electrodes, and this for two minutes at a sampling rate of 20 000Hz.

6.1 Summary of the obtained Biological Network Properties

In previous work [1] we analyzed the relationship between the spike trains of the neurons (obtained through spike sorting) and their effective connectivity properties (obtained from the calculated autoregressive kernels). Below, we summarize the conclusions in a qualitative way.

[1] *Connecting Neurons: A Stochastic Model of Information Flow in Cultured Networks of Human Neurons*, master thesis physics (unpublished).

Spontaneous Activity

All the neurons exhibit spontaneous activity, i.e., the neurons spontaneously produce action potentials, typically at a relatively low average rate (10 or less spikes per second). This spontaneous activity is very irregular, but not entirely random either: if the neuron produced a spike, the probability of a new spike spontaneously occurring shortly after this previous spike is highly reduced. This 'generalized' refractory period presumably corresponds to the recovery period of the stochastic mechanisms underlying the spontaneous action potential generation.

It is insightful to visualize this type of activity in a raster plot, i.e., a plot where each spike for each neuron is represented as a dot in function of time. Figure 6.1 shows such a raster plot, obtained from spike sorted recordings covering a region where the neurons are not significantly

connected (the latter is important, since, as we will explain in the next section, the firing pattern changes when the neurons are connected.)

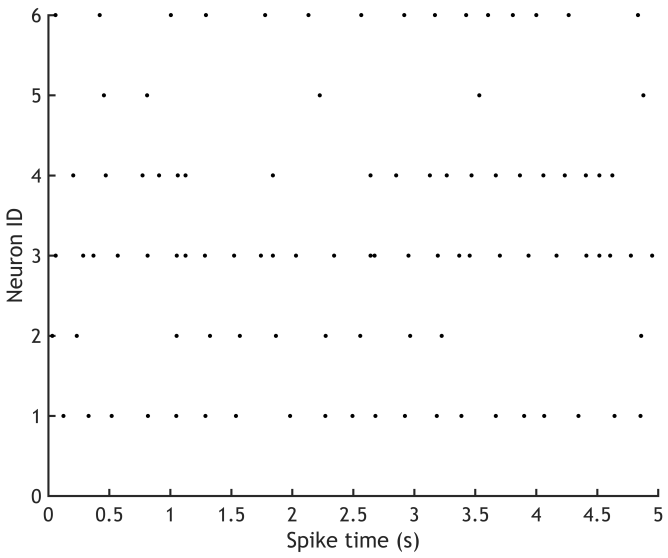


Figure 6.1: Raster plots showing 5s of spontaneous activity for 6 different neurons, each producing spikes at a different but relatively slow average rate.

Synchronization

The typical values for the transmission delays between the signals we found are much smaller ($\approx$ 0.001s) than the typical timescales of the spontaneous activity ($\approx$ 0.1s). This means that the activity in highly connected regions will appear synchronized: whenever one of the neurons in such a region spontaneously produces a spike, the action potential is transferred within a couple of *ms* to the post-synaptic neurons, which in their turn can produce a spike. If one bins the data over a couple of ms (e.g., a binning of 10ms), it will thus appear as if the connected neurons spiked at the same time. Given the much slower rate of the spontaneous spike generation, it might then again take a relatively long time before the next 'synchronized spike' appears. Characteristic firing patterns for neurons in a significantly connected region are shown in Fig. 6.2.

Effective Connectivity

We will discuss a single, distinctive example of connectivity properties we derived from two significantly connected neurons in our recordings. Figure 6.3 shows a snippet of their spike trains. A close inspection of (a) shows that the spike train of the second neuron (bottom trace) correspond to the sum of its own spontaneous activity plus the spikes of the first neuron. When we zoom in on the spike trains (Fig. 6.3 (b)), it seems to be the case that the spikes of the first neuron always precede the spikes of the second neuron. Upon inspecting the computed kernels (shown in Fig. 6.4), we can see that this information is exactly captured by the kernels: the first neuron sends spikes to the second neuron (direction of the transfer) with a transmission delay of around 0.5*ms* (which might indicate a transfer via a chemical synapse).

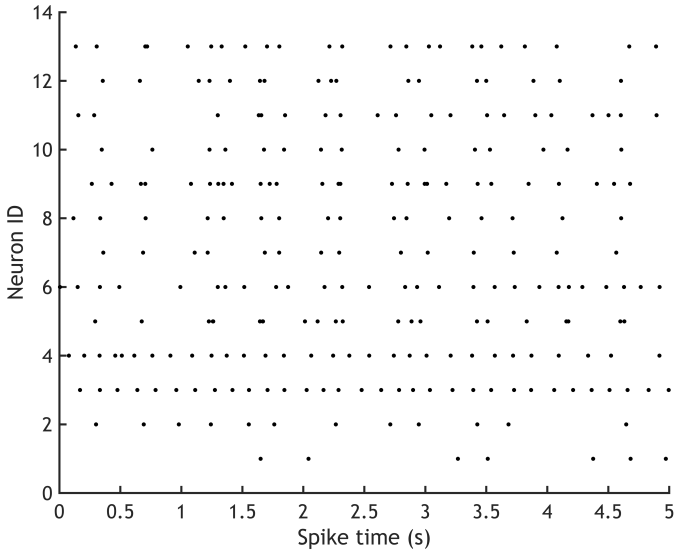
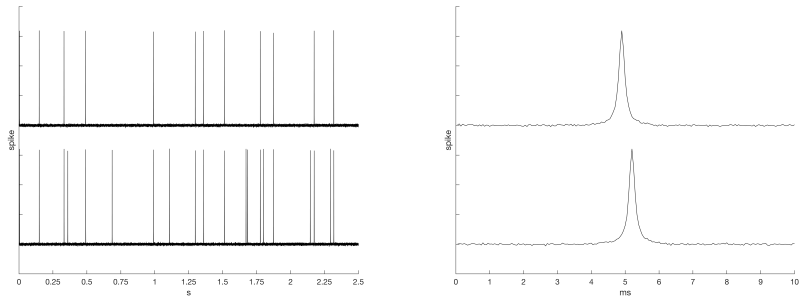
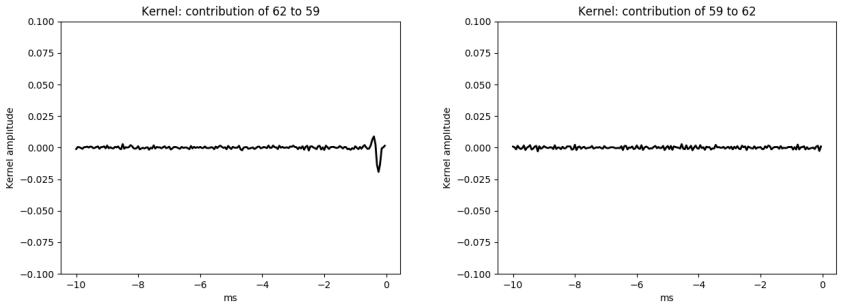


Figure 6.2: Raster plots showing 5s of activity for 13 different neurons, each producing spontaneous spikes and receiving post-synaptic potentials induced by the neurons to which they are connected.



(a) (b)

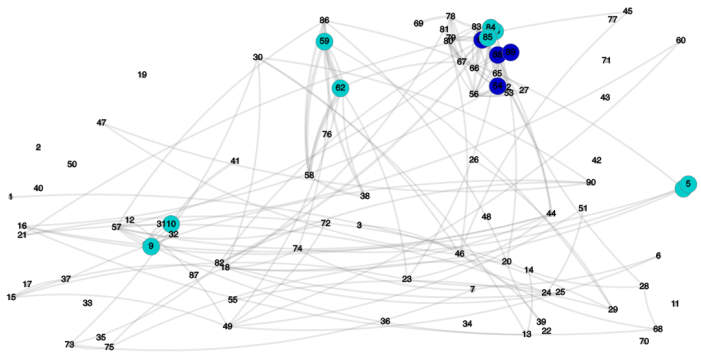
Figure 6.3: Spike train snippets of two closely connected neurons. (b) is a zoomed in version of (a).



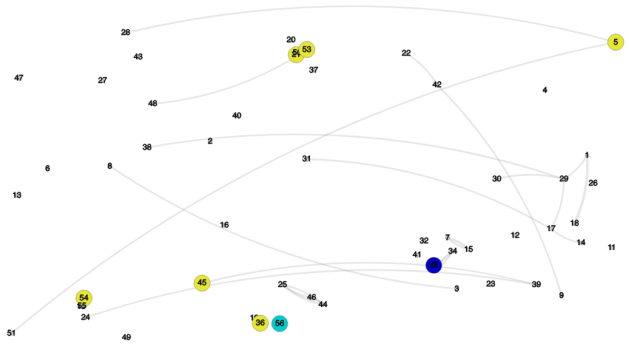
(a) (b)

Figure 6.4: Resulting kernels. (a) a significant contribution from neuron 1 to neuron 2; (b) the kernel corresponding to the other direction is not significant.

Finally, we show the complete graphs of information flow as derived from the significant kernels in Fig. 6.5, for the two earlier considered regions of low and high effective connectivity. (a) corresponds to the region of much larger effective connectivity, with even a local cluster of super high connectivity visible; (b) on the other hand is a region where connectivity has not developed beyond some sparse connections.



(a)



(b)

Figure 6.5: Graphs of the significant information flows. (a) shows the significant connections for a region of dense effective connectivity. (b) shows the significant connections in a much sparser connected region. The small, gray nodes correspond to axonal structures, the gray lines to significant connections. The somatic structures are colored according to their degree of connectivity (yellow for low connectivity, green for intermediate connectivity, and blue for large connectivity.)

6.2 *Relating the Results to a Network of Spiking Neurons*

Spontaneous activity. Without spontaneous activity, a network without external input would be completely silent. But given a certain input, spontaneous activity might not seem too important—were it not for the fact that a certain level of background spontaneous activity is needed to *sustain* the computations on the input (which has been shown both in biological neural networks [2] as in simulated models thereof [3]). From our analysis of the spike trains of unconnected neurons [4], we can estimate all the properties of the typical spontaneous activity: we showed that it can be modeled by a stochastic process characterized by a higher order gamma function with a given average spike rate. We can also estimate the distribution of those average spike rates over all unconnected neurons. An SNN where the neurons have a realistic spontaneous activity profile could then be constructed from using these estimated statistics: assign to each artificial neuron an average spike rate sampled from the estimated distribution of spike rates, and let the neurons subsequently emit spikes spontaneously with a probability given by the corresponding gamma process.

Network parameters. In spiking neuron networks, both the synaptic weights and the transmission delays make up the parameters of the network; however, computations can be performed by only using one of

[2] *Resonance or integration? Self-sustained dynamics and excitability of neural microcircuits*. Muresan and Savin,2007.

[3] *Large-scale model of mammalian thalamocortical systems*. Izhikevich and Edelman,2008.

[4] *Connecting Neurons: A Stochastic Model of Information Flow in Cultured Networks of Human Neurons*, master thesis Physics

these two different parameter types, and by reducing the other one to trivial values.

As we have noted before, using variable transmission delays is the most efficient approach in terms of needed computational units. In that case we can very straightforwardly translate our results to SNN model parameters: significant connections become weights with a value of 1, and 0 otherwise; and we can use the same canonical response function for each neuron, only shifted according to the time delays obtained from the computed kernels.

Another option is to trivialize the time delays. This means that we assume the input signals all arrive simultaneously—this is in general not a very biologically plausible assumption, but since our network exhibits synchronization, it is not a bad mapping of the biological model to an ANN model in this particular case. We can, e.g., bin the timesteps over 10 ms (thus 200 samples per bin). It is interesting to note that if we treat the output of each neuron as binary (1 when a spike is emitted in a period of 10ms, and 0 when there is no spike in this period), the biological network is in fact equivalent to a perceptron network.

The question then remains how the weights for this perceptron network relate to the effective connectivity measure we computed. The effective connectivity measure is in fact a measure of the directed information flow between two neurons; it is as such influenced by the spike rate of the neurons it connects. This is merely a reflection of the fact that we assume the information is encoded in the spike times, and the more transferred spikes, the more information is exchanged—thus the higher the measure of effective connectivity (this property can be mathematically derived [5]). It is not only a function of the spike rates, however, but also contains factors related to the actual biological synaptic strength and the transferral probability. Intuitively, one would say that the dependence on spike rate is an 'artifact' of the chosen method, and one should therefore appropriately normalize the measure to retain only the factors related to the actual synaptic parameters (this could be done by normalizing the correlation functions by the average spike rates before calculating the kernels). However, it is also quite interesting to note that in this way an encoding of the spike rate could be introduced, next to the temporal code. We noted before that in case of SNN, it has been shown that the synaptic weights can take the same role as computational units in more conventional ANN using rate codes. We leave the further insight and development of these ideas to future work.

[5] *Connecting Neurons: A Stochastic Model of Information Flow in Cultured Networks of Human Neurons*, master thesis physics (unpublished).

6.3 *Conclusion*

In this thesis, we have implemented a suite of methods to analyze large amounts of high-resolution neuronal activity recordings. These methods

allowed us to compute the detailed activity and connectivity properties of a cultured network of human cortical neurons. We then went on to describe how these computed properties can be translated into parameters of artificial networks of spiking neurons.

Our general results are meant to be a starting point for future research exploring the relation between biological and artificial neural networks. Through developing these methods, we hope to be able to—at some point in the future—increase insight in the computational properties of biological neural networks, as well as in the biologically inspired principles that could further advance the field of artificial neural networks.

Bibliography

David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski. A learning algorithm for boltzmann machines. *Cognitive science*, 9(1):147–169, 1985.

Adela-Diana Almási, Stanisław Woźniak, Valentin Cristea, Yusuf Leblebici, and Ton Engbersen. Review of advances in neural networks: Neural design technology stack. *Neurocomputing*, 174:31–41, 2016.

Theodore Wilbur Anderson. *An introduction to multivariate statistical analysis*, volume 2. Wiley New York, 1958.

Ehsan Arabzadeh, Stefano Panzeri, and Mathew E Diamond. Deciphering the spike train of a sensory neuron: counts and temporal patterns in the rat whisker pathway. *Journal of Neuroscience*, 26(36):9216–9226, 2006.

Wyeth Bair and Christof Koch. Temporal precision of spike trains in extrastriate cortex of the behaving macaque monkey. *Neural computation*, 8(6):1185–1202, 1996.

Robert E Baker, Michael A Corner, and Jaap van Pelt. Spontaneous neuronal discharge patterns in developing organotypic mega-co-cultures of neonatal rat cerebral cortex. *Brain research*, 1101(1):29–35, 2006.

William Bialek, Fred Rieke, RR De Ruyter Van Steveninck, and David Warland. Reading a neural code. *Science*, 252(5014):1854–1857, 1991.

Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4-5):175–308, 2006.

Sander M Bohte, Joost N Kok, and Han La Poutre. Error-backpropagation in temporally encoded networks of spiking neurons. *Neurocomputing*, 48(1-4):17–37, 2002.

Romain Brette, Michelle Rudolph, Ted Carnevale, Michael Hines, David Beeman, James M Bower, Markus Diesmann, Abigail Morrison, Philip H Goodman, Frederick C Harris, et al. Simulation of networks of spiking neurons: a review of tools and strategies. *Journal of computational neuroscience*, 23(3):349–398, 2007.

Kenneth H Britten, William T Newsome, Michael N Shadlen, Simona Celebrini, and J Anthony Movshon. A relationship between behavioral choice and the visual responses of neurons in macaque mt. *Visual neuroscience*, 13(1):87–100, 1996.

Nicolas Brunel. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of computational neuroscience*, 8 (3):183–208, 2000.

Michela Chiappalone, Marco Bove, Alessandro Vato, Mariateresa Tedesco, and Sergio Martinoia. Dissociated cortical networks show spontaneously correlated activity patterns during in vitro development. *Brain research*, 1093(1):41–53, 2006.

Michela Chiappalone, Alessandro Vato, Luca Berdondini, Milena Koudelka-Hep, and Sergio Martinoia. Network dynamics and synchronous activity in cultured cortical neurons. *International journal of neural systems*, 17(02):87–103, 2007.

Carson C Chow and John A White. Spontaneous action potentials due to channel fluctuations. *Biophysical journal*, 71(6):3013–3021, 1996.

George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

Jeffrey Dean and Sanjay Ghemawat. Mapreduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1):107–113, 2008.

Dominique Debanne, Emilie Campanac, Andrzej Bialowas, Edmond Carlier, and Gisèle Alcaraz. Axon physiology. *Physiological reviews*, 91 (2):555–602, 2011.

Alain Destexhe. Oscillations, complex spatiotemporal behavior, and information transport in networks of excitatory and inhibitory neurons. *Physical Review E*, 50(2):1594, 1994.

Scott N Deyo and William W Lytton. Inhibition can disrupt hypersynchrony in model neuronal networks. *Progress in neuropsychopharmacology & biological psychiatry*, 21(5):735–750, 1997.

Ricardo Dolmetsch and Daniel H Geschwind. The human brain in a dish: the promise of ipsc-derived neurons. *Cell*, 145(6):831–834, 2011.

Julia H Downes, Mark W Hammond, Dimitris Xydas, Matthew C Spencer, Victor M Becerra, Kevin Warwick, Ben J Whalley, and Slawomir J Nasuto. Emergence of a small-world functional network in cultured neurons. *PLoS computational biology*, 8(5):e1002522, 2012.

Chris Eliasmith and Charles H Anderson. *Neural engineering: Computation, representation, and dynamics in neurobiological systems*. MIT press, 2004.

Wulfram Gerstner and Richard Naud. How good are neuron models? *Science*, 326(5951):379–380, 2009.

Wulfram Gerstner, Richard Kempter, J Leo van Hemmen, and Hermann Wagner. A neuronal learning rule for sub-millisecond temporal coding. *Nature*, 383(6595):76, 1996.

Wulfram Gerstner, Henning Sprekeler, and Gustavo Deco. Theory and simulation in neuroscience. *Science*, 338(6103):60–65, 2012.

Wulfram Gerstner, Werner M Kistler, Richard Naud, and Liam Paninski. *Neuronal dynamics: From single neurons to networks and models of cognition*. Cambridge University Press, 2014.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. `http://www.deeplearningbook.org`.

AL Hodgkin. Evidence for electrical transmission in nerve. *The Journal of physiology*, 90(2):183–210, 1937.

Alan L Hodgkin and Andrew F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500–544, 1952.

John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

John J Hopfield. Pattern recognition computation using action potential timing for stimulus representation. *Nature*, 376(6535):33, 1995.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2 (5):359–366, 1989.

David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106–154, 1962.

Eugene M. Izhikevich. Simple model of spiking neurons. *IEEE transactions on neural networks*, 14 6:1569–72, 2003.

Eugene M Izhikevich and Gerald M Edelman. Large-scale model of mammalian thalamocortical systems. *Proceedings of the national academy of sciences*, 105(9):3593–3598, 2008.

Ian Jolliffe. Principal component analysis. In *International encyclopedia of statistical science*, pages 1094–1096. Springer, 2011.

Maciej Kamiński, Mingzhou Ding, Wilson A Truccolo, and Steven L Bressler. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological cybernetics*, 85(2):145–157, 2001.

Eric R Kandel, James H Schwartz, Thomas M Jessell, Department of Biochemistry, Molecular Biophysics Thomas Jessell, Steven Siegelbaum, and AJ Hudspeth. *Principles of neural science*, volume 4. McGraw-hill New York, 2000.

Adam Kepecs, Xiao-Jing Wang, and John Lisman. Bursting neurons signal input slope. *Journal of Neuroscience*, 22(20):9053–9062, 2002.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

Pawel Kudela, Piotr J Franaszczuk, and Gregory K Bergey. Changing excitation and inhibition in simulated neural networks: effects on induced bursting behavior. *Biological cybernetics*, 88(4):276–285, 2003.

Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

Jun Haeng Lee, Tobi Delbruck, and Michael Pfeiffer. Training deep spiking neural networks using backpropagation. *Frontiers in neuroscience*, 10:508, 2016.

Michael S Lewicki. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network: Computation in Neural Systems*, 9(4):R53–R78, 1998.

Henry W Lin, Max Tegmark, and David Rolnick. Why does deep and cheap learning work so well? *Journal of Statistical Physics*, 168(6):1223–1247, 2017.

Heiko J Luhmann, Anne Sinning, Jenq-Wei Yang, Vicente Reyes-Puerta, Maik C Stüttgen, Sergei Kirischuk, and Werner Kilb. Spontaneous neuronal activity in developing neocortical networks: from single cells to large-scale interactions. *Frontiers in neural circuits*, 10:40, 2016.

Wolfgang Maass. An efficient implementation of sigmoidal neural nets in temporal coding with noisy spiking neurons. 1995.

Wolfgang Maass. Bounds for the computational power and learning complexity of analog neural nets. *SIAM Journal on Computing*, 26(3): 708–732, 1997a.

Wolfgang Maass. Networks of spiking neurons: the third generation of neural network models. *Neural networks*, 10(9):1659–1671, 1997b.

Wolfgang Maass and Michael Schmitt. On the complexity of learning for spiking neurons with temporal coding. *Information and Computation*, 153(1):26–46, 1999.

Eisaku Maeda, HP Robinson, and Akio Kawana. The mechanisms of generation and propagation of synchronized bursting in developing networks of cortical neurons. *Journal of Neuroscience*, 15(10):6834–6845, 1995.

S Marĉelja. Mathematical description of the responses of simple cortical cells. *JOSA*, 70(11):1297–1300, 1980.

Jessica Mariani, Maria Vittoria Simonini, Dean Palejev, Livia Tomasini, Gianfilippo Coppola, Anna M Szekely, Tamas L Horvath, and Flora M Vaccarino. Modeling human cortical development in vitro using induced pluripotent stem cells. *Proceedings of the National Academy of Sciences*, 109(31):12770–12775, 2012.

Henry Markram, Joachim Lübke, Michael Frotscher, and Bert Sakmann. Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275(5297):213–215, 1997.

Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5 (4):115–133, 1943.

Paul A Merolla, John V Arthur, Rodrigo Alvarez-Icaza, Andrew S Cassidy, Jun Sawada, Filipp Akopyan, Bryan L Jackson, Nabil Imam, Chen Guo, Yutaka Nakamura, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 345(6197):668–673, 2014.

Jan Müller, Marco Ballini, Paolo Livi, Yihui Chen, Milos Radivojevic, Amir Shadmani, Vijay Viswam, Ian L Jones, Michele Fiscella, Roland Diggelmann, et al. High-resolution cmos mea platform to study neurons at subcellular, cellular, and network levels. *Lab on a Chip*, 15(13): 2767–2780, 2015.

Raul C Muresan and Cristina Savin. Resonance or integration? self-sustained dynamics and excitability of neural microcircuits. *Journal of neurophysiology*, 97(3):1911–1930, 2007.

Arnold Neumaier and Tapio Schneider. Estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Transactions on Mathematical Software (TOMS)*, 27(1):27–57, 2001.

Sergiu P Paşca, Thomas Portmann, Irina Voineagu, Masayuki Yazawa, Aleksandr Shcheglovitov, Anca M Paşca, Branden Cord, Theo D Palmer, Sachiko Chikahisa, Seiji Nishino, et al. Using ipsc-derived neurons to uncover cellular phenotypes associated with timothy syndrome. *Nature medicine*, 17(12):1657, 2011.

Quentin Pauluis, Stuart N Baker, and Etienne Olivier. Emergent oscillations in a realistic network: the role of inhibition and the effect of the spatiotemporal distribution of the input. *Journal of computational neuroscience*, 6(1):27–48, 1999.

Michael C Quirk, Kenneth I Blum, and Matthew A Wilson. Experience-dependent changes in extracellular spike amplitude may reflect regulation of dendritic action potential back-propagation in rat hippocampal pyramidal cells. *Journal of Neuroscience*, 21(1):240–248, 2001.

Michael W Reimann, Max Nolte, Martina Scolamiero, Katharine Turner, Rodrigo Perin, Giuseppe Chindemi, Paweł Dłotko, Ran Levi, Kathryn Hess, and Henry Markram. Cliques of neurons bound into cavities

provide a missing link between structure and function. *Frontiers in computational neuroscience*, 11:48, 2017.

Hernan Gonzalo Rey, Carlos Pedreira, and Rodrigo Quian Quiroga. Past, present and future of spike sorting techniques. *Brain research bulletin*, 119:106–117, 2015.

Fred Rieke and David Warland. *Spikes: exploring the neural code*. MIT press, 1999.

Alex Rodriguez and Alessandro Laio. Clustering by fast search and find of density peaks. *Science*, 344(6191):1492–1496, 2014.

Edmund T Rolls. Brain mechanisms for invariant visual recognition and learning. *Behavioural Processes*, 33(1-2):113–138, 1994.

Edmund T Rolls and Martin J Tovee. Processing speed in the cerebral cortex and the neurophysiology of visual masking. *Proc. R. Soc. Lond. B*, 257(1348):9–15, 1994.

Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

Cyrille Rossant, Shabnam N Kadir, Dan FM Goodman, John Schulman, Maximilian LD Hunter, Aman B Saleem, Andres Grosmark, Mariano Belluscio, George H Denfield, Alexander S Ecker, et al. Spike sorting for large, dense electrode arrays. *Nature neuroscience*, 19(4):634, 2016.

Staurt J Russell and Peter Norvig. *Artificial Intelligence (A Modern Approach)*. Prentice Hall, 2010.

Seref Sagiroglu and Duygu Sinanc. Big data: A review. In *Collaboration Technologies and Systems (CTS), 2013 International Conference on*, pages 42–47. IEEE, 2013.

Anil K Seth. Neural coding: rate and time codes work together. *Current Biology*, 25(3):R110–R113, 2015.

Belle A Shenoi. *Introduction to digital signal processing and filter design*. John Wiley & Sons, 2005.

Larry Squire, Darwin Berg, Floyd E Bloom, Sascha Du Lac, Anirvan Ghosh, and Nicholas C Spitzer. *Fundamental neuroscience*. Academic Press, 2012.

A Strassberg and DeFelice. Limitations of the hodgkin-huxley formalism: effects of single channel kinetics on transmembrane voltage dynamics. *Neural computation*, 5(6):843–855, 1993.

Steven H Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. CRC Press, 2018.

Amirhossein Tavanaei, Masoud Ghodrati, Saeed Reza Kheradpisheh, Timothee Masquelier, and Anthony S Maida. Deep learning in spiking neural networks. *arXiv preprint arXiv:1804.08150*, 2018.

Simon J Thorpe and Michel Imbert. Biological constraints on connectionist modelling. *Connectionism in perspective*, pages 63–92, 1989.

Jaap van Pelt, Ildiko Vajda, Pieter S Wolters, Michael A Corner, and Ger JA Ramakers. Dynamics and plasticity in developing neuronal networks in vitro. *Progress in brain research*, 147:171–188, 2005.

Vladimir Naumovich Vapnik. An overview of statistical learning theory. *IEEE transactions on neural networks*, 10(5):988–999, 1999.

Gilbert Thomas Walker. On periodicity in series of related terms. *Proc. R. Soc. Lond. A*, 131(818):518–532, 1931.

Mati Wax and Thomas Kailath. Efficient inversion of toeplitz-block toeplitz matrix. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 31(5):1218–1221, 1983.

Paul John Werbos. *The roots of backpropagation: from ordered derivatives to neural networks and political forecasting*, volume 1. John Wiley & Sons, 1994.

Pierre Yger, Giulia LB Spampinato, Elric Esposito, Baptiste Lefebvre, Stéphane Deny, Christophe Gardella, Marcel Stimberg, Florian Jetter, Guenther Zeck, Serge Picaud, et al. A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo. *ELife*, 7:e34518, 2018.

Yuguo Yu, Yousheng Shu, and David A McCormick. Cortical action potential backpropagation explains spike threshold variability and rapid-onset kinetics. *Journal of Neuroscience*, 28(29):7260–7272, 2008.

Y Zuo, H Safaai, G Notaro, A Mazzoni, S Panzeri, and M Diamond. Spike timing and spike rate make complementary contributions to perceptual decisions in rat s1 and s2 cortex. *Current Biology*, 25:357–363, 2015.