Imperial College
London

BIOLOGICALLY-INSPIRED MACHINE LEARNING APPROACHES TO LARGE-SCALE NEURAL DATA ANALYSIS

PAMELA HATHWAY

Thesis submitted for the degree of Doctor of Philosophy

Supervised by DR. DAN GOODMAN

Neural Reckoning
Department of Electrical and Electronic Engineering
Imperial College London

September 2020

COPYRIGHT DECLARATION

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution Non-Commercial No Derivatives license (CC BY-NC-ND).

Under this licence, you may copy and redistribute the material in any medium or format on the condition that; you credit the author, do not use it for commercial purposes and do not distribute modified versions of the work.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

ORIGINALITY DECLARATION

I hereby declare that this thesis and the work herein detailed, was composed and originated by myself, except where appropriately referenced and credited.

London, September 2020

Pamela Hathway

ABSTRACT

Recent progress in recording techniques now allows researchers to record from hundreds and even thousands of neurons simultaneously. New, scalable methods need to be developed to handle such large data sets. Ideally these methods should not only analyse the multi-dimensional data, but also provide results which could be interpreted by the brain.

One key problem in neuronal data analysis is to identify neuronal assemblies i. e. groups of neurons displaying coordinated neuronal activity. Currently available methods either search for assemblies whose neurons participate in repeated spike sequences or search for assemblies whose neurons display similar firing rate modulations. In this thesis, I present two approaches to the search for neuronal assemblies.

I investigate whether a spiking neural network equipped with biologically plausible synaptic learning rules can provide a biologically interpretable way of finding repeating spike patterns in neuronal data. Due to the similarities between spiking neural networks to how brains function, this might be very close to how the brain itself detects such repeated activity.

Furthermore, I present neural topic modelling – a new data analysis method for large neuronal data sets. Based on a machine learning method from text mining, neural topic modelling is scalable and produces interpretable results. By including multiple features of neuronal spike trains and even other data types such as local field potentials into the analysis, I can expand the definition of neuronal assemblies to any type of neuronal activity features which are co-modulated. The application of neural topic modelling to neuronal recordings reveals interactions between features of neuronal activity which have previously not been identified.

Since both approaches are biologically plausible, the results from both methods can be used to generate hypotheses about how the brain processes information and may reveal hitherto unknown information processing pathways.

ACKNOWLEDGEMENTS

This PhD is the most challenging project I have ever attempted. I am grateful for having been given the opportunity to learn and grow on this journey. It would have been impossible without the support of so many people.

From the bottom of my heart, I would like to extend the sincerest thank you

to Dan
for your expert guidance on any subject.
for allowing me room to explore my own ideas and your endless curiosity.
for your unwavering belief in me and my project.

to Claudia
for excellent advice and supervision during parts of my first year.

to my examiners Tim Constandinou and Thomas Nowotny
for valuable feedback and stimulating discussion during my viva

to Imperial College and the EEE Department
for awarding me with a scholarship for 7/8th of my PhD.
for valuable resources such as Graduate School and Counselling Centre.
for making SCR food worse every year so we could complain about it.

to the Neural Reckoning group
for entertaining group lunches and fun group meetings.
in particular to Lotte for reading, liking and improving my introduction.

to Yang, Jonny and Jean
for making me smile every day for four years.
I couldn't have asked for better PhD companions.

to the office crew: Yang, Jonny, Jean, Juil, Daphne, Sara and Razvan
for making the office a welcoming and *warm* place to work.
for improving the most stressful of days with coffee, pints, games & pizza.
in particular to Daphne, my partner in crime for all front row endeavors

to my PhD thesis writing group
for making thesis writing something to look forward to every day.
for mutual support throughout the toughest part of this process.

to the squad from IC Squash
for being an extraordinary group of hilarious and big-hearted people.
for filling my life with (Squash) games, so much laughter and joyous memories.

to all the interesting people I met in EEE
in particular Cat, Tobi, Ben, Guillermo, Binod and Al-Amin for elevator chat, interesting discussions and fights over croissants and macaroons.

to Catherine
for your support – I could have not written this thesis without you.

to Marga
for your friendship full of scheduled joy, laughter and encouragement.
for having exactly the right amount of ambition during boardgames.

to friends abroad: Nadine, Anna, Clare, Steffen, Liza, Lukas, Delwen
for your belief in me and your encouragement.
for friendships that endure separation by sea and time zones.

to the ANDA decoding geniusses
for braving the cold and daily potatoes to learn about neural decoding.
in particular to Amit and Mattia for valuable last minute advice.

to the organisers of ASPP
for teaching me invaluable Python skills in the most fun way possible.
for providing me with an opportunity to pass on the knowledge I gained.

to Colin and Penny
for making it possible to enjoy my time in London to the fullest.

to my family
in particular to Jenny whose wisdom saved me (and thereby this thesis) more than once.
in particular to my parents whose love and generosity and patience supported me throughout all of these years.

to Jui
for every single gadget and LaTeX or Mac tool you persuaded me to get (after I initially refused), all of which turned out to be indispensible.
for absolutely everything, from the tiniest to the largest acts of support.
for your love and for making me laugh when I need it the most.

to myself
for being brave.
for not giving up.

CONTENTS

1 GENERAL INTRODUCTION 17
1.1 (Neuro)science – a brief history of an ongoing process 17
1.2 Introduction 19
1.3 Contributions 22
1.4 Thesis Roadmap 24

2 NEURAL CODE 25
2.1 Single neurons 27
2.2 Multiplexed coding 30
2.3 Neuronal assemblies 31
2.3.1 Definitions of assemblies 32
2.3.2 Spike sequences 34
2.3.3 Detection of spike sequences 36
2.3.4 Covariations in firing rates 38
2.3.5 Detection of covariations of firing rates 39
2.4 Extension to current state-of-the-art 41
2.4.1 Restriction to two types of neuronal activity 42
2.4.2 Multi-memberships 42
2.4.3 Limiting assemblies to a finite set of neurons 43
2.4.4 Scalability 44
2.4.5 Biological interpretability 45
2.5 Conclusion 45

3 NEURAL TOPIC MODELLING – CONCEPT AND IMPLEMENTATION 47
3.1 Inspiration 47
3.2 Pipeline 50
3.2.1 Neural documents: time windows 52
3.2.2 Neural words: event filtering 52
3.2.3 Topic model 58
3.2.4 Clustering 62
3.2.5 Thresholding 66
3.3 Discussion 68
3.4 Conclusion 74

4 NEURAL TOPIC MODELLING – VALIDATION 75
4.1 Intention . . . 75
4.2 Artificial data . . . 78
4.2.1 Input generation . . . 78
4.2.2 Measuring performance . . . 79
4.2.3 Run parameters . . . 84
4.3 Comparison to Diana et al., 2019 . . . 85
4.4 Further explorations . . . 88
4.5 Detecting non-obvious correlations . . . 92
4.6 Realistic assembly assignment . . . 95
4.7 Discussion . . . 98
4.8 Conclusion . . . 104

5 NEURAL TOPIC MODELLING - APPLICATION TO BIOLOGICAL DATA SETS 105
5.1 Allen institute data . . . 106
5.1.1 Recording and preprocessing . . . 106
5.1.2 Experiments . . . 108
5.2 Parameters . . . 109
5.3 Neural words . . . 112
5.3.1 Instances per word type . . . 113
5.3.2 Influence of firing rates . . . 114
5.4 Analysis . . . 117
5.4.1 Relationship of neural words to experimental variables . . . 118
5.4.2 Visualisation explanation . . . 121
5.4.3 Detecting known assembly structures . . . 124
5.4.4 Unknown interactions between word types . . . 127
5.4.5 Comparison across experiments . . . 132
5.5 Discussion . . . 133
5.6 Possible algorithmic improvements . . . 138
5.7 Future works . . . 141
5.8 Conclusion . . . 143

6 A SNN WITH STDP FOR SPIKE SEQUENCE DETECTION 145
6.1 Background . . . 145
6.2 Spiking neural network and learning rule . . . 146
6.3 Validation of the original study . . . 150
6.3.1 Pattern finding and robustness . . . 150
6.3.2 Implementation parameters . . . 151
6.4 Application to real data sets . . . 153

6.5 Discussion 156
6.6 Conclusion 156

7 CONCLUSION 159
7.1 Overview and Contributions of the Thesis 159
7.2 Limitations of applying machine learning to neural data 160
7.3 Future Directions 161
7.4 Concluding remarks 162

A AUTHOR'S PUBLICATIONS 165

BIBLIOGRAPHY 167

LIST OF FIGURES

Figure 1.1 Thesis roadmap . . . 23
Figure 2.1 Timeline of electrophysiological data set size . . . 25
Figure 3.1 Neural topic modelling pipeline . . . 51
Figure 3.2 Neural word creation from spike data . . . 54
Figure 3.3 Neural word creation from LFP data . . . 57
Figure 3.4 LDA graphical model . . . 60
Figure 3.5 Performance improvement through cluster selection . . . 63
Figure 3.6 Influence of k on detecting ground truth assemblies. . . . 64
Figure 3.7 Cluster selection filters out noise. . . . 65
Figure 3.8 Choice of threshold for assembly assignment . . . 67
Figure 4.1 Input matrix with default settings . . . 81
Figure 4.2 Influence of number of iterations on performance . . . 85
Figure 4.3 Influence of asynchrony on performance . . . 87
Figure 4.4 Influence of number of assemblies on performance . . . 88
Figure 4.5 Influence of number of neural words on performance . . . 89
Figure 4.6 Influence of assembly activity on performance . . . 90
Figure 4.7 Influence of additional noise words on performance . . . 91
Figure 4.8 Influence of multi-membership on performance . . . 92
Figure 4.9 Input with non-obvious correlations . . . 93
Figure 4.10 Performance on input with non-obvious correlations . . . 94
Figure 4.11 Assemblies with complex correlational fine structures . . . 96
Figure 4.12 Performance using shaped assembly structures . . . 97
Figure 5.1 Locations of electrodes. . . . 107
Figure 5.2 Visual stimuli shown . . . 108
Figure 5.3 Input matrices . . . 113
Figure 5.4 Instances per neural word. . . . 114
Figure 5.5 Influence of firing rate on word instances. . . . 116
Figure 5.6 Extraction of neural words from LFP signal. . . . 117
Figure 5.7 Two examples of neural words. . . . 120
Figure 5.8 Annotated cluster visualisation. . . . 122
Figure 5.9 List of brain structures and substructures. . . . 123
Figure 5.10 List of word types extracted from data set. . . . 124
Figure 5.11 Firing rate modulated clusters. . . . 125
Figure 5.12 Hippocampus csp and visual cortex csp neural words. 128

Figure 5.13 Visual cortex sc- and subcortical lfp neural words. . . . 130
Figure 5.14 Visual cortex csp and subcortical lfp neural words. . . 131
Figure 5.15 Cooperation between structures. 132
Figure 5.16 Difference in cooperativity across experiments. 134
Figure 6.1 Effect of different learning rules on synaptic weight . . 149
Figure 6.2 Latency 151
Figure 6.3 Robustness 152
Figure 6.4 Choice of EPSP 154
Figure 6.5 STDP network on electrophysiological data 155

LIST OF TABLES

Table 3.1 Translation of concepts from text mining to neural data 49
Table 3.2 Description of parameters for LDA 61
Table 4.1 Description of parameters for artificial data generation. 77
Table 4.2 Parameters for artificial data generation. 80
Table 5.1 Neural topic modelling parameters for neuronal data 111
Table 5.2 Dataset sizes for different experiments 112
Table 6.1 Parameters for neuron model and STDP synapses . . . 148
Table 6.2 Success-relevant implementation details 154

ACRONYMS

ATA	All-To-All
EPSP	Excitatory Postsynaptic Potential
ICA	Independent Component Analysis
ISI	Inter-Spike Interval
KL	Kullback-Leibler
LDA	Latent Dirichlet Allocation
LFP	Local Field Potential
NN	Nearest Neighbor
PCA	Principle Component Analysis
RNN	Reduced Nearest Neighbor
STDP	Spike Timing Dependent Plasticity
VB	Variational Bayes

1

GENERAL INTRODUCTION

The core motivation of this thesis is to gain an understanding of what is going on inside our brains. In particular, I am interested in understanding how animals (including humans) are able to display such a vast amount of behaviours and cognitive abilities and how their brains manage such incredible complexities.

Before I get into the specifics of the research conducted for this thesis, I want to zoom out and place this work into the context of neuroscientific research to date.

1.1 (NEURO)SCIENCE – A BRIEF (AND EXAGGERATED) HISTORY OF AN ONGOING PROCESS

Understanding the brain is like a very, very large puzzle, in which some pieces are missing, many are redundant and the motive is so complex that it is not always obvious whether a piece actually fits or not.

To the first people dissecting animals, the brain must have seemed like any other tissue in the body - complex and layered, but ultimately silent. It wouldn't be until the eighteenth century until it was discovered that the nervous system could elicit electrical signals.

So some researchers thought "Is it possible to record activity from a neuron?". To investigate, they stuck wires into the brain.

It must have been an exciting day when those researchers found out that electrodes were able to record the electrical impulses sent out by neurons in the brain (Renshaw et al., 1940). They not only could record these signals, but also note down their timings to try to figure out their meaning. Sure enough, they soon found that the activity of individual neurons reflected part of the experience of the animal or one aspect of their outside world e. g. moving objects in their field of view (Hubel and Wiesel, 1961). With this piece of information, the first puzzle piece had been laid out. Hundreds of

researchers turned their gaze towards the study of individual neurons to understand the brain and to uncover the intricate computations single neurons are capable of. Everything pointed into the same direction: **the neuron is the information processing unit of the brain!** It was clear that understanding the behaviour of individual neurons would allow researchers to understand the brain.

So they thought: "If only we could record more neurons!". To investigate, they came up with tetrodes and recorded from up to ten neurons at the same time (Marg and Adams, 1967).

Fascinatingly, the timings of spikes from different neurons were sometimes coordinated at millisecond precision one after the other, like clockwork (Abeles and Gerstein, 1988). This neuronal behaviour would often occur in response to some event in the outside world. More puzzle pieces were being added one at a time. This was an exciting new direction: **not individual neurons, but multiple neurons involved in spike sequences are the information processing unit of the brain!** It was clear that understanding the precise coordination of spikes would allow researchers to understand the brain.

So they thought: "If only we could record more neurons!". To investigate, they came up with electrode arrays and recorded from 50-150 neurons at a time (Jones et al., 1992).

Researchers found that those precise spike sequences could be much larger than initially thought - sequences of up to 100 spikes were not uncommon (Torre, Quaglio, Denker, Brochier, Riehle and Grün, 2016). At the same time, neurons also responded to external stimulation with coordinated raising and lowering of their firing rates. These findings of coordinated activity of these groups of neurons replaced some of the older puzzle pieces, but it was much less clear how the new pieces fit together in the bigger picture. Now there were multiple directions: **not only neurons participating in spike sequences, but also groups of neurons with similar firing rate modulations are the information processing unit of the brain!** It was clear that understanding the precise coordination of groups of neurons would allow researchers to understand the brain.

So they thought: "If only we could record more neurons!". To investigate, they came up with even larger electrode arrays and recorded more than a thousand neurons at the same time (Lopez et al., 2016).

The findings reveal that neurons coordinate across brain regions and across multiple locations (Juavinett et al., 2019). A lot of the activity of neuronal spike trains that was previously dismissed as noise in the system was found to have meaning (Insanally et al., 2018). And as before, even larger groups of neurons showed coordinated activity and that previously ignored aspects of the animal's experience such as movement were also represented in neuronal activity patterns. The idea that groups of neurons involved in coordinated activity – neuronal assemblies – are the main information processing units of the brain is currently a prominent hypothesis about how the brain works (Yuste, 2015). [1]

But as always, researchers are wondering what they could uncover if they could only record from more neurons with the same millisecond precision...

Each improvement in recording technology will allow researchers a more detailed and intricate insight into the inner workings of the brain. Until the next progress in recording methods further increases the number of recorded neurons, we have to slowly assemble the puzzle pieces with the information we can find in the amount of data that we have.

Just as during every other time in neuroscience history, researchers are coming up with more and more complex ways to detect coordinated activity in the brain. The work presented in this thesis is one of them. This is my piece of the puzzle – a small contribution to the overall picture. However, who knows whether it will turn out to be an edge piece or one connecting two larger already assembled sections.

1.2 INTRODUCTION

Action potentials – their timing, their frequency and the context of their emission – are the language of the brain. They represent the transfer of information between individual neuron or between groups of neurons or any combination thereof. Information can be conveyed via multiple channels e. g. through the exact timing of individual spikes (Hahnloser et al., 2002), through the number of spikes or the relative frequency of spikes (Zelden-

1 It is of course very likely that in a few years or decades – when we can record spikes from whole brains – neuroscientists will proclaim that neuronal assemblies could have never been the information processing units of the brain and that researchers believing in this theory had been terribly naive and unimaginative. They will hold up [insert a larger and more complex structure of the brain such as network dynamics here] and declare that they will not solve the brain until they figured out [the structure mentioned before].

rust et al., 2018), through the frequency of spikes in a population of active neurons (Panzeri et al., 2010), through synchrony between multiple neurons (Roy and Alloway, 2001) or though larger spike sequences involving multiple neurons (Prut et al., 1998). At the same time it is becoming clear that groups of neurons – also called neuronal assemblies – may show emergent behaviour which go well beyond the capabilities of single units (Yuste, 2015).

In line with the growing knowledge surrounding the activity of neurons, the techniques with which to record from the brain are also steadily improving (Cunningham and Yu, 2014). Advances in electrode design and production allow simultaneous recordings of hundreds or even thousands of neurons in rodents (Lopez et al., 2016; Stringer et al., 2018). This would allow researchers to look at the multitude of neuronal responses and how they interact with each other across multiple brain regions. The bottleneck for analysing neuronal data has now shifted to the data analysis methods used to process these large recordings. Many previously developed methods for neuronal data analysis are already unusable because they are not scalable to larger data sets of more than 100 neurons (e. g. brute force methods used in Abeles and Prut (1996)), since the number of possible interactions between any subset of neurons increases exponentially (Prut et al., 1998; Quaglio et al., 2017).

Neuronal assemblies are said to be a fundamental computational unit of the brain (Sheng et al., 2019), which is the reason why many new methods for the identification of neuronal assemblies are currently in development (Mölter et al., 2018). Researchers agree that neuronal assemblies are groups of neurons which display coordination in some aspect of their activity – any combination of for example timed spikes, coordinated firing rates changes, synchronised spikes, bursting, shared silences (Deolindo et al., 2018). However, this broad and inclusive definition is not mirrored in the methods for the detection of neuronal assemblies, which search for assemblies in one of two ways. One approach is to find repeated spike sequences in data sets and declare all neurons participating in these spike patterns to be part of an assembly (Abeles and Prut, 1996; Quaglio et al., 2017). The other approach is to group neurons into an assembly which display the same modulations of their firing rates (Lopes-dos Santos et al., 2011; van der Meij and Voytek, 2018).

Ideally, new methods for neuronal assembly detection should be flexible enough to include more activity features than either firing rates or spike sequences. This would enable researchers to study interactions between activ-

ity patterns in an unprecedented way. Additionally, they need to be scalable to be usable for current data sets and potentially for future data sets which will be even larger.

Recent developments in machine learning methods have been successfully applied in other fields of research, including to biological data sets (Liu et al., 2016). They often rely on complex mathematical frameworks and data transformations, but promise scalability and unprecedented insights into large data sets (Doshi-Velez and Kim, 2017). Machine learning is already being applied to large neuronal data sets for complex problems such as the detection of neuronal assemblies or large spike sequences (Mackevicius et al., 2018; van der Meij and Voytek, 2018). It seems unlikely that the brain actually computes e. g. Fourier transforms or non-negative matrix multiplication to extract meaning from spikes. While these methods achieve promising results in terms of decoding abilities (Gerwinn, 2009; Musall et al., 2019; Stringer et al., 2018), their goal is not to tell us how the brain itself could make sense of the activity patterns.

If we want our data analysis models to potentially teach us about how the brain processes neuronal signals, we need to select methods which search for activity in a similar way to what the brain is capable of. In this thesis, I attempt to detect neuronal assemblies in large neuronal data sets using brain-inspired methods from machine learning to generate insights into how the brain communicates. Additionally, in order to be usable for longer than only a few years, these methods need to be highly scalable since the technology around electrophysiological recordings is bound to evolve further.

In this thesis I introduce neural topic modelling – a new data analysis method for neuronal recordings that might present a step towards tackling these challenges. It is based on topic modelling, which is a scalable algorithm used in text mining to find underlying topics (i. e. groups of words) in collections of documents. An analysis of all the articles in a newspaper, for example, might find groups of words across all articles referring to certain themes such as politics (words: Tories, government, failure), sports (words: national, team, championships, equal pay) and technology (words: algorithm, bias, data, discrimination). Neural topic modelling embraces the broad definition of neuronal assemblies and extracts multiple features from neuronal data (in contrast to extracting firing rate as the only feature): increases and decreases in firing rates, synchronous spiking and power in the local field potential (LFP) signal. As a result, neural topic modelling identifies similar neuronal assemblies which would have been found by other methods (if run

separately), such as groups of neurons in the visual cortex which increase their firing rates simultaneously or neurons in the hippocampus which show a high amount of synchronous spiking. However, it additionally identifies neuronal assemblies which could not have been found by existing methods: for example a neuronal assembly which suggests that power in the subcortical local field potential of is related to decreases in firing rates in neurons in the visual cortex.

Generally speaking, there is a wide range of features that could be extracted from neuronal recordings. In neural topic modelling, the selected features are known to carry meaning in certain circumstances and could in principle be detected by the brain itself. This makes neural topic modelling a brain inspired and biologically interpretable data analysis method.

In this thesis, another scalable and biologically interpretable data analysis method is applied to neuronal data: a spiking neural network with biologically plausible synaptic learning rules. This thesis presents the replication of the neural network developed by Masquelier et al. (2008) and investigates its usefulness for the detection of repeated spike sequences.

This section has provided an overview over the main themes of this thesis. The next section outlines the specific contributions of this thesis.

1.3 CONTRIBUTIONS

This thesis makes the following main contributions:

- It introduces the design and implementation of **neural topic modelling – a new data analysis method for the detection of neuronal assemblies** in large scale neural data sets. Other neuronal assembly methods look for assemblies by either identifying neurons participating in the same spike sequences or by finding neurons whose firing rates are modulated in the same way. Contrastingly, neural topic modelling does not limit assemblies to either of these two activity patterns, but includes firing rate modulations, synchronous spiking and even other neuronal signals such as local field potentials into the analysis. This approach accounts for neurons using multiplexed coding (neurons adopting different encoding strategies for different stimuli), which is something no other method takes into account.
- It demonstrates the **performance of neural topic modelling when tested on artificial data** for which the underlying assembly structure is

known. The artificial data used is based on a previous publication comparing different assembly detection methodologies and neural topic modelling, for the most part, surpasses the performance of the previously developed algorithms.

- It shows the **application of neural topic modelling to large neuronal data sets** of more than 1600 neurons. It demonstrates that neural topic modelling can recover similar neuronal assemblies as would have been found with other methods as well as neuronal assemblies which can only be found by taking into account multiplexed coding.
- It describes the use of a **spiking neural network to identify repeating spike sequences** in large neuronal data sets. This includes making the code open source and easy to use in Python, validating the method on artificial data and exploring the use of different learning rules and integration methods. It then also shows the results from applying this method to biological data sets.

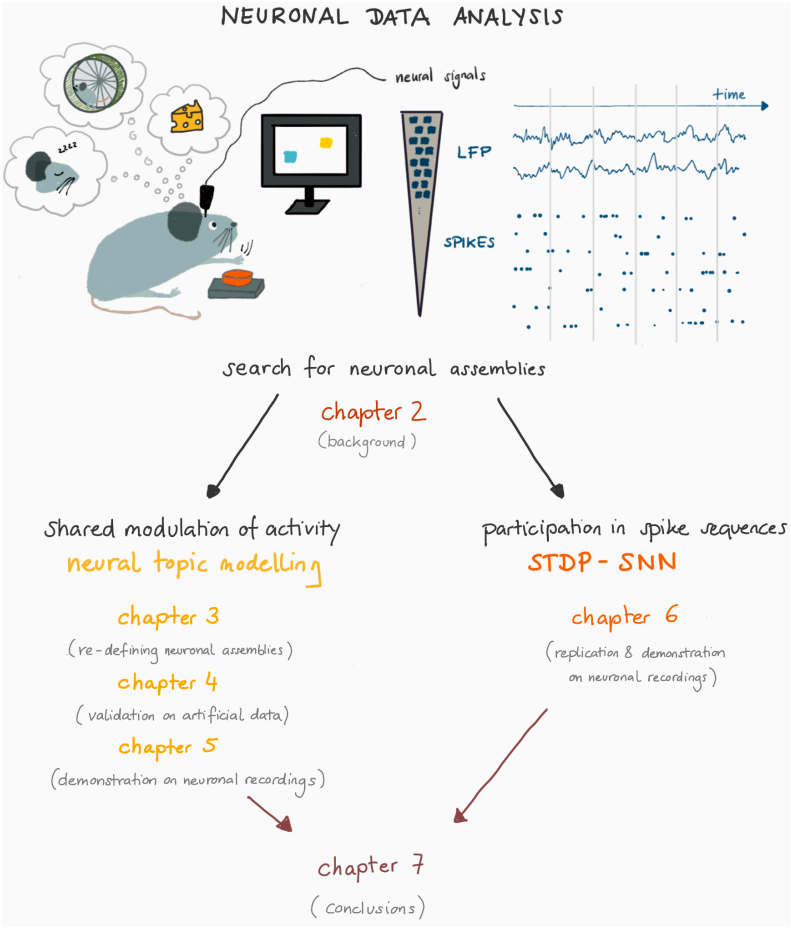


Figure 1.1: **Thesis Roadmap.**

1.4 THESIS ROADMAP

This thesis is organised in seven chapters and one appendix. A visualisation of the connections between chapters in this thesis is shown in Figure 1.1.

- **Chapter 2** outlines what is known about how the brain conveys information both on the level of individual neurons and of groups of neurons (neuronal assemblies). It describes the currently available approaches to detecting neuronal assemblies (via spike sequences and comodulation of firing rates) and the challenges for methods approaching this complex problem.
- **Chapter 3** introduces neural topic modelling as a new data analysis method and gives an overview over the steps involved. It also discusses the choice of default parameters.
- **Chapter 4** shows how neural topic modelling was tested using artificially created data and compares its assembly detection performance to results from previously published methods. The influence of a variety of parameters is investigated and discussed. Additionally, neural topic modelling is put to the test using complex neuronal assembly structures which would not easily be picked out by other methods.
- **Chapter 5** demonstrates the application of neural topic modelling to a large neuronal data set and discusses the neuronal assemblies that were found. It shows some possible analyses and suggests directions for future work.
- **Chapter 6** describes the replication of a spiking neural network for the detection of repeated spike sequences. The network is applied to a large neuronal data set and its future applicability for neuronal assembly detection is discussed.
- Finally, **Chapter 7** summarises the findings outlined in this thesis. It discusses the main contributions of the work presented, its limitations and points to potential future directions.
- **Appendix A** contains a list of all peer-reviewed publications.

This chapter introduced the aim of identifying neuronal assemblies in large-scale neuronal data analysis and outlined the contributions made in this thesis towards this goal. The next chapter will detail relevant background knowledge on neuronal data analyses and review related works.

2

NEURAL CODE

Our brain activity determines our experience of the world and our reactions to it. The brain receives information pouring in from our senses and processes those inputs. Any reaction to those experiences – such as moving towards or away from an object, focussing our eyes, emitting of sounds – are also elicited by our brains. The key question in neuroscience is how the brain goes from collecting information to complex behaviours such as movement, communication, mating or problem solving.

To this end, scientists have been keen to investigate the cells that relay information in the brain: neurons. They communicate via sharp changes in membrane voltage called action potentials or "spikes". All sensory information from skin, eyes, ears, noses is translated into action potentials, since this is the main format the brain uses to encode information. Similarly, all commands from the brain to muscles, eyes or internal organs are also encoded as a succession of action potentials.

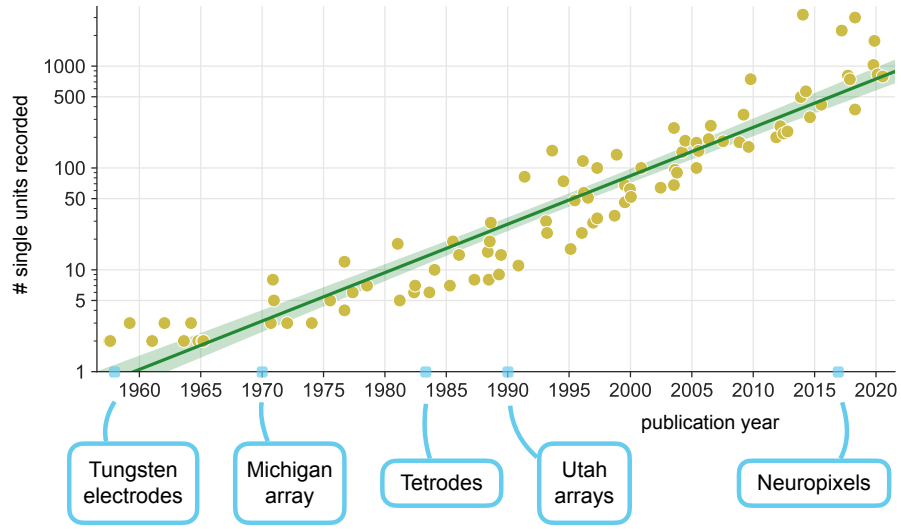


Figure 2.1: **Timeline of electrophysiological data set size.** The number of single units in 93 publications are shown against their publication date. Data and source code are available on github under the CC-BY licence (Stevenson, 2015). Additional information from (Hong and Lieber, 2019)

Since the first time action potentials had been recorded, researchers have tried to decode their meaning. This has been a slow process as the tools to record from single neurons have only been around for about eighty years (Renshaw et al., 1940). The investigation has been particularly difficult because the number of neurons from which it was possible to record had hovered around a handful to maybe one hundred until a few years ago as visualised in Figure 2.1. Advances in recording equipment and electrode design now allows recording thousands of neurons over multiple hours and days (Lopez et al., 2016; Stringer et al., 2018). While thousands of neurons are still a small number compared to the 100 million neurons in the brain, they still represent an immense leap in the field of neuroscience.

The current challenge is to find ways to analyse such vast amounts of data and to find out how neural activity relates to environmental inputs and behaviours. Traditionally, researchers have investigated the behaviour of single neurons and found that they encode various signals such as sensory information (Hubel and Wiesel, 1961), face recognition (Kreiman et al., 2000) and motivational states (Allen et al., 2019). Furthermore, multiple neurons may coordinate their firing in multi-neuronal spike sequences (Quaglio et al., 2017). Recording larger numbers of neurons allows researchers to explore groups of neurons, also called neuronal assemblies (Gerstein and Aertsen, 1985). Groups of neurons in networks may show emergent behaviour which go way beyond the capabilities of single units (Yuste, 2015). The main difficulty is that the current analysis tools were not developed with such large numbers of neurons in mind and are therefore often not scalable or limited in the reach of interactions they can measure e. g. are limited to pairwise correlations of neuronal firing rates.

This thesis revolves around the question of finding new ways to analyse large neural data sets and in particular how to identify neuronal assemblies which may hold the key to understanding brain function (Yuste, 2015).

In this chapter, I will review the current knowledge on how single neurons transfer information (see Section 2.1) and what encoding strategies they use (see Section 2.2). I then discuss the role attributed to neuronal assemblies in neuroscience in Section 2.3. I detail the two main ways of defining assemblies (spike sequences, see Section 2.3.2 and firing rate modulations in Section 2.3.4) and how researchers detect them (spike sequences, see Section 2.3.3 and firing rate modulations in Section 2.3.5). Section 2.4 then identifies limitations in the ways researchers define and look for neuronal assem-

blies and describes how the work in this thesis will contribute to overcoming those limitations.

2.1 SINGLE NEURONS

Decades of research have established that single neurons are the structural, functional, perceptual and behavioural units of the brain. This section will describe the evidence found and what this means for the field of neuroscience.

Single neurons have been recognised as the physical units of the brain since Cajal accurately observed at the end of the 19th century that neurons were physically separated from each other via synapses (Gerstein et al., 1989; Kandel Eric R., Schwartz James H., Jessell Thomas M., Siegelbaum Steven A., 2013). These findings would spark a focus on single neurons as the building blocks of the brain and inspire widespread analyses of neuronal wiring diagrams and neuronal cell compartments and their functions (Yuste, 2015).

The functional role of neurons would only start to be discovered a few decades later when Renshaw et al. (1940) developed a method to record single cell activity via tungsten electrodes. This technological development made it possible to investigate the functional properties of neurons in practice. The idea that neurons are the functional units of the nervous system – the so called "neural doctrine" – had been around for a while time beforehand. Sherrington proposed in 1906 that a scratch reflex is responsive to a particular area of skin, in essence describing a neuron with a sensory receptive field (Yuste, 2015).

Indeed, researchers soon found such clear examples of neurons increasing or decreasing their firing rates (the number of spikes emitted per second) in response to stimuli that were presented to the animal. Hubel and Wiesel (1961) demonstrated that single neurons in the cat primary visual cortex exhibited such receptive fields as proposed by Sherrington and responded preferentially to edges and shapes of particular sizes and orientations. Researchers started to describe neurons by their tuning curves, which indicated response rates for a particular set of sensory stimuli e. g. lines or gratings with different orientations (Hardcastle et al., 2017; Hubel and Wiesel, 1961). In parallel to neurons in the visual system, neurons in the auditory cortex also showed a preferential tuning for certain frequencies (Ehret, 1975; Hardcastle et al., 2017; Merzenich et al., 1975), neurons in the sensory cortex were

found to be sensitive towards particular stimulation in clearly defined areas of the skin (Panzeri et al., 2010; Pruszynski and Zylberberg, 2019) and neurons in the olfactory cortex responded only to a particular set of olfactory cues (Wilson, 2013).

Receptive fields have not only been identified in the context of direct sensory information, neurons in the hippocampus have been found to be sensitive to a wide variety of behavioural parameters. Grid cells, place cells and border cells in the medial enthorhinal cortex of rats increase their firing rates in response to the animal's location in space (Hafting et al., 2005; O'Keefe and Dostrovsky, 1971; Solstad et al., 2008; Wilson and McNaughton, 1993). Other cells encode the speed at which an animal is moving (Sargolini et al., 2006) or the direction in which the head of an animal is facing (Kropff et al., 2015). Of course one neuron is not enough to encode all the possible states when it comes to spatial navigation or body position, but such cells are often embedded in a network. Individual neurons in such networks are tuned towards different states e. g. head directions and taken together their responses cover the full 360 degrees of possible head orientations (Munn and Giocomo, 2020). Similar networks have been identified in fruit flies (Green et al., 2018; Kim et al., 2017). Such internal maps can come in the form of 1D indicators (e. g. light intensity), 2D compasses (e. g. head direction) or 3D representations of space (e. g. a network of grid cells) (Munn and Giocomo, 2020).

Researchers also found evidence that single neurons can act as perceptual units. The mammalian visual system sends visual information through a multitude of different cortical areas and their layers. At the end of this processing streams, single cells have been identified in the inferior temporal cortices of humans and monkeys which respond selectively to faces (Desimone et al., 1984; Kreiman et al., 2000; Tanaka, 1996). Even though the myth of the extremely specific so-called "grandmother" cells (neurons which would respond exclusively to the face of a particular person e. g. one's own grandmother) has been debunked (Lopes-dos Santos et al., 2011), neurons in the higher visual system present an impressive degree of specificity e. g. to the concept of one particular famous person (Quiroga et al., 2005).

Similarly, the activation of a small number of neurons can elicit or change the behavioural pattern of an animal. In rats, whisker movements are controlled by the spike trains of individual neurons (Brecht et al., 2004). This suggests that some single neurons may indeed have a significant amount of influence on an animal's decision making and behaviour (Allen et al., 2019).

As mentioned above, the first type of neuronal activity identified to carry information about the animal's environment were changes in firing rate. Therefore firing rate modulations were the main focus of early neuroscientific research. However, it soon became clear that neurons also encode information in other ways, namely through the timing of their spikes.

Temporal codes have been demonstrated in neurons across multiple sensory cortices. The time between the presentation of a stimulus and the first spike (latency) is highly informative in neurons conveying visual information in the primary visual cortex (Butts et al., 2007; Gollisch and Meister, 2008) and in the thalamus (Reinagel and Reid, 2000). Unpredictable stimuli are followed by millisecond precise spike timing in monkey cortical neurons (Bair et al., 2002) and track stimulus fluctuations in rat neurons (Mainen and Seinowski, 1995). The timing of spikes has been shown to encode information about stimulus location and edge-orientation in in primary somatosensory neurons (Mackevicius et al., 2012; Panzeri et al., 2001; Pruszynski and Johansson, 2014) and about whisker stimulation in rats (Arabzadeh et al., 2006). Even motor neurons show temporal coding strategies despite being heavily firing rate-modulated (Pruszynski and Zylberberg, 2019). The first spikes reaching muscles crucially affect contraction patterns (Srivastava et al., 2017). Whisker movement in rats depend on the rhythmic activity of action potentials of single neurons and the timing reveals movement direction and amplitude (Brecht et al., 2004).

Researchers also analysed other aspects of spike trains of individual neurons (Perkel et al., 1967). The distribution of inter-spike intervals (ISIs) and count variability (Coefficient of Variation, Fano Factor) have been shown to encode stimulus information (Buzsáki, 2004; Insanally et al., 2018; Nawrot, 2010; Reinagel and Reid, 2000).

Bursting – when a single neuron emits a rapid succession of action potentials – is another way in which neurons can relay information. The advantage of bursts is that they can interact very little with other timed responses and thereby optimising information transmission (Naud and Sprekeler, 2018). Indeed, bursts of spikes are common in hippocampal neurons (O'Keefe and Dostrovsky, 1971) and have more recently been shown to drive singing in songbirds (Buzsáki, 2010).

Statistical relationships between spike trains have been investigated in detail, most commonly through cross-correlation based methods (Pastore et al., 2018). A large number of methods attempt to characterise the distance between two spike trains to uncover how spikes of two neurons relate to each

other (van der Meij and Voytek, 2018; Van Rossum, 2001; Victor and Purpura, 1996).

Researchers have traced back all these different aspects of neural spike trains to environmental variables measured during the experiment[1]. Now the question is whether each neuron adopts exactly one or multiple ways of encoding information. While this can vary from neuron to neuron and depends on neuron type (König et al., 1996) and context (Yamins et al., 2014; Zylberberg and Strowbridge, 2017), we do see that neurons make use of multiple encoding strategies simultaneously (Pruszynski and Zylberberg, 2019).

2.2 MULTIPLEXED CODING

A close look at neuronal activity reveals that neurons code for multiple variables at once (Hardcastle et al., 2017; Reinagel and Reid, 2000; Stringer et al., 2018). This makes a lot of sense intuitively, since using different aspects of the spike train to convey separate pieces of information increases the total amount of information that can be transferred (Naud and Sprekeler, 2018; Reinagel and Reid, 2000).

More specifically, neurons encode different information in their spike trains on different temporal scales. This phenomenon is called multiplexing (Pruszynski and Zylberberg, 2019). For example, a neuron can simultaneously encode one aspect of a sensory stimulus with its firing rate and another aspect through timed spikes (Hardcastle et al., 2017; Harvey et al., 2013; Stringer et al., 2018).

Neurons in monkey primary sensory cortex convey vibrational amplitude with their firing rates and frequency with phase-locked timed spikes to the population activity (Harvey et al., 2013). The same behaviour was demonstrated in zebra fish olfactory bulbs (Friedrich et al., 2004) and sensory cells of leeches (Pirschel and Kretzberg, 2016), indicating that multiplexing occurs in neurons across species. Multiplexing increases when tasks get more complex or when the animal has to encode information faster e. g. when a rat is running at a higher speed (Hardcastle et al., 2017).

Interestingly, multiplexed coding has also been established in grid cells in the postsubiculum and entorhinal cortex of rats – cells which have been

1 Of course it is entirely possible that these codes are used for conveying information about the internal state of the animal, in which case these codes would still be used but we would have no way of measuring or proving this. Just because we cannot prove that multiplexing takes place does not mean that it is not.

thought to be extremely rate code-driven (Hafting et al., 2005). While the firing rates consistently encoded the location of the animal, the direction of movement was encoded in temporal aspects of the spike trains (Munn and Giocomo, 2020). Multiplexing in these higher order areas is thought to contribute to the emergence of cognitive behaviours (Sheng et al., 2019). In another fascinating discovery, Leutgeb et al. (2005) demonstrated that hippocampal cells could remap part of their multiplexed coding strategy while keeping the other maps stable: the neurons adapted their activity for a stimulus cue without disturbing their current location mapping. This indicates that neurons are able to transmit information about one feature in a reliable way even while they are changing their tuning for others.

Traditionally, multiplexing has been thought to occur on very different time scales – precisely timed spikes on the order of milliseconds and the modulation of firing rates on the order of tens to hundreds of milliseconds. Evidence is mounting that the multiple levels of multiplexed coding might not need be on completely different orders of magnitude with respect to time. Spike bursts exist on a similar time scale to timed spikes and have been shown to encode complementary information (Naud and Sprekeler, 2018). In addition, neurons can also take part in concerted activity with other neurons – which will be addressed in the upcoming section – which also occurs on the same very short time scale.

2.3 NEURONAL ASSEMBLIES

As has been laid out in the previous section, single neurons display a large repertoire of strategies to encode salient information in their spike trains. At the same time, recording techniques have advanced and made it possible to record from multiple neurons at the same time – tens of neurons at first (Prut et al., 1998), to a few hundred (Quaglio et al., 2017) until a few years ago. Neuropixel electrodes made it possible to record spike trains from thousands of neurons at the same time (Lopez et al., 2016). The spotlight of neuroscience started to shift to the interactions between groups of neurons (Smith et al., 2018). Co-active groups of neurons have been hypothesised by Hebb and many after him to be the basic functional unit for information processing in the brain (Buzsáki, 2010; Harris, 2005). With the advent of large neural data sets, the search for groups of neurons with coordinated spiking behaviour – neuronal assemblies – received a big boost.

Neuronal assemblies are groups of neurons which share some characteristic in the way they receive or send information or in their connections. With a definition this broad it is no wonder that a large variety of neuronal phenomena fall under the heading of neuronal assemblies. In the first part of this section, I will describe the different types of assemblies which are commonly investigated (Section 2.3.1). I will then focus on two types of neural assemblies – neurons taking part in spike sequences (Section 2.3.2) and neurons with coordinated firing rates (Section 2.3.4). I will cover their relevance for brain functioning and Sections 2.3.3 and 2.3.5 will provide an overview over the methods that are used to detect them respectively.

2.3.1 *Definitions of assemblies*

Neuronal assemblies are groups of neurons which are linked together through their input, the information they process or their output. Gerstein et al. (1989) distinguish four main types of neuronal assemblies: motor pools, cortical columns, Hebbian assemblies and correlation of firing.

Motor pools are groups of neurons with a common target (the muscle) or output. One defining characteristic of motor pools is that the effect of assembly activity will be larger than the sum of the effects of the individual neurons.

Cortical columns are groups of neurons in close proximity to each other with a common input (in the case of cortical columns: visual cortex). This type of assembly is also referred to as functional assemblies since place cells and grid cells also fall under this category but are not located in the cortex (Roudi et al., 2015).

Motor pools and cortical columns are defined through their shared input or output. They show a high level of co-activation in response to sensory stimulation, but also in the absence of relevant input or output (Carrillo-Reid et al., 2016)[2]. It should be noted that for these two types of assemblies, no assumptions are made about the interactions or connections between the neurons, only about the incoming or outgoing signals (Gerstein et al., 1989).

Hebbian assemblies are groups of neurons exhibiting significant co-activation due to strong synaptic connectivity between neurons within the assembly

2 Sometimes these two types of assemblies are also referred to as neuronal *ensembles* (Carrillo-Reid et al., 2016). Others call neurons recorded from the same tetrode neural ensembles (Trouche et al., 2019), so the difference between assemblies and ensembles are not clearly defined. In this thesis, I have chosen to only use the word "assemblies" to avoid confusion.

(and weaker connections to neurons outside the assembly) (Buzsáki, 2010; Roudi et al., 2015). This type of assembly is thought to be the result of repeated overlapping activation from common input sources after Hebbian principles of synaptic associations (Gerstein et al., 1989). Indeed, stable neural assemblies can be induced by forcing neurons to spike together through repeated optogenetic stimulation (Carrillo-Reid et al., 2016). Newly formed assembly activity did not disrupt pre-existing assembly dynamics and even displayed spontaneous assembly activity and pattern completion. The resulting attractor dynamics of Hebbian assemblies which includes pattern completion are believed to be a main contributor to complex behaviours and a cornerstone of many cognitive functions (Roudi et al., 2015).

A group of neurons does not need to be synaptically connected (or the exact connectivity does not need to be known) in order to be called a neuronal assembly. Neuronal assemblies of the correlation of firing type have been found to span prefrontal, visual, somatosensory and parietal areas during a tactile task (Deolindo et al., 2018). Tasks and movements modulate nearly all brain regions as multiple studies have demonstrated Deolindo et al. (2018); Juavinett et al. (2019); Musall et al. (2019). This type of assembly exhibits co-activations in their firing rates which are above chance (Roudi et al., 2015). This similarity in behaviour does not necessarily depend on common input, although it is often the reason for the coordinated behaviour (Gerstein et al., 1989).

Hebbian assemblies and correlation of firing assemblies are characterised by the interactions between the neurons in them (Gerstein et al., 1989). Both of these types of assemblies do not require spatial proximity and have been shown to be distributed across the brain (Harris, 2005).

Neuronal assemblies are a main feature of neuronal organisation in the brain (Carrillo-Reid et al., 2016). In the hippocampus of mice, Hebbian assembly activity is the default organisation (Dragoi, 2020), suggesting that assembly activity comprises the basis of hippocampus information encoding (Dragoi and Buzsáki, 2006). Their ubiquity in other parts of the brain emphasises their importance information processing (Carrillo-Reid et al., 2016; Dragoi, 2020; Quaglio et al., 2017). In rodents, assemblies form in the first postnatal weeks in the hippocampus of mice and demonstrate consistent offline assembly activity by week three (Dragoi, 2020). New assemblies can be formed at any time (Carrillo-Reid et al., 2016) and were observed to emerge when rodents were allowed to explore their environment for a long time (Wilson and McNaughton, 1993).

The definition of neuronal assemblies (in particular the ones based on correlation of firing) has been widened to encompass neurons which exhibit any type of above-chance interaction (Deolindo et al., 2018; Dragoi, 2020). Nevertheless, many researchers still base their search for neuronal assemblies on the co-modulation of firing rates (Deolindo et al., 2018; Lopes-dos Santos et al., 2011, 2013).

In parallel, a lot of research has gone into the search for neuronal assemblies which consist of sets of neurons which take part in repeating spike sequences (spatiotemporal spike patterns) (Prut et al., 1998; Quaglio et al., 2017). The following section will describe neuronal assemblies which are based on the neurons taking part in repeated spike sequences, their relevance and the methods with which researchers are trying to find them.

2.3.2 *Spike sequences*

Coincident spiking between pairs of neurons has been shown to play a role in cortical processing (Riehle et al., 1997; Roy and Alloway, 2001). Furthermore, Prut et al. (1998) discovered that cortical neurons (2-8 neurons) repeatedly exhibited precise firing sequences (with delays between spikes in the order of hundreds of milliseconds) which were associated with the monkey's movement. The small number of neurons involved in the spatiotemporal spike patterns is mostly caused by limitations in the recording techniques at the time (six microelectrodes in total). Improvements in recording techniques were accompanied by improvements in software for the identification of fine temporal spike patterns, so today spike patterns involving more than one hundred neurons can be identified (Quaglio et al., 2017).

Repeated spatiotemporal spike patterns have been suggested to enable the neuronal networks to process and transfer information in a robust way (Grün et al., 2008). Multiple network models have been proposed for the readout of such spike patterns. The synfire chain model poses that synchronous spikes elicited by one group of neurons may reach another group of neurons simultaneously through identical delays (Abeles, 1982). The second group of neurons acting as coincidence detectors may then elicit their own synchronous spikes and thereby relay the information. A generalisation of this work, the synfire braid model, eliminated the need for identical delays due to variability in spike timings and propagation delays (Bienenstock, 1995). Building on those ideas, Izhikevich (2006) suggested that a randomly connected spiking neural network with spike timing dependent

plasticity (STDP) synapses would be able to propagate spike sequences with fixed delays reliably. These theoretical models underline the ability of the brain to send and read such spatiotemporal spike patterns and give experimental findings of such patterns a framework with which to tie them to higher processes such as cognition and attention (Pauli et al., 2018).

Small and large spike sequences have been shown to carry relevant information. In the hippocampus of rodents, their occurrences and reliability have been found to relate to learning and memory processes (Buzsáki, 2004; Sasaki, 2019). Indeed, plasticity in hippocampus neurons during memory consolidation is dominated by temporal coding and in particular spike sequences, whereas frequency modulations are not very involved (Farooq et al., 2019; Liu et al., 2019). The same is true for motor learning – spatiotemporal spike patterns emerge during motor learning and remain stable for some time after the task (Sheng et al., 2019). Interestingly, the neurons involved in the spike patterns do not have to be spatially close, and instead can be quite spaced out within the brain structure (Sheng et al., 2019). Even in cortical areas, spike sequences are highly relevant and can be used to predict the grasp type in a center-out reach task in monkeys, greatly surpassing prediction performance for firing rate alone (Smith et al., 2018).

Dragoi (2020) shows that spatiotemporal patterns in the hippocampus are important for memory and learning. In fact, coordinated spikes in cell assemblies during slow wave sleep can be traced back to patterns that were present during the spatial navigation task the rats were completing pre-sleep (Pastalkova et al., 2008; Wang et al., 2015). It is assumed that this repetition of activity during sleep represents the animal processing and replaying the task activity during sleep. Tellingly, when the replay of these assembly activity patterns were disrupted during sleep, the animal demonstrated a recall deficit for the mazes (Gridchyn et al., 2020). This strongly indicates that spike patterns in neuronal assemblies are a meaningful and important processing unit in the hippocampus of rodents (Dragoi, 2020).

The larger the neuronal recordings become, the harder it becomes to find all the precise temporal spike sequences present (Quaglio et al., 2017). The number of potential spike patterns grows exponentially with the number of neurons in the recording. Moreover, the larger the recordings have become, the larger the identified spike patterns have been, suggesting that the possibility of very large spike sequences should be taken into account by these methods.

While some methods have already been retired since they were not able to cope with even one hundred neurons, others have been developed to meet this challenge. The next section will give an overview over some of these newer methods.

2.3.3 *Detection of spike sequences*

As spike sequences have been found to carry highly relevant task information, the search for them has brought forward a large number of methods. Some of the most recent ones will be described here, demonstrating the need for powerful methods to approach this complex problem.

The first methods to detect repeated spike sequences in recorded spike trains were brute force approaches such as template matching (Abeles and Prut, 1996; Prut et al., 1998). With the growing number of neurons in neuronal recordings, these methods became computationally too expensive to run when analysing more than about 50 spike trains. The same is true for several methodologies based on information geometry (Shimazaki et al., 2012), pairwise correlations (Gansel and Singer, 2012) or even more complex analytical methods such as Formal Concept Analysis (Yegenoglu et al., 2015). Other approaches limited the information they sought after to keep computational resources manageable e. g. identifying synchronous spikes without specifying the neurons from which the spike sequences originated (Grün et al., 2008; Staude, 2010).

In contrast, a number of machine learning-based methods have been introduced in recent years, which promise to be more scalable and therefore more useful in this time of ever larger neuronal recordings.

One methodology currently in use is a method called synchronous pattern detection and evaluation (SPADE) which extracts spike patterns which occur at a high enough frequency to be of interest and tests their statistical significance (Quaglio et al., 2017). SPADE is based on frequency item set mining and is an extension of earlier methods (Torre et al., 2013; Yegenoglu et al., 2015). It improves on computational stability and adds statistical corrections. Its usefulness has been demonstrated in recordings of up to 150 monkey cortical neurons where SPADE identified spike patterns relevant to motor behaviour and task variables (Torre, Quaglio, Denker, Brochier, Riehle and Grün, 2016).

Mackevicius et al. (2012) focus on the sparsity of neuronal spike data and apply seqNMF, a dimensionality reduction method based on non-negative matrix factorisation to extract sequences of neuronal events. Convolutional non-negative matrix factorisation is routinely applied to audio recordings of speech. They successfully identify spike sequences in hippocampal spike data and demonstrate the possibilities for usage on even larger datasets.

The approach SPOTDisClust by Grossberger et al. (2018) combines the computations of a measure of transport costs between spike trains with clustering to arrive at repeated spike patterns. Both parts of this methods have roots in machine learning – the algorithm to compute the transport cost is based on optimal transport theory often used in Computer Vision. The density-based clustering algorithm applied in SPOTDisClust is highly popular in pattern recognition and image analysis. The authors tested their method on a dataset of only 64 neurons but declare that it is scalable to a much larger number of neurons.

The development towards computationally efficient spike pattern detection algorithms is in full bloom. There is no doubt that as machine learning methodologies improve, more and more methods from other computer science-based fields will be applied to neural data. While this is the only way we can deal with this issue of the exponentially growing problem of finding spatiotemporal patterns, we should not forget that the brain also does this task. The ubiquitous presence of spike sequences tells us that the brain is creating and reading out this information seemingly effortlessly.

With that in mind, Masquelier et al. (2008) created a temporal pattern detection methodology which attempts to recreate the way the brain might be finding large spatiotemporal patterns. It is a well known fact that neurons can function as coincidence detectors and this ability might be the readout mechanism for spike sequences (König et al., 1996; Roy and Alloway, 2001). Therefore, Masquelier et al. (2008) built a spiking neural network with a coincidence detecting neuron whose activity became specific to the presence of a certain spike pattern. The authors report robustness to noise, good scalability (circumvention of the explosion of possible pattern combinations) and propose that this method can inform researchers about possible mechanisms of how neurons in the brain might make use of these repeated events. In Chapter 6, I investigate the robustness of this spiking neural network in detail and assess the applicability for spatiotemporal pattern detection. The idea of finding repeating spike sequences in a seemingly brain-like manner is currently being explored further: in a similar spirit to the network men-

tioned above, Montangie et al. (2020) developed a method to find neuronal assemblies via triplets and STDP.

The participation in spike sequences is one way of identifying neuronal assemblies as was illustrated in this section. The next section will introduce the search for assemblies which are based on correlated modulations in firing rate.

2.3.4 *Covariations in firing rates*

Neuronal assemblies have been identified through the repeated spike sequences they take part in as explained in the previous section. The other method of detecting neuronal assemblies is to find them through correlations in their firing rate modulations. In this section, I will present evidence for the relevance of such neuronal assemblies for brain function as well as give an overview over some of the current methods for detecting them.

Multiple studies have linked the firing rate-related activity of groups of neurons to animal behaviour (Dejean et al., 2016; Romano et al., 2015) while other have shown that groups of neurons have superior classification power compared to single neurons when it comes to decoding task variables (Deolindo et al., 2018; Lopes-dos Santos et al., 2013; Russo and Durstewitz, 2017). Firing rates of cortical and some subcortical areas are significantly modulated by the task the animals are given (Deolindo et al., 2014), but also by movements carried out during the completion of the task e. g. fidgeting (Musall et al., 2019).

As mentioned in the previous section, spike patterns may emerge during the learning of a task and then remain stable. Similarly, correlated firing rate dynamics may emerge and disappear over the course of a task (Deolindo et al., 2018) or stay stable over multiple days (Malvache et al., 2016). This emphasises their importance as possible functional units for information processing for encoding and retrieving experiences (Malvache et al., 2016).

Neuronal assemblies consisting of neurons linked by their firing rate dynamics have been shown to increase input selection (Tsitoura et al., 2020) and improve reliability of signal propagation (Buzsáki, 2010). They are also able to carry rhythmic activity in areas such as the hippocampus (Buzsáki, 2010) and in the olfactory bulb (Tsitoura et al., 2020).

2.3.5 *Detection of covariations of firing rates*

Neuronal assemblies have been found to carry highly relevant task information and their detection is therefore a much worked on subject in neuroscience research. It is a highly complex problem due to the exponential possibilities of combinations of neurons which could belong to an assembly. Additional difficulties arise since neurons might be part of multiple assemblies, that neuronal activity includes more spikes than just assembly activity and that assembly activity does not always predict neuronal spike trains (Mölter et al., 2018).

The first methods developed for the identification of neuronal assemblies were based on cross-correlations (Moore et al., 1966; Strangman, 1997). They remain the core principle underlying a large number of methodologies today although they have been significantly improved and extended.

Lopes-dos Santos et al. (2011) apply Principle Component Analysis (PCA) on the binned and z-scored spike trains and assemblies are identified from the resulting covariance matrix. An additional step involves an Independent Component Analysis (ICA) on the principle components in order to distinguish significant assemblies from spurious correlations and to allow neurons to be part of more than one assembly. The method gives one vector per assembly, which indicates each neuron's affinity to the assembly, which are thresholded to create a finite set of neurons per assembly. In a follow-up study, the authors improved the clustering step (Lopes-dos Santos et al., 2013).

A similar PCA-based method has the same starting point as Lopes-dos Santos et al. (2011), but swap the ICA clustering with a Promax oblique rotation (Romano et al., 2015). The method includes adaptations for calcium imaging (adds a pre-filtering of the calcium signals) and adds more statistical tests to distinguish significant assemblies from noise.

Instead of using PCA for the initial dimensionality reduction step on the covariance matrix, Peyrache et al. (2010) apply ICA on the covariance matrix (calculated from z-scored and binned spike trains) in their ICA-MP method. Instead of the circular shifts, meaningful assemblies are estimated by applying the Marchenko-Pastur distribution (Peyrache et al., 2010). As an optimisation, Mölter et al. (2018) selected assembly vectors based on a thresholding with Kolmogorov-Smirnov.

One other cross-correlation-based method to highlight is the work by (Pastore et al., 2018). As many others before them, they initially calculate cross-correlation histograms, but they take into account excitatory and inhibitory links in their recordings. The network topology of the data is estimated using a graph theoretical approach.

One major limitation of PCA-based methods is their linearity and that their performance deteriorates when assemblies overlap in time and space (Strangman, 1997). While many methods try to overcome the latter limitation (Lopes-dos Santos et al., 2011, 2013; Peyrache et al., 2010), the highly complex and non-linear neuronal dynamics slightly question the suitability of PCA-based models for the analysis of neuronal recordings (Deolindo et al., 2018).

Some researchers take a different approach and do not look at each spike train separately, but take into account the activity of all the neurons in the recording at the same time. Such methods estimate the state of the system at selected points in time and are particularly useful for calcium imaging data, since fine temporal resolution of spikes is not available (Mölter et al., 2018).

One approach is frequency item set mining. This method first specifically selects time windows in which there is a lot of co-activity between neurons. Only from those time windows sets of neurons which are active together with a certain threshold frequency are derived (Picado Muiño et al., 2013). The method has been successfully applied to electrophysiological spike recordings to find sequential spike sequences (Quaglio et al., 2017; Torre, Quaglio, Denker, Brochier, Riehle and Grün, 2016), but can also be applied on calcium imaging data (Mölter et al., 2018). Surrogate data is used to distinguish assemblies from noise.

Avitan et al. (2017) developed a method called spectral clustering for the identification of neuronal assemblies. The first step involves searching for time windows in which neurons synchronously increased their firing rates to above double the standard deviation of a uniformly spiking model, and then groups those neurons according to their co-activity. The groups of neurons with such correlated firing rate modulations are fed into a graph to identify community structures. Neurons are assigned an affinity to assemblies, which are thresholded to delineate sharply between neurons belonging to an assembly and the ones that do not.

In the assembly detection method CORE, assembly detection is based on binned time windows of the spike train and which neurons are active in each (Miller et al., 2014). Only instances of time windows in which the neurons were co-active reliably and to a high degree are included in the analysis. Sig-

nificance of the found assemblies was determined by permuting the assemblies and selecting only assemblies surpassing the 95th percentile threshold. Mölter et al. (2018) reported that co-activity patterns found were highly similar and had to be additionally clustered. Again, the affinity of each neuron was thresholded to achieve a binary mapping of which neurons are part of and which neurons are not part of an assembly.

In similar manner to the spectral clustering and CORE, singular value decomposition (SVD) is based on the co-activation patterns of neurons in a binary matrix which holds the activity of the neurons per time window (Carrillo-Reid et al., 2015). Only co-activation patterns which included a large number of neurons and which occurred often were included in the analysis. The relevance of each neuron to an assembly was assessed using a term frequency-inverse document frequency algorithm. After removing redundancies, the singular value decomposition is performed on this matrix to identify the assemblies from the singular values above the chance level.

Instead of focussing only on time window in which some neurons exhibited large firing rate increases, Diana et al. (2019) include all time windows in which there was some at least some activity into their Bayesian inference algorithm. Their generative hierarchical model of synchronous activity represents a clustering of neural activity with minimal assumptions about the data (variable number of assemblies, various levels of assembly activity). Their inference algorithm also provides and estimate of the uncertainty for each assembly. The main drawback is that while the authors claim that Bayesian inference is suited for cases in which neurons belong to multiple assemblies, their performance drops significantly for that particular scenario.

Humphries (2017) approach the issue from a completely different angle and use a network approach for the detection of neuronal assemblies. Their main disadvantage is that they require each neuron to be part of an assembly, which is unlikely to be the case in neuronal recordings (Deolindo et al., 2018).

The next section will identify the some of the limitations in these methods that arise because of the assumptions underlying them and outline how this thesis will contribute to the advancement of the field.

2.4 EXTENSION TO CURRENT STATE-OF-THE-ART

It is evident from the long (but in no way extensive) list of methods in the previous sections, that there are many ways to find neuronal assemblies either

based on their participation in spike sequences or based on co-fluctuations in neuronal firing rates. In this section I want to address certain assumptions about neuronal assemblies made by all of the presented methods and how the work of this thesis intends to challenge some of these assumptions.

2.4.1 *Restriction to two types of neuronal activity*

Currently, neuronal assemblies are searched for either through their participation in spike sequences or their correlated firing patterns. Methods in both groups have made a lot of progress in the last years. Spike patterns have been extended from originally synchronous spikes to spike patterns with latencies between the spikes (Quaglio et al., 2017). The search for neurons with correlated firing rate dynamics has seen a number of approaches which do not reduce whole spike trains to one covariance (Peyrache et al., 2010), but which take into account dynamics on small time scales (Diana et al., 2019). While some still clearly limit their assembly definition to firing rate increases (Carrillo-Reid et al., 2015; Deolindo et al., 2018; Miller et al., 2014), other adopt a more flexible stance and also include firing rate decreases (Pastore et al., 2018).

However, neurons may interact with each other as a neuronal assemblies in more ways than just through those two behaviours (Buzsáki, 2010; Harris, 2005). As explained in Sections 2.1 and 2.2, neurons exhibit many different encoding strategies at the same time (Reinagel and Reid, 2000; Yuste, 2015). Lamentably, this treasure trove of neuroscientific knowledge is not utilised in the search for neuronal assemblies.

In Chapter 3 of this thesis, I will present neural topic modelling, which is able to extract multiple aspects of neuronal spike trains for the detection of neuronal assemblies and how the neurons in each assemblies interact with each other. Neural topic modelling is so generalisable that other data types from the recording, e. g. local field potentials (LFPs) can also be included in the analysis as demonstrated in Chapter 5.

2.4.2 *Multi-memberships*

Neuronal populations forming neuronal assemblies overlap substantially and an individual neuron may be part of multiple assemblies (Allen et al., 2019). Neurons have been found to participate in different assemblies encod-

ing different motor behaviours (Sheng et al., 2019), reward choice (Pfarr et al., 2018) or task cues (Carrillo-Reid et al., 2016). This applies to both assemblies exhibiting spike sequences and firing rate co-modulations. While all modern methods acknowledge that neurons may belong to multiple assemblies and that this needs to be taken into account, they often struggle to allow such multi-memberships in their models (Lopes-dos Santos et al., 2013). 'other algorithms exhibit a drop in performance as a direct result of the presence of such neurons participating in multiple assemblies (Diana et al., 2019).

The more neurons are present in a recording, the higher the chances are that overlapping assemblies are present (recording from multiple assemblies in a recording of only a few neurons is quite unlikely). Therefore any new method trying to identify neuronal assemblies should have neurons with multi-memberships in assemblies as an assumption in its model.

Neural topic modelling, the assembly detection algorithm presented in Chapter 3 has this assumption of multi-membership built in due to its origin in text mining. In a collection of documents, groups of words frequently overlap (Blei, Ng and Jordan, 2003). In Chapter 4, I put this assumption to the test and demonstrate the robustness and limits of neural topic modelling. Indeed, neural topic modelling handles multi-memberships of neurons better than other methods (as demonstrated in Section 4.3).

2.4.3 *Limiting assemblies to a finite set of neurons*

Neuronal assembly detection methods based on co-modulations of firing rates aim to identify a finite set of neurons which are in the assembly while all other neurons are not part of it. For the most part, this is a simplification which makes it easy to obtain a clean and easy to visualise image of different assemblies found in the model. However, this does not reflect the reality of the results from these methods. Many approaches actually yield one vector per assembly attributing a weight to each neuron as a measure of how tightly that neuron is bound to that particular assembly. This distribution of weights is then thresholded to obtain a binarised version of the vector. Common thresholds are two standard deviations above the mean weight (Lopes-dos Santos et al., 2013) or a data-driven threshold (Romano et al., 2015). At the same time, in a review of six different assembly detection methods, Mölter et al. (2018) report that automatic thresholding often did not give satisfactory results and had to be done by hand.

None of the papers asks the question whether thresholding is actually a good idea. It has been suggested before that neuronal assemblies are not a set of neurons which all uniformly belong to an assembly in the same way, but that some neurons might participate in assembly activity more than others (Buzsáki, 2010).

The assembly detection method developed during the course of this thesis, neural topic modelling, avoids taking such a harsh binarisation as described in Chapter 3. The advantage of holding off thresholding is further demonstrated in Chapter 4 where I show that complex hierarchies of neurons within an assembly can be recovered from the data.

2.4.4 *Scalability*

The size of neuronal recordings is steadily increasing (Musall et al., 2019). Scalability will be a large influence on whether a method can be used for years to come or whether it will need to be rapidly replaced by other methods which are computationally more efficient.

In a very positive development, research groups are making their large data sets publicly available (Allen Brain Observatory, 2019; Musall et al., 2019; Stringer et al., 2018). Thanks to open science efforts, the analysis of massive data sets is no longer the privilege of a few laboratories, but any person may analyse those datasets given they have the right tools to their disposal. Many papers have also made their code available and a large number of open source software tools have been released to analyse spike trains and other electrophysiological brain signals: Elephant (Python) (Yegenoglu et al., 2015), allensdk (Python) (Allen Brain Observatory, 2019), neo (Python) (Zhang et al., 2017), Spyke Viewer (Python) (Pröpper and Obermayer, 2013) SPIKY (Python) (Kreuz et al., 2015), Brainstorm (Matlab) (Tadel et al., 2011), Chronux (Matlab) (Partha and Bokil, 2008) and FieldTrip (Matlab) (Oostenveld et al., 2011).

Chapter 3 outlines neural topic modelling, a new open-source method for neuronal assembly detection, which is scalable to large datasets and has great potential for optimisations.

In turn, Chapter 6 describes the replication of a spiking neural network which proposed a computationally cheap way to identify spike sequences specifically. This open source implementation in Python is a replication of

a study for which the original code was written in Matlab and not publicly available.

2.4.5 *Biological interpretability*

Most of the methods identifying neuronal assemblies by their firing rates mentioned in this literature review do not have the explicit goal to arrive at their conclusions in a biologically interpretable way. Given the complexity of the problem – finding spike sequences or co-activations of variable size, length and delay – this is not surprising. Nevertheless, it would be very helpful to generate hypotheses about how the brain might achieve such an impressive feat.

Neural topic modelling, as presented in Chapter 3, aims to extract simple, biologically interpretable features from spike trains without dimensionality reductions that make interpretation more difficult (e. g. the second principle component of the covariance matrix). Not only that, but the detection of the neuronal assemblies is done in a transparent and interpretable way through a generative model.

The spiking neural network described in Chapter 6 presents an approach which attempts to identify spike sequences in a spiking neural network using coincidence detection, which is an ability known to exist in neurons in the brain (Roy and Alloway, 2001).

2.5 CONCLUSION

Large recordings have already started to provide a large number of insights we have never been able to access before (Musall et al., 2019; Stringer et al., 2018). The search for organisational structures such as neuronal assemblies needs to keep up with the improvements in the field of recording techniques.

This chapter has provided an overview over what kind of information single neurons are able to encode and how they employ multiplexing to encode multiple variables at once. It described the two main ways in which neural assemblies are currently defined and searched for: through spike sequences or correlations in spike rate time series. For each approach, popular methods for the detection are delineated. Finally, I make suggestions for improving the search for neural assemblies by widening what aspects of neural activity

can be used for defining neural assemblies and by softening the hard split between assembly neurons and non-assembly neurons.

3

NEURAL TOPIC MODELLING – CONCEPT AND IMPLEMENTATION

In this chapter, I describe neural topic modelling – a novel neural data analysis method developed during the course of this thesis. I will introduce a more general definition of a neuronal assembly and show that a technique from text mining, topic modelling, can be used for neuronal data analysis.

In its current implementation, neural topic modelling extracts four different types of neural activity events from electrophysiological recordings. From the spike trains of the neurons, time windows with particularly high and particularly low firing rates are identified (see Section 3.2.2.1) as well as time windows in which pairs of neurons exhibited a higher-than-expected number of synchronous spikes (see Section 3.2.2.2). From the local field potential (LFP) signal, the time windows with the highest amount of power in five different frequency bands are extracted for each brain region on each electrode (see Section 3.2.2.3). Next, the probabilistic generative model Latent Dirichlet Allocation (LDA) (see Section 3.2.3) is applied to the resulting neuronal event-time window matrix to uncover hidden topics or themes in the data i. e. groups of neuronal events which co-occur across time windows. To increase the quality of the topics found, the results from multiple random initialisations are pooled to filter out noise and to extricate topics which are likely to represent neural correlational structures (see Section 3.2.4).

3.1 INSPIRATION

The concept of groups of neurons being physically or functionally organised in neuronal assemblies is very old on the relative time scale of neuroscience research (Hebb, 1949). Initially, limitations on the number of neurons per recording made the search for such assemblies all but impossible. When recording techniques improved and enabled the recording of larger numbers

of neurons, the limitations shifted to the lack of methods with which to detect assemblies as outlined in Section 2.3.1 of the previous chapter.

The larger data sets are forcing neuroscientists to look beyond their own field to come up with ways to cope with these large amounts of data. There is a desperate need for scalability. I therefore looked beyond the neuroscience horizon and searched for suitable methods in fields which routinely deal with such large amounts of data. Text mining is such an area where data sets might include all the words in whole books, the entirety of Wikipedia or the millions of tweets with the hashtag *#blacklivesmatter* or *#womeninstem*. Researchers in other fields have treaded this path before me and applied text mining tools – and in particular the topic model LDA – to genomic data (Liu et al., 2016), medical records (Doshi-Velez and Kim, 2017; Hughes et al., 2018) and fMRI signals (Koyejo et al., 2013). I therefore decided to take advantage of the scalability and the interpretability of LDA and figure out how to apply it to neural datasets.

LDA is best explained through an example of how it works on text data. Let's assume LDA is applied to a set of documents:

- Document 1: Blossom is the leader of the Powerpuff Girls, but Buttercup is my favourite.
- Document 2: To complete a PhD in computational neuroscience one needs a powerful computer.
- Document 3: Powerpuff Girls airs every Monday at 9 am on CN.
- Document 4: To better deal with the high computational demand in the Neural Reckoning research group, three new computers were bought to act as servers.
- Document 5: The PhD students in the lab chose Blossom, Bubbles and Buttercup as names for the new servers.
- Document 6: A PhD student from the Neural Reckoning lab gave a talk on computational neuroscience at an all girls' school.

In this example, LDA might find two topics – groups of words – that could be labeled as "related to cartoons" and "related to computers" even though LDA does not provide the label. Each topic is a probability distribution over the words in the dataset, so the found topics might resemble the following:

- Topic 1: Powerpuff(0.23), Blossom(0.2), girl*(0.16), Buttercup(0.1), ...
- Topic 2: PhD(0.22), comp*(0.19), server(0.17), girl*(0.1), neur*(0.05), ...

It is important to note that the word "girl*" is part of both topics since it appears in both contexts (computers and cartoons). The topic model will therefore assign the word "girl*" a relatively high weight in both topics. Additionally, each document will be assigned a number of how much of each topic seems to be represented in the document:

- Documents 1 and 3: 92% topic 1; 8% topic 2
- Documents 2 and 4: 1% topic 1; 99% topic 2
- Document 5: 60% topic 1; 40% topic 2
- Document 6: 11% topic 1; 89% topic 2

There are a number of ways the concepts from text mining (words, documents, topics) can be mapped to neural data. Table 3.1 juxtaposes the concepts from text mining and how they will be used in this thesis for neural data analysis. When applied to text data, topic modelling uses a vector of how many times each word appears in each document as its input. Here we generalise this to a document being a window in time and a "neural word" being a feature of neural data such as an elevated firing rate in one neuron, synchronous spikes in pairs of neurons or high power in the beta frequency band in the LFP signal. The step of extracting such neural words from neuronal recordings is referred to as "event filtering" throughout this thesis.

Table 3.1: **Translation of concepts from text mining to neural data analysis**

text mining	neural data	description
document	neural document	time window of recording
word	neural word	feature of the spike train such as "neuron A has a high firing rate"
topic	neural topic	neuronal assembly in which neurons interact with some part of their activity

This event filtering for the creation of the neural words opens up a lot of possibilities. Any feature of a spike train can become a neural word in this way: firing rate modulations, bursting, timed spikes with other neurons and even abstract concepts such as inter-spike interval (ISI) variability. Moreover, other kinds of data can also be converted into such a format, e. g. aspects of the LFP signal.

As a consequence of the flexibility of neural topic modelling, the method is not strictly bound to the current definitions of neuronal assemblies as either a group of neurons who participate in repeated spike sequences (Prut et al., 1998; Quaglio et al., 2017) or as a group of neurons who exhibit comodulations of their firing rates (Avitan et al., 2017; Lopes-dos Santos et al., 2011; Mölter et al., 2018). Researchers acknowledge that neuronal assemblies should be defined in a broader sense to include neuronal assemblies in which neurons interact through other means, but the assembly detection methods are slow to adapt to this (Dragoi, 2020). Why neuronal assemblies have traditionally been defined in these two particular ways and why current detection methods are limited to them is outlined in Section 2.3.1 of Chapter 2.

Topic modelling results in a number of topics, which are probability distributions over the words in the input matrix. In other words, each topic ascribes each word a weight for how important that word is for this particular topic. For neural data, this means that every neural topic includes a weight mapped to each activity pattern extracted from the data. These neural topics can be used to gather ideas about assemblies in the data.

In this chapter, I will outline the steps involved in neural topic modelling. Section 3.2.2 describes how the neural data is divided into time bins and goes into detail about how certain activity patterns are extracted from the binned spike trains. The subsequent topic modelling and clustering of the results are mapped out in Sections 3.2.3, 3.2.4 and 3.2.5. Finally, I will discuss some of the advantages and disadvantages of the components of neural topic modelling and mention which shortcomings of current methods it addresses and the challenges it faces in Section 3.3.

3.2 PIPELINE

Neural topic modelling aims to identify neural assemblies in large-scale neural datasets. These neuronal assemblies can include the neurons whose firing rates are correlated (as many assemblies which have been identified before), but can also extract multiple spike train features to broaden the idea of how activity from different neurons relates to each other.

Neural topic modelling involves multiple steps: dividing the recording into time bins, event filtering to generate an input matrix, running multiple initialisations of fitting the input data using a generative model and clus-

tering the results. Each of the steps will be described in more detail in the following sections.

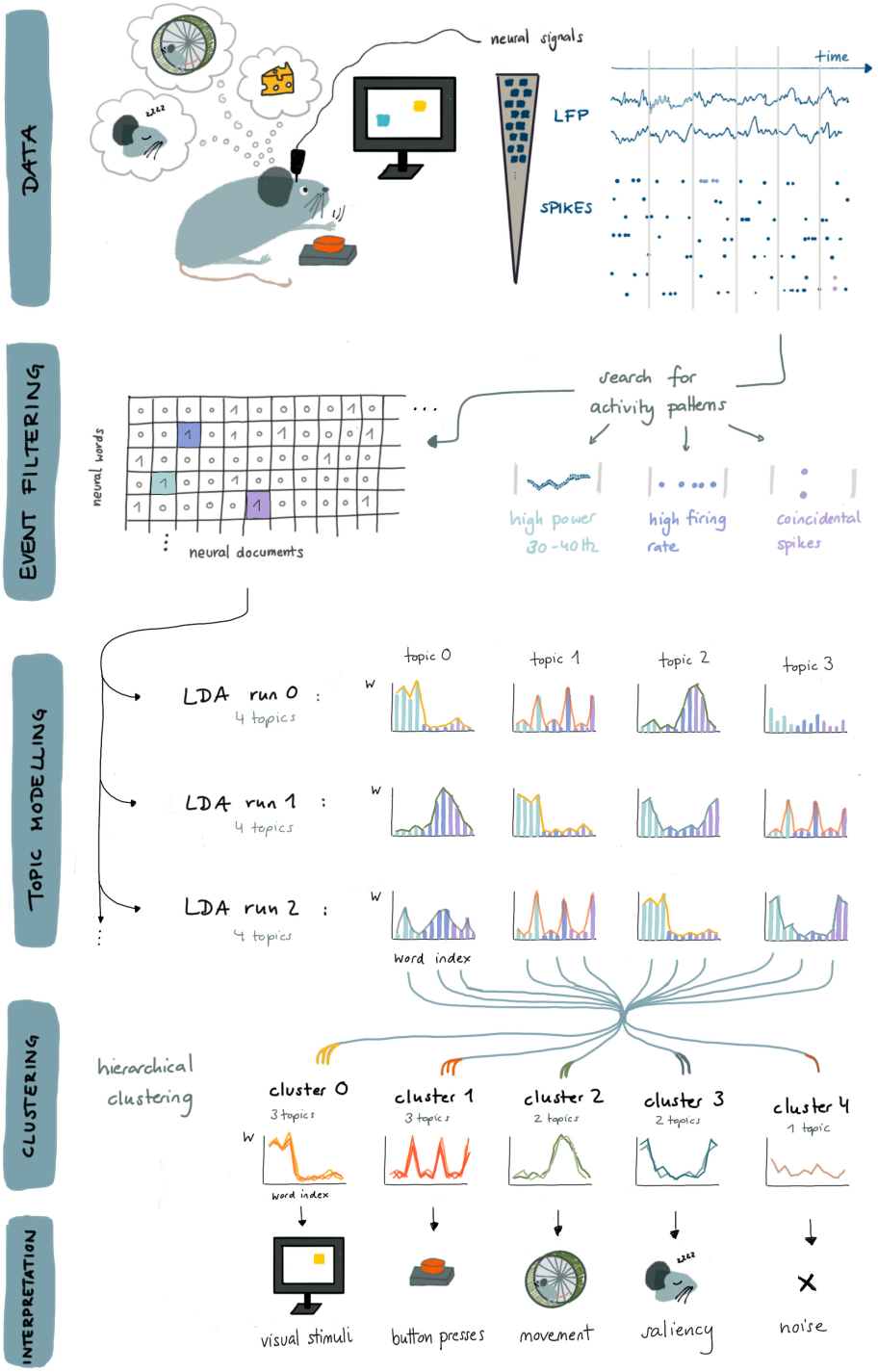


Figure 3.1: **Neural topic modelling pipeline.** Visualisation of the steps involved in neural topic modelling: the event filtering of the neural data, the topic modelling, the clustering of the found topics, and the interpretations of the clusters. LFP = local field potential; LDA = Latent Dirichlet Allocation

3.2.1 *Neural documents: time windows*

To translate neural data into a format that can be used in topic modelling, the recording time needs to be divided into time windows (which are equivalent to different documents in text mining). These time windows then form the basis on which spike features are calculated e. g. spike count per spike window (see Section 3.2.2.1), number of coinciding spikes between pairs of neurons (see Section 3.2.2.2) or power in the alpha frequency band in the LFP signal (see Section 3.2.2.3). The time windows can be aligned to the start of a stimulus block, but do not have to be. Time windows are chosen to be of the same length throughout so as to guarantee comparability between different time bins across the whole recording.

The length of the time window can be set depending on the nature of the data. The length needs to represent a good balance between capturing fast changing activity and allowing spike train features to unfold in the time bin while allowing activity features to be slightly delayed relative to each other (Ozden et al., 2008). Spike train features such as firing rate modulations and sequential spiking usually occurs on shorter timescales (tens of milliseconds) favouring shorter time bins (Butts et al., 2007; Dragoi, 2020), while other data types such as LFP prefer longer time windows since they include slower dynamics.

In the following section I will describe how different spike train features were derived for each time window for each neuron.

3.2.2 *Neural words: event filtering*

As many studies have demonstrated, neurons encode different internal or external variables with different aspects of their spike trains. This is not limited to primary sensory neurons, but also applies to neurons in higher processing areas in which multiple inputs – motor activity, behavioural or motivational states – need to be processed at the same time. There is a lot of evidence that neurons encode different variables using different encoding strategies, a phenomenon called multiplexing (Hardcastle et al., 2017; Stringer et al., 2018). For a more detailed description of the literature on multiplexing, please refer to Section 2.2.

Extracting these different types of events from the same spike train makes it possible to study their interactions in neural activity with neural topic

modelling. This section will describe the event filtering – which in the context of neural topic modelling represents the creation of neural words – for three different types of spike train features: a particularly high or particularly low firing rate (see Section 3.2.2.1) and coincidental spiking (see Section 3.2.2.2). Since neural words do not have to be based on the spike train themselves, but can also be extracted from different types of neural signals, I present the power in different frequency bands in the LFP signal as a fourth example for a neural word (see Section 3.2.2.3).

3.2.2.1 *High or low firing rate*

Modulations of their firing rate is an important way in which neurons convey information (Allen et al., 2019; Harris, 2005; Hubel and Wiesel, 1961). This is the reason why most assembly detection algorithms are based on the firing rates of neurons or the pairwise correlations thereof. Neurons display preferences for some stimuli, which they express with a high firing rate, or significantly lower their firing rate for other stimuli they are not attuned to. Therefore I am aiming to extract time windows in which neurons exhibited a particularly high or a particularly low firing rate compared to the spike counts in the other time windows. This should filter out time windows with high information content while ignoring the rest.

The simplest way to estimate the firing rate of a neuron in a time bin is to count the spikes which occurred in that time bin. Other assembly detection methods also start out with this step (Lopes-dos Santos et al., 2011, 2013).

To create the neural word of "high firing rate" (this will be called *sc+* in Chapter 5, short for "spike count +"), I first calculate the number of spikes in each window for each neuron. Then I have to set a threshold to determine during which time windows each neuron had a "high firing rate". Only windows with a spike count higher than this threshold will be recorded as having a high firing rate. These thresholds are based on percentiles, which can be adjusted as a parameter. During the course of this thesis, I use a percentile threshold of 75th percentile for the high firing rate neural word i. e. I only registered the event "neuron 1 high firing rate" if the spike count is higher than in 75% of all time windows of neuron 1. Similarly, I decided on a 25th percentile for the low firing rate neural words i. e. I only registered the event "neuron 1 low firing rate" if the spike count is lower than in 75% of all time windows. Registering an event in this context means placing a 1 into the row of that word and column of that time window in the neural word-document matrix. To be clear, I did not include 25% of all time windows of

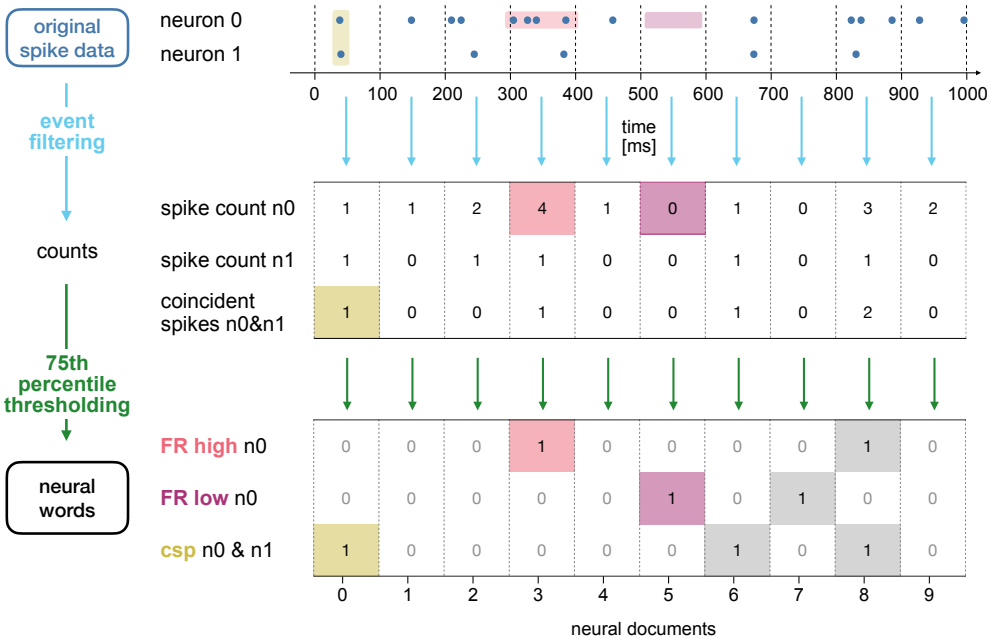


Figure 3.2: **Neural word creation from spike data.** The time of the recording is divided into time bins (neural documents) – in this example the windows are 100 ms long. In the first step, certain aspects of these spike trains are calculated per time bin e. g. the spike count per time window per neuron or the number of synchronous spikes between each pair of neurons per time window. Finally, a thresholding step is applied, so that whenever an event (e. g. the spike count) happened above a threshold (e. g. spike count was above 2 spikes, where 2 spikes represent the 75th percentile threshold), then a 1 was recorded for that neural word in those documents. The example neural words used here were: high firing rate (pink), low firing rate (purple) and coincidental spikes (yellow). no = neuron 0, n1 = neuron 1, FR = firing rate, csp = coincidental spiking

each unit in the analysis. Rather, only events that occur at most in 25% of time windows are included. Therefore no events would be recorded from perfectly tonically firing neurons.

This threshold selection has two effects: it ensures that I only capture the firing rates at either end of the spectrum and it creates a hard upper bound by restricting the number of 1's to a maximum of 1/4 of all documents. The former makes sure that each neuron is judged relative to its own activity. The latter reduces the noise in the system by only taking into account neural words which are likely to convey information. In text data, noise reduction is highly important and so-called stop words – words that are not specific to a theme e. g. word such as "to", "be", "or", "not", "two", "but" – are routinely excluded from analysis.

3.2.2.2 *Coincidental spiking*

Another feature of neural spike trains known to convey information in the brain are coincidental spikes between two or more neurons. As an example, motor neurons encode information through spike synchrony which is not recoverable by their firing rates only (Cohen and Newsome, 2008; Perez-Orive et al., 2004; Riehle et al., 1997). Even short spike sequences involving 2 – 8 spikes carry task-relevant information (Prut et al., 1998). The non-trivial search for synchronous spikes between more than a few neurons has brought forward a number of different methods (Mackevicius et al., 2018; Masquelier et al., 2008; Quaglio et al., 2017; Torre, Canova, Denker, Gerstein, Helias and Grün, 2016). For neural topic modelling I restrict myself to coincidental spikes between pairs of neurons as the simplest multi-neuron neural word.

It is up to the user how strict or how lenient she[1] wants to define coincidence in this context. She needs to provide the neural topic modelling algorithm with a time window (in the order of milliseconds) representing the longest time between two spikes which count as coincident. Additionally the user determines a percentile threshold similar to the high / low firing rate neural word described in the previous section. The percentile calculation works slightly different here, since the number of coincident spikes in a certain time window is not compared to the rest of the time windows as in the previous section, but is compared to the expected number of coincident spikes, had all the spikes in the time window occurred at random. This served the purpose of only recording an event when the number of coincidences surpasses the e. g. 75th percentile of expected coincidences when these spikes occur at random times. The expected number of coincidences is pre-calculated by going through the possible pairs of numbers of spikes in a time window (0 to 100 for each neuron) and placing those numbers of spikes into a 100 ms time window uniformly. This process is repeated 10000 times and the 75th percentile is saved for later comparison in a lookup table.

For the sake of neural topic modelling, the algorithm goes through the two spike trains of a pair of neurons and counts the number of coincidences

1 In this thesis, I have decided to use "she" when referring to individuals instead of "he" or "they". While "they" is an attempt to be more gender-neutral and has many merits as a singular pronoun, its use can lead to ambiguities. The use of "she" is not gender-neutral, but when contemplating that the also non-gender-neutral pronoun "he" has been in use for centuries and is still currently the norm, the explicit use of "she" may serve as a tiny attempt to create some balance. If the use of "she" highlights the unfamiliarity of seeing the female pronoun being used in this way and a feeling that this pronoun does not represent all individuals, then this might lead the reader to contemplate sexism in the use of language – a goal as important as neural data analysis.

within each 100 ms time window. This number is then compared to the expected number of coincidences based on the pre-calculated coincidence count threshold described above. For example the expected number of coincidental spikes at the 75th percentile when one neuron spikes once and the other neuron spikes three times is zero. If the spike from neuron 0 did indeed happen within 4 ms of one of the spikes in neuron 1, then this time window is given a 1 in the row of this neural word and the column of that particular time window in the neural word-document matrix.

3.2.2.3 *LFP power spectral density*

The feature from the LFP signal I selected to use as a neural word is the power spectral density in different frequency bands for different brain regions.

In the data set used in this thesis, LFP signals are available from every 4th Neuropixel electrode (Allen Brain Observatory, 2019). Each electrode therefore yields about 90 LFP time courses. LFP time courses of channels not inside the brain are discarded. The remaining time courses exhibit high correlations with each other when the channels were located in the same brain structure. To avoid highly correlated neural words from the LFP signal, I calculated an average time course per brain region on the electrode. These regions may be given by the data (e. g. the Allen Institute data includes the substructure in which a channel is located such as CA1 or dentate gyrus for neurons in the hippocampus) or can be set by the user by segmenting the time courses by hand. The goal is to achieve time courses per region which are as de-correlated as possible to not introduce artificial correlations into the topic model.

In the next step, the power spectral density of different frequency bands is extracted for each time window by calculating a spectrogram. More specifically, a Fourier Transformation is applied to the LFP signal in different time windows which correspond to the time bins used to extract the spike train features described in the previous two sections. The time window size of 100 ms used for the spike train feature extraction is relatively small for an LFP signal analysis. The standard procedure in a situation such as this is to calculate the Fourier Transformation on overlapping time windows (Worrell et al., 2012). Therefore, I decided to extend the time window (of 100 ms) by 100 ms before and 100 ms after and calculate the Fourier Transformation on that 300 ms window which is centred on the 100 ms time window of interest. To simplify the calculation of the power spectral density, I resampled the

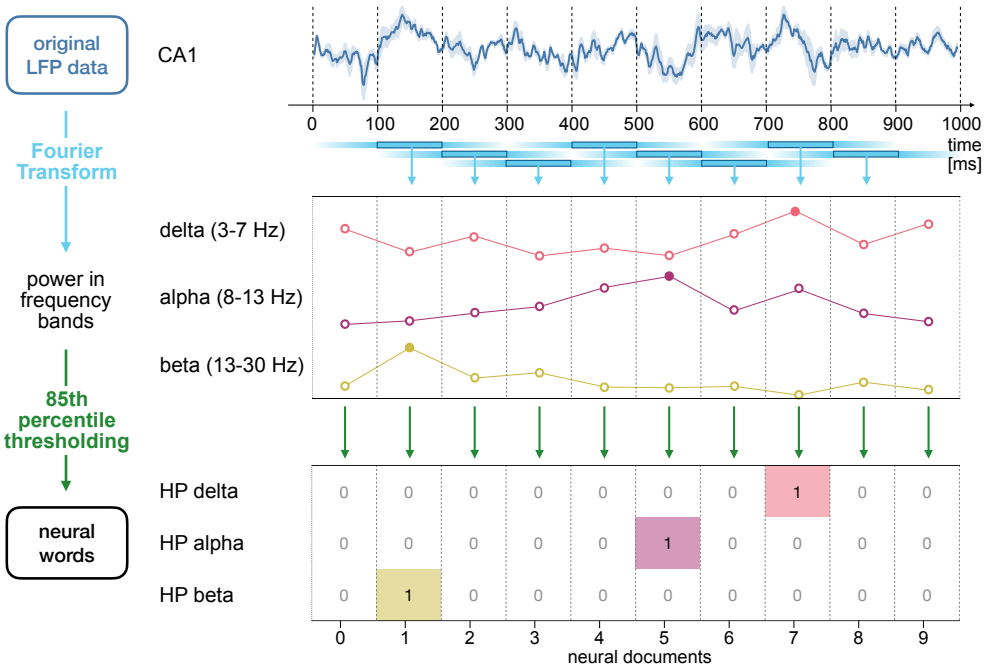


Figure 3.3: **Neural word creation from LFP data.** The time of the recording is divided into time bins (neural documents) – in this example of 100 ms length. A Fourier Transform is applied to time windows of 300 ms length, each of which is centred on the 100 ms neural documents (light blue stripes). From the resulting spectrogram, the power spectral densities in five frequency bands is averaged (shown here are only three: delta (pink), alpha (yellow) and beta (yellow)). Finally, a thresholding step is applied, so that whenever the power spectral density in a certain time window was above the 85th percentile threshold of that frequency band, then a 1 was recorded for that neural word in those documents (filled circle in the frequency band time courses in pink, purple and yellow). As an example, the LFP signal from CA1 in the hippocampus is shown and example time courses for the power in the different frequency bands are shown. HP = high power

LFP signal from 1250 data points per second to 2500 data points per second before the calculation of the Fourier Transformation.

Applying a Fourier Transformation on the LFP signal gives a spectrogram – an indication of the power spectral density per frequency in the original signal. Traditionally in LFP signal analysis, power is averaged over certain the frequency bands: delta (3-6 Hz), alpha (6-13 Hz), beta (13-30 Hz), high beta (30-40 Hz) and gamma (40-70 Hz)[2]. This separation into frequency bands here is very crude because of the low frequency resolution of the spectrogram, so the frequency bands should only be taken as concepts, not at their exact values. The cutoff at 70 Hz was chosen here because extracting higher

2 The choice of boundaries for frequency bands can be a hot topic for debates. The values chosen here should be seen as guidelines and not as absolute values, since the frequency resolution is not high enough to make distinctions e. g. between 8-14 Hz and 8-15 Hz for the alpha frequency band.

frequencies is unreliable due to the resampling mentioned above and uninteresting since very high frequency bands in the LFP signal contain mostly spikes, which have already been processed in the spike trains. In this case, the frequency resolution of the spectrogram was 3.33 Hz. The power spectral density per frequency bands was therefore extracted from the average time courses in each frequency band.

This analysis results in a relatively continuous signal per region per frequency band with one value per 100 ms time window. To determine in which time windows the neural word of high power spectral density occurred, neural topic modelling again relies on a percentile given by the user. I decided on the threshold being all values above the 85th percentile power spectral density per frequency band time courses. Time windows with a value higher than this threshold are marked as having that neural word occurring and receive a 1 instead of a 0 in the input matrix.

3.2.2.4 *Combining neural word types*

One neuron can give rise to multiple neural words i. e. high firing rate and low firing rate and coincidental spiking with all other neurons. Therefore the resulting neural word-document matrix that is the input for the topic modelling will most likely be larger than the number of neurons in the dataset. The rows of the matrix each represent one neural activity pattern with 0's in the documents (time windows) in which this pattern did not occur and 1's in the documents in which it did.

To slim down this input matrix and to reduce runtime of the algorithm, it is advisable to throw out neural words which occur less than 0.5% of all time windows. If a large number of neural words remain, it is possible to restrict the input to the top 5000 or 10000 neural words which occur with the highest frequencies. Since I had restricted the number of coincidental spiking words as mentioned above, I did not need to apply such a global restriction on the number of words to include.

3.2.3 *Topic model*

Topic modelling was originally developed to discover latent variables in large bodies of texts – uncovering the themes which underlie the data. Examples of texts that have been used for topic modelling are diverse: the entire history of articles in the New York Times, all Wikipedia articles (Hoff-

man et al., 2013) or twitter posts about flu to model health trends (Paul and Dredze, 2014). Topic models infer topics present in the data in an unsupervised way without the need of specifying them in advance.

Probabilistic generative models – to which topic models belong – approximate lower dimensional representations of the given data (here: topics) by simulating the process of generating the data from the hidden variables. The following section will go into more details about this process and how hidden variables are inferred.

Because of their simplicity and the interpretability of the results, topic models have become very popular outside the field of text mining. For example, in Computer Vision topic models are used to find useful features of images (Luo and Maunsell, 2015). Other fields of research such as biology and medicine have caught on to the usefulness of this familiy of models. Topic models are used for the analysis of gene expression data (Caldas et al., 2009), DNA sequencing (Chen et al., 2010, 2012), protein sequences (Pan et al., 2010) and protein interactions (Rogers et al., 2005). Patterns in much higher level health data have also been discovered using topic models e. g. predictions of responsiveness to anti-depressants (Hughes et al., 2018) and behavioural markers in patient monitoring (White, 2017). Moreover, researchers have converted time series data such as Magnetic Resonance Imaging (Castellani et al., 2010; Koyejo et al., 2013) and electroencephalography (Lara et al., 2014) into forms which can be analysed with topic models.

Latent Dirichlet Allocation

LDA is an example of a topic model and a probabilistic generative model (Blei, Ng and Jordan, 2003). It was developed as an improvement of probabilistic latent semantic analysis (Hofmann, 2001), which in turn was an improvement of the (non-probabilistic) latent semantic indexing (Deerwester et al., 1990). Its goal is the discovery of latent topics in a dataset of text documents in an unsupervised manner and is therefore particularly effective for large datasets of discrete data (Blei, Ng and Jordan, 2003). Probably the most commonly used topic modelling algorithm is Latent Dirichlet Allocation developed by Blei, Ng and Jordan (2003), which is also the one used in neural topic modelling.

LDA receives a "bag of words" as input – a matrix with the counts of each word in each individual document. Neither the order of the words nor the order of the documents are taken into account. LDA assumes that there

exist hidden variables which explain the observed frequencies of words in the documents. It further assumes that multiple topics are present in each document and that the words in a document reflect this. The goal of LDA is to estimate the topics which give rise to the observed words in the documents by assuming that each topic is a probability distribution over words. It also estimates which topics are present in each document. LDA therefore infers two variables: the topic representation of each document (θ) and the distribution over words in each topic (β). The notation for LDA was chosen to match the original publication (Blei, Ng and Jordan, 2003) and all subsequent publications by the same author. For a table of the variables included in LDA see Table 3.2.

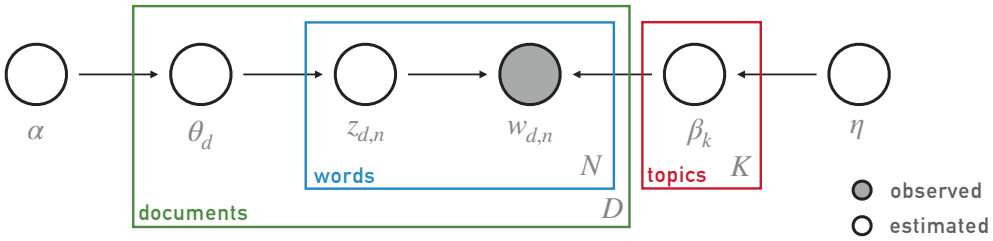


Figure 3.4: **LDA graphical model.** LDA aims to generate the given data through a generative process based on hidden (white nodes) and observed (grey nodes) variables. See Table 3.2 for a description of all parameters in this figure. Boxes indicate repeated processes and the letter in the right corner of each box gives the number of repetitions. Arrows signify dependencies.

Since LDA is a generative model, it infers these variables through a generative process – it tries to re-create the data from the estimated hidden variables and adjusts the hidden variables in order to maximise the (marginal) log likelihood of the data. The probabilistic graphical model of LDA is shown in Figure 3.4 and visualises the generative process. Importantly it visualises the dependencies of the variables, which are the key assumptions at the basis of LDA.

The generative process can also be described in words:

- Pick the number of words N to be in the document according to a Poisson process ξ
- Choose a topic mixture to be present in the document according to a Dirichlet distribution θ over a set of K topics (the number of topics K is fixed)
- For each word in the document do the following:
  - Choose a topic k according to the multinomial distribution (θ) from the last step

– Choose a word w from this topic according to the topic's multinomial probability β

The priors for the distribution over topics in each document and for the distribution over words for each topic are assumed to be Dirichlet distributions which depend on the parameters α and η respectively. In LDA, both α and η are initialised as small values (i. e. as $1/K$) which leads to sparse probability distributions of θ and β. In other words, the model initially assumes that only a few topics are present in each document and only a few words dominate each topic. It is worth noting though that the Dirichlet distribution will never assign a probability of 0 even though it may come close (Blei, 2012).

Table 3.2: **Description of parameters for LDA.**

parameter	description
α	Dirichlet prior for θ
θ	per document topic distribution
z	document-specific topic
w	topic-specific word
β	per topic word distribution
η	Dirichlet prior for β
K	number of topics to be found
N	number of words in the data
D	number of documents in the data

In order to be able to estimate the hidden variables through a generative process, a joint probability distribution over the observed and the hidden variables is formulated.

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D}) = \prod_{i=1}^{K} p(\beta_i) \prod_{d=1}^{D} p(\theta_d)(\prod_{n=1}^{N} p(z_{d,n}|\theta_d)p(w_{d,n}|\beta_i, z_{d,n})) \tag{3.1}$$

This joint distribution is used to calculate the conditional distribution (posterior distribution) of the hidden variables given the observed variables (Blei, 2012).

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}|w_{1:D}) = \frac{p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, w_{1:D})}{p(w_{1:D})} \tag{3.2}$$

The conditional probability given in equation 3.2 cannot be computed and needs to be estimated since the denominator represents the summed joint distribution over all possible hidden topic structures – a number much too large to compute (Blei, 2012). Multiple approximation methods have been developed to estimate the denominator and with that estimating the posterior distribution. For neural topic modelling, I chose to stick with the deterministic Variational Bayes (VB) approach which poses the estimation as an optimisation problem: VB finds the best estimator from a set of parameters (Blei, Ng and Jordan, 2003; Hoffman et al., 2010; Jordan et al., 1999; Wainwright and Jordan, 2008). Other estimation algorithms include sampling-based methods such as Gibbs sampling (Steyvers and Griffiths, 2007). While there is much debate about which inference algorithm should be used for the estimation, there is evidence that the results do not differ much (Asuncion et al., 2009). The topic modelling LDA algorithm used for neural topic modelling is the implementation in the Python Scikit learn package (Pedregosa et al., 2011), in which the VB algorithm is already embedded.

Regardless of which inference algorithm is used, LDA gives two distributions: the distribution over topics in each document (θ) and the distribution over words in each topic (β). The generative process by which it arrives at these distributions is transparent and comprehensive. Another point which makes topic modelling stand out is the assumption that each document is a mixture of different topics, which is the case for many real datasets (Liu et al., 2016). For these reasons, topic modelling is valued for its potentially interpretable results – especially in contrast to other "black-box"-y machine learning techniques.

3.2.4 *Clustering*

Since the exact results of LDA depend on the randomness introduced during training and inference (Blei, Ng and Jordan, 2003), two runs of LDA may lead to two sets of topics which have a number of topics in common but where several topics are only found by either run. Additionally, the more overlap there exists between the assemblies, the less likely it is that one run finds the optimal solution. To circumvent these issues, I decided to run LDA on the same input multiple times (20) with different random seeds. There will be some topics which appear in each of the 20 runs, while others come up in half of the runs and others arise maybe once or twice. I assume that the topics which appear more often in different runs are more likely to be based on an

underlying correlational structure of the neural words. The number of times a topic appeared can provide an idea of the robustness of the cluster. The performance of neural topic modelling for identifying assemblies increased significantly compared to the best result from a single run.

To demonstrate this, neural topic modelling was applied to an artificial dataset which included five known assemblies. As a measure of performance, I assessed the similarity between the assemblies found and the ground truth assemblies using a Jaccard similarity (the intersect of two sets divided by their union). A performance of 1 indicates that all neurons were assigned to their correct assemblies and a performance of 0 indicates that the found assemblies did not resemble the ground truth assemblies at all. Figure 3.5 shows the performance of neural topic modelling as a function of how many words are in the artificial dataset. The "selected" (i. e. clustered and thresholded) performance is always above the best single result. The best result from any single result is included in all figures in Chapter 4 as grey lines.

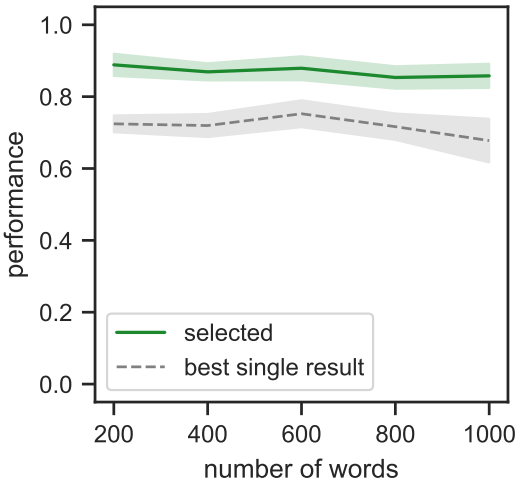


Figure 3.5: **Performance improvement through cluster selection.** Running LDA multiple times and then selecting the clusters which appear most consistently ("selected", green) improves performance in comparison to selecting the best performance of the runs ("best single result", grey).

Pooling the results from runs with different random seeds has the additional advantage of making the choice of number of topics to find (k) less important. The number of topics k is a main parameter in topic modelling and its choice is debated (Blei, 2012). Choosing k so that it is smaller than the number of ground truth assemblies in the data will by definition not lead to the discovery of all the assemblies present. Choosing k so that it is much

larger than the number of ground truth assemblies in the data will lead to the identification of many topics likely to represent noise.

The number of topics to be found is not known in unsupervised learning settings, and is certainly not available for neuronal data sets. Running neural topic modelling multiple times and only including clusters which appeared consistently results in a good performance (identifying all ground truth assemblies in the data) as long as k is similar to the correct number of assemblies as can be seen in Figure 3.6. The number of topics to be found does not necessarily have to be greater than the number of assemblies (i. e. 20 topics are still reliably identified at k = 15) in the data since robust topics will be found consistently enough in multiple runs with different random seeds. There seems to be no direct downside to k being even four times larger than the real number of assemblies since performance with k = 20 is consistently high.

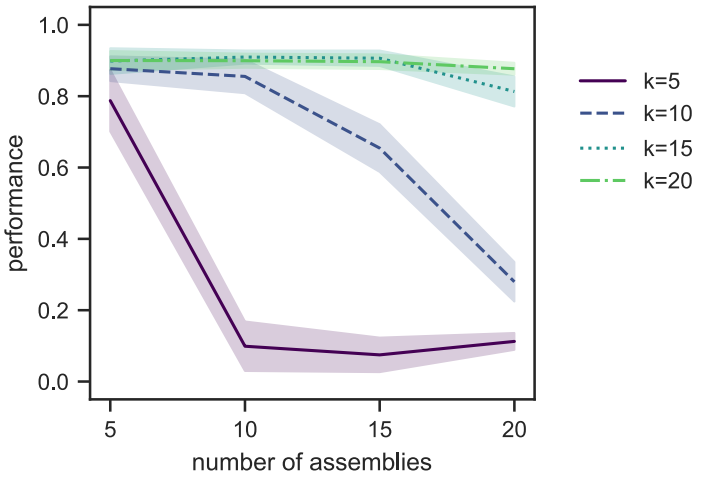


Figure 3.6: **Influence of k on detecting ground truth assemblies.** The influence of the number of topics to be found (k) on assembly detection performance. Performance was measured by the Jaccard similarity between the found and the ground truth assemblies (for details see Section 4.2.2). Lines represent mean for different k and shaded regions indicate standard deviation from 20 repetitions.

The topics found during each of the runs will not be the exactly the same and also not in the same order as in the other runs. To obtain which topics showed up consistently in each run, I cluster the topics in each result using an hierarchical clustering algorithm, which has several advantages. Firstly the most similar topics will be clustered together first, which means that the more similar the topics are (which they should be if they represent an assembly with a strong correlational structure), these clusters will be formed early in the clustering process. Secondly, there is no need to set the exact number of clusters to be found as in e. g. K-means clustering algorithms. Instead,

hierarchical clustering can produce a variable number of clusters because it takes a maximum distance as a parameter for when to stop clustering.

The choice of what maximum distance to define is up to the user and could also be done by hand. For neural topic modelling I automatically calculate a maximum distance threshold based on the distribution of all the pairwise Kullback-Leibler (KL) divergences between all topics found and set the threshold at the 33rd percentile of all the pairwise divergences. This seems to work well and result in some clusters which include around 20 topics (the number of runs) and more clusters with only a few topic in them. The larger clusters represent topics which have appeared consistently and clusters with few topics most likely represent noise.

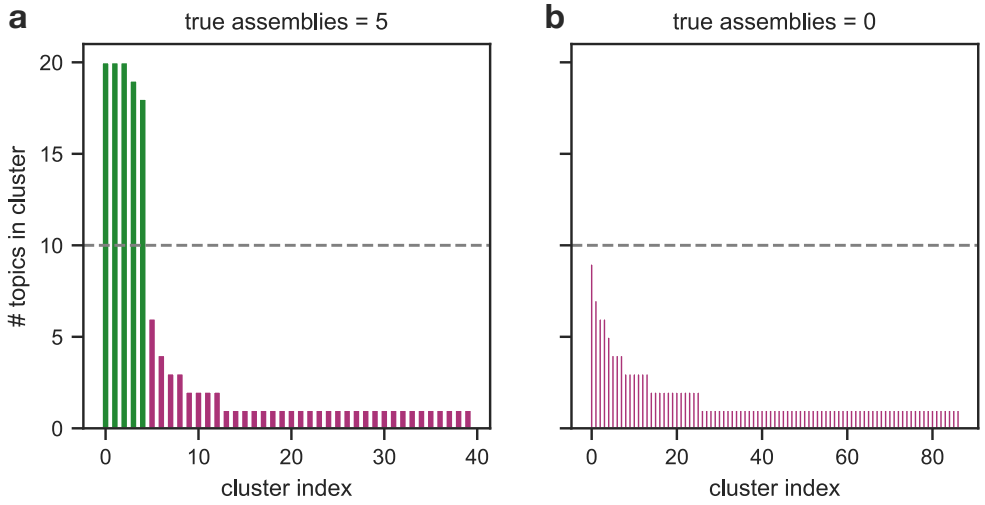


Figure 3.7: **Cluster selection filters out noise.** LDA is applied to the same artificial input data 20 times and the resulting topics are clustered. **(a)** When the input contains five assemblies, then the selection of clusters will disregard all clusters apart from the five assemblies which were present. **(b)** If no correlational structure is injected into the data, then no clusters pass the threshold of appearing at least 10 time (half of the number of runs) and all clusters are flagged as noise.

To select the clusters which most likely represent assemblies, I throw away clusters which include fewer than 10 topics (half the number of runs) and more than 30 topics (more than 1.5 times the number of runs). Only these clusters are entered into further analyses (e. .g. green clusters in Figure 3.7 (a)). This approach works very well on the artificial data the algorithm was tested on, especially considering that this is an automatic threshold which does not require any human intervention.

A good test of the usefulness of the clustering steps is to see what happens when this pipeline is applied to input with no correlational structure at all and the neural words appear in randomly chosen documents. One single re-

sult may come up with likely candidates of topics according to the random correlations in the data. Since the correlations are quite weak, different runs of applying LDA to the same randomly generated input is more likely to come up with a different set of topics every time. Therefore the clustering step should yield smaller clusters than when there is a strong correlational structure in the data (visualised in Figure 3.7 (b)). Ideally, the cluster selection process described in the last paragraph (disregarding all clusters that appear in less than half of the runs) will exclude all topics if the input is indeed random and only shows very weak correlations.

Indeed, this is what I see when random input is used. Sometimes a cluster with 10-15 topics in it will appear even when the input is random, flagging that cluster as an "interesting" cluster to analyse further even though the input was only noise. While this is not ideal, I will not apply stricter selection criteria since neural topic modelling is meant as an exploratory tool, so I would rather overestimate the number of "interesting" clusters in the data than underestimate it.

3.2.5 *Thresholding*

As in other algorithms for neural assembly detection, neural topic modelling results in distribution of weights across neural words (in other cases these would be neurons). To define an assembly, one needs to divide the distribution into two: neural words that are part of the assembly and the others which are not. Determining this threshold is not straightforward and very often is done by hand (Mölter et al., 2018).

For the artificial data created for testing and validating neural topic modelling (see Chapter 4), there existed a "correct" answer because the data was created in such a way where some neural words were part of the ground truth assembly and others were not. Therefore in this case, I wanted a clear threshold to validate how accurate neural topic modelling recovered the ground truth assemblies.

A popular choice of threshold among other methods is using two standard deviations above the mean as a cutoff value (Mölter et al., 2018). I found that a slightly different threshold gave more flexibility and worked better for neural topic modelling: defining the threshold as the midway point between the median value and the highest recorded weight.

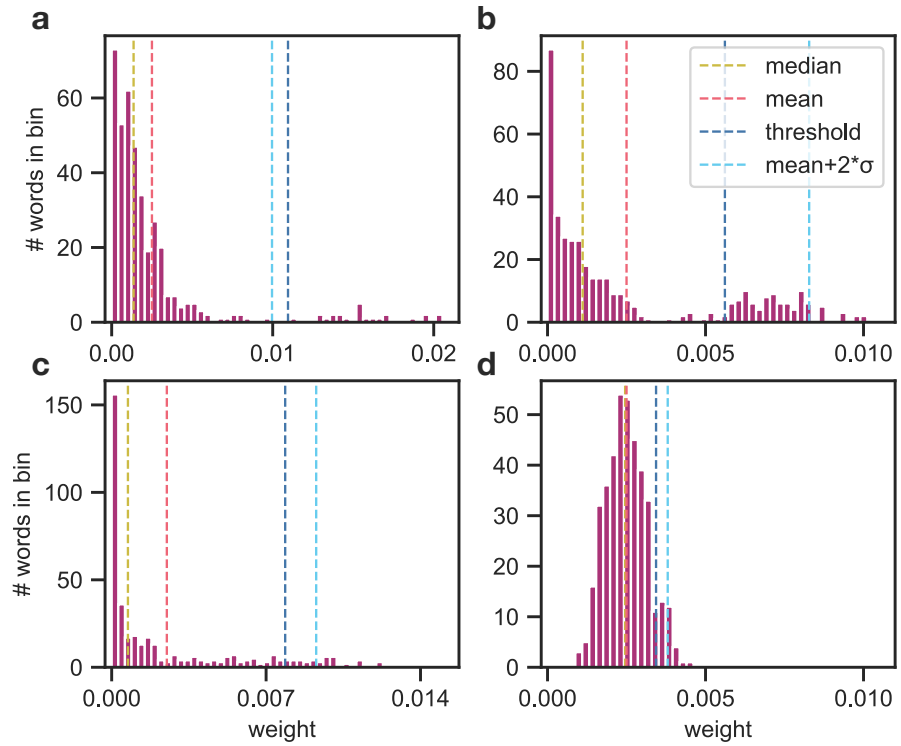


Figure 3.8: **Choice of threshold for assembly assignment.** To validate the performance of neural topic modelling, each assembly has to be assigned a number of neural words in order to compare these to the ground truth. The assignment of neural words to an assembly is done by selecting a weight threshold and assigning all neural words with weights above that threshold to the assembly. Here, two different thresholds are compared: "mean+2*standard deviation" (light blue) and "median+0.5*(max-median)" (dark blue), labelled as "threshold" in the legend. To put the values into perspective, the median (yellow) and the mean (red) are also shown. **(a)** The weight histogram for a small assembly (20 out of 400 neural words). **(b)** The weight histogram for a large assembly (80 out of 400 neural words). **(c)** The weight histogram for a large assembly (80 out of 400 neural words) when neural words are noisy (active outside of the assembly activity). **(d)** The weight histogram for an assembly which was not given in the ground truth, but instead groups neural words which are active when no assembly is active.

In some cases, both thresholds would give a satisfactory result, particularly when many small assemblies were present as seen in Figure 3.8 (a). However, for artificial datasets with different parameters, the mean+2std threshold is too restrictive as the resulting weights vector is a very skewed probability distribution towards zero. Selecting a threshold of twice the standard deviation would exclude a lot of neural words with high weights as can be seen in Figure 3.8 (b). In cases where there is no clear minimum in the weights histogram as in Figure 3.8 (c), neither threshold does a particularly impressive job. Neural topic modelling sometimes also reliably identifies as-

semblies which are related to the time windows when no assembly is active. The weight distribution of one such assembly is not as skewed and shown in Figure 3.8 (d). Note that in this case any thresholding would be rather meaningless.

For testing the algorithm, having a clear threshold is necessary to compare the found assemblies to the ground truth assemblies. On the contrary, for neural data, implementing such a hard threshold may be too simplistic and ignores the hierarchy of weights which is given by the topic model.

Therefore for neural recordings, instead of a hard threshold based on a weight value, I include all neural words in an assembly whose weights sum to at least 75% of the total weight.

3.3 DISCUSSION

Neural topic modelling was developed to address the challenges the neuroscience community faces when it comes to analysing large neural recordings. It makes use of topic modelling – a scalable machine learning algorithm developed for use on extremely large text corpora – to identify neural assemblies. The analysis pipeline of neural topic modelling includes three parts: the translation of neural data into a text-equivalent format, the topic model and the clustering and assessment of the found topics.

In its current state, neural topic modelling extracts four different "neural words" which are different types of neural activity. From the spike trains of the neurons, time windows with particularly high and particularly low firing rates are identified (see Section 3.2.2.1) as well as time windows in which pairs of neurons exhibited a higher-than-expected number of synchronous spikes (see Section 3.2.2.2). From the LFP signal, the time windows with the highest amount of power in five different frequency ranges are extracted for each brain region on each electrode (see Section 3.2.2.3). Next, the probabilistic generative model LDA is applied to the resulting neural word-time window matrix to uncover hidden topics or themes in the data i. e. groups of neural words which co-occur across time windows. To increase the quality of the topics found, the results from multiple random initialisations are pooled to filter out noise and to extricate topics which are likely to represent neural correlational structures.

Neural topic modelling addresses a number of open issues in the assembly detection field. However, as every other method, it also has its drawbacks.

Both limitations and strengths of neural topic modelling will be discussed in this section.

In text data, LDA works best with strong correlational structures i. e. when words are specific to topics instead of appearing in all topics at the same probability. The same approach should be taken for neural words – they should be chosen so that the neural words accurately mirror the specificity of neurons. Therefore the neural words in neural topic modelling are chosen so that they only include time windows with unusual or extreme activity (e. g. occurrences of very high firing rates) and disregarding all instances in which that particular neural word was only average (e. g. occurrences of average firing rate). This means that neural activity is thresholded and with it information is potentially lost. This is not unique to neural topic modelling, as other neural assembly detection algorithm include some kind of thresholding step (Diana et al., 2019; Mölter et al., 2018). This cutoff seems to work well enough to discover assemblies despite not carrying all the information. The key to also including some of this information might be in the improvement of the translation of spike counts into neural words, which will be discussed in the following section (5.6).

Using a percentile as a cutoff value and binarising the activity to 0's (spike pattern not present) and 1's (spike pattern present) is an attempt to normalise the activity between neurons which by default have very different activity characteristics such as firing rate. Other methods achieve this by using the z-score of the estimated firing rate (Lopes-dos Santos et al., 2011, 2013; Mölter et al., 2018; Panzeri et al., 2010; Stringer et al., 2018). For neural topic modelling, I not only wanted to create a balance between neurons, but also ensure that the word counts are comparable to each other, which is why I chose the percentile cutoff. Nevertheless, neurons with low firing rates will often not give rise to as many neural words as neurons with high firing rates because they cannot lower their firing rate below 0 spikes per window or because some spike statistics cannot be calculated for them (if for example a Coefficient of Variation or an ISI distribution is used as a neural word).

LDA was developed for text data and therefore makes certain assumptions about the data, in particular that the documents are independent. This is the main problem with using topic models on non-text data, since the samples in very few biological or other datasets are truly independent, including neural data. One could argue though that also documents in text data are not truly independent and that independence is a necessary simplification for the model. Researchers have successfully applied LDA to time series

data regardless (Koyejo et al., 2013; Lara et al., 2014). Additionally, variants of LDA have been developed which soften this requirement for document independence and to take into account the order of the documents (Blei et al., 2010; Liu et al., 2016). For the first version of neural topic modelling, I decided to stick with the basic version of LDA which was also readily available as open source software. However, picking a variant of LDA better suited to time course data might give more precise results.

LDA is a successful model for finding groups of related words in large text data. So as a topic model, LDA will find clusters which resemble clusters found in text data. Due to the different nature of neural data, this may not be the ideal case, since neural assemblies may come in all shapes and sizes, which might not overlap with the structure of text topics. For example, the topics in text often include a a large number of words which are related to them, but neural assemblies might be comprised of a small number of neurons. As will be discussed in the following Chapter (Section 4.4), detection rates drop when assemblies are smaller than 5% of the total number of neural words. This might limit the usefulness of a topic model in finding neural assemblies which are very small. However, the sparse Dirichlet prior and the clustering step in the pipeline make it more likely also for small assemblies to be identified.

Due to the limitations mentioned above, neural topic modelling is unlikely to find all possible neuronal assemblies in the data – just as any other assembly detection method out there. The performance of neural topic modelling when it comes to identifying assemblies from artificial test data sets is comparable to other methods (for details, see Section 4.3 in Chapter 4). At the same time, neural topic modelling is able to discover additional assembly structures to the currently available methods. Expanding the features of neuronal activity to multiple aspects of the spike train can generate a completely different set of assemblies which other methods would not be able to find.

While topic modelling might not be the solution to all of the challenges currently discussed in neural assembly detection, it does take steps towards addressing some of them.

Currently, there exist multiple methods for identifying neural assemblies, which have brought forward interesting results and assemblies whose activity related to experimental variables (Mölter et al., 2018). Many of them start with estimating the time course of the firing rate of each unit and the correlation between all pairs of units (Mölter et al., 2018; Panzeri et al., 2010). The

resulting connectivity matrix forms the basis of all further data explorations (Stringer et al., 2018). Others directly apply dimensionality reduction techniques on the estimated firing rate time courses of the neurons (Cunningham and Yu, 2014; Lopes-dos Santos et al., 2011). A completely different approach is taken by methods focussing on defining assemblies as neurons involved in repeated sequential spike patterns (Mackevicius et al., 2018; Masquelier et al., 2008; Prut et al., 1998; Quaglio et al., 2017; Torre, Canova, Denker, Gerstein, Helias and Grün, 2016).

What both approaches completely leave out is that neurons might do both at the same time. The fact that neurons multiplex i. e. code for different experimental variables with different aspects of their activity is well known (Hardcastle et al., 2017; Huk and Hart, 2019; van der Meij and Voytek, 2018), but this knowledge is not included into the methods for searching for neural assemblies. Neural topic modelling aims to allow for multiplexing in neurons by extracting different aspects of the spike trains and other neural data separately. This allows for the combined analysis of multiple spike train or LFP features – something that has not been tried yet in the search for neural assemblies.

Moreover, neural topic modelling also takes into account that one neuron or one feature of a neural spike train might participate in multiple assemblies. While some methods allow for this (Carrillo-Reid et al., 2015, 2016), sometimes the detection performance degrades when neurons are part of multiple assemblies (Diana et al., 2019). Other methods assign each neuron at most one assembly (Lopes-dos Santos et al., 2013), especially the ones based on PCA since principle components are by definition orthogonal. Neural topic modelling assumes overlapping assemblies from the start, due to it being developed for text data. Words in any language can be used in different contexts. The word "score" for example can come up in a variety of topics such as "sports", "sheet music" and "exams". Indeed, the firing rate of neurons have been shown to play a role in multiple assemblies (Carrillo-Reid et al., 2015). Neural topic modelling might be able to uncover not only how the firing rates of neurons relate to each other in assemblies, but also how firing rates relate to other features of neural activity (e. g. synchronous spiking, ISI variability, bursts and any other spike train feature). Therefore, neural topic modelling might be able to detect assembly structures and find interactions between spike patterns which other methods would not be able to find.

All methods apply a hard threshold in the final steps of the search for assemblies to clearly define which neurons belong to an assembly and which ones do not (Lopes-dos Santos et al., 2013; Mölter et al., 2018; Peyrache et al., 2010; Romano et al., 2015). For neural topic modelling, I want to challenge this binary definition of an assembly where firstly a neuron is either in or out of the assembly and secondly where all neurons within it are equally important. The shape of the probability distribution over neural words given in the topics might hold interesting information (see section 4.6). It intuitively makes sense that some neurons are more faithful to the overall activity of an assembly (Buzsáki, 2010), and it might be interesting to see what is driving this discrepancy.

The choice of LDA – a text analysis tool – for neural data has another interesting implication. A topic model assumed that in each document there are multiple topics present. This characteristic of LDA is the main reason why topic modelling has been popular for biological and medical datasets since this assumption often holds true in data from those fields (Liu et al., 2016). This assumption fits very neatly with neural data since at any point in time in every part of the brain, neurons integrate and process multiple inputs. In this regard, topic models are suitable models to use for the analysis of neural data.

Another aspect which makes topic models a convenient tool for neural data is their scalability, which is a huge challenge with the ever-growing data sets that are now being recorded (Allen Brain Observatory, 2019; Stringer et al., 2018). Already we see analysis methods being retired since they are not scalable to such large numbers of neurons e. g. the brute-force approach to identify repeated spike sequences by Abeles and Prut (1996). The number of possible assemblies grows exponentially with the number of neurons recorded (Blei, 2012). New methods for detecting neuronal groups need to recognise this fact and provide a better scaling factor. There exist highly optimised inference algorithms – the bottleneck in terms of computational time – which allow the analysis of millions of documents or even real-time LDA and are available as open-source software (Blei, 2012). While my implementation is not one of those highly optimised ones, there is no reason why they can't be utilised for neural data. This would ensure that neural topic modelling stays relevant despite the impressive growth of neural recordings.

Interpretability is also a big goal for the data analysis of neural data. Many complex mathematical models provide interesting low-dimensional representations of large neural data sets, but the question remains how the brain

could arrive at such a representation and whether it can actually make use of that information. Nevertheless, dimensionality reduction techniques have been applied with great success to neuronal data (for a review see Cunningham and Yu (2014)) and enabled researchers to identify neural mechanisms for sensory input selectivity (Mante et al., 2013), odour discrimination (Mazor and Laurent, 2005), and premotor responses (Kaufman et al., 2014). Popular amongst these methods are Principle Component Analysis (PCA) and its derivatives such as demixed PCA (Kobak et al., 2016), Gaussian-Process Factor Analysis (Yu et al., 2009), and sequential non-negative matrix factorisation (Mackevicius et al., 2018). Current methods which search for neural assemblies also rely heavily on complex data transformations such as PCA in combination with a Promax oblique rotation (Romano et al., 2015), singular value decomposition (Carrillo-Reid et al., 2016) or spectral graph clustering (Avitan et al., 2017). The lower dimensional representations from these highly sophisticated methods are difficult to interpret biologically and unlikely to be the way in which the brain processes this information.

In neural topic modelling, I tried to develop a method which does not rely on complex transformations of the data. On the contrary, neural topic modelling is based on searching the data for unlikely events (e. g. exceptionally high firing rates or a exceptionally high number of coincident spikes in a pair of neurons) and then grouping these activity patterns by their co-occurrences. When the events chosen are activity patterns which could be found by neurons themselves, the interpretation of the groups of activity patterns is spectacularly easy and at the same time the groupings are the same that could be detected in exactly the same way by the brain. It can therefore serve not only as a neural assembly detection tool, but it also offers a representation of the data with reduced dimensionality. Even though topic modelling does not give a topic a label or a biological interpretation, the biological interpretability of the neural words in each topic make is much easier to relate the results from topic modelling back to biological variables (Liu et al., 2016). In addition, the insights gained from neural topic modelling might be used to generate new hypotheses about how the brain processes information from what we learn about the interactions between different neural activity patterns.

3.4 CONCLUSION

Neuronal assemblies are coordinated groups of neurons, which exhibit similar activity over time. There is much evidence for the existence and relevance of these assemblies, suggesting that much of neural activity is organised into these assemblies and that they hold the key to understanding brain function (Yuste, 2015). Finding assemblies in large-scale neural data is difficult and the focus of much ongoing research.

Neural topic modelling aims to identify neural assemblies in a more inclusive fashion than previous methods and to address some of the gaps in current analysis methods. It extracts multiple neural activity features from one spike trains and thereby takes into account that neurons multiplex. This aspect sets neural topic modelling apart from all other assembly detection algorithms which purely focus on the firing rate of neurons. The base assumption that during each time window multiple neural topics are active and contribute to the neural words in the document make it a suitable model for neural activity, since neurons also process multiple inputs at one point in time. As an additional innovation for neural data analysis, neural topic modelling gives the option of including LFP signals in the analysis.

Its scalability and interpretability makes neural topic modelling a promising method for the exploration of large neural datasets and for generating hypotheses for how neuronal activity patterns interact with each other.

4

NEURAL TOPIC MODELLING – VALIDATION

This chapter describes how neural topic modelling was tested and validated and compared against existing methods.

Identifying neuronal assemblies in large brain recordings is a difficult problem in neural data analysis because there are no data sets for which the ground truth about neuronal assemblies is known. Additionally, the larger the number of neurons per recording becomes, the more important it is to have scalable and reliable assembly identification algorithms.

The work shown in this chapter demonstrates that neural topic modelling is a suitable algorithm for the detection of neuronal assemblies and its performance is equal to or surpasses currently available algorithms (see Section 4.3). I demonstrate this by applying neural topic modelling on artificially created data for which the true underlying correlational structure is known (see Section 4.2). Moreover, I show that neural topic modelling can recover assemblies with non-linear correlations, which are also not detectable with other methods (see Section 4.5). Finally, neural topic modelling also allows for a more detailed and careful analysis of neuronal assemblies by assessing how aligned a neural word is to an assembly instead of reducing this to a binary choice of "part of the assembly" or "not part of the assembly" (see Section 4.6).

4.1 INTENTION

In order to confirm the ability of neural topic modelling to identify correlated neural activity of assemblies in brain recordings, I tested the algorithm in a variety of ways.

The previous chapter described how different activity patterns are extracted from neural spike trains to (see Section 3.2.2). Importantly, I showed how to allow for the possibility of a multiplexed neural code in the recorded neurons by extracting multiple features of neural activity (activity patterns

such as high or low firing rates) per spike train. To identify only the time windows in which the activity pattern was the most pronounced, I only use a certain percentage of the time windows in which that spike pattern has been detected e. g. if the activity pattern is "high firing rate", I only record a "high firing rate" event for the time windows in which the firing rate is higher than in 75% of all other time windows. This acts as a filter to make it more likely that only significant events are included in the analysis. This filtering is done for each activity pattern separately. The analyses presented in this chapter are based on the event filtered time series of the activity patterns and not the original spike trains.

Since neural topic modelling is based on a method developed for text data, the input data it was developed for are documents and the words appearing in each document. In the context of neural recordings, a document is actually a time window and a word corresponds to whether a certain neuronal activity pattern was present during this time window. In this chapter I will refer to time windows rather than documents, but to neural words rather than neurons or activity patterns, since a neural word might be anything – an actual spike pattern or a spike train characteristic such as inter-spike interval variability or even something like the power of a frequency band in the local field potential (LFP) signal.

As a first step, in Section 4.3 I compare the results from neural topic modelling to the assembly detection algorithm of Bayesian inference (Diana et al., 2019), which itself was compared favourably to a few other state-of-the-art assembly identification methods. The artificial data used here (and its notation) was modelled on the artificial data used in the original publication of Bayesian inference to ensure comparability. The performance of neural topic modelling is evaluated for a variable number of assemblies, for the amount of noise activity (asynchrony), and the number of neural words in the data set.

To further investigate the limits of neural topic modelling, in Section 4.4 I also vary some additional parameters to see for which assembly structures neural topic modelling is successful and which are more difficult to identify. To this end, I vary the activity of the assembly, add additional neural words which do not adhere to any assembly activity (noise), and I investigate how large assemblies need to be in order to be found by the algorithm.

Then I go a step further and in Section 4.5 investigate non-linear correlated assemblies which would not be easily picked up on by other traditional assembly detection algorithms. I investigate the case when a group of neu-

ral words shows coordinated activity with multiple other groups of neural words, creating neural words which are part of multiple different assemblies. Smith et al. (2018) showed that interactions between neuronal assemblies are task-specific and accounting for these different assembly activations for different tasks increased their encoding (predicting spikes) and decoding (predicting task type) performance significantly. It seems logical that neurons are part of multiple different assemblies in the brain, e. g. depending on the external stimuli arriving or the current behaviour.

Table 4.1: **Description of parameters for artificial data generation.**

parameter	description
assembly parameters	
# words	number of words belonging to at least 1 assembly
# docs	number of documents (time windows)
# assemblies	number of assemblies
% active	percentage of time windows in which assemblies are active
% multi-membership	percentage of words belonging to multiple assemblies
neural word parameters	
synchrony	probability of a word occurring when assembly is active
asynchrony	probability of a word occurring when assembly is inactive
max # memberships	maximum number of assemblies one word belongs to
noise parameters	
# noise words	number of words with random activity (Poisson)
% active	percentage of time windows in which the noise words are active
input parameters	
min # words/doc	minimum number of events per time window for inclusion

Note that for brevity reasons the "neural" from the expression neural words was dropped

As the last step, in Section 4.6 I propose that the way researchers have traditionally thought about neural assemblies is too simplistic – the view that an assembly membership is a binary mapping and a neuron is either within or outside of an assembly. Instead, I propose that assemblies are likely made up of neurons with very different characteristics and some of them might adhere more to the assembly activity than others. This means that the boundary of assembly membership is more fluid than previously thought and that investigating how much each neuron is bound to an assembly can give researchers more and potentially interesting information. Looking at the shape

of an assembly (how much different neurons in the assembly adhere to its activity) will probably give us a more realistic picture of what is going on inside the brain than trying to draw sharp lines delineating which neurons belong to an assemblies and which ones do not.

In the next section, I will detail the artificial data used in this chapter, how it was created and which parameter ranges were investigated.

4.2 ARTIFICIAL DATA

Neural topic modelling as a neural assembly detection method was evaluated using artificial data sets for which the underlying assembly structure was known. The form of the artificially generated data was based on the artificial input used to test the algorithm by (Diana et al., 2019).

4.2.1 *Input generation*

The default artificial input includes 400 neural words divided equally into 5 assemblies (80 neural words per assembly), which can be active or not active over 1000 time windows. The process of the input generation is described in a procedural way in the box below:

1. **For** each ground truth assembly (# *assemblies*)
   determine in which documents it is active according to % *active*.
2. **Initialise** the document-word matrix (size: # *documents*, # *words*)
3. **For** each neural word
   assign it to a number of assemblies according to *max # memberships*
   a) **For** each document
      i. **if** one of the assemblies it is part of is active,
         record a 1 in the document-word matrix at a probability of *synchrony*
      ii. **else** (no assembly it is part of is active),
         record a 1 in the document-word matrix at a probability of *asynchrony*

Since neurons are able to multiplex, they can take part in multiple assemblies. To reflect that some neural words in a data set might be part of more than one assembly, one fifth of all neural words were allowed to belong to

two or three assemblies. A neural word which belongs to two assemblies will then have a high probability of also being active (synchrony of 0.5) when either of the assemblies it is part of is active. This resulted in 10% of neural words belonging to two assemblies and an additional 10% of neural words belonging to three assemblies (synchrony of 0.5 was kept the same for each assembly). No additional asynchronous activity was added to these neural words belonging to multiple assemblies. Assemblies were allowed to be active during the same time windows.

In an input created with the default parameters, each neural word is active in accordance with its assemblies 25 times and active 95 times when its assemblies are not active. An overview of the parameters used to create the artificial data can be found in Table 4.1 and an example for an input matrix with default settings (as stated in Table 4.2) can be seen in Figure 4.1.

Some parameters for the creation of the artificial input were varied to test the robustness of the assembly detection across different parameter ranges. As in Diana et al. (2019), I created artificial inputs of different sizes (varying number of words), with different number of ground truth assemblies and with asynchrony levels (how much individual neural words are active outside of assembly activity). The parameter ranges used in the different figures can be found in Table 4.2.

For further investigations of the robustness of neural topic modelling, I also modified the assignment of neural words to assemblies in the input - creating overlapping assemblies or gradients of assembly memberships (neural words could participate in assemblies to varying degrees). Details are available in each section.

This section delineated the parameters used for generating the artificial input to neural topic modelling. In the next section, I describe how I measured the success of neural topic modelling and why I chose to add a different metric to the one used in Diana et al. (2019).

4.2.2 *Measuring performance*

The performance of an assembly detection algorithm on artificial data can be assessed in multiple different ways. One could assess whether the correct number of assemblies was found (Mölter et al., 2018). Then the quality of the assemblies found could be investigated i. e. how closely the found as-

semblies resemble the ground truth assemblies (Diana et al., 2019; Mölter et al., 2018; See et al., 2018).

In the case of neural topic modelling, I chose to focus more on the quality of the assemblies identified. Since neural topic modelling is a tool to find previously unknown relationships in neural data, it seems more important that the ground truth assemblies are recovered than whether "only" the ground truth assemblies are identified. Neural topic modelling is more likely to not detect all "correct" assemblies than to identify a non-ground truth assembly. So the over-estimation of assemblies by neural topic modelling does not come at the cost of the assembly quality. Moreover, some assemblies identified by neural topic modelling which are not part of the ground truth set of assemblies turned out to represent some interesting information: a group of neural words which were active in time windows when none of the ground truth assemblies were active. This kind of finding is potentially interesting to researchers, so I did not want to exclude such assemblies on principle.

Measuring the quality of the found assemblies is similar to comparing the results of two clusterings. However, the main comparison metrics available for clustering rely on the fact that groups or assemblies are exclusive, i. e.

Table 4.2: **Parameters for artificial data generation.**

	Default value	Variation range	Figure
assembly parameters			
# words	400	200-1000	4.5
# docs	1000		
# assemblies	5	5-20	4.4
% active	0.05	0.025-0.3	4.6
% multi-membership	0.2		
neural word parameters			
synchrony	0.5		
asynchrony	0.1	0.02-0.3	4.3
max # memberships	3	1-3	4.8
noise parameters			
# noise words	0	0-2400	4.7
% active	0	0.125	4.7
input parameters			
min # words/doc	13		

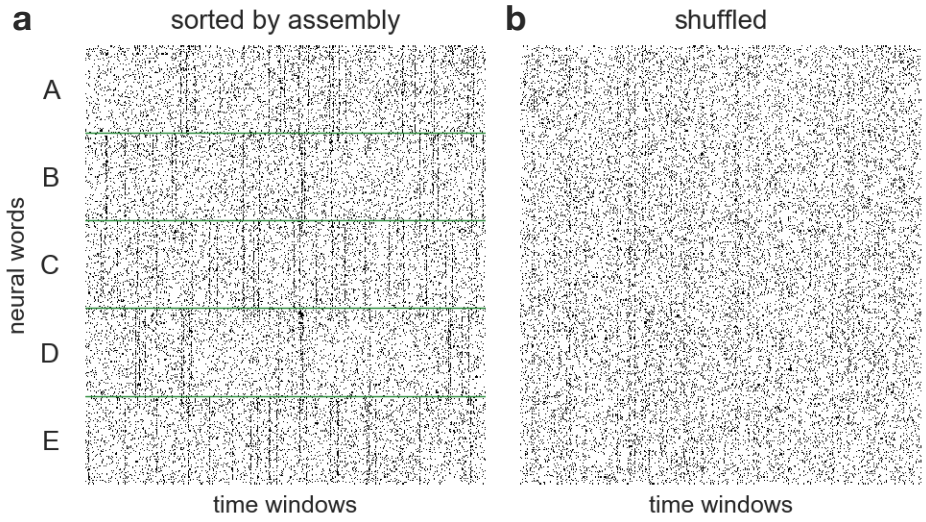


Figure 4.1: **Input matrix with default settings.** Each neural word's activity is visualised as one row in which activity (1) is denoted by a black square and inactivity (0) is denoted by a white square. **(a)** When the input matrix is sorted by assemblies (A-E, boundaries denoted by a green line), the concerted assembly activity becomes visible. In each assembly, the neural words which belong to multiple assemblies are sorted to the top and their slightly higher activity levels can be made out. **(b)** The assembly activity disappears when the rows of the matrix are shuffled randomly.

that there exist no items which are part of more than one group, making them unhelpful in this case (Goldberg et al., 2010).

To measure the quality of the found assemblies, I employed two measures: a correlation and a Jaccard score. Firstly, because this was the measure of success used by Diana et al. (2019), I calculated a correlation between the ground truth pairwise assignment matrix $I_{ij}^{(orig)}$ and the resulting pairwise assignment matrix from the Bayesian inference $I_{ij}^{(\tau)}$ for an assignment map $i \to \tau(i)$ for neurons $i = 1, ..., N$. The correlation score gives one score for each result, making it a simple and easy to compute measure.

$$\rho_\tau = \frac{1}{\frac{N(N-1)}{2}} \sum_{i<j} I_{ij}^{(\tau)} I_{ij}^{(orig)} \tag{4.1}$$

The assignment matrix $I_{ij}^{(orig)}$ has a value of 1 if a pair of neurons i and j are within one assembly and -1 if they are not. While still representing the general assembly structure, this measure is not sufficiently sensitive to catch possible mis-assignments and particularly useless when the data includes a lot of non-assembly neurons. For example, when no assemblies are identified, the found correlation matrix is a matrix with only -1's, but still gives

a performance of 0.86 for 5 assemblies and 0.90 for 20 assemblies. The high number of entries and in particular the high number of -1's in the ground truth assignment matrix masks wrong assignments quite well. The correlation score should therefore be accompanied by another score or the number of detected assemblies. Another potential problem is that the authors expand this definition to the case when neurons are part of multiple assemblies by setting I_{ij} to one when two neurons have at least one assembly in common. This makes the correlation matrix very imprecise when it comes to overlapping assemblies. It does however give some indication of whether assemblies were found which are not part of the ground truth assemblies, which would cause a drop in the correlation score.

Additionally, I measure the quality of the identified assemblies by using a Jaccard similarity, which is the size of the intersection of two arrays divided by the union of two arrays, between all true assemblies and clusters from neural topic modelling.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \tag{4.2}$$

For each ground truth assembly GT I calculate the Jaccard score to all found assemblies Found and assign the highest score. To obtain one number per result, I calculate the mean of the highest Jaccard scores for each ground truth assembly which were above a value of 0.5 to only include pairs of assemblies which have a high similarity. Since the ground truth assemblies do not overlap much, this will rule out counting one found assembly towards two or more ground truth assemblies.

$$score = \frac{1}{\#GT} \sum_{i=1}^{\#GT} \max_{j=1}^{\#Found} J(GT_i, Found_j) \tag{4.3}$$

It is possible to go back and investigate the performance on each ground truth assembly, giving a more detailed insight of how well each true assembly was identified. When no assemblies are found, the Jaccard score is 0, providing a better understanding of the closeness between the found and the ground truth assemblies than the correlation without having to also verify the number of found assemblies. The Jaccard score allows for overlapping assemblies and weights all assemblies equally regardless of their size. While I do argue that the Jaccard score is a more suitable measure of performance for assembly detection, it also has its drawbacks as any performance measuring metric. It does not take into account found assemblies which are not

related to the ground truth assemblies. As expressed above, such assemblies might hold value in their own right - there might be correlations in the artificial data which I did not purposefully put in there. Neurons A and B may be active in the same time windows by chance even though they are not in the same ground truth assembly. I would not want to penalise a result where neurons A and B where put into a separate assembly in addition to the ground truth assemblies that were recovered. Therefore I am using a score which is a strong indicator of whether all ground truth assemblies were identified while allowing for the possibility of additionally identified assemblies.

Mölter et al. (2018) also use a variant of the Jaccard score in their "Best Match" distance, but at the same time the authors put great importance on the number of found assemblies being correct. Therefore I am not using their metric, but report the simple averaged Jaccard score which conveys the same information in a less restrictive way.

In practice, the two measurements gave very similar scores for runs where the correct number of assemblies was found, but only the Jaccard score indicated when less or no assemblies were identified. In turn, a drop in the correlation score but not in the Jaccard score indicates the presence of additionally detected topics, which might not be directly related to the groud truth assemblies. Because both scores can provide information and to ensure maximal comparability to Diana et al. (2019) both scores were calculated and will be shown in the upcoming results.

The performance scores shown in the upcoming sections are based on the multiple runs of each parameter combination and subsequent clustering of the results as described in Section 3.2.3. As a baseline, I added the performance from the best individual runs from the repetitions.

In Section 4.6, I need to measure the similarity between the shape of two vectors. To this end, I calculated the cosine distance d, which describes the similarity of two vectors u and v by calculating the cosine of the angle between them.

$$d_{cos} = 1 - \frac{u \cdot v}{||u||||v||} \tag{4.4}$$

This has the advantage that the scale of the vectors does not matter, and it instead becomes a measure of how closely the two vectors (which are ordered in exactly the same way) resemble each other. A value close to 0 means perfect match and a value closer to 1 indicates a distinct shape.

This section laid out how the results from neural topic modelling will be assessed. The upcoming section will provide details on the run parameters used.

4.2.3 *Run parameters*

As explored in the previous chapter, neural topic modelling needs to receive two key parameters: the number of topics to be found k and the number of iterations. I showed in Section 3.2.3 that the performance of neural topic modelling remains stable as long as k is chosen to be sufficiently large such as 10 for a ground truth of 5 assemblies. For most runs k was therefore set to 10, except for the runs where the number of assemblies was varied, for which I set k to 25.

For each parameter combination, 10 different input matrices were generated which differed in their (randomly-generated) assembly timings. Each of these input matrices were run through neural topic modelling 20 times with a different random seed for the LDA algorithm every time. The reason for the repeated runs with different random seeds is to select topics which appear consistently, thereby being more likely to represent strong correlational structure in the data. For more details about this choice please refer to Section 3.2.4.

The number of iterations parameter controls how often the inference algorithm traverses through the data. In general, the higher the number of iterations that higher the likelihood is that the algorithm has converged to a local minimum. To determine a satisfactory choice for the number of iterations, I ran neural topic modelling on 10 individual input matrices an increased the number of iterations from 10 to 1000 as can be seen in Figure 4.2.

After about 100 iterations the performance of the topic model saturates and does not improve further with more iterations as shown in Figure 4.2 (b). The result at 100 iterations is not much different to the result at 1000 iterations in terms of Kullback-Leibler (KL) distance and the performance does not improve further in terms of the Jaccard score and the correlation to the true assembly memberships. Even though performance does not increase much after even 10 iterations, the KL distance to the result at 1000 iterations only flattens out at 100. Similarly, the performance of the result does not improve after 50 iterations and instead wobbles around a score of about 0.8 both with the Jaccard score and the correlation score (correlation between

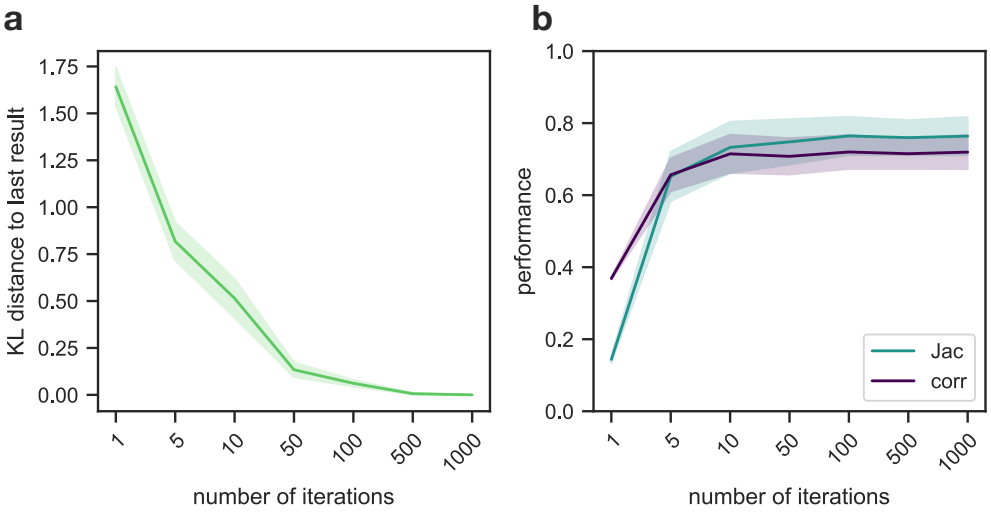


Figure 4.2: **Topic model performance depends on number of iterations. (a)** The distance to the result obtained with 1000 iterations decreases steeply until about 100 iterations, when the decrease in distance seems to plateau. **(b)** The number of iterations has an effect on the similarity of the result to the ground truth assemblies as measured by Jaccard score (Jac) and correlation (corr). Shaded areas represent the standard deviation of the results from 10 runs with different artificial inputs. KL = Kullback-Leibler

the ground truth and the result from neural topic modelling, which will be discussed in the upcoming section) as seen in Figure 4.2. So 100 iterations seems to be a good choice for the artificial data and allows for a little bit of leeway for less clear-cut assembly structures.

The upcoming section will present the results from various test data sets and investigate the strengths and limitations of neural topic modelling.

4.3 COMPARISON TO DIANA ET AL., 2019

In this section, I test the assembly detection capability of neural topic modelling and juxtapose them with the performances of four algorithms reported in Diana et al. (2019): Bayesian inference (the author's own contribution), spectral clustering, PCA+varimax and K-Means (previously suggested methods). The algorithms are judged on artificial data with the same specifications, varying the same parameters and using the same and an additional performance metric.

Diana et al. (2019) show that their method, Bayesian inference, is equal to or slightly better that two other assembly identification algorithms they implemented – PCA+varimax and spectral clustering – and drastically better

than K-Means alone. In a different publication, Mölter et al. (2018) compare multiple assembly detection algorithms and find that spectral clustering and two other methods very similar to PCA-varimax perform best of the ones they tested. Therefore by demonstrating that neural topic modelling performs better than Bayesian inference, I reason that neural topic modelling will also be comparable or more successful than the other PCA-based methods, since I am testing neural topic modelling on equivalent artificial data.

To test their neuronal assembly detection algorithm, Diana et al. (2019) vary three key parameters of their input data: asynchrony, number of assemblies, number of neural words (in the original publication they use "number of neurons" because in their method each neuron only gives rise to one activity pattern). They measure their performance using a correlation of the pairwise assembly memberships as described in the previous section. Their result of in-or-out assembly memberships is based on the assumption that each neural word is either part or not part of an assembly.

The first parameter which was varied was the asynchrony of the neural words: the probability of a neural word being recorded outside of time windows in which the assemblies it belongs to are active. I varied asynchrony between 0.1 and 0.3 i.e. in time windows in which none of the assemblies the neural word belongs to are active, the neural word has a 10% or 30% chance of being active. Figure 4.3 shows that performance decreases with higher asynchrony with all methods, even though the neural topic modelling has a steeper slope – starting at a higher performance at asynchrony of 0.1 and decreasing to a lower value at an asynchrony of 0.3. Tellingly, the Jaccard score indicates that the quality or number of found assemblies is so poor that the Jaccard score performance is close to a score of 0.1 at high asynchrony (0.3), which is not reflected by the correlation score.

The next parameter which was varied in the input was the number of assemblies. In Figure 4.4 I show that the number of ground truth assemblies does not seem to influence the performance hugely up to 15 assemblies in the 400 neural words (each assembly consists of on average 27 neural words). Only when the number of assemblies increased beyond that (and the assembly size falls to 5% of the whole recording) does the Jaccard-based performance decrease slightly. It is worth noting that the performance of neural topic modelling is higher than the performance by Bayesian inference and the other algorithms in every case. Additionally, another interesting point is that the correlation seems to increase with the higher number of assemblies regardless of the decrease in performance in the Jaccard score. This is

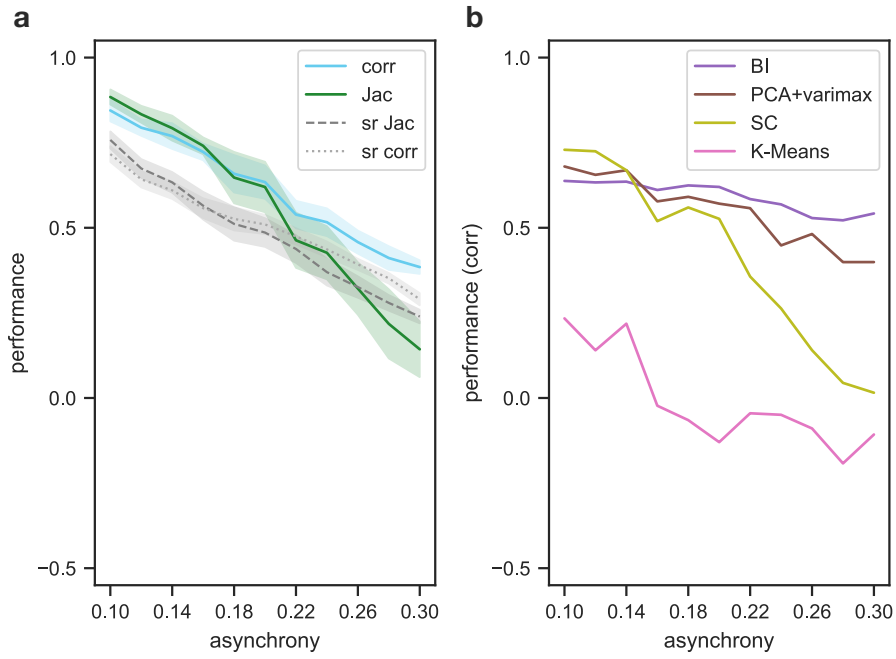


Figure 4.3: **Influence of asynchrony on performance.** Asynchrony was varied from 0.1 to 0.3 in steps of 0.02. **(a)** Performance of neural topic modelling on inputs with different levels of asynchrony as measured by correlation (blue) and Jaccard score (green). The grey lines represent the baseline of the performance achieved by calculating the performance of each individual, single results (sr) i. e. the results without running multiple initialisations and clustering. Shaded areas represent the standard deviation of the results from 10 runs with different random artificial inputs. **(b)** Results of the algorithms run in Diana et al. (2019) (estimated from original publication via gauging). Lines represent mean performance across 15 different inputs. The blue line in (a) (neural topic modelling) represents the correlation score and is directly comparable to all the results in (b). corr = correlation score; Jacc = Jaccard score; sr = single result; BI = Bayesian Inference; SC = Spectral Clustering

likely due to the high number of -1's in the correlation matrix as discussed in Section 4.2.2.

Lastly, the size of the input matrix was varied in Diana et al. (2019) by increasing the number of neural words from 200 to 1000. The input size did not have an influence on the performance with either measure as can be seen in Figure 4.5. Since the number of assemblies was kept constant at 5, the assembly size also increased from 40 to 200, which did not seem to impact performance.

Neural topic modelling shows a comparable performance to Bayesian inference and even surpasses the latter in most of the parameter settings mentioned in Diana et al. (2019). In the following section, I show further explorations of the limits of neural topic modelling.

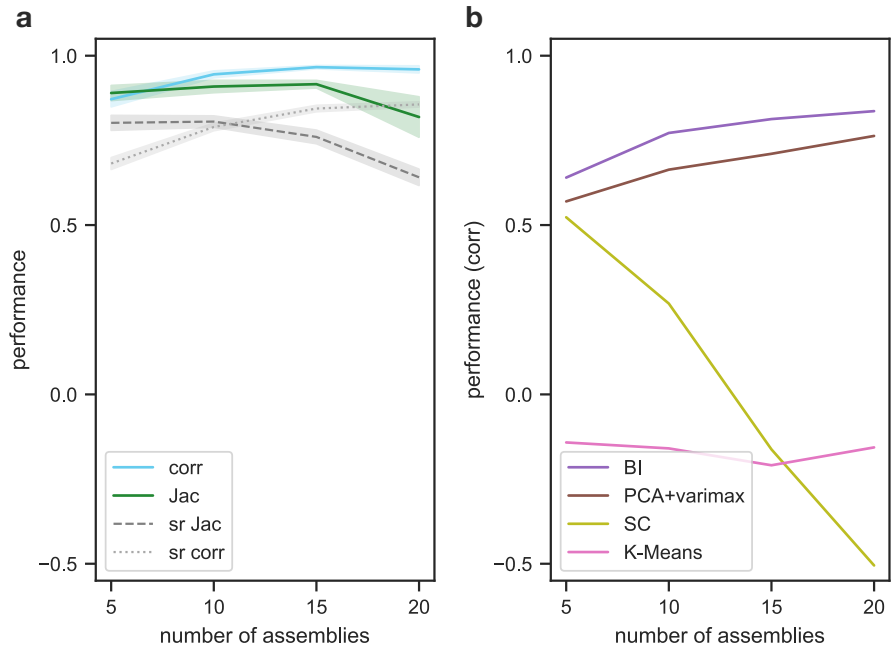


Figure 4.4: **Influence of the number of assemblies on performance.** The true number of assemblies was varied between 5 and 20. **(a)** The performance of neural topic modelling in response to different numbers of assemblies was measured by correlation (blue) and Jaccard score (green). The grey lines represent the baseline of the performance achieved by calculating the performance of each individual, single results (sr) i. e. the results without running multiple initialisations and clustering. Shaded areas represent the standard deviation of the results from 10 runs with different random artificial inputs. **(b)** Results of the algorithms run in Diana et al. (2019) (estimated from original publication via gauging). Lines represent mean performance across 15 different inputs. The blue line in (a) (neural topic modelling) represents the correlation score and is directly comparable to all the results in (b). corr = correlation score; Jacc = Jaccard score; sr = single result; BI = Bayesian Inference; SC = Spectral Clustering

4.4 FURTHER EXPLORATIONS

After confirming that neural topic modelling performs to a similar standard as Bayesian inference on the parameter variations they tested, I present some further parameter explorations to investigate the strengths and limitations of neural topic modelling. The parameters explored were the activity of the assemblies (in percentage of total documents), the presence of additional randomly active neurons (additional noise), and the effect of letting neural words be part of multiple assemblies.

Assembly detection was assessed in relation to the activity of the assemblies. In the default settings, the assembly activity is set to 0.05 of all time windows, leading to on average 50 active time windows per assembly. This parameter was varied between 0.025 (25 time windows in which the assem-

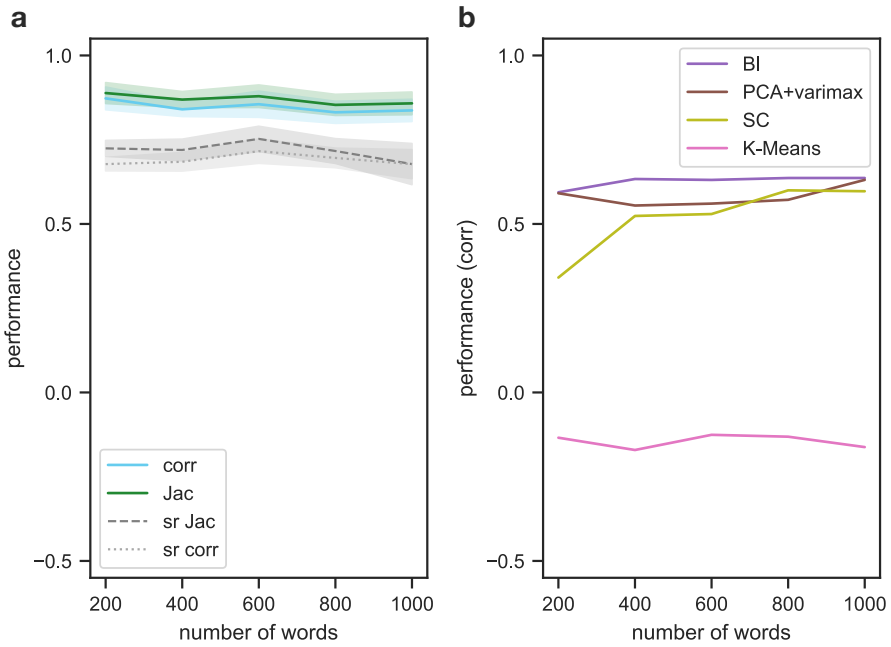


Figure 4.5: **Influence of the number of neural words on performance.** The input size was varied to include between 200 and 1000 neural words. **(a)** Performance of neural topic modelling on inputs of different sizes as measured by correlation (blue) and Jaccard score (green). The grey lines represent the baseline of the performance achieved by calculating the performance of each individual, single results (sr) i. e. the results without running multiple initialisations and clustering. Shaded areas represent the standard deviation of the results from 10 runs with different random artificial inputs. **(b)** Results of the algorithms run in Diana et al. (2019) (estimated from original publication via gauging). Lines represent mean performance across 15 different inputs. The blue line in (a) (neural topic modelling) represents the correlation score and is directly comparable to all the results in (b). corr = correlation score; Jacc = Jaccard score; sr = single result; BI = Bayesian Inference; SC = Spectral Clustering

bly is active) and 0.3 (300 time windows in which the assembly is active). Figure 4.6 shows that when assemblies are active at a level of 0.075 (during at least 75 time windows), the performance is close to perfect. Below an activity level of 0.05 (50 or less active time windows), the performance decays. It should be noted that the default parameter value is 0.05 or 5% i. e. when the algorithm does not perform at its best. It is also interesting that the baseline results (best result for a single run) increases to 1 for the Jaccard index, but not for the correlation. This indicates that the additional topics found during each run – which are discarded during the clustering – lower the score despite all assemblies having been correctly identified (Jaccard score at 1). At higher numbers of active time windows per assembly, the likelihood that multiple assemblies are active during one time window increases. Topic

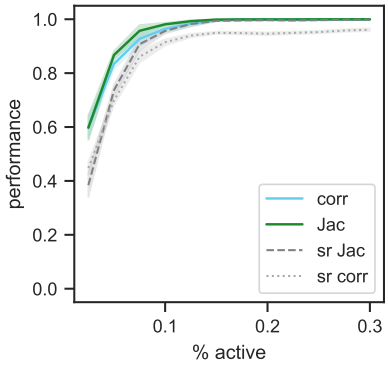


Figure 4.6: **Influence of assembly activity on performance.** The assembly activity controls the number of time windows in which each assembly is active (given as a percentage of all time windows in the data set). The asynchrony (probability of activity of an individual neural word in a non-assembly time window) was set to 0.1 for all conditions. Performance of neural topic modelling was measured by correlation (blue) and Jaccard score (green). The grey lines represent the baseline of the performance achieved by calculating the performance of each individual, single results (sr) i. e. the results without running multiple initialisations and clustering. Shaded areas represent the standard deviation of the results from 10 runs with different random artificial inputs.)

modelling seems to be capable of separating assemblies effectively, since performance does not decrease with higher assembly activity.

Real datasets are likely very noisy i. e. include many neurons with no correlations to other neurons in the data. For example, recordings from Neuropixel electrodes can include 250-1000 neurons of which 40-70% may not correlate highly with the other neurons in the recording (Diana et al., 2019; Stringer et al., 2018) (they might of course be active in concert with other neurons not in the recording). Any assembly detection method should therefore be largely insensitive to neurons in the data which are not part of any assembly.

To assess neural topic modelling performance for this more realistic case in which not all neurons in the recording a part of an assembly, I added randomly active "noise words" with the same total activity levels as the neural words belonging to at least one assembly. Adding such noise words to the correlated activity neural words does not seem to affect the performance of neural topic modelling until there are 4x as many noise words as assembly words in the data (1200 additional noise words) as can be seen in Figure 4.7. The decrease in performance when 1600 noise words are added might be attributable to the fact that the assembly size falls below 5% of the total

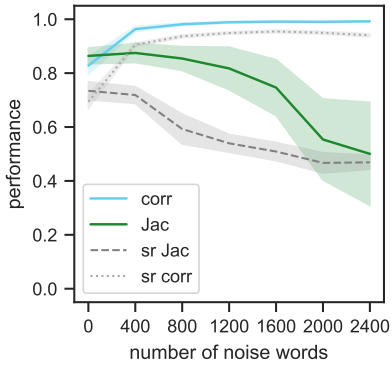


Figure 4.7: **Influence of additional noise words on performance.** Performance of neural topic modelling was measured by correlation (blue) and Jaccard score (green). The grey lines represent the baseline of the performance achieved by calculating the performance of each individual, single results (sr) i. e. the results without running multiple initialisations and clustering. Shaded areas represent the standard deviation of the results from 10 runs with different random artificial inputs.

number of neural words: at a total data set size of 2000 the size of each assembly consists of 4% of all neural words. It is worth noting that we can only tell from the Jaccard score that the performance decreases, since the correlation score increases to nearly 1 regardless. This is due to the fact that at 2000 neural words including 1600 noise neural words most entries are -1's. When no assemblies are detected, the relatively small number of incorrect entries in the matrix will make less of a difference.

As stated in Section 2.2, neurons are able to code for multiple stimuli at once, suggesting that each neuron may be part of more than just one assembly (Carrillo-Reid et al., 2016). This needs to be taken into account by assembly detection algorithms. The algorithms suggested in Diana et al. (2019) all struggle when one fifth of neural words are allowed to belong to multiple assemblies: the performance decreases from close to 1 to around 0.7. Neural topic modelling on the other hand is based on text analysis in which one word (here: neural word) can be part of multiple topics (here: assemblies) – for example the word "live" may belong to topics covering medicine, concerts, broadcasts etc. This is likely the reason why neural topic modelling performance does not decay when neural words are members of multiple assemblies as can be seen in Figure 4.8.

This section has shed some light on the performance of neural topic modelling under difficult conditions for assembly detection. In the following sec-

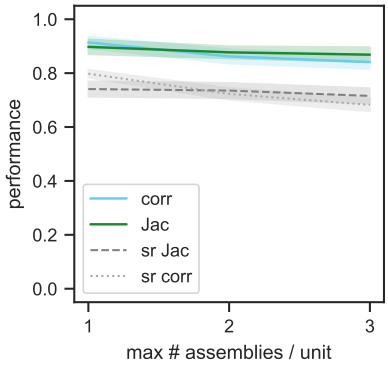


Figure 4.8: **Influence of multi-assembly memberships on performance.** 20% of all neural words were allowed to belong to 0, 1 or 2 additional assemblies (resulting in 1, 2 or 3 assemblies per unit). Performance of neural topic modelling was measured by correlation (blue) and Jaccard score (green). The grey lines represent the baseline of the performance achieved by calculating the performance of each individual, single results (sr) i. e. the results without running multiple initialisations and clustering. Shaded areas represent the standard deviation of the results from 10 runs with different random artificial inputs.)

tion I will expand on the detection capabilities under more extreme multi-membership scenarios.

4.5 DETECTING NON-OBVIOUS CORRELATIONS

The first step of many assembly detection algorithms is to calculate the pair-wise correlations between each pair of neurons on the basis of (often z-scored) firing rates (Humphries, 2017; Lopes-dos Santos et al., 2013; Peyrache et al., 2010; Romano et al., 2015; See et al., 2018). Thereafter these algorithms focus on finding structure in the correlation matrix or use it for further calculations. However, some correlational structures do not appear in the correlation matrix. This section presents a case for which the pair-wise correlation matrix does not offer the most relevant information about the assemblies, but where neural topic modelling is still able to recover the ground truth assemblies.

Let's consider three groups of neurons A, B and C. They do not form three separate assemblies, but instead their activities are correlated in a circle or network (an example of a network of four groups of neurons is shown in the top row of Figure 4.9). Group A and group B are active together (e. g. when a certain stimulus is shown), group A and C are active together (e. g. when

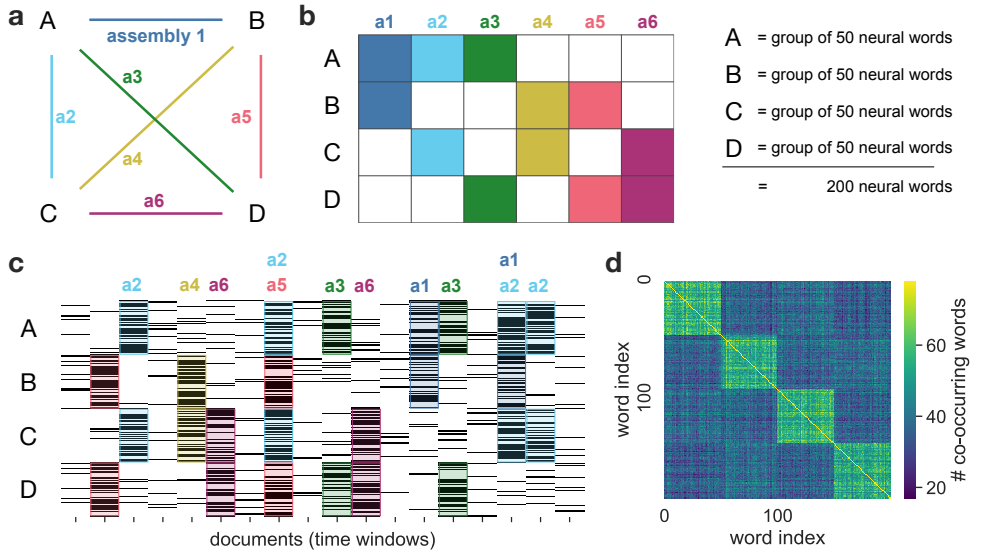


Figure 4.9: **Input with non-obvious correlations.** The input shown is a network of four groups of neurons. **(a)** The groups of neural words (50 per group) and the connections forming the assemblies. **(b)** Schematic of which neural words are expected to be active in a time window in which a specific assembly is active. Groups are never active just by themselves, only in a pair. **(c)** The first few time windows of an input matrix. A black line represents activity of that neural word in that time window, whereas a white line represents non-activity. **(d)** The correlational structure does not show up in the matrix of the pairwise number of co-occurring words.

a different stimulus is shown) and groups B and C are active together (e. g. when another separate stimulus is shown). An example of such activity for four groups can be seen in Figure 4.9 (c). The "correct" assemblies that we would want an assembly detection method to find are an assembly consisting of neural words from groups A and B (a1 in Figure 4.9 (b)), neural words from groups A and C (a2 in Figure 4.9 (b)) and neural words from groups B and C (a6 in Figure 4.9 (b)). A correlation-based assembly detection method or one that does not do well with multi-memberships however will probably identify either groups A, B and C as separate assemblies or will group all groups of neurons into one large assembly since there is no structure in the pairwise correlation matrix. An example of such a correlation matrix with four groups can be seen in Figure 4.9 (d).

To test neural topic modelling for such a case, I construct input which includes such small networks of groups of neurons (40 or 50 neurons each) in which each pair of groups form one assembly. I vary the network size from 3 groups of neurons to 5 groups of neurons. If the network size is 3, each group of neurons is part of 2 assemblies (and each neuron is part of 2 assemblies). If the network size is 4, each group of neurons is part of 3 assemblies (and each neuron is part of 3 assemblies). If the network size is 5,

each group of neurons is part of 4 assemblies (and each neuron is part of 4 assemblies). In each case, the correlational fine structure within the network is not obvious from the correlational matrix.

One note on the use of the number of co-occurring words as a proxy for a correlation matrix: correlation matrices as used in other assembly detection algorithms are the result of pairwise correlations of the time courses of all neuron pairs in the data. The binary neural word activity vectors of the input I use here do not represent a normally distributed time course of neural activity (in fact it the vector only indicates the presence of peaks in the neural activity), making correlation an inadequate measure to use. Since the input vectors are binary vectors (only include 0's and 1's), the summed dot product is an adequate proxy of the correlation.

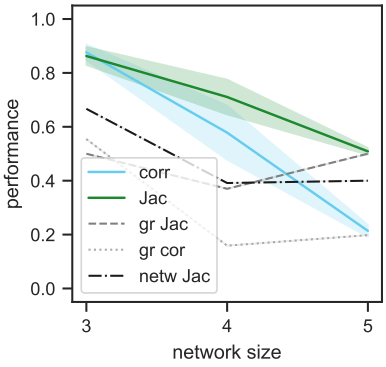


Figure 4.10: **Performance on input with non-obvious correlations.** The assembly detection by neural topic modelling correlates with the network size and thereby also with the number of assemblies each neural word is part of. Networks were constructed of 3 to 5 groups of neurons of 40 or 50 neurons each. To achieve a similar data set size and similar assembly size (80-100 words per assembly) to other test data sets in previous sections, a data set consisted of 3 (when network size was 3) or 2 networks (when network size was 4 or 5). Performance of neural topic modelling was measured by correlation (blue) and Jaccard score (green). The grey and black lines represent the baseline of the performance: the Jaccard scores for the case when only the individual groups or only the networks are identified as assemblies and the correlation score when only the individual groups are identified. The correlation score for the identification of the networks was dropped because it misleadingly remains at 1 since the correlation matrix is not fit for representing multiple assembly memberships as detailed in Section 4.2.2. Shaded areas represent the standard deviation of the results from 10 runs with different random artificial inputs.

Neural topic modelling is able to detect the "correct" assemblies i. e. find the different combinations of the small groups as seen in Figure 4.10. The network size was increased from three to five and as a result the number of

assemblies each neural word is a member of increases from two to four. I constructed input which includes two or three networks, which are all-to-all connected as visualised in Figure 4.9. While performance is still reasonably high when neural words are part of up to three assemblies, is drops when the number of memberships per neural word increases beyond that. Indeed, the performance drops to the scores expected when only the individual groups are identified, and further enquiry confirmed that neural topic modelling will also fall back to this option.

This section gave an insight into the behaviour of neural topic modelling for a highly specific case in which all neural words belong to multiple, overlapping assemblies. This specific artificial data is not a realistic case that would be found in neural recordings, but it induces confidence that neural topic modelling would be able to detect such assembly interactions in real data should they exist there. In the next section, I discuss more realistic cases of assemblies and argue that we should re-think the way we have looked at neural assemblies.

4.6 REALISTIC ASSEMBLY ASSIGNMENT

Traditionally, neuronal assemblies have been seen as these groups of neurons that have coordinated behaviour. An assembly is a specific subset of the neurons in a recording; each neuron may be either part or not part of the assembly. I argue that this forced binary is not necessarily an accurate description of neuronal activity. In reality, neurons tend to not fire exactly according to an assembly they might belong to and some neurons might adhere to assembly activity more than others.

I therefore propose to look at assemblies slightly differently: neurons might belong to an assembly with different "affinities" or "weights". This means that there might be a few neurons which closely follow the assembly activity and other neurons being more loosely connected e. g. displaying concerted activity in some contexts and not others. This dimension of assembly structure is lost when looking at neurons in possible assemblies in an all-or-nothing way and might mean that we loose out on some interesting fine structure of neuronal activity because of our simplistic approach.

Some algorithms for the detection of neuronal assemblies are perfectly able to recover a certain fine structure since their results include an affinity or weight of each neuron to each assembly (Lopes-dos Santos et al., 2013;

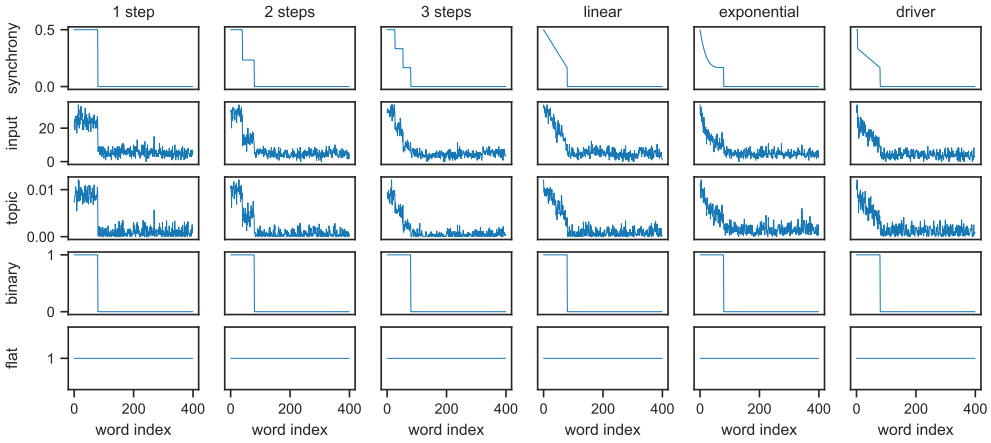


Figure 4.11: **Creation of more realistic assemblies with varying strengths of assembly memberships.** This grid exhibits an example for each base shape (1 step, 2 steps, 3 steps, linear, exponential, driver) that was used in the following analyses. The first row depicts the schematic of the shape the synchrony was based on. The second row shows the synchrony in the input (document-word-matrix) as a number of active documents when the assembly is active. The third row shows the actual weights neural topic modelling assigned to each word. The fourth row visualises the traditional view of how these assemblies would be viewed by any other assembly detection algorithm though the either-in-or-out lens. The fifth row shows an additional baseline case in which all words are assigned the same weight in the assembly.

Peyrache et al., 2010; Romano et al., 2015). However, they all threshold this affinity to achieve either-in-or-out assemblies (Mölter et al., 2018).

Neural topic modelling provides the weights of how much neurons belong to certain assemblies. So the question I wanted to answer is whether neural topic modelling can recover certain "shaped" fine structures of assembly memberships. To demonstrate the reproduction of these fine structures, I generated artificial data for which neural words in an assembly had different affinities (realised by giving them different synchronies). In previous sections, all the neural words in one assembly had the same synchrony. As a result, I tested six different "shapes" of assembly membership: 1 step, 2 steps, 3 steps, linear, steps, exponential, drivers, see Figure 4.11. The choice of shapes was loosely based on assembly fine structure found when running neural topic modelling on actual brain recordings.

To measure the performance of neural topic modelling in recovering the shape of the assembly affiliations, I calculated the cosine distance, which describes the similarity of two vectors by calculating the cosine of the angle between them. This has the advantage that the scale of the vector does not matter, and it instead becomes a measure of how closely the two vectors (which are ordered in exactly the same way) resemble each other. A value

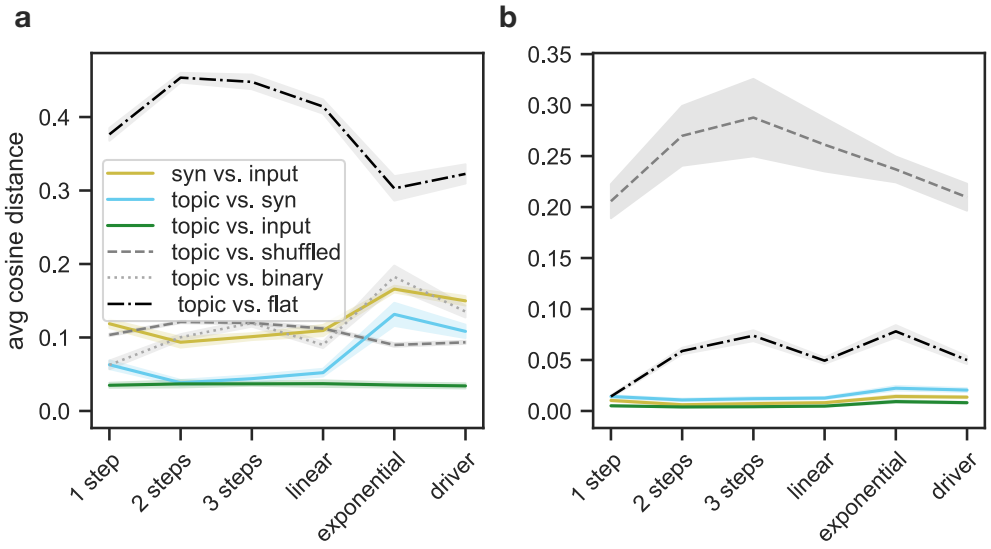


Figure 4.12: **Performance using shaped assembly structures.** Neural topic modelling reconstructs the affinities of the neural words in the assemblies quite well. **(a)** The cosine distance of the found assemblies to the ground truth shape (synchrony) and actual overlap of activity between the neural word and the assembly (real) was calculated. **(b)** To assess whether the low distance by the shuffled input was due to the large number of neural words outside of the assembly, the cosine distance was re-calculated using only the neural words in the assembly. The grey and black lines represent the baseline of the cosine distances: the shuffled vector, a step function as a representation of an either-in-or-out assembly structure and a flat line. The step function and the flat line are also depicted in Figure 4.11. Shaded areas represent the standard deviation of the results from 10 runs with different random artificial inputs.

close to 0 means perfect match and a value closer to 1 means a dissimilar shape.

I calculated the cosine distance from the resulting topic to the ideal shape of the synchrony (see Figure 4.11 row 1) and to the actual co-occurrance structure of the words to the assembly activity (see Figure 4.11 row 2). In addition, I calculated three baseline cases: shuffling the weights in the topic, a step function as would have been the perfect result from another assembly detection algorithm (see Figure 4.11 row 4), and a flat line representing the case where no assembly structure was found (see Figure 4.11 row 5). To put the distance between the topic and the synchrony into context, I also calculated the average cosine distance of the synchrony (see Figure 4.11 row 1) to the actual input (see Figure 4.11 row 2).

Figure 4.12 (a) shows that the cosine distances between the topic found and the actual input used is very small. The topic also resembles the synchrony more than the actual input (blue line is below yellow line). For three shape types (2 steps, 3 steps, linear), the distance between the found topic

and the synchrony is also smaller than for the shape type "1 step", which represents the traditional view of neuronal assemblies (all neurons in an assembly contribute to it to the same degree). Neural topic modelling is actually better at recovering the weight structure in an assembly when neurons do not show the exact same amount of affiliation with the assembly. For the exponential and the driver shapes, shuffled weights actually also achieve a very small cosine distance. This is probably due to the fact that many weight in these scenarios are low and that also many neural words which are actually in the assembly do not have a much higher synchrony than some of the non-assembly neural words. To prove that this drives the good performance of the shuffled input, I calculated the cosine distance for only the neural words in the assembly, see Figure 4.12 b). The cosine distance between the found topic and both the ideal shape of the synchrony and the actual fine structure of the input is very close to zero and significantly below shuffled and flat baselines.

I have shown in this section that neural topic modelling is capable of recovering complex and distinct assembly structures. In the next section I will summarise the findings of the performance of neural topic modelling in comparison to existing methods and will outline the impact the additional information provided by neural topic modelling can have on neural data analysis.

4.7 DISCUSSION

Neural topic modelling is an assembly identification algorithm which performs better or to the same standard as previously suggested algorithms (Diana et al., 2019) on the same artificial data under the same performance metric. In this chapter, I have illuminated strengths and limitations as well as presented the case that neural topic modelling can be used for further analyses compared to most other assembly detection algorithms.

As a first step, I confirmed that the performance of neural topic modelling is comparable to the results of four state-of-the-art methodologies obtained by Diana et al. (2019) with regards to the same parameter variations. Neural topic modelling performs better than Bayesian inference – the best of the four tested – when it comes to different data set sizes and number of assemblies. Further probing into the capabilities of neural topic modelling revealed that assemblies are detected at a better rate when the assemblies were active at least 5% of the total number of time windows. Assemblies

were also identified more accurately when the assembly activity made up more than 15-20% of the total number of neural words in the data, which is a realistic estimate for currently recorded data sets (Stringer et al., 2018). The results from neural topic modelling were only minimally influenced by neural words belonging to more than one assembly.

Next, I created assemblies in which every neural word was part of more than one assembly and where the assembly structure was not obvious from the pairwise correlations of the neural words. Neural topic modelling was able to identify the correct assembly structures as long as neural words belonged to three or fewer assemblies. Finally, I demonstrated that neural topic modelling is able to recover more complex assembly structures than the traditional binary representations of neurons being either in or not in an assembly. When neural words in an assembly possess different synchronies (have different probabilities whether they will be active when the assembly is active), the shape of these synchronies is recovered in the assembly weights.

Neural topic modelling seems to be more sensitive than Bayesian inference and PCA+varimax to high levels of asynchrony (the number of active time windows that does not align with the assembly activity). The performance of neural topic modelling dips beneath the performance of Bayesian inference and the PCA-based method when asynchrony increases beyond 0.25. At this level of asynchrony each neural word is active 10 times as often when its assembly is not active than when its assembly is active. The good performance of Bayesian inference at such high asynchrony levels in that plot are impressive. However, this high performance is slightly surprising since in a different figure in the paper, the performance of Bayesian inference at an asynchrony of 0.25 and higher is given as much lower. So it might be good to keep this in mind, that there might be a difference in the artificial data generation which was not reported clearly in the paper. In any case, the first step of neural topic modelling is an event filtering step with the explicit goal of limiting the activity of a neural word to time windows in which the activity pattern is most likely to be related to meaningful activity. This step hopefully reduces the asynchrony of the neural words to levels which allow a good performance of neural topic modelling.

Most of the presented artificial data sets do not represent realistic assembly activity. Assemblies in actual electrophysiological recordings or calcium imaging data sets will most definitely not display the same average activity levels, assembly affinities and sizes as in the test data sets. Using these uniform data sets was a choice to make the results as comparable as possible to

previous publications on the subject (Diana et al., 2019) and to control the input statistics. By adding noise and by varying the synchrony, I attempted to generate input which is more similar to the data that the method will be used on with good assembly detection performance. I have also shown that neural topic modelling is very successful for a range of activity levels of assemblies and despite increasing overlap of assembly activity in time windows. Since experimental protocols often include multiple repetitions of a certain stimulus, these activity levels are hopefully obtainable in practice (Lopes-dos Santos et al., 2013; Peyrache et al., 2010; Stringer et al., 2018). The results from these tests induce confidence that neural topic modelling will reliably identify meaningful assemblies in recorded neural data sets.

While the proposal to look at assemblies in a more non-binary way is a new suggestion, other methods are capable of providing such an estimate. It is difficult to comment on whether neural topic modelling's results are closer to the ground truth since no other publication has provided this analysis. Therefore I can only point out the high similarity between the found assemblies and their underlying synchrony structure and argue that neural topic modelling will identify assembly structures that are grounded in data.

Neural topic modelling does not dampen the weights of neural words which are slightly but not closely associated in the assembly to very low values. Instead of looking for closed sets of neural words as for example in frequency item set mining (Picado Muiño et al., 2013) and excluding neural words which do not meet specific criteria, neural topic modelling gives a more raw result, which can be used for further investigation. These raw results do represent the data accurately, since neural words are giving a weight according to how much activity they share with the assembly. Further investigations will be needed to find out whether the order and hierarchy of these weights are meaningful or might of course occur by chance.

Neural topic modelling is likely to be less accurate than for example Bayesian inference on estimating the number of assemblies in the data. Diana et al. (2019) do not report the numbers of assemblies found so that it is not clear whether their lower scores stem from detecting the wrong number of assemblies or from incorrectly identifying the assemblies. It is therefore not easy to compare the results. However at default settings neural topic modelling does tend to find one or two additional assemblies which represents a hierarchy of neural words which are active during time windows when no assembly is active. The method is more likely to identify an additional assembly when the ground truth assemblies have also been found. So

the over-estimation of assemblies by neural topic modelling does not come at the cost of a lower performance as long as the number of topics is chosen to be sufficiently high. Since neural topic modelling is a tool to find previously unknown relationships in neural data, it seems more important that the ground truth assemblies are recovered than whether "only" the ground truth assemblies are identified.

There seems to be a minimum size for an assembly to be reliably detectable by neural topic modelling. When an assembly includes less than 5% of the whole recording, then the algorithms does not perform as well as for larger assemblies. There are a few possibilities to remedy this, but in general this might present a systemic problem. It is possible to expand the analysis to assemblies which appear in a lower number of different random initialisations (currently this threshold is set to 50% of initialisations). This indeed increases the chances of additionally identifying correct assemblies, but will also increase the number of assemblies not based on the ground truth. There is also the hope that a small but very consistent assembly might have multiple more loosely associated assemblies which increase the chances of detection. Whether this is a problem in other assembly detection algorithms is not clear to me since this information is not included in the respective publications. This is definitely a point to keep in mind. But generally, the trend is going into the direction of larger and larger brain recordings, which will hopefully bring larger networks of correlated neuronal activity to the surface, making neural topic modelling a good tool to analyse such larger networks.

For all the limitations mentioned above, neural topic modelling also has some specific advantages which make it a unique tool in the pool of assembly detection methods.

Many current models of assembly detection are based on the pairwise correlations between all z-scored or otherwise normalised spike trains (Lopes-dos Santos et al., 2013; Perkel et al., 1967; Smith et al., 2018). As mentioned above and shown in Section 4.5, this correlation matrix may not reveal the full picture. Other methods have moved away from the correlation matrix because of the same concerns (Diana et al., 2019; Mölter et al., 2018; Picado Muiño et al., 2013). They instead work on the binary matrix which denotes the presence or absence of a peak in firing rate per neurons – a similar concept to the event filtering in neural topic modelling. The main difference is however that neural topic modelling can extract multiple different activity patterns from one spike train (1 neuron $\rightarrow$ multiple activity patterns). The

current methods ignore the influence of spike patterns other than firing rate and operate under the assumption that 1 neuron → 1 activity pattern of firing rate. Moreover, neural topic modelling can be applied on any type of data that can be event-filtered and on any appropriate time scale. It would for example be possible to combine spike patterns and local field potentials in the same analysis. So one major contribution of neural topic modelling is to allow multiple facets of neural activity to be included in the analysis (see Chapter 3).

Another feature setting apart neural topic modelling from Bayesian inference and other methods is the consistently high performance despite some neural words being part of multiple assemblies. The supplemental material of Diana et al. (2019) shows that performance of Bayesian inference decreased from close to 100% to 60-80% when only 20% of neurons belong to two or three assemblies. A similar dramatic performance drop was observed for spectral clustering and PCA+varimax (Lopes-dos Santos et al., 2011; Romano et al., 2015). This makes sense for PCA-based methods since principle components are pushed towards being orthogonal. Methods whose performance does not degrade with multi-memberships (Avitan et al., 2017; Von Luxburg, 2007) were nevertheless mostly tested on pairwise overlap (Mölter et al., 2018), so more complex overlap scenarios might also pose a problem. The performance of neural topic modelling, on the other hand, stays the same despite the same percentage of neural words being part of multiple assemblies. Even when all neural words belong to two or three assemblies, the assembly detection is high. This is likely due to the fact that Latent Dirichlet Allocation (LDA) – the generative model neural topic modelling is based on the assumption that words may appear in different contexts (e. g. the word "score" can appear in texts about sports, film music or school reports). The assumption that the same word may be present in different contexts is basically built into the algorithm from the start and comes in handy since it probably reflects the reality of neural activity. The traditional assumption that 1 neuron → 1 assembly neglects to take into account the whole breadth and variability that neurons display constantly (Hardcastle et al., 2017). Neural topic modelling allowing for such interactions of one neural word with many other combinations of neural words may give researchers additional insights beyond the findings from other methods.

Another suggestion I am making is to look at the found assemblies in more detail and not to dismiss the weights of each neural word in each assembly. The ability to do this is not exclusive to neural topic modelling, since other assembly detection algorithms also end up with a similar measure of

affinity of each neuron to an assembly (Lopes-dos Santos et al., 2011, 2013; Mölter et al., 2018; Peyrache et al., 2010; Romano et al., 2015). The standard procedure of all those methods is to choose a cut-off value and declare all neurons above the threshold to be in the assembly and all others not to be. Thresholding these weights somewhere without examining the fine structure seems somewhat like a waste of information. Neural topic modelling is a data exploration tool and its goal is to generate new hypotheses about the data and pointing to areas where to base subsequent analyses. Taking these complexities of assembly membership into account might be an interesting way of thinking about neural data and cooperation between neurons.

There are a few directions which might be pursued to further explore the capabilities of neural topic modelling.

Similarly to many other methods, neural topic modelling creates a binary matrix indicating in which time windows the neural words were active. At the moment this matrix is binary. However the topic modelling algorithm can accept any integer values instead of only 1's and 0's. When applying this on text data, this matrix indicates the number of times a word occurred in each document (equivalent to the time windows in neural topic modelling). It would therefore be possible to quantify for example the firing rate increase of a certain neuron and populate its vector with 1's, 2's and 3' depending on how high the firing rate increase actually was. Assigning a fair and balanced tiering system for different neural activity patterns might be a challenge since it is unclear what the equivalent of a doubling of the firing rate is for the neural word of coincidental spiking. These are ideas and questions for future works, but they might significantly improve the accuracy of neural topic modelling.

There also exists the possibility of replacing the LDA algorithm with a different one. Non-negative matrix factorisation (NMF) for example has been shown to find similar groupings as LDA. In addition there exist several more sophisticated versions of topic modelling algorithms. Once example is semi-supervised LDA, which has been successfully applied to other non-text data sets (Hughes et al., 2018, 2017) and also implementations which determine the optimal number of topics to find. The latter would be a good innovation, since it eliminates the need to specify the parameter k. Future expansions of neural topic modelling should verify that the automatic estimation of the number of topics works as well as the current implementation.

4.8 CONCLUSION

Neural topic modelling is an assembly detection algorithm that shows a, for the most part, better performance on artificial data than current state-of-the-art methodologies. On top of that, neural topic modelling encourages a change in how we think about and how we attempt to identify neural assemblies: it allows for neurons to perform multiplexed coding by enabling the separation of different activity patterns and it also allows these activity patterns to be part of multiple assemblies without a drop in performance. The flexibility neural topic modelling allows for is unique in the pool of neuronal assembly detection algorithms, making it an ideal starting point of finding previously unknown relationships in large-scale neural data. At the same time, it is the user's choice whether to make use of these flexible options – it is also possible to use neural topic modelling for finding neuronal assemblies only based only firing rate correlations as done by other methods.

The goal of this chapter was to demonstrate the value and novelty neural topic modelling can offer to neural data analysis. Most currently available assembly detection method find assemblies of neurons which show co-modulations in their firing rates, even though the algorithms will arrive at slightly different assemblies (Mölter et al., 2018). Neural topic modelling is a useful tool for exploratory analyses with few restrictions, great flexibility and the ability to uncover unknown activity structures in large-scale neural data.

5

NEURAL TOPIC MODELLING - APPLICATION TO BIOLOGICAL DATA SETS

The previous two chapters have outlined the steps involved in neural topic modelling and explored its neuronal assembly detection capabilities on artificial data sets. In this chapter, I present the results of applying neural topic modelling on large neuronal recordings as a demonstration of what kind of new insights can be gained from this method.

Firstly in Section 5.1, I provide details about the data that was used in this chapter: the electrophysiological recordings of a mouse brain and the visual stimuli that were shown during the recording. The parameters used for the analysis in neural topic modelling are outlined in Section 5.2. In Section 5.3 I detail how the neural words extracted from the data relate to the firing rates of the units and how the neural words from the local field potential (LFP) signals are processed. Additionally, I present evidence for multiplexing in units in the recording in Section 5.4.1.

I present the results from running neural topic modelling on this large neuronal recording in three parts. The neuronal assemblies resembling assembly structures already identified by other methods are shown in Section 5.4.3. Section 5.4.4 presents the assembly structures which have not been detected by other methods because these do not take into account all the aspects of neuronal recordings that neural topic modelling does. It is also possible to compare the cooperation between different brain structures across different experiments in a higher level analysis as demonstrated in Section 5.4.5. Finally, I will discuss the results and outline the limitations and weaknesses with respect to applying neural topic modelling to neuronal recordings in Section 5.5. Future directions of this research are outlined in Section 5.7.

5.1 ALLEN INSTITUTE DATA

This section gives details on the data used to demonstrate the application of neural topic modelling to real data. The Allen Institute provides large electrophysiological data sets through their *Visual Coding – Neuropixels* project which are of high quality, well documented and were acquired in line with current ethical guidelines for animal experiments. Thanks to their commitment to open science, many researchers now have access to large-scale neuronal data sets furthering the development of new analysis tools for neuroscientific data.

The following section will describe the experimental setup used to acquire the data.

5.1.1 *Recording and preprocessing*

The information about recording procedures and experimental setups in this section is based on the Technical White Paper published by the Allen Institute Allen Brain Observatory (2019).

The Visual Coding data sets used in this chapter are *in vivo* electrophysiological recordings from six Neuropixel electrodes (Lopez et al., 2016) implanted into a brain of a wild-type C57BL6/J mouse (Allen Institute session ID: 767871931).

The experimental setup for recording such an elaborate data set is complex and a full description of the setup is given in Allen Brain Observatory (2019). To ensure similar entry points and angles of the probes across experiments, the six probes are placed on a custom-built rig from which their location in space can be controlled. The six electrodes are called A, B, C, D, E, F with probe A in the anteriomedial position (top right in the plots shown here) and subsequent letters are assigned to electrodes in a clockwise manner (see Figure 5.1). Each probe enters a different part of the visual cortex – chosen to target separate functionally defined visual areas (Juavinett et al., 2019) – and also extends to subcortical structures such as the hippocampus, thalamus and the midbrain. The positioning of the rig depends on the retinotopic map of each mouse. A visualisation of the physical locations of the electrodes can be found in Figure 5.1 (b). The electrodes sometimes appear curved because the position in space was registered to the CCF template brain to standardise the labelling of the brain regions. Similarly, electrodes seem to be of different

lengths, which is due to the viewing angle. In reality, all electrodes used were straight and of the same length during the recordings.

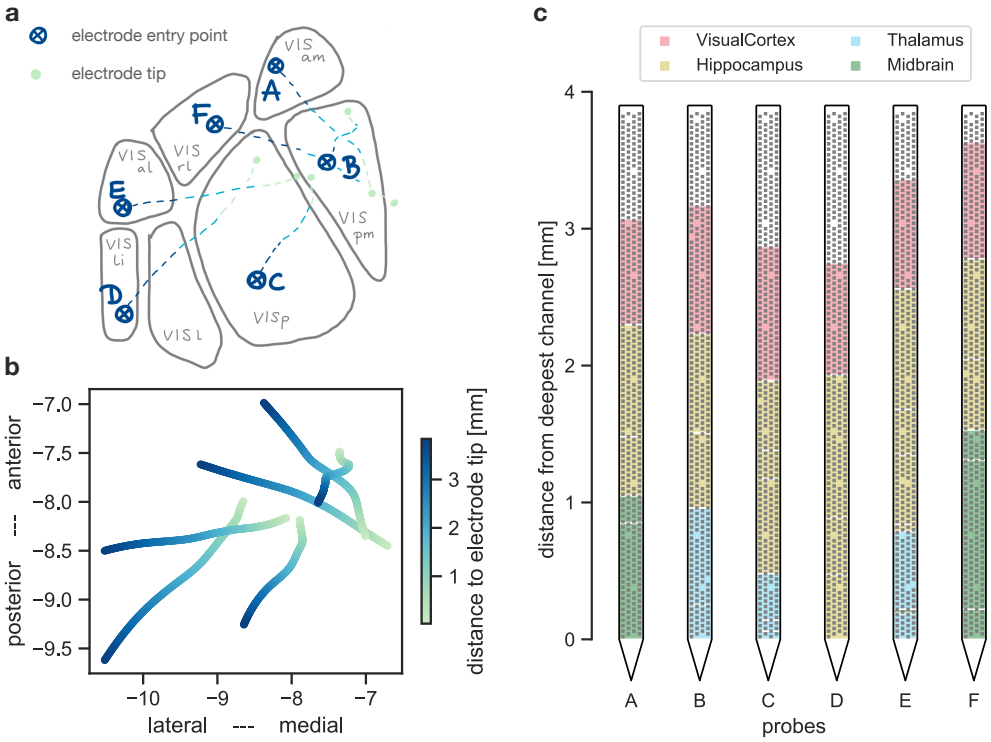


Figure 5.1: **Visualisation of the positioning of the electrodes used to record the data. (a)** The locations of the six entry points of electrodes A-F are shown. **(b)** A visualisation of the paths of the electrodes in standardised coordinates. **(c)** The brain structures each electrode traverses. VISam = anteromedial visual cortex, VISpm = posteriomedial visual cortex; VISp = primary visual cortex; VISli = laterointermediate visual cortex; VISal = anterolateral visual cortex; VISrl = rostrolateral visual cortex

During the recording the head-fixed mouse stood on a disc-shaped wheel on which it could walk or run. The velocity of the wheel was recorded and is also openly accessible.

The six Neuropixel electrodes contain either 374 (Neuropixels 3a) or 383 (Neuropixels 1.0) channels (Jun et al., 2017) from which extracellular signals were recorded at a sampling rate of 30 kHz (spikes) and 2.5 Hz (LFP). Preprocessing included a 500 Hz high pass filter for the spike band and a 1000 Hz low pass filter for the LFP band (Allen Brain Observatory, 2019). To extract spike times of individual units[1] from the signals, the spike sorting software Kilosort2 was applied (Stringer et al., 2018). Even though LFP was recorded

1 Since it cannot be guaranteed that all spikes assigned to one unit actually were emitted from one individual neuron (e. g. spikes from equidistant neurons might be assigned to one unit), the Allen Institute documentation used the term "units" instead of "neurons". This terminology was adopted in this chapter.

from every channel, it was downsampled so that only every fourth channel was included in the data set. LFP signals were also high pass filtered at 0.1 Hz and noise common to all electrodes was subtracted by re-referencing to channels outside of the brain.

The whole data set including spikes, LFP and behavioural data is available in the Neurodata Without Borders data format and is available for processing and download via the open source AllenSDK Python package (https://github.com/AllenInstitute/AllenSDK). The particular data set used in this thesis included 1604 units and LFP signals from 461 channels (all the channels located inside the brain of the mouse).

5.1.2 *Experiments*

The mouse whose brain activity was recorded was shown visual stimuli for passive viewing. The stimuli were presented on a monitor 15 cm in front of the mouse's right eye and the images shown were warped to account for the proximity of the mouse to the monitor.

For the extent of this thesis, four experiments and one time period without visual stimulation were analysed. The timeline of the experiments and an example image for each are presented in Figure 5.2.

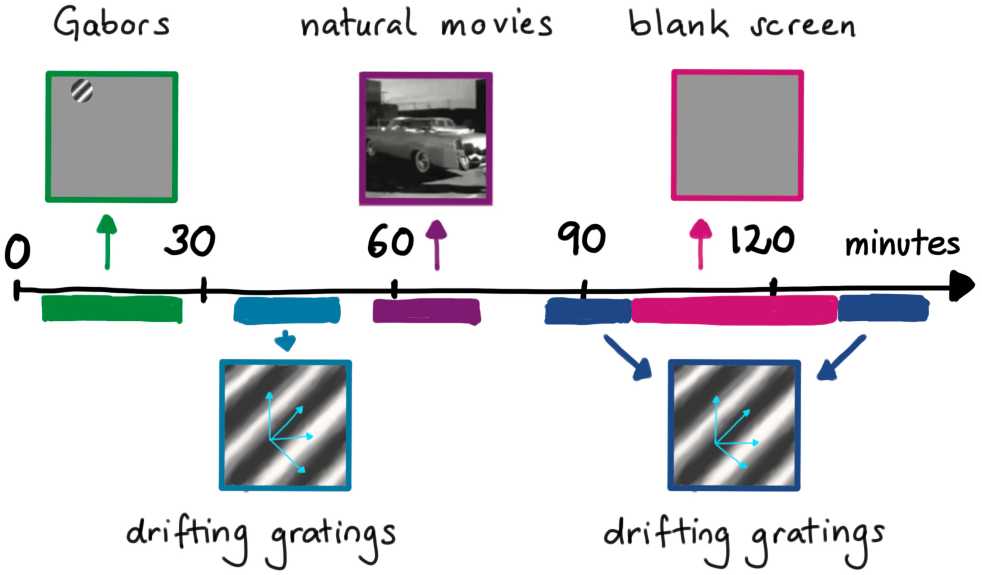


Figure 5.2: **Experiments** The visual stimuli presented to the mouse.

The experiments included in the analyses were:

- **Gabors (gab):** Gabor patches (small circles of drifting gratings at 2 Hz, see Figure 5.2) were presented at one of 81 locations (on a 9 x 9 grid) for 0.25 s each.
- **Drifting gratings contrast (dgc):** Drifting gratings were presented across the entire screen. The gratings could be moving in one of four directions (0^{o}, 45^{o}, 90^{o}, 135^{o}, clockwise where 0^{o} represents movement from left to right) and one of nine different contrasts (0.01, 0.02, 0.04, 0.08, 0.13,0.2, 0.35, 0.6, 1.0) for 0.5 seconds each.
- **Drifting gratings repeats (dgr0, dgr1):** Similarly to the drifting gratings contrast described above, drifting gratings were shown to the mouse at the same four directions (0^{o}, 45^{o}, 90^{o}, 135^{o}) but only at two different contrasts (0.1 and 0.8) for two seconds each. This paragigm was shown twice and both instances were included in the analysis.
- **Natural Movies (nat):** The "natural" movie shown in this condition is a 30 s greyscaled and contrast-normalized clip from the opening scene of the movie Touch of Evil (1958), which was repeated 60 times.
- **Spontaneous activity (spont):** Between the presentation of the two drifing grating repeat experiments there was a time period where no stimulus was shown to the mouse. This recording was also included in the analysis.

5.2 PARAMETERS

The previous sections described the large data set used in this chapter to demonstrate that neural topic modelling can obtain interesting results which could not be found with any other single method. Neural topic modelling was applied to the data in the way described in Chapter 3. Four neural words were extracted from the spike trains and the LFP recordings: high firing rate (high spike count, *sc+*), low firing rate (low spike count, *sc-*), coincidental spiking (*csp*) and the power in five different LFP frequency bands (*lfp*, delta (3-6 Hz), alpha (6-13 Hz), beta (13-30 Hz), high beta (30-40 Hz) and gamma (40-70 Hz)).

The parameters used for the main steps of neural topic modelling are listed in Table 5.1.

During the testing phase of neural topic modelling, I tried window lengths of 50, 75, 100, 150, 200 ms. The results were not overwhelmingly different,

suggesting that picking a time window within such a range will lead to interesting results regardless. I settled on a window length of 100 ms for the work presented in this thesis.

Furthermore I made the decision to let the time windows overlap by 50%. Firing rate is often calculated in non-overlapping documents. However this may lead to effects where bursts of spikes are separated by a time bin border and are therefore not detected by counting spikes per time bin. Having the time windows overlap reduces of this effect. During the development of neural topic modelling I also tried out different degrees of overlap between time windows (0%, 50%, 75%) and since the results were very similar, I arrived at the conclusion that the amount of overlap will not impact the results heavily and choosing a 50% overlap will give the benefits of not separating spike train features by bin borders and creating a large enough number of time windows for topic modelling to be effective.

For the analysis of how the visual stimuli relate to the found assemblies, I needed to relate the experimental and behavioural data that was available back to the time windows. In the case of visual stimuli shown, I labelled a time window with the presence of a particular stimulus if the presentation of that stimulus and the time window of the document intersected at least 75% (at an overlap of 50%).

For each individual word, another threshold has to be set – the percentile threshold basically determining what counts as an extraordinary enough event to be included in the analysis (see Table 5.2). The exact meaning of the threshold for every single neural word was described in Chapter 3 (*sc+* and *sc-*: Section 3.2.2.1, *csp*: Section 3.2.2.2, *lfp*: Section 3.2.2.3).

For the work presented in this thesis I chose the 75th percentile, i. e. I only registered the event "neuron 1 high firing rate" if the spike count in a time window is higher than in 75% of all time windows of that neuron. The number of spikes in the x-th percentile window + 1 becomes the minimum number of spikes that have to occur in a time window in order for a "high firing rate" neural word to be recorded. I chose the threshold for "low firing rate" to be the 25th percentile in a complementary fashion.

For the coincidental spikes neural words, I chose a window of 4 ms to catch precisely timed coincident spikes between pairs of neurons. Since the data sets used in this thesis are recorded from visual cortex, hippocampus, thalamus and midbrain, this relatively large time window should catch interesting synchronous spiking activity (Panzeri et al., 2010). To make the maximum number of word occurrences of the *scp+*, *scp-* and *csp* neural words

Table 5.1: **Neural topic modelling parameters for neuronal data.**

parameter	value
Document creation	
time bin length [ms]	100
overlap of documents	50%
Event filtering	
sc+ percentile threshold	75
sc- percentile threshold	25
lfp percentile threshold	85
csp percentile threshold	75
csp smallest ISI to count as coincident [ms]	4
csp maximum number of words to include	2000
Topic modelling	
number of initialisations per data set	20
number of topics to be found k	50
number of iterations	500

comparable, I chose the same 75th percentile threshold for the *csp* neural word.

Since the number of pairs of neurons grows exponentially, not all pairs of neurons can be included in the analysis. There is no ideal way of deciding on which pairs to select and which pairs to leave out. In text mining this issue is solved by only selecting the x number of words which occur most often (Blei, Ng and Jordan, 2003; Hughes et al., 2018). The exact number of words may differ and may depend on the number of documents available. In general, cutoff values of 5000 or 10000 are commonly used. For neural topic modelling I decided to restrict the number of coincidence spiking words to roughly 2000, which is comparable to how many words were generated from all other word types combined.

For the *lfp* neural word, it is also necessary to choose a percentile threshold, which was set to the 85th percentile, so slightly higher than for the neural words which are based on spiked. The reason for choosing a higher percentile for this neural word is that the nature of the underlying signal is slightly different. The power spectral density in one frequency band is a continuous variable, so choosing a 75th percentile threshold is equal to including all of the top 25% values. Since the other neural words are based on discrete counts, a 75th percentile most often leads to word occurrences in

Table 5.2: **Dataset sizes for different experiments.**

	gab	dgc	dgro	dgr1	nat	spont
number of documents	18240	10800	17996	17996	18016	36051
number of words	3741	3790	3880	3905	3841	3975

a much lower percentage of the time windows (on average around 13% for high firing rate, 16% for low firing rate and 10% for coincidental spiking). To make sure that the number of word occurrences of the LFP neural word are comparable to the other words, I decided to set the percentile threshold for this neural word to the 85th percentile.

The real number of topics in the data is of course unknown, so I decided to look for 50 topics in each of the 20 random initialisations (runs) and let the algorithm run for 500 iterations (which gave good results for the artificial data shown in Chapter 4). The concept of running topic modelling multiple times and clustering the results is described and discussed in detail in Section 3.2.4 in Chapter 3.

The event filtering results in a binary matrix with the dimensions *number of documents* x *number of neural words* with 1's representing that a particular neural word (neural activity event as described in Chapter 3) had occurred in that particular time bin and 0's representing the absence of this neural word in that time bin. Since the experiments were of different lengths, the resulting matrices are also of different sizes as can be seen in Table 5.2.

5.3 NEURAL WORDS

The first step of neural topic modelling is to apply an event filtering step during which the algorithm goes through the spike trains and local field potentials to search for particular activity patterns. This is a step not done in other assembly detection methods although it would not be uncommon in other neuroscientific analyses to find correlates of task or behavioural variables. The analysis of the event filtering step could be a whole chapter by itself, since many interesting insights can be gained by the detailed analysis of neural signals. Since the thesis revolves around neural topic modelling, I will confine myself to briefly presenting the neural words that were extracted from the neuronal recordings.

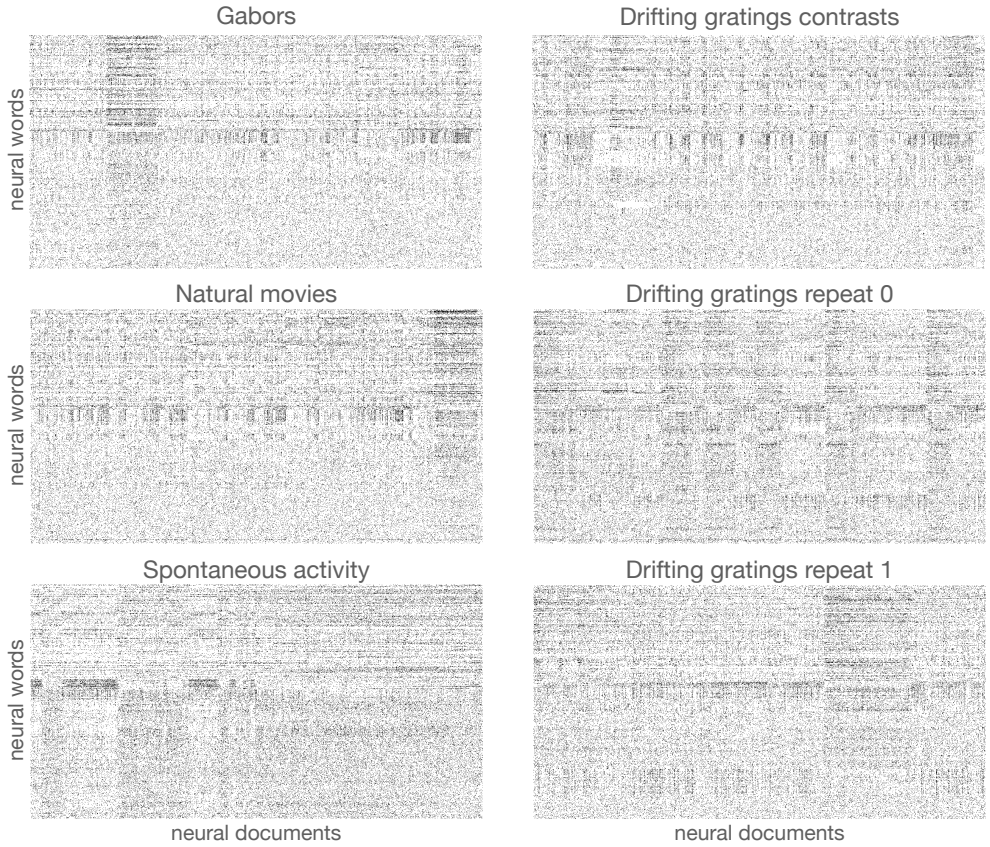


Figure 5.3: **Input matrices for the six different experiments.** Each input matrix is of a different size which can be found in Table 5.2.

This part of the analysis is likely to be data set specific – it depends heavily on how much data is available and what data is included in the recording. Therefore I want to give a few insights into the neural words extracted for this particular data set and their relationship to the firing rates of the units that gave rise to them. I will also discuss the specific preprocessing involved in the *lfp* neural words.

5.3.1 *Instances per word type*

In general, I have tried to keep the instances of each word (the number of times that words was recorded to occur across all documents i. e. 1's in the time window-neural word-matrix) balanced. The number of times a neural word occurs will directly influence the number of co-occurrences that neural word can have with others. Therefore, neural words with much higher word instances will be much more likely to be assigned to multiple

topics. I attempted therefore to ensure that the number of word instances per word type (*sc+*, *sc-*, *csp*, *lfp*) were roughly the same.

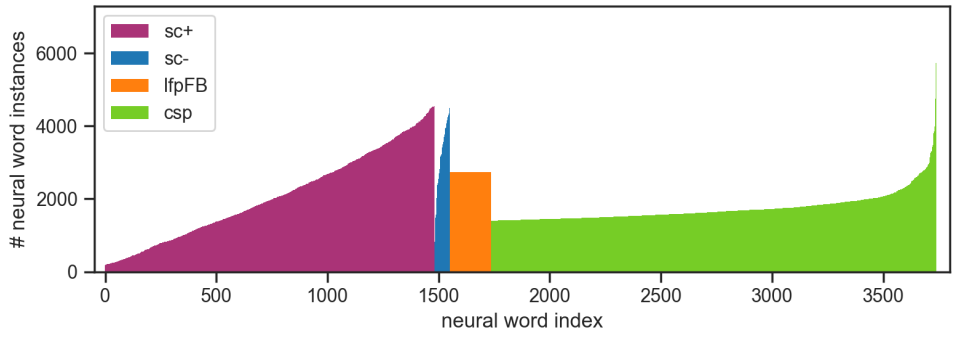


Figure 5.4: **Instances per neural word.** The word parameters – in particular the percentile thresholds – were chosen to make the number of instances broadly comparable across word types.

5.3.2 *Influence of firing rates*

Ideally, all units in a neuronal recording should be treated the same by a data analysis algorithm regardless of the firing rate of the neuron. In practice, this is difficult because the firing rates of units in the recording differ greatly (0.01 - 80 Hz). Another issue arises from the fact that spike counts are measured in discrete small integers. For the creation of neural words it is therefore important to try to compensate for the different activity levels of the units as much is possible.

The *sc+* neural word records an instance of a neural word when the number of spikes in a time window is greater than in 75% of the other time windows at a 75 percentile threshold (for details on the creation of neural words, please see Section 3.2.2 in Chapter 3). This means that for units with firing rate up to 2.5 Hz, practically every time they emit a spike, this will be considered as a high spike count and an instance of a neural word will be recorded in that time window. This steep initial increase in word instances as a result of increasing firing rates until 4 Hz can be seen in Figure 5.5 (a). For firing rates above 4 Hz there also seems to be a firing rate dependence, meaning that units with higher firing rates often result in higher number of word instances. However, the 75th percentile threshold effectively regulates the maximum number of word instances a unit can create (at most, only the top 25% of time windows are given an instance of the *sc+* neural word).

A unit with a low firing rate of around 2 Hz can generate as many word instances as a unit with 80 Hz, lessening the bias towards high firing rates.

The neural word of low firing rate, *sc-*, is of course inherently biased towards units with higher firing rate, since it denotes time windows in which a unit showed fewer spikes than in 75% of all time windows. This is only possible for units with relatively high firing rates since for units with low firing rates, having zero spikes in a 100 ms time window will be the norm. Having said this, units with a firing rate of around 10 Hz already display high word counts of the low firing rate word and there seems to be no bias towards units of any particular firing rate above 10 Hz (see Figure 5.5 (b)).

The neural word of coincidental spikes, *csp*, is based on the synchronous spiking behaviour of pairs of neurons. Figure 5.5 (c) visualises the word instances of coincidental spiking words of unit combinations with different firing rates. Since there are many *csp* neural words with a similar number of word instances (also see Figure 5.4, green bars), most of the dots have a similar green-ish colouring on the lower end of the colour map spectrum. Only a small number of words show very high word instances indicated by the blue colour. Interestingly the unit combinations with the highest numbers of word instances do not stem from units with high firing rates and instead originate from units with firing rates between 5-30 Hz. Another observation would be that single units seem to form a multitude of coincidental spikes with many other units (vertical and horizontal "stripes" of dots).

As much as possible, biases towards units with higher firing rates have been reasonably well regulated, apart from the unavoidable ones due to spike counts being given as small integer values.

5.3.2.1 *Handling of LFP signals*

As mentioned earlier, the LFP signal was available for every fourth channel. LFP signals from channels which are in close proximity are very likely to be correlated (Herreras, 2016), which is also the case in this data set. Calculating the power from each available LFP trace will only result in highly correlated neural words. Therefore, I initially decided to apply the event filtering only to the average LFP trace for different brain region on each electrode.

However, the borders suggested by the labels provided in the data set did not seem to correspond well with the visually perceivable borders in the LFP signal as can be seen in Figure 5.6 (a). This does not imply that the labels are inaccurate – the LFP signal is just not a very localised sig-

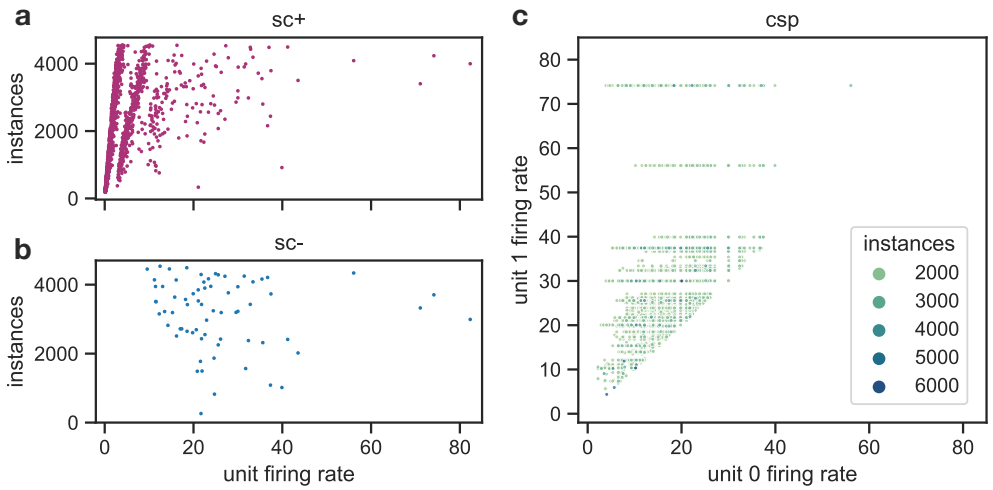
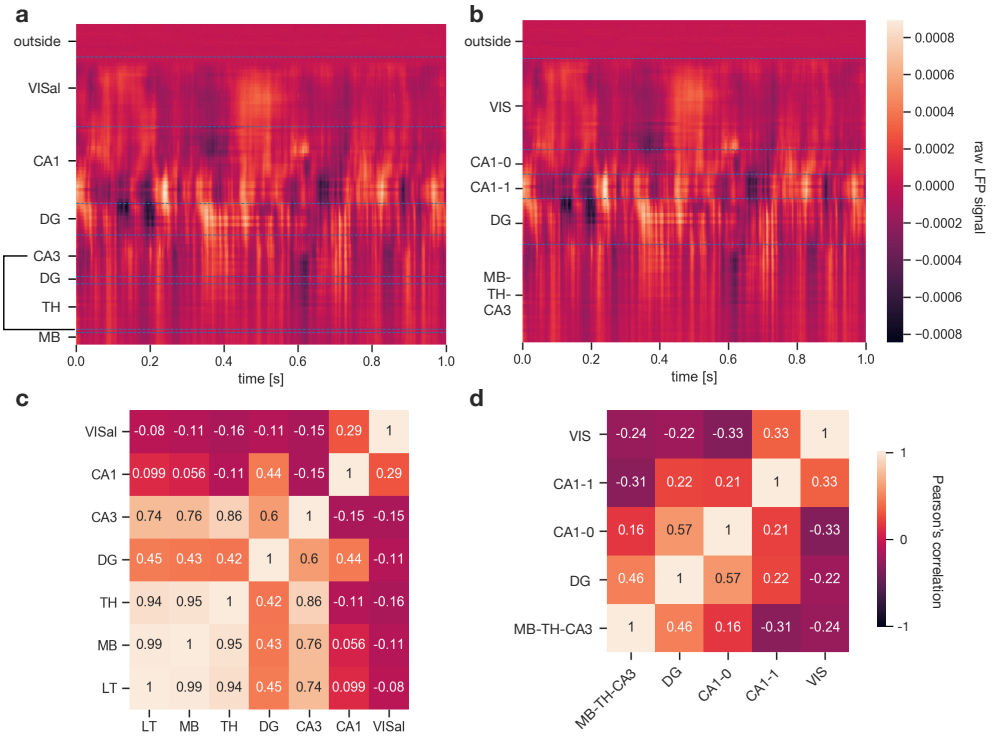


Figure 5.5: **Influence of firing rate on word instances.** The number of instances of each neural word is shown in relation to the firing rate of the unit for the word types **(a)** *sc+* and **(b)** *sc-*. **(c)** Since the word type *csp* combines two units in the neural word, the number of instances of each neural word is shown in relation to the firing rates of both units . This particular plot shows the firing rates taken from the experiment "Gabors".

nal that adheres to strict brain region borders. Additionally, brain structures are often sub-structured into multiple layers which display different activity characteristics and units on the borders between brain regions are additionally influenced by each other. Averaging across brain regions as defined by the Allen Institute channel labels (borders visualised in Figure 5.6 (a)) results in highly correlated time courses (see Figure 5.6(c)), and is therefore not suitable for this analysis.

Therefore, I decided to manually segment the visibly correlated time courses and average the time courses of the signals within each newly defined region as can bee seen in Figure 5.6 (b). This additional processing step of the LFP signals resulted in time courses which were much less correlated (Figure 5.6 (d)). I assigned a brain substructure to each of the time courses according to which originally labelled brain regions provided the highest channel count in each segment. The brain region labels given to these LFP segments should therefore be taken with a grain of salt because they often include multiple brain regions (especially midbrain and thalamus structures). For a discussion on how to improve the LFP segmentation, see Section 5.5.

This section gave a brief overview of the event filtered input generated in the first step of neural topic modelling. In the next section I will present the neuronal assemblies found by applying the remaining steps of the neu-



(a) Original borders

Figure 5.6: **Extraction of neural words from LFP signal. (a)** Visualisation of the brain substructure labels (provided by the Allen Institute) for each channel and how the borders between brain regions would be applied on the LFP signal. **(c)** Calculating the average time courses per substructure as given in (a) results in highly correlated time courses. **(b)** For this reason, new segments were defined on the basis of visual inspection of the LFP signal. **(d)** The time courses for the segments specified in (b) are less correlated than before.

ral topic modelling pipeline to the input shown here (topic modelling and clustering).

5.4 ANALYSIS

There exists a myriad of interesting analyses that could be done on the data set, on the neural words that were extracted and on the resulting assemblies. Because this thesis focusses on the insights gained by applying neural topic modelling on such a dataset, I will confine myself mainly to analyses which highlight the unique findings only neural topic modelling can provide. In this section, I will show examples of multiplexing by individual units. Then I will focus on the identified assemblies (the result from the clustering and thresholding analysis described in Section 3.2.4) and contrast them to assem-

blies which could have been found by other methods. In particular, I present neuronal assemblies which could not have been identified with any other current method. Lastly, I show that the assemblies found during different experiments may be used to generate insights into functional brain connectivity.

5.4.1 *Relationship of neural words to experimental variables*

In the Gabors experiment, the spikes of about 1500 individual units were translated into neural words. Of these units, 233 gave rise to more than one neural word e. g. one *sc+* and two *csp* neural words. As stated in previous chapters, neural topic modelling explicitly allows for the possibility of multiplexed code by extracting multiple activity features per unit. If units in the data set indeed use multiplexing, it should be possible to find units which give rise to neural words which code for fundamentally different experimental or behavioural variables.

In order to relate the clusters to experimental or behavioural variables, a time course for each cluster needs to be compared to the time courses of such external variables. Besides giving an estimate of the assemblies in the data, neural topic modelling also provides an estimate of how much each assembly was active during each time window. The resulting vector indicates how much each cluster contributed to the activity in that time window and can be used as a proxy for activity of each assembly over time.

For the relation to the wheel velocity, the time courses of the clusters were correlated (Pearson's correlation) with the time course of the wheel velocity (downsampled to one data point per time window). To avoid spurious correlations, only clusters with a correlation coefficient above 0.8 were deemed as being related to movement of the animal. The choice of threshold here is not important since most clusters had low correlation coefficients (between -0.3 and 0.2) and only few clusters showed a very high correlation with the disc velocity time series. For the relation to visual stimuli shown, the weights in each time window per cluster were simply multiplied with a matrix indicating the location of the stimulus during each time window. The resulting matrix represents the contribution of each cluster for each stimulus location.

As an example, of the 44 clusters which appeared consistently during the Gabors experiment (after clustering and thresholding, as explained in Section 3.2.4), 12 showed a visual receptive field and 2 were related to the

wheel velocity. In order to show multiplexing in units, I would expect a unit to take part in multiple clusters which are related to different stimuli with high weights in each cluster. This creates a need to determine which neural words in a cluster participate at a "high weight".

A neural word appears in a cluster if it contributes to the top 75% of weight in the cluster (starting at the highest weight). For some clusters, this still means that a few hundred neural words are included in this assembly, since it assigned many neural words a similar weight. For other clusters this means that only five neural words might be included in the assembly. The 20 highest weights in each assembly were deemed "high weights" or for assemblies with more than 200 neural words, a weight in the top 10% of all neural words was also labelled as high.

Figure 5.7 shows two units that display multiplexing. The first example of interesting cluster participation is a unit in the Thalamus which gives rise to 8 neural words (1 *sc+* and 7 *csp* neural words, see Figure 5.7 (a)). Its *sc+* neural word participates in three clusters with a high weight: one cluster with a negative receptive field as shown in Figure 5.7 (c) and two clusters which are both related to the movement of the animal Figure 5.7 (d). This unit encodes multiple variables, namely movement of the animal and visual information in its increases in firing rate. Since neural topic modelling is able to handle neural words being part of multiple assemblies, both of these relationships could be recovered. Other methods less suited for multi-memberships might have not been able to establish this relationship.

The second example is a unit in the visual cortex which gives rise to 17 neural words Figure 5.7 (b). Its *sc+* neural word exhibits a high weight in one of the clusters strongly related to the movement of the animal Figure 5.7 (d). Interestingly, one of its *csp* neural words shows a very high weight in a cluster which is preferentially active when the Gabor stimulus is located in the middle of the screen Figure 5.7 (e). This is an example of two relationships which could have not been discovered in the same analysis with currently available methods. A firing rate based method might have found the relationship to movement and a method for timed spike detection might have found the relationship to visual areas, but not at the same time.

The fact that spike trains are influenced by movement and visual stimulation is not surprising (Juavinett et al., 2019; Stringer et al., 2018). In order to show more interesting examples of mulitplexing with cognitive processes or complex behaviours, more external variables would need to be available in the data set (only visual and movement information were easily accessible

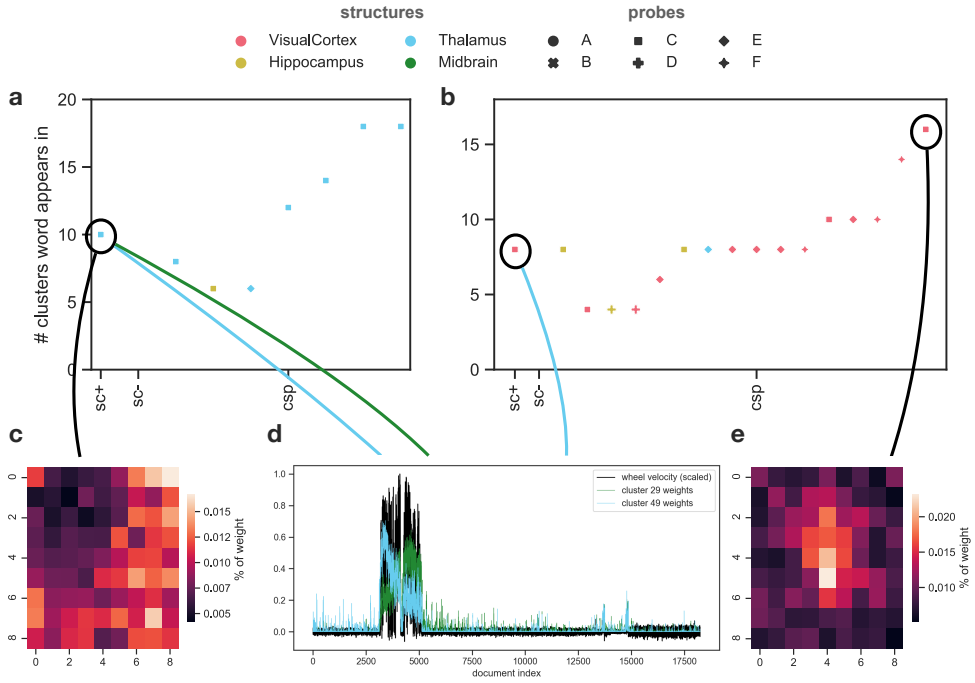


Figure 5.7: **Two examples of neural words participating in clusters related to experimental variables. (a)** A unit located in the thalamus which gives to 8 neural words. Its *sc+* neural word (increases in firing rate) participates at a high weight in clusters which are strongly related to **(c)** visual stimulus location or **(d)** wheel velocity. **(b)** The second example is a unit in the visual cortex whose *sc+* neural word is also correlated to the movement of the animal (participation in cluster 49, blue line in (d)). Its synchronous spikes with another unit in the visual cortex contribute a high weight in another cluster exhibiting a receptive field **(e)**. (a) - (b) The colour and shape of the leftmost dot (at *sc+*) indicate the brain region the unit was located in and which probe it was recorded from. For the *csp* words (all but the dots at the first two positions), the shape and the colour indicate the location and region of the neuron it exhibited coincidental spiking with.

in this case). Rather than being ground-breaking discoveries, these examples indicate it is useful to allow units to be part of multiple different assemblies which might code for fundamentally different stimuli. Additionally, they indicate that extracting multiple activity patterns from spike trains can provide information which was not accessible by any single method before. It would be great if neural topic modelling could save researchers the time and effort to do separate analyses.

This section showed an example of a unit which encodes multiple variables in its activity and a unit which encodes different variables through different aspects of its spike train. The next section will introduce the way I visualise assemblies.

5.4.2 *Visualisation explanation*

Since the following two sections will consist mainly of visualisations of clusters, I would like to explain the plot in detail here to avoid repetition later. The following figures will refer back to this section for details on the plots. The visualisation of one cluster is done in one row of three subplots as can be seen in Figure 5.8. The experiment in which a cluster was found will be indicated in the caption.

The left and the middle plots indicate the spatial origin of the neural words. The coordinates follow the orientation and arrangements of the electrodes in Figure 5.1 (a). The limits of all cluster visualisations are kept the same for easier comparison of the locations between clusters. For better visibility of the separate locations of neural words on each electrode, the x and y coordinates of the data points are jittered. The jitter applied to the coordinates was the same for both plots so that the points are in the same location for both plots.

The space in which the locations of the neural words are visualised are the channel locations given in the original data in the lateral $\rightarrow$ medial (x-axis) and posterior $\rightarrow$ anterior (y-axis) directions. The units of the axes are given in millimetres and denote the distance from the lambda (a physical marking on the mouse skull to which all recordings are referenced) (Allen Brain Observatory, 2019).

The **left plot** shows the location of the neural words in space and indicates the larger brain structure by colour and the weight in the cluster by transparency. The **middle plot** shows the location of the neural words in space and indicates the word type by colour and the weight in the cluster by transparency. Note that no colour bar or equivalent reference for the weight-transparency relationship is given. The transparency is quantitatively proportional to the weight in the cluster, but it is extremely difficult to make quantitative judgements from the transparency of a dot in a scatter plot with up to a few hundred, overlapping data points with up to four different colours. These plots should rather be taken as a qualitative impression of the distribution of the cluster in space.

I chose to indicate only the larger structure (midbrain, thalamus, hippocampus, visual cortex) as the colour in the left plot even though the original data offers twenty different substructures (APN, CA1, VISp etc.). A list of all substructures present in the data, the corresponding larger structure they belong to and the full name can be found in Figure 5.9. The main rea-

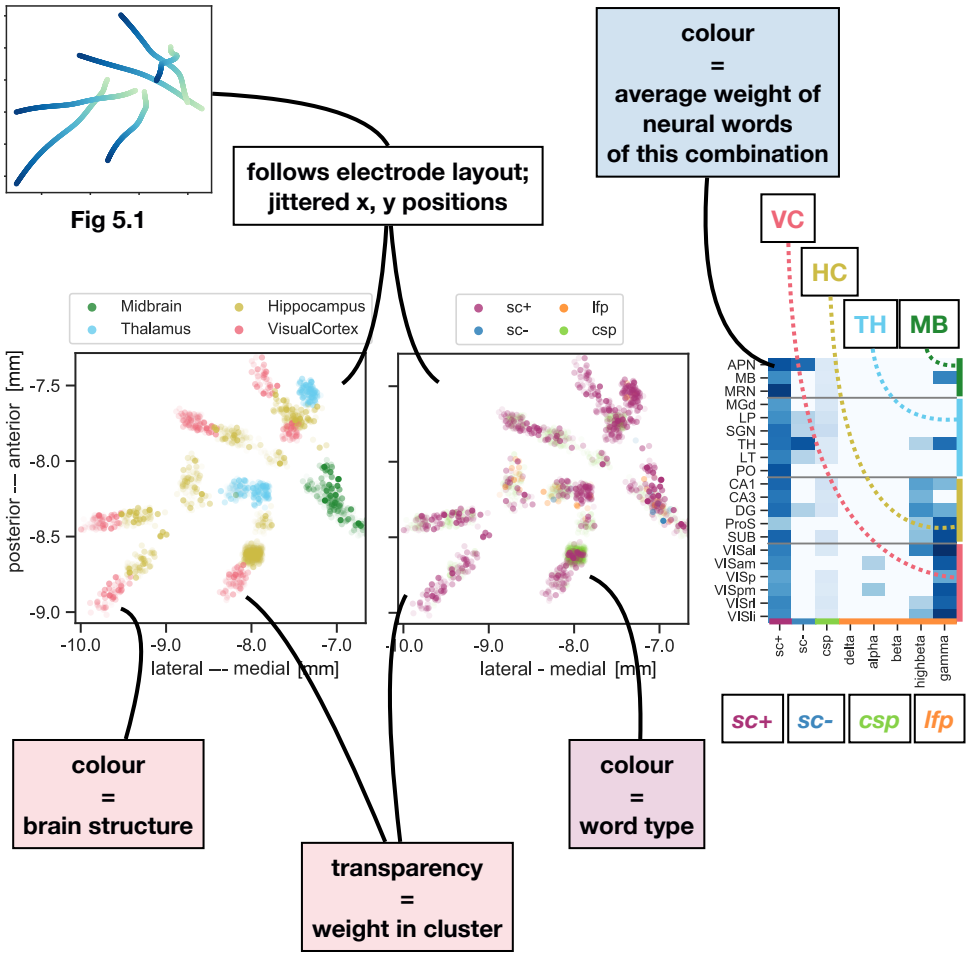


Figure 5.8: **Annotated cluster visualisation.** Each cluster is visualised in three plots placed in a row. The **left plot** depicts the locations of the neural words contributing to this cluster as projected onto a space spanning the anterior-posterior and lateral-medial axes (units: millimeters). The colour indicates the brain structure of each neural word and the transparency signifies the weight in the cluster. The **middle plot** uses the same space and transparency as the left one but uses a colour code for the word type. The locations in the left and middle plots have been jittered in x and y to increase visibility in the plot. The general locations of the dots are equivalent to Figure 5.1 (b). The **right plot** shows the distribution of total weight in the cluster per substructure-word type combination. For a list of substructures and their acronyms please refer to Table 5.9. In the caption of the figure, the experiment during which the cluster was identified will be given e. g. **(a)** (natural movies) indicating that this cluster was found during the presentation of natural movies. Experiments are detailed in Section 5.1.2. VC = visual cortex; HC = hippocampus; TH = thalamus; MB = midbrain; *sc+* = high firing rate neural word; *sc-* = low firing rate neural word, *csp* = coincidental spiking neural word; *lfp* = high power in LFP signal neural word

son for this choice is that the use of twenty different colours with different transparencies makes a plot confusing. To include the information about which exact substructures and which neural words (including which LFP frequency bands) contributed to the weight in the cluster, the **right plot** de-

structure	acronym	full name
Midbrain	APN	Anterior pretectal nucleus
	MB	Midbrain (general)
	MRN	Midbrain reticular nucleus
Thalamus	LP	Lateral posterior nucleus of the thalamus
	MGd	Medial geniculate complex, dorsal part
	SGN	Suprageniculate nucleus
	TH	Thalamus (general)
	LT	Lateral terminal nucleus of the accessory optic tract
	PO	Posterior complex of the thalamus
Hippocampus	CA1	Field CA1
	CA3	Field CA3
	DG	Dentate Gyrus
	ProS	Prosubiculum
	SUB	Subiculum
Visual Cortex	VISal	Anterolateral visual area
	VISam	Anteromedial visual area
	VISp	Primary visual area
	VISpm	Posterolateral visual area
	VISrl	Rostrolateral area
	VISli	Laterointermediate area

Figure 5.9: **List of brain structures and substructures the in data set.**

picts the average weight of the neural words in a cluster across the different substructure-word type combinations. Grey borders in the heatmap indicate the borders between different brain regions: (from top to bottom) midbrain, thalamus, hippocampus, visual cortex.

It should be noted that the depiction of the contributions per structure-word type combination in the rightmost row of all the following plots in this section is unavoidably not ideal. Different combinations represent very different counts of neural words i. e. there are 262 CA1-*sc*- neural words but only 1 VISp-*lfp* (alpha band) word since only one probe traverses the VISp region and only one VISp average trace of the LFP signal on that probe is entered into the analysis. Averaging the weight over the number of neural words contributing to the combination will give a biased number for combinations in which only one neural word is available. For example, let's assume there are multiple CA1 *sc*+ neural words in a cluster which are given a range of weights in a cluster (including the highest three and the lowest

word type	description
sc+	high firing rate in one unit
sc-	low firing rate in one unit
csp	coincidental spiking in a pair of units
delta	high delta (3-6 Hz) power in averaged LFP signal per structure
alpha	high alpha (6-13 Hz) power in averaged LFP signal per structure
beta	high beta (13-30 Hz) power in averaged LFP signal per structure
highbeta	high high beta (30-40 Hz) power in averaged LFP signal per structure
gamma	high delta (40-70 Hz) power in averaged LFP signal per structure

Figure 5.10: **List of word types extracted from data set.**

three). A single VISp-*lfp* (alpha band) neural word with a slightly higher than average weight will look more important in the cluster than CA1-*csp* neural word, even though the neural words with the three highest weights are of that combination. There is no perfect way of normalising the contributions of groups when one group has one member and another group has 250 members. The normalisation per contributing neural words seemed the best choice and represented the assembly structures reasonably well.

On average there are around 3800 neural words in each analysis (each experiment). Not all the neural words and their weights are shown in each cluster visualisation or are used for the average weight calculations. For every cluster only the neural words which contribute to 75% of the weight (summed up starting at the highest weights) are included in the plots. Therefore the number of data points shown in each cluster differ.

5.4.3 *Detecting known assembly structures*

In general in neuronal data analysis, researchers investigate only one aspect of neuronal activity e. g. methods for assembly detection are either based on firing rate co-modulation or on participation in spike sequences. Researchers have shown that units interact with each other when separately investigating firing rates, spatiotemporal patterns and LFP (for an overview, please see Section 2.3).

So there is reason to expect that neural topic modelling will bring forward assemblies which group together neural words of the same word type. Since one main direction in the search for neuronal assemblies is based on the pairwise correlations of neuronal firing rates (Lopes-dos Santos et al., 2011;

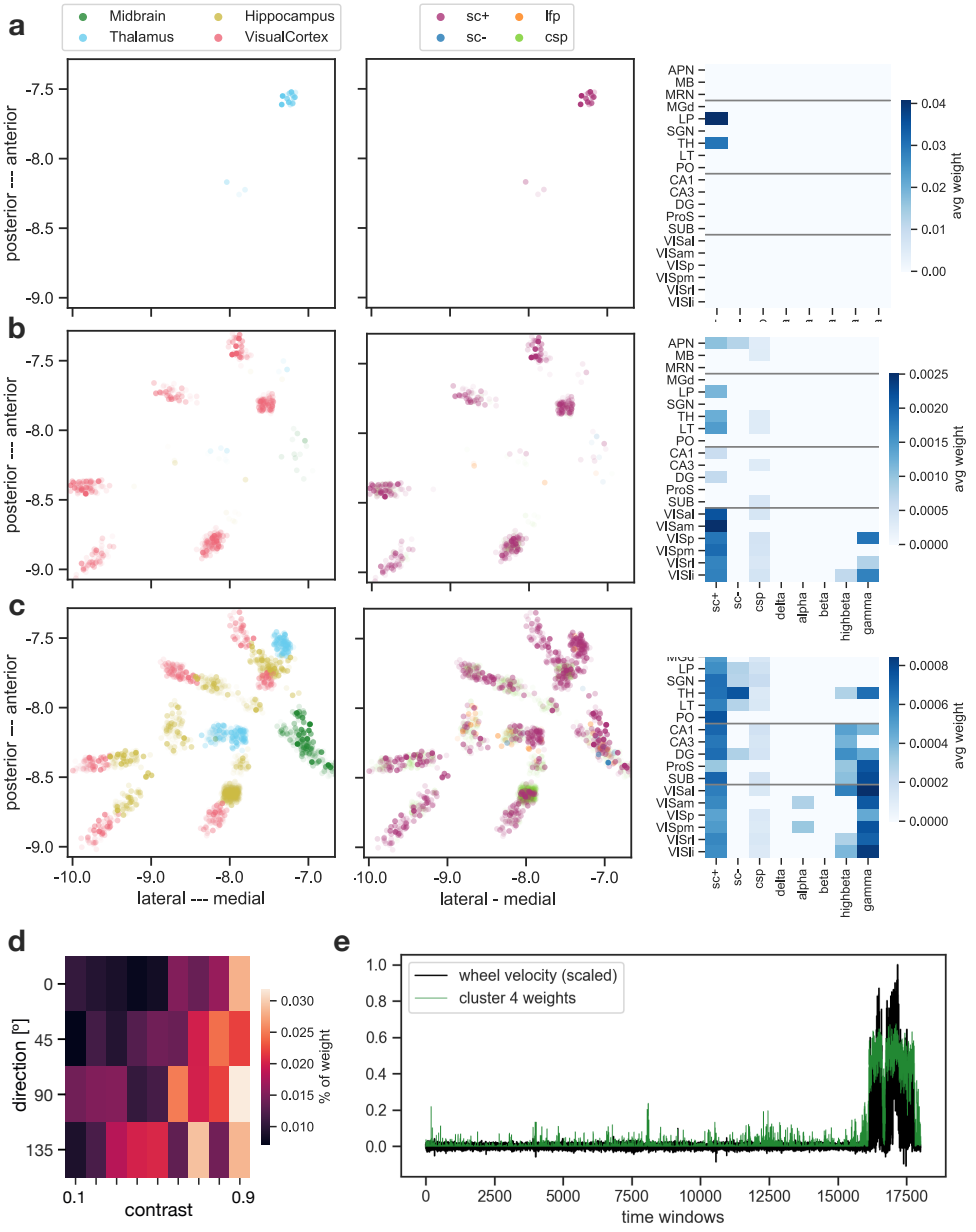


Figure 5.11: **Examples of firing rate modulated clusters. (a)** (natural movies) A cluster involving units in the lateral posterior nucleus of the thalamus on one particular probe. **(b)** (drifting gratings contrast) A cluster involving units in visual cortex regions across all probes. **(c)** (natural movies) A cluster involving neural words in all structures and across all electrodes. **(d)** The cluster of *sc+* neural words across the visual cortex in (b) shows a preference for drifting gratings with higher contrasts. **(e)** The wheel velocity during the recording is shown together with the weights of the cluster in (c) per time window (an indication of when a cluster was "active"). (a) - (c) Left and middle panels: location of the neural words in the cluster; transparency = weight in cluster. Right panel: the distribution of total weight in the cluster. For a detailed explanation, please refer to Section 5.4.2.

Peyrache et al., 2010; Romano et al., 2015), neural topic modelling should find clusters which consist of only *sc+* neural words. Indeed, there are many examples across all data sets (each data set is the recording during one visual experiment) of such *sc+* clusters.

Among the neuronal assemblies found there were small assemblies of *sc+* neural words which are located on the same electrode; see Figure 5.11 (a) for an example of a cluster of thalamus *sc+* neural words. Similar local clusters could be found for all other structures.

Clusters did not need to be confined to one probe but could also represent correlations of high firing rates across multiple structures on different electrodes. The assembly shown in Figure 5.11 (b) displays a cluster which involves *sc+* words across all recorded visual cortex regions. This particular cluster also displayed a preferential response to drifting gratings (the visual stimulus shown in that experiment) with high contrasts (see Figure 5.11 (d)).

Other clusters involved *sc+* neural words across all probes and brain structures; an example can be seen in Figure 5.11 (c). The cluster activity can be compared to behavioural variables such as the wheel velocity. The cluster time course is given by neural topic modelling since the algorithm provides an estimate of how "active" each cluster was during each time window. For the particular cluster shown in Figure 5.11 (c), the time course very much resembled the time course of the wheel velocity of the disc on which the animal was moving as evident in Figure 5.11 (e). Multiple papers have shown whole brain modulation of firing rates due to movement (Juavinett et al., 2019; Musall et al., 2019; Stringer et al., 2018), so this is in line with what is known about movement-related activity modulation in the brain.

The clusters described above could have been identified by other methods based on firing rate modulations to define assemblies. The other approach to identify neuronal assemblies is through timed spiking behaviour in neurons. Such spike patterns are commonly found in cortical areas (including sensory cortices) and in particular in the hippocampus (Buzsáki, 2004; Sasaki, 2019) (for an overview, see Section 2.3.2). It is therefore not surprising to find assemblies which consist mainly of *csp* neural words and are confined to the hippocampus. Indeed, multiple hippocampus-*csp* assemblies are detected per experiment. In each one, the distribution of the neural words in space looks slightly different, suggesting that there are multiple subassemblies in the hippocampus which potentially process separate streams of information.

Additionally, clusters which consist of only *lfp* words are also present in the found assemblies. This is not surprising since the LFP signal is a wide-

reaching signal which is often correlated across different electrodes (Herreras, 2016). However, this information is usually only revealed in a separate analysis focussing on the correlation of the LFP time courses. The advantage of neural topic modelling is that information such as found here – that LFP signals in particular subcortical regions are correlated – is included in the results without extra work. There are multiple assemblies consisting of mainly LFP neural words per experiment and mostly involve *lfp* neural words in subcortical regions and are based on the power in lower frequency bands (delta, alpha). Neural topic modelling can be used as a good starting point for further analyses of LFP signals.

Neural topic modelling recovers assemblies that are comprised of units with shared firing rate modulations and also assemblies which group units with coincidental spiking behaviour. These are well-known features in neural data, so this inspires confidence in the capability of neural topic modelling to find salient neuronal assemblies. While this is a good confirmation of the performance of neural topic modelling, these results are not necessarily surprising or ground breaking. The clear advantage of neural topic modelling lies in the fact that multiple aspects of neuronal activity are entered into the same analysis. The next section will show assemblies which consist of neural words of multiple word types and which therefore could not have been found by any other assembly detection method.

5.4.4 *Unknown interactions between word types*

Neural topic modelling can provide interesting insights into how neural words interact with each other across regions and word types. These are the results which currently cannot be found by any other assembly detection method – one would have to explicitly look for those interactions in separate analyses. This section will give a few examples of assemblies in which neural word types appear to interact in ways that are not commonly thought of in the context of neuronal assemblies.

A large number of the assemblies identified by neural topic modelling show some sort of mixture of word types. Each one of them is potentially interesting to researchers depending on the questions they might want to ask of the data. There are therefore hundreds of clusters I could show here. As a demonstration of the kind of unknown assembly structures neural topic modelling can uncover, I present three types of assemblies with character-

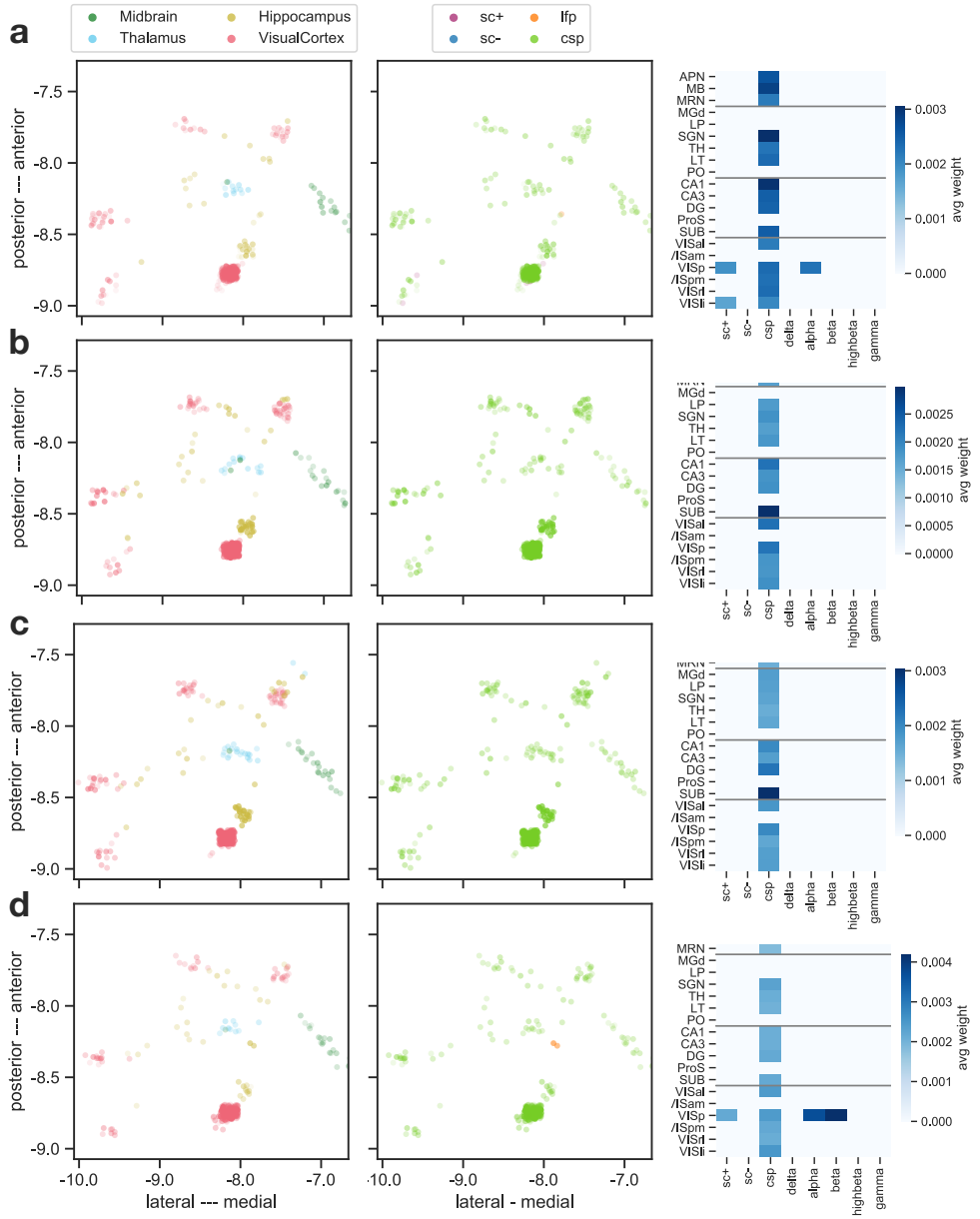


Figure 5.12: **Hippocampus *csp* and visual cortex *csp* neural words.** Coincidental spikes between neural words in the visual cortex and the hippocampus come up consistently across multiple experiments: **(a)** drifting gratings contrast, **(b)** drifting gratings repeats (block 0), **(c)** drifting gratings repeats (block 1) and **(d)** Gabors. Left and middle panels: location of the neural words in the cluster; transparency = weight in cluster. Right panel: the distribution of total weight in the cluster. For a detailed explanation, please refer to Section 5.4.2.

istic word type interactions which were consistently found across multiple experiments.

One cluster appearing repeatedly was a cluster consisting of *csp* words across all brain structures, but mainly in the primary visual cortex and the

hippocampus (subiculum and CA1). The appearances of this cluster across multiple experiments is shown in Figure 5.12. The average weights in the right plot show the weights to be well distributed across all substructures, while the left plots suggest the cluster to be localised in the primary visual cortex and the hippocampus on the same electrode. This is one of the issues that arises with the normalisation per combination (of word type and substructure), because the counts in each combination differ hugely. Neither plot represents the data perfectly. Interestingly, this interaction between the hippocampus and the visual cortex seems more pronounced in the experiments in which drifting gratings were shown (Figure 5.12 (a), (b), (c)) while the interaction between visual cortex and hippocampus was less strong when Gabor stimuli were presented to the mouse (Figure 5.12 (d)). In the latter case, instead of hippocampus activation, there was a high weight contribution of low frequency *lfp* neural words from the visual cortex.

Another cluster identified by neural topic modelling in multiple experiments was a combination of decreased firing in visual cortex areas and subcortical *lfp* neural words as seen in Figure 5.13. In particular the delta, alpha and beta frequency bands were involved in this cluster. This cluster was not only found during visual stimulation (Figure 5.13 (a), (b), (c)), but also during no stimulation (Figure 5.13 (d)). The contribution of *lfp* neural words is increased during spontaneous activity compared to experiments in which visual stimuli were presented.

The third consistently appearing cluster pattern is the interaction between coincidental spiking events in multiple areas in the visual cortex and the *lfp* neural words in visual and hippocampus areas. What makes this cluster interesting is that in the case of the Gabor stimulus (Figure 5.14 (a)) there is a more localised contribution of *lfp* neural words from the primary visual cortex. Additionally, from the spatial maps it seems like the primary visual cortex does not interact as much with the other visual areas as during the drifting gratings (Figure 5.14 (c), (d)) and natural movie (Figure 5.14 (b)) experiments.

The three examples mentioned in this section are not the only consistently appearing clusters. Other include a combination of visual cortex *sc-* neural words and subcortical *sc+* and *csp* neural words (hippocampus, thalamus and midbrain). There are also many examples of clusters consisting of mainly *csp* neural words in different locations of the hippocampus or clusters consisting of only *lfp* neural words in either lower (delta and alpha) or higher (beta, highbeta, gamma) frequency bands.

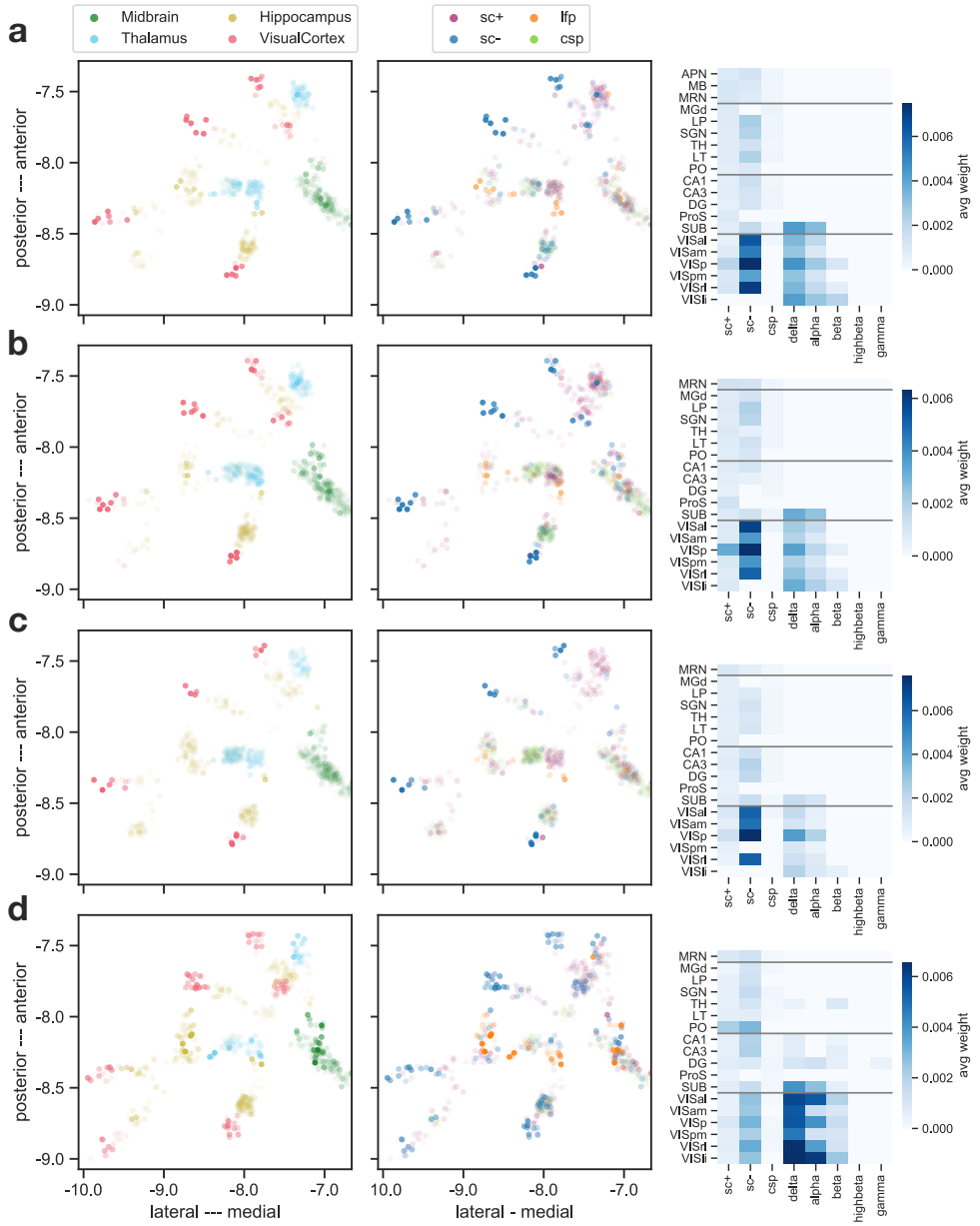


Figure 5.13: **Visual cortex *sc-* and subcortical *lfp* neural words.** Decreases in firing rates in units in the primary visual cortex seem to interact with high power in lower frequency bands of the LFP signal in the hippocampus. These clusters were found in multiple experiments: **(a)** drifting gratings repeats (block 0), **(b)** drifting gratings repeats (block 1), **(c)** natural movies and **(d)** spontaneous activity. Left and middle panels: location of the neural words in the cluster; transparency = weight in cluster. Right panel: the distribution of total weight in the cluster. For a detailed explanation, please refer to Section 5.4.2.

However, a cluster does not need to come up consistently in order to be interesting - clusters which represent activity related to experimental procedures are likely to only appear in the analysis of that specific experiment.

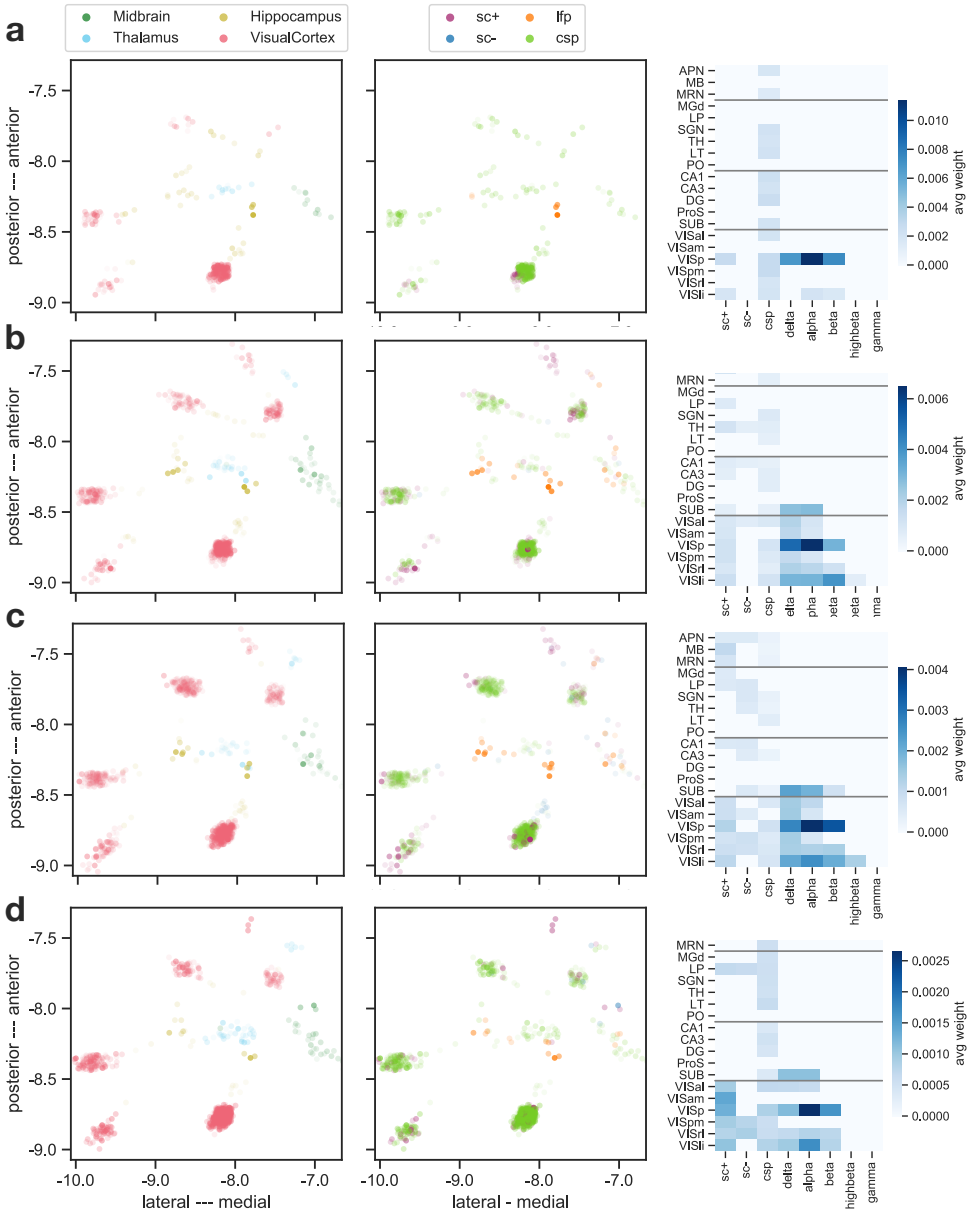


Figure 5.14: **Visual cortex *csp* and subcortical *lfp* neural words.** Coincidental spikes across multiple areas in the visual cortex seem to interact with high power in the LFP signal in subcortical areas. These clusters were found in multiple experiments: **(a)** Gabors, **(b)** natural movies, **(c)** drifting gratings repeats (block 0) and **(d)** drifting gratings repeats (block 1). Left and middle panels: location of the neural words in the cluster; transparency = weight in cluster. Right panel: the distribution of total weight in the cluster. For a detailed explanation, please refer to Section 5.4.2.

The fact that clusters appear consistently across activity patterns might indicate that these are general interactions between brain regions rather than stimulus specific activations.

Applying neural topic modelling to neuronal recordings can provide an insight into how activity patterns in units in different regions interact with each other. This section showed a few examples of such interactions which could have not been found by other currently available assembly detection methods. The next section will give an idea of the subsequent analyses that can be done when multiple recordings were done in the same animal.

5.4.5 *Comparison across experiments*

Neural topic modelling gives a number of concrete assemblies which can be analysed as shown in the previous section. It is also possible to look at assemblies as a measure of how much two brain structures cooperate with each other – if neural words from a region A very often appear in the same cluster as neural words of a region B, then this points towards those regions being functionally connected.

For each pair of substructures, I calculated the ratio between the number of clusters in which both substructures contributed at least 5% of the total weight and the number of clusters in which neural words from either one or the other substructure contributed at least 5% of the total weight. This gives a Jaccard score as an indicator for how consistently two regions cooperated with each other in one experiment. The resulting Jaccard scores for two experiments Gabors (gab) and drifting gratings repeats (block 1) (dgr1) are displayed in Figure 5.15.

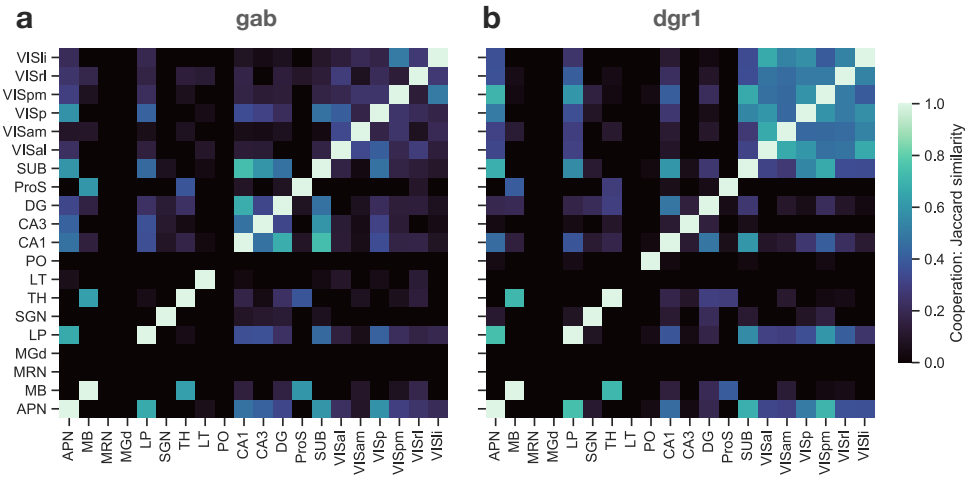


Figure 5.15: **Cooperation between structures.** A Jaccard score was computed to indicate how often two substructures participated in a cluster together. For a list of the abbreviations of the brain substructures, please refer to Figure 5.9.

While many elements of the cooperativity patterns remain the same across these two experiments (e. g. little cooperation between midbrain and thalamus regions, no cooperation within the thalamus substructures, cooperation within hippocampal and visual cortex substructures, cooperation between the APN and LP and visual areas), there are also some interesting differences. The substructures in the visual cortex cooperate at a much higher level in the drifting gratings experiment while the hippocampus substructures are more cooperative during the presentation of the Gabors.

Such cooperativity matrices can be calculated for each experiment and inspected to detect differences across cooperation patterns. To gain a very coarse understanding of how cooperativity changes across experiment, I averaged the cooperativity score of all the substructures per pair of structures. As an example, the average cooperativity between the visual cortex and the hippocampus will be calculated as the average between all pairs of visual cortex substructures (VISal, VISam, VISli, VISp, VISpm, VISrl) and hippocampus substructures (CA1, CA3, DG, ProSUB, SUB). The resulting cooperativity matrices can be compared to each other by subtracting one from the other across experiments as shown in Figure 5.16.

What stands out immediately in Figure 5.16 is that the cooperativity within the visual cortex during the Gabor stimuli is lower than in any other experiment. Visual cortex internal cooperativity (multiple visual cortex substructures appearing in multiple clusters together) is highest during the presentation of drifting gratings. Interestingly, natural movies did not stimulate internal visual cortex cooperativity as much as the drifting gratings and is actually quite close to what was found during no visual stimulation (spont). All the drifting gratings conditions seem to elicit similar cooperativity patterns, which is what would be expected.

This section showed an example of a higher level analysis of the findings from neural topic modelling. The next section will discuss the findings and analyses presented.

5.5 DISCUSSION

The aim of this chapter was to show how neural topic modelling could be applied to actual neuronal recordings. To demonstrate this, I applied the analysis pipeline described in Chapter 3 to a large data set consisting of spike trains of around 1500 units and corresponding LFP data. The mouse

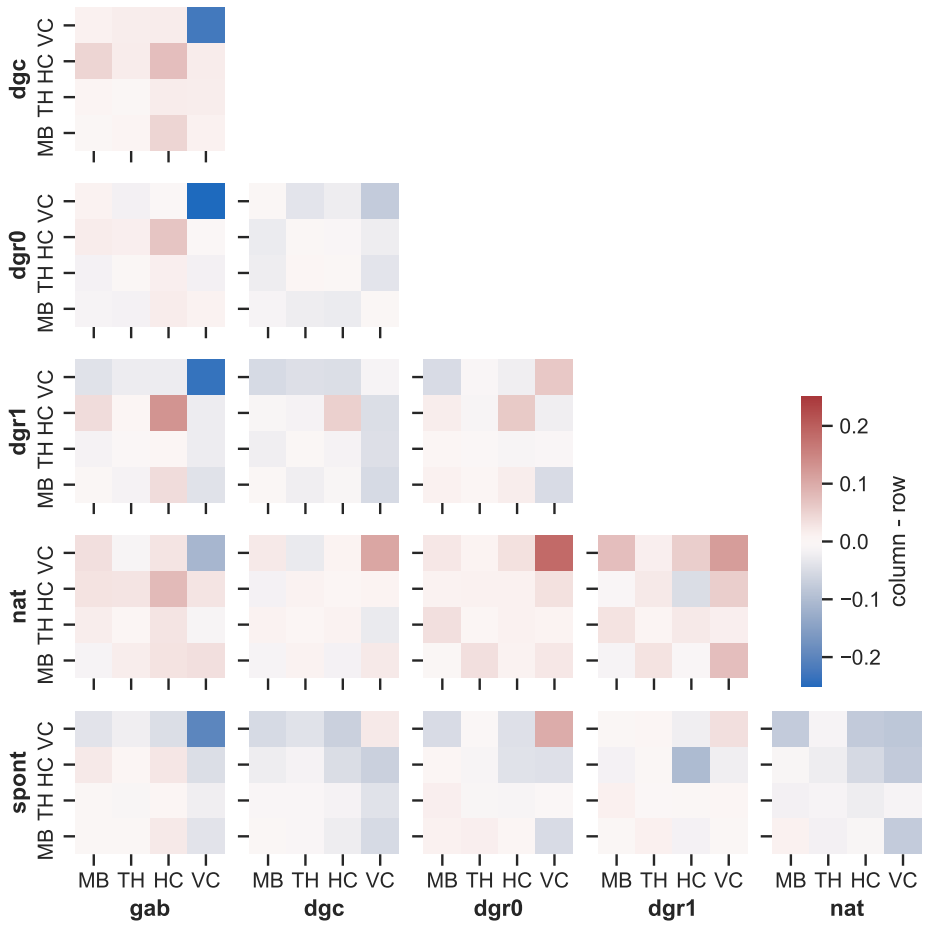


Figure 5.16: **Difference in cooperativity across experiments.** The colours indicate the difference calculated from the experiment named in each column to the experiment named in each row. MB = midbrain, TH = thalamus, HC = hippocampus, VC = visual cortex

from which the data was recorded was shown a variety of visual stimuli of which five experimental paradigms and one period without visual stimulation were selected for analysis.

The first step in neural topic modelling is to apply event filtering for specific neuronal activity events to create the neural words. To demonstrate how flexible these neural words can be, I included four word types in the analysis: high firing rate (*sc+*), low firing rate (*sc-*), coincidental spiking in pairs of units (*csp*) and high power in LFP frequency bands (*lfp*). This demonstrates that neural words can be extracted from single neuron activity, from neuronal activity which is shared across multiple units and from non-spike data types such as LFP.

As the presented results show, neural topic modelling identifies clusters which are dominated by a shared modulation of firing rates and therefore resemble the ones found by other methods (Juavinett et al., 2019; Mölter et al., 2018). Shared modulation of coincidental spiking events in the hippocampus and visual cortex are also expected (Sasaki, 2019) as well as correlations in LFP signals, in particular among the lower frequencies (Herreras, 2016). At the same time, neural topic modelling also identified clusters which represent the coordination of activity patterns which were previously not associated with assembly activity. Such combinations include amongst others for example coincidental spiking involving primary visual cortex & hippocampus, synchronous spiking in visual areas & subcortical LFP signals or low firing rates in visual areas & subcortical LFP signals. These were examples picked for their consistent appearance across experiments. Many other examples of interesting combinations were present in single clusters. In addition to the analysis of single clusters, it was shown that it is possible to do higher level analyses with the information about which type of neural words participated in each cluster. Cooperation between substructures of the visual cortex and the hippocampus depended on the visual stimulation that was shown. Drifting gratings invoked the highest cooperation between substructures of the visual cortex while Gabor stimuli showed higher cooperation among hippocampal substructures.

As with every other analysis method, neural topic modelling has limitations. I will now discuss some of the drawbacks and which steps were taken to overcome them.

As shown in Section 5.3, the incidence rates of *sc+* and *sc-* neural words are somewhat dependent on the firing rate of the unit involved. This is impossible to prevent in its entirety since very low firing rates cannot produce more events while not spiking. In reality though, units with very low firing rates do contribute high weights to assemblies – units with firing rates of below 1 Hz gave rise to *sc+* neural words which were assigned the highest or second highest weights in their respective clusters. So even neural words with low incidence rates (as low as 120 events in 900 seconds of recording) could achieve high weights in clusters. 100 or more events might still be quite high for precisely timed spiking sequences which can be quite rare. For the analysis of such events, there are other methods which are much more specialised towards this specific goal and neural topic modelling is not going to replace these methodologies. However, it would be possible to enter the precisely timed spike sequences identified by a specialised method into neural topic modelling as a neural word. This would allow the analysis of

the interactions between exact spike sequences with other activity patterns – a fascinating prospect.

The other data type included in the analysis are the LFP signals recorded during the same sessions. LFP signals are difficult to compress into the format of neural words since they are highly correlated across all channels and since they are usually processed across longer time windows. The former issue of the high correlations between LFP traces was addressed by drawing new borders to delineate less correlated regions and to average the traces per newly defined segment. As demonstrated in Section 5.3.2.1, this was partially successful, but there were still a large number of clusters in each experiment which consisted of only LFP signals. A likely explanation is that the averaged traces still shared a lot of signal modulations, particularly in the low frequency bands. This procedure can definitely be improved on, which I will comment on in the following section.

One of the neural words used in the analysis is the one of coincidental spikes between a pair of neurons. While the presence of higher than expected near-synchronous spikes does give an idea of how accurately spikes are timed within a pair of neurons, it cannot provide the researchers with more details e. g. how precisely the spikes were timed and in what order they occurred. This neural word also does not look for spike sequences or other spatiotemporal patterns in which precisely timed spikes are further apart than the specified time window (here: 4 ms). Other methods which search for neuronal assemblies on the basis of spatiotemporal spike patterns will give much more fine grained assemblies and the specific spike patterns in which they were involved. This is not possible with neural topic modelling without further analysis. As with the clusters discussed previously, the advantage of neural topic modelling might lie in the fact that these *csp* assemblies are found alongside assemblies which are based on firing rates – something unique to neural topic modelling. The method proposed here might also indicate an interesting subset of units to further scan for repeating spike patterns with methods which do not scale easily with increasing number of units in the data set.

As with every other unsupervised algorithm applied to a data set without a ground truth, there is no way of validating the found result. In the case of neural data it is even more difficult since we don't even know how a "meaningful" assembly should be defined – meaningful to the researchers who want to decode brain signals or meaningful to the brain processing the information. Indeed, this is a problem with neuronal assembly detection

methods, since all of them come up with slightly different assemblies when applied to the same dat set (Mölter et al., 2018). This is an indication that every assembly detection method brings with it a bias towards certain types of assemblies or assemblies with specific features and characteristics (be it size, activity level, amount of overlap with other assemblies etc.). I hope to have created an analysis pipeline which has a bias towards biologically relevant assemblies which are similar to how the brain processes information, but it is impossible to know this. One way of estimating whether a method has found sensible assemblies is to assess their relation to known experimental variables. Multiple clusters show a clear preference for stimulus location or display a time course which is aligned to the movement of the mouse. It therefore seems likely that the clusters found do carry biological meaning. This does not imply that other assemblies with no corresponding experimental measure are meaningless - on the contrary, they might represent external variables we are not recording such as sounds or internal variables such as hunger, cognitive processes etc. Finding out what those assemblies are attuned to should be the next goal in neural topic modelling and ways to achieve this will be discussed in the following section.

Neural topic modelling has very clear advantages which distinguish it from every other neuronal assembly analysis method currently available.

I showed that there were many clusters which represented neuronal assemblies in which very different activity patterns interacted with each other. This is a finding which could have not been found by any other assembly detection method. Additionally, assemblies which resemble known assembly structures (firing rate co-modulations, synchronous spiking) are also found in the same analysis. The presence of such "unsurprising" assemblies can be taken as confirmation that neural topic modelling is producing sensible results. This makes the "surprising" assemblies with the word type interactions even more intriguing. Neural topic modelling can hopefully function as a springboard for further analyses of previously unknown interactions of neuronal activity patterns.

The fact that very similar clusters appear across multiple experiments also points towards neural topic modelling picking up on some consistent interactions in the neuronal activity. This might indicate that these are general interactions between brain regions rather than stimulus specific activations. These clusters, to my knowledge, represent interactions in the data which have not been identified previously. Such finding will be very valuable to

the neuroscience community because they might represent different channels through which brain regions communicate.

Another capability unique to neural topic modelling is that it takes into account that neurons encode multiple variables at the same time and sometimes apply different encoding strategies to do so. As was shown in this chapter, units do seem to encode multiple variables at once with one neural word (e. g. movement and visual stimulation) or encode multiple variables using different neural words. This neuronal behaviour is not captured well or even at all by other methods which struggle with multi-memberships and only extract either timed spikes or firing rates from neuronal activity. Neural topic modelling is able to account for both of these scenarios and therefore uncover many more insights into neuronal assembly dynamics than currently available methods. By providing all of this information in a single analysis, neural topic modelling could save researchers the time and effort to do separate analyses.

5.6 POSSIBLE ALGORITHMIC IMPROVEMENTS

Since neural topic modelling is a new data analysis tool and (to my knowledge) the first time a topic model is being applied to neural data in this way, there are obviously many things to improve upon and to add.

There is room for additions to neural topic modelling, in particular in the definition of the neural words.

The firing rate and synchronous spikes are certainly interesting features of neural spike trains, but they are by far not the only ones. Other spike train statistics such as inter-spike interval (ISI) and count variability (Coefficient of Variation, Fano Factor) have been shown to encode stimulus information (Nawrot, 2010). Spike distances (Victor and Purpura, 1996), onset responses (Reinagel and Reid, 2000) and bursts of spikes (Buzsáki, 2010; Naud and Sprekeler, 2018) are also good candidates for spike train features to extract. Another interesting suggestion would be to include cooperative silences into the list of possible neural words since they have been implicated in information processing (Gärtner et al., 2017).

Even more potential for an expansion of the neural word repertoire is given for neural words involving the LFP signal. Event related potentials and spindles are features in the LFP signal which could be turned into neural words. The slope of the LFP spectrogram has also been shown to

include information about the balance between excitatory and inhibitory inputs present (Haller et al., 2018). Additionally, the relationship between spikes and LFP signal is analysed through phase coupling e. g. place neurons in the hippocampus exhibit phase locking to the local LFP signal. While the spikes convey the position of the animal, the degree of phase locking encodes information about the distance of objects (López-Madrona et al., 2020) – another fascinating example of why analysing different features of neural signals is a good idea.

Apart from new additions, there are also many options to optimise the current implementation.

The algorithm for finding coincidental spikes which was applied here is not an optimised implementation[2] and is only meant as a proof of concept. Since many extremely successful algorithms for the search of synchronous or sequential spike patterns exist already (Mackevicius et al., 2018; Quaglio et al., 2017; Torre, Canova, Denker, Gerstein, Helias and Grün, 2016), it would make sense to incorporate the search for spike patterns through those channels. Sequential spike patterns are often more informative than pairwise synchronous spikes (Quaglio et al., 2017) and synchronous spikes are often slightly shifted in time (Ozden et al., 2008). So using such algorithms would allow for shifts and make it possible to include larger ISIs than is currently possible.

Currently, neural activity features are binarised before being analysed by the topic modelling algorithm. However, Latent Dirichlet Allocation (LDA) is capable of accepting integer values in its input, not just 0's and 1's. It would therefore be possible to represent the degree to which that activity pattern was present e. g. how much the firing rate increased or how many coincidental spikes were found. How exactly one would define those increments, especially when faced with different word types (e. g. spikes and LFP), is up to whoever will implement this. However, it is a possible change which might make the assemblies found more sharply tuned to the correlational structure in the data.

Combining of different data types such as spike trains and LFP signal arrives with its own set of challenges. The need to binarise continuous data such as the power in frequency bands in the LFP data is problematic. Any percentile threshold will result in all neural words from LFP having nearly

2 This is mainly because existing optimised algorithms are difficult to incorporate since they are either written in C or require a different data format and the conversion between neurodata formats is not straightforward – looking at you, neo.

the exact same number of word occurrences. For future versions, a better solution might be found to this problem, but in the meantime it should be enough to make sure that the number of word occurrences of neural words from continuous signals is comparable to those from discrete signals.

One way to improve the current neural topic modelling algorithm would be to use a different version of LDA. There have been numerous extensions to the basic version, some of which might be well-suited for neural data. Hughes et al. (2018) show that a semi-supervised version gives besser anti-depressant recommendations than the unsupervised version. Other groups have focussed on identifying correlated topics, for which hierarchical Latent Dirichlet Allocation (hLDA) (Blei, Griffiths, Jordan and Tenenbaum, 2003), Pachinko allocation model (PAM)[3] (Li and Mccallum, 2006) and correlated topic model (CTM) (Blei and Lafferty, 2007) were developed. Many more hierarchical adaptions to LDA have been suggested in particular for biological data (Liu et al., 2016). Other methods also use the order of the documents to inform their topics e. g. the dynamic topic model (Blei and Lafferty, 2006). Moreover, there exist more computationally efficient implementations of LDA. In the current state, the need to run multiple instances of the topic modelling on the input and the rather long runtime (1-10 hours depending on the size of the input) makes running neural topic modelling difficult without a High Performance Computing cluster. More optimised implementations might enable running neural topic modelling without large computational resources.

Using a semi-supervised or a supervised topic model in neural topic modelling might result in more task-related assemblies. In my opinion, this is not what neural topic modelling was intended for, since a supervisory signal imposes onto the model what researchers *think* is a relevant label. Instead, it was developed as an agnostic method for exploratory data analysis and hypothesis generation. The main goal would be to find interactions between neural activity patterns which researchers did not know about and did not expect.

Neural topic modelling relies on the clustering of the topics from multiple instantiations with different random seeds. Because each topic is theoretically supposed to be unique within one result, the goal of the clustering would be to achieve as many clusters as possible of a particular size (the number of runs, in this case: 20). This is an unusual goal in clustering

3 this particular topic model is abbreviated as PAM, which is why it had to definitely be included in this thesis :).

and there exists no standard solution. While the current automatic threshold works well, it is possible to strive towards a clustering algorithm which is more data-driven. Pons and Latapy (2006) propose an algorithm (*Walkdrop*) which determines the optimal threshold for hierarchical clustering by comparing the increase in variance during every linkage to the previous one. According to the authors, their approach increases the quality of the community structures found. Implementing such an algorithm for determining the best clustering threshold might also increase the quality of the clusters found in neural topic modelling.

A substantial innovation would be to not rely on the pre-defined neural words presented here. To truly achieve unbiased insights into the brain, we would need to find a way of defining the neural words automatically and from the data instead of the event filtering presented here. Such a step would bring us one step closer to identifying directions into which no human researcher has thought to look. At the same time, such an automatic neural activity pattern detection algorithm would merit its own PhD thesis and is beyond the scope of this one. The interpretation of the automatically detected neural words might also present a problem, counteracting the interpretability of the topic model. It would nevertheless be an interesting project that could expand our knowledge beyond the currently known firing patterns and assembly structures.

5.7 FUTURE WORKS

There are almost too many directions to list for future work with neural topic modelling. I will make a few suggestions based on the analyses I conducted for this thesis, but it very much depends on the data set that is to be analysed and what questions are being asked.

Specifically for this data set, there are a few additional analyses which could be conducted.

There are many more details in the data which have not been included in the analysis. The "shapes" of the assemblies with respect to how sharp the histogram of the ordered weights per neural word declines have not been analysed further. That neural topic modelling can recover such shapes of assemblies in artificial data was demonstrated in Section 4.6 of the previous chapter. It could be interesting to see whether those shapes align with brain regions, word types or distribution across electrodes. The data itself

also includes more information than was used here, such as the videos of the mouse's body movements or the pupil size. Other features of the units recorded could also be included in the analysis such as whether they have a receptive field. Furthermore, it might be interesting whether such characteristics of the units influence the order in which neural words appear in a cluster.

Another open question revolves around the clusters in which *sc-* neural words are cooperating with another word type. Do the same units giving rise to the *sc-* neural words also interact with the other neural words as *sc+* neural words? Or is it exclusively the decrease in firing rate which is central to the cluster? If it's the latter, then this would mean that these are findings likely unique to neural topic modelling, since most other assembly detection methods are based on whole spike train correlations or only extracting time periods with high firing rates, and do not consider firing rate decreases independently of firing rate increases.

While the current implementation of neural topic modelling already produces interesting results, there are various ways in which the analysis pipeline could be improved. Many suggestions have already been covered in Chapter 3, Section 5.6, so I will only mention a specific improvement possibility in the analysis of LFP signals here.

The LFP signals are available from every fourth channel in the data set presented here. Many of the traces are highly correlated, especially within traces of the same brain region. Therefore it is necessary to reduce the dimensionality of the LFP signals somehow while retaining the significant information. The way this was implemented here was to segment the time courses into new segments based on visible correlations in the traces. The time courses of the traces in the new segments were then averaged. While the de-correlation seemed to work reasonably well and the *lfp* neural words showed interesting interactions with other neural word types as shown in Section 5.4.4, there might be a better way to do this. One option would be to extract orthogonal LFP traces from the data e. g. by applying a PCA or Independent Component Analysis (ICA) to the LFP time courses. Another option might be to extract the power in all the frequency bands of all region-averaged traces and then before applying the topic model, combining *lfp* neural words which show a high overlap in time windows in which events were recorded.

I hope to have shown that neural topic modelling generates interesting results in the current implementation. Applying neural topic modelling to

different data sets will help highlight the usefulness of the method and the possibilities it offers with respect to large scale neural data analysis.

It would be intriguing to use neural topic modelling on data sets of freely behaving animals or during more complex tasks. Neuronal assemblies are established during learning of tasks (Sheng et al., 2019) and assembly dynamics might indicate successful learning during higher cognitive loads (Dragoi, 2020). Therefore, applying neural topic modelling to such a data set recorded during memory or decision making tasks would be highly interesting. Comparing assemblies identified throughout different stages of learning could give some interesting insights into the formation of memories and the learning process. It would be particularly interesting to investigate how the weights within an assembly change in an assembly that is present throughout multiple stages of an experiment. Additionally, a more complex task with more known experimental variables might shine a light on the role of all the identified assemblies which were not related to visual or behavioural information. So analysing a data set with more behavioural variables (e. g. more details on movement) and cognitive variables (decision making, cognitive load) might shine a light onto what other kinds of information the other neuronal assemblies encode.

5.8 CONCLUSION

Neural topic modelling is a new neuronal data analysis method to detect neuronal assemblies. It has been shown to identify neuronal assemblies which share the characteristics of assemblies consistently found by other methods. It goes beyond existing methods by allowing units to use multiplexed coding and extracting multiple activity patterns from each spike train or from the LFP signal. By including all of these activity patterns as potential contributors to neuronal assemblies, neural topic modelling can detect neuronal assemblies consisting of combinations of neural activity features which cannot be found by any other current method.

All of these aspects point towards the usefulness of neural topic modelling as a neuronal data analysis method. The results presented here indicate that neural topic modelling has the potential to uncover many currently unknown relationships in neuronal data.

6

A SPIKING NEURAL NETWORK WITH SPIKE TIMING DEPENDENT PLASTICITY FOR SPIKE SEQUENCE DETECTION

The method proposed in this chapter addresses the challenge of finding ever larger temporally precise spike sequences in large electrophysiological brain recordings. Moreover, it proposes to identify spike patterns in a biologically-plausible way i. e. through spike integration processes (coincidence detection) which biological neurons in the brain also possess. The spiking neural network presented here is based on a spiking neural network published in Masquelier et al. (2008).

From a computational point of view, finding repeating sequences of spikes becomes exponentially more expensive the more units the recording encompasses. Even in data sets of a few hundreds of neurons brute force algorithms will take days to complete due to the combinatorial explosion that occurs when having to take into account every possible combination of spike sequences (Buonomano and Maass, 2009).

The replication of the spiking neural network by Masquelier et al. (2008) in this chapter has been previously published in Hathway and Goodman (2018).

6.1 BACKGROUND

As described in more detail in Section 2.3.2, neurons exhibit repeating precisely timed spike sequences. These spike patterns have been linked to behavioural variables and external stimuli (Fellous et al., 2004; Prut et al., 1998; Torre, Quaglio, Denker, Brochier, Riehle and Grün, 2016). The first method developed to find these spike patterns was a brute force algorithm searching for all different combinations of temporal patterns (Abeles and Prut, 1996; Prut et al., 1998). Tellingly, this method was developed during a time when neuronal recordings included only up to ten neurons per session due to

recording limitations of the electrodes and spike patterns found included at most three different units (Prut et al., 1998). In recent years more sophisticated methods were developed to detect spike sequences which are computationally much more efficient (Grossberger et al., 2018; Mackevicius et al., 2018; Mishne et al., 2016; Quaglio et al., 2017; van der Meij and Voytek, 2018; Watanabe et al., 2019).

While ever longer repeating spike sequences are being found by ever improving detection methods, there is the question of whether the brain can actually detect and make use of such large patterns. The goal was therefore to build a spike pattern detection method that finds spike sequences in the same or a similar way in which the brain is detecting them. It is known that neurons can act as coincidence detectors and use that to do complicated calculations in a very efficient way (Abeles, 1982). Masquelier et al. (2008) show that a leaky integrate-and-fire neuron which receives input from 2000 input neurons via excitatory synapses can reliably become exclusively active in response to specific repeated spike sequences. They show that this simple feedforward network is robust against various amounts of noise, jittering of the spike times and will identify spike sequences in neurons which are indistinguishable by other characteristics e. g. their firing rates.

Firstly, I replicated their findings using the spiking neural network simulator Brian (Goodman and Brette, 2009; Stimberg et al., 2014) in Python, whereas the original study implemented the simulations in Matlab, with the main functions being computed in C/C++ through mex files. In addition, I examined some of the implementation details and parameters and investigated whether they were essential to the success of the algorithm. Subsequently I applied the artificial neural network to spike trains from actual electrophysiological recordings from mice to attempt to find large-scale spike patterns in the recordings.

6.2 SPIKING NEURAL NETWORK AND LEARNING RULE

This section describes the spiking neural network and the parameters involved in more detail.

6.2.0.1 *Artificial and electrophysiological data*

The artificial spike data created for the original study consists of 2000 neurons spiking at an average firing rate of 60 Hz, although the firing rate was

allowed to fluctuate between 10 and 90 Hz. The spike trains are then additionally made more noisy by jittering the pattern spikes, adding random noise spikes to all neurons, and ensuring a constant population firing rate as well as equal overall firing rates across all input neurons. Half of the input neurons spike according to a spike pattern of 50 ms length for about 25% of the time, which is the pattern that the network is supposed to find and become selective for.

The electrophysiological brain recording used in Section 6.4 is a publicly available data set provided by the Cortex Lab at University College London: "dualPhase" available at http://data.cortexlab.net/dualPhase3/. A384-channel "Phase3" Neuropixels electrode array was inserted into the brain of an awake, head-fixed mouse at the posterior position. The electrode traverses the visual cortex, hippocampus and some parts of the thalamus up until a maximum depth of about 3.8 mm. In the dualPhase data set, an additional electrode recorded from motor cortex and thalamus at the same maximum depth (Lopez et al., 2016). Neurons were identified via automatic and additional manual spike-sorting using Kilosort and Phy. The recordings last for about 90 minutes, during which three different visual paradigms are shown to the animal: horizontal and vertical bars, drifting grating, white or black squares were presented at random locations on a 9x34 grid. The visual stimuli were followed by a long rest period in the dark.

6.2.0.2 *Network architecture and learning rule*

The artificial neural network is a simple feedforward 2-layer network. The input neurons in layer 1 (input layer) spike at pre-determined times and are all connected to one leaky integrate-and-fire neuron in layer 2 (output layer). Spikes in the input layer increase the voltage of the output neuron according to the parameters listed in Table 6.1. The output neuron therefore spikes according to the incoming signals it receives.

The input layer consists of a large number of neurons with excitatory connections to the output neuron. In the case of the artificial data, 2000 neurons in the input layer spike according to spike trains which were injected with specific spike sequences 25% of the time. In Section 6.4, when using the dualPhase data, 407 input neurons spiked according to the recorded spike trains. The voltage of the output neuron is governed by equation 6.1 which is composed of a presynaptic excitatory postsynaptic potentials (first term) and a negative afterpotential after a spike (second term).

$$\frac{du}{dt} = \frac{Xx - u}{\tau_m} + \frac{Aa}{\tau_s} \tag{6.1}$$

In the event of a presynaptic spike at neuron i, x is increased by the respective synaptic weight w_i ($x \leftarrow x + w_i$), which initiates the voltage increase in the postsynaptic neuron.

$$\frac{dx}{dt} = -\frac{x}{\tau_{syn}} \tag{6.2}$$

When the postsynaptic voltage reaches the threshold T, a spike occurs: the voltage u is set to twice the threshold ($u \leftarrow 2T$), all excitatory postsynaptic potential (EPSP) are flushed (all x set to 0, $x \leftarrow 0$) and the negative afterpotential is set into motion (a is set to 1, $a \leftarrow 1$).

$$\frac{da}{dt} = -\frac{a}{\tau_s} \tag{6.3}$$

Parameters are listed in Table 6.1. Due to the smaller number of neurons in the electrophysiological recordings and their lower firing rates, the threshold of the output neuron was adjusted (T = 30 [a.u.] instead of T = 500 [a.u.]) so that the output neuron was active at a high enough firing rate to ensure learning.

The weights of the connections between the two layers are governed by a spike timing dependent plasticity (STDP) learning rule. STDP is a learning rule that describes the effect of relative spike timings on synaptic weight (Bi and Poo, 1998). In this network, when an input neuron spikes before the output neuron, the synaptic weight of this connection increases, reinforcing possible causal relationships between spikes (Masquelier et al., 2008). Con-

Table 6.1: Parameters for neuron model and STDP synapses

	Neuron model							
	τ_m [ms]	τ_s [ms]	τ_{syn} [ms]	T [a.u.]	X	A	Δx	Δa
artificial	10	2.5	2.5	500	$\frac{\tau_s}{\tau_m}^{\frac{\tau_m}{\tau_s - \tau_m}}$	-3T	1	1
recorded	10	2.5	2.5	30	$\frac{\tau_s}{\tau_m}^{\frac{\tau_m}{\tau_s - \tau_m}}$	-3T	1	1
	STDP synapses							
	τ_+ [ms]	τ_- [ms]	a_+	a_-	w_{min}	w_{max}		
art/rec	16.8	33.7	2^{-5}	$0.85a_+$	0	1		

trastingly if an input neuron spikes just after the output neuron, the synaptic weight will decrease, suggesting that this connection does not contribute much to the output of the second neuron.

There is much experimental evidence for the existence of STDP-like rules in the brains of various animal species (Meliza and Dan, 2006; Young et al., 2007; Zhang et al., 1998). It is important to note that there is no consensus on the exact version of STDP – which might even vary across brain regions – and that it is not the only mechanism of synaptic plasticity found (Lisman and Spruston, 2005; Shouval et al., 2002; Zenke et al., 2017). Using STDP as a learning rule however does imitate a plasticity rule present in the brain and therefore suggests that spike patterns found via STDP could also have been found by neurons in the brain, making this a brain-inspired approach to neural data analysis.

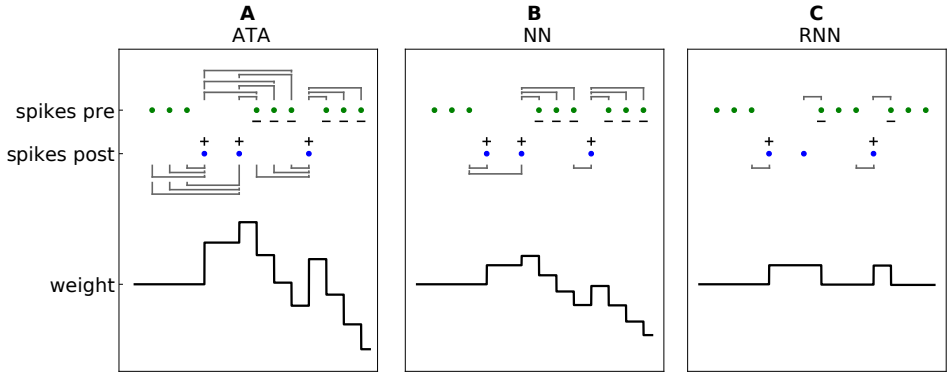


Figure 6.1: **Effect of different learning rules on synaptic weight.** Three different versions of the STDP learning rule are compared for a case in which the presynaptic neuron (green dots) spikes more often than the postsynaptic neuron (blue dots): **(A)** the All-To-All (ATA) version, **(B)** the Nearest Neighbor (NN) version and **(C)** the Reduced Nearest Neighbor (RNN) version. Grey lines indicate which spike pairs are taken into account for the calculation of the synaptic weight change; the sign denotes a corresponding increase (+) or decrease (-) in weight at that spike. The synaptic weight in ATA (A) and (B) experience more weight changes and also a net decrease in weight after only a few spikes. Under the RNN rule (C), there are fewer changes and – in this example – no visible net weight change. Figure was adapted from Hathway and Goodman (2018).

There are various versions of STDP, which differ mainly in how many spikes they take into account for the calculation of the change in synaptic weight (see Figure 6.1). The original study by Masquelier et al. (2008) uses the Reduced Nearest Neighbor (RNN) learning rule – a more restrictive version of the Nearest Neighbor (NN) STDP rule (Morrison et al., 2008); a visualisation is provided in Figure 6.1 (C). In the standard NN rule – visualised in Figure 6.1 (B) – every spike causes a weight change (with the amount de-

pending on the timing of the last spike of the other neurons), while for RNN a weight change only happens for the first postsynaptic spike immediately following a presynaptic spike (or vice versa). This means that for NN, there can be a series of potentiations or depressions, whereas for RNN potentiations and depressions strictly alternate. Both rules stand in contrast to the ATA STDP rule, which is shown in Figure 6.1 (A), where all spikes are taken into consideration.

6.3 VALIDATION OF THE ORIGINAL STUDY

I could reproduce the ability to find repeating patterns as well as the robustness of the model towards different types of noise (Hathway and Goodman, 2018). The pattern finding capabilities of the network rely heavily on some implementation parameters, namely simulation time step, learning rule, EPSP shape.

6.3.1 *Pattern finding and robustness*

The success of the neural network in finding a repeating spike sequence is measured in a stable spike time of the output neuron in relation to the start of the pattern, called latency. As in the original publication, at first the latency of the output neuron's spikes is randomly distributed across the time of the pattern presentation and outside of the pattern (represented as a latency = 0). I successfully reproduced the stabilisation of the latency i. e. the output spike honing in on a certain delay relative to the start of the pattern (usually around 5-15 ms). For better comparison, I used the same spike train as input for both the original code and in the replication implementation using the standard parameters found in Table 6.1). The latency development looks similar in both implementations. The time until a stable state arises (number of discharges until no spikes happen outside of the pattern at latency = 0) depends on the simulation time step and is therefore longer in the replication (1400 discharges instead of 700 or 20 s instead of 13.5 s of simulated time at a simulation time step of 10^{-4}. A more detailed inspection of the differences in performance due to the time step can be found in the following section on *Implementation details*.

The reason for the development of a stable latency is that the synaptic weights that are maximally potentiated (close to $w_{max} = 1$) belong exclu-

sively to neurons involved in the pattern, whereas weights from neurons not involved in the pattern are depressed nearly completely. Neurons that spike at the beginning of the pattern are potentiated, causing the postsynaptic neuron to spike at a low latency.

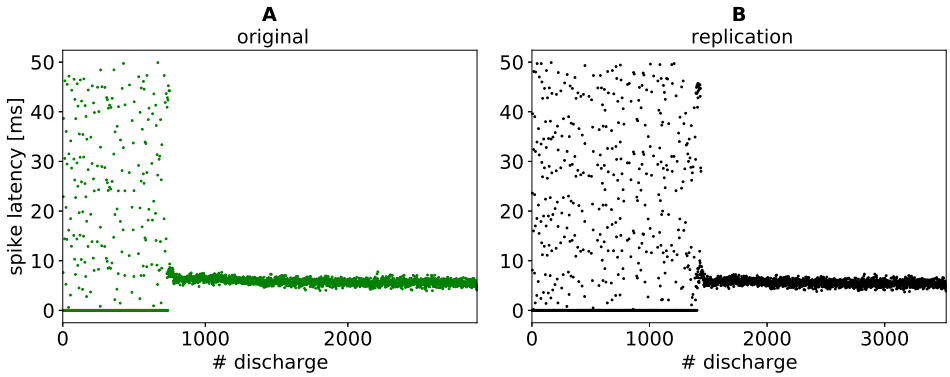


Figure 6.2: **Latency.** Comparing the latency development of **(A)** the original study and **(B)** the replication implementation. In both implementations, the timing of the output spike in relation to the start of the pattern (vertical axis, set to latency = 0 if output spike happens outside the pattern) is random at first, but becomes selective after a few hundred discharges. From then on the postsynaptic neuron always spikes only within the pattern (no spikes at latency = 0). The number of discharges until this happens is smaller in the original paper than in the replication implementation. Figure was adapted from Hathway and Goodman (2018).

Our simulations showed largely the same resilience to degradations as in the original paper: for different amounts of jitter (Fig. 6.3 (B)), various proportions of neurons involved in the pattern (Fig. 6.3 (C)), different initial weights (Fig. 6.3 (D)) and when a percentage of spikes within the pattern are deleted (Fig. 6.3 (E)) However, at a high pattern repetition frequency of 0.5 (Fig. 6.3 (A)), when the pattern is presented every 100 ms for 50 ms, the performance of the replication differs from the original: at larger time steps of 10^{-4} s the replication version does not perform as well as the original, but shows good results at smaller time steps (10^{-6} s).

6.3.2 *Implementation parameters*

Certain – not necessarily obvious – implementation details can affect the pattern finding success of the network significantly as I discuss in the following section. The parameter values referred to in this section can be found in Table 6.1 and a summary of the implementation details which resulted in successful and unsuccessful spike pattern detection are listed in Table 6.2.

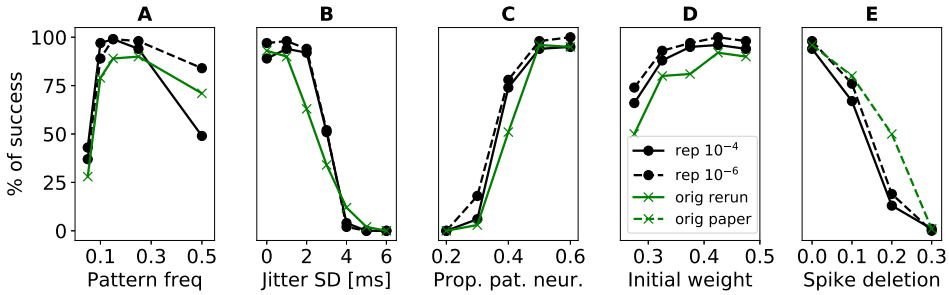


Figure 6.3: **Robustness.** Pattern finding abilities of the system were tested against **(A)** different pattern presentation frequencies, **(B)** jitter, **(C)** number of neurons spiking according to the pattern, **(D)** initial weights and **(E)** deletion of spikes in the pattern. For each parameter combination 100 simulations were run and the number of successful runs are reported as percentage of success. The replication implementation was run at two different time steps: 10^{-4} s (solid black lines) and 10^{-6} s (dashed black lines). The results from the original study were calculated using the source code provided (solid green lines in (A), (B), (C), (D)) or estimated from the original publication (dashed green line in (E)). Figure was adapted from Hathway and Goodman (2018).

The size of simulation time step affects the time until the pattern is found: the larger the time step, the later the output neurons becomes selectively active for the pattern presentation (10^{-6} s < 10^{-5} s < 10^{-4} s). However at 10^{-3} s – a common choice for spiking network simulations – no specificity emerges at all. This delay in finding can be attributed to the use of time bins in the network simulations in Brian 2. The original study calculated the effect of spikes on a spike-to-spike basis which made it nearly impossible for two spikes to happen simultaneously. Using binned time steps – especially larger ones – leads to output neuron spikes and input neuron spikes happening in the same time bin (time difference between the spikes of $\Delta t = 0$), where the STDP rule is undefined. Brian 2 treats all of the input neuron spikes in this time bin as if they happened just before the output neuron spike ($\Delta t < 0$, due to the scheduling of events in Brian) and will therefore increase all those weights instead of increasing some and decreasing others. This higher number of potentiations makes it more difficult for the system to systematically depress unimportant weights in order to become selective to the pattern. If the learning rule is modified so that the change in synaptic weight reflects the fact that on average half the input neurons spike before the output neuron and the other half afterwards (by adding the mean of LTP and LTD traces), the pattern is found at a time close to the performance for smaller time steps (10^{-6}s) and the original paper.

Similarly crucial is the choice of the RNN learning rule. The RNN rule results in stable posystnaptic firing and reliable finding of the pattern. The use of other learning rules does not result in this stable learning. The reason other learning rules fail to result in a pattern-specific output neuron lies in the firing rate characteristics used in the original publication (on average 60 Hz per neuron). Such a high firing rate will effect that input neurons fire multiple spikes in the time between output neuron spikes. The RNN learning rule will only take into account the latest spike of an input neuron in the calculation of the weight change. How many spikes occurred in the input neuron between output neuron spikes is irrelevant to this learning rule. On the contrary, the ATA and NN learning rules consider all spikes that happened between output neuron spikes for the weight change calculation. Refer to Figure 6.1 for a visualisation of this effect. Input neurons will experience a decrease in weight a lot more often (every time an input neuron spikes) than an increase in weight (every time the postsynaptic neuron spikes). This leads to a strong overall depression of the synaptic weights in the first few seconds of the simulation. With the parameters specified in Masquelier et al. (2008) and an ATA or a conventional NN learning rule, the output neuron stops firing after a few seconds because the output neuron voltage does not reach the voltage threshold necessary to evoke an output neuron spike anymore. When no output neuron spikes occur, learning cannot take place and no specificity can emerge.

The exact shape of the excitatory postsynaptic potentials is also crucial to the pattern finding abilities of the model. The kernels used in the original study simulate a gradual increase of the postsynaptic neuron voltage. Other studies sometimes also model the effect of the presynaptic spike as an immediate jump in postsynaptic voltage instead of that gradual increase. The latter, using an immediate increase in postsynaptic voltage, does not lead to stable pattern finding. This might have to do with the fact that the kernel shape and the immediate increase exhibit slightly different spike times as seen in Fig. 6.4.

6.4 APPLICATION TO REAL DATA SETS

After verifying that the artificial neural network with STDP synapses can find the start of a repeating pattern in artificial data (albeit only under certain circumstances), I applied the same network to the spike times of electrophysiological data. To account for the lower firing rates (6.4 Hz instead

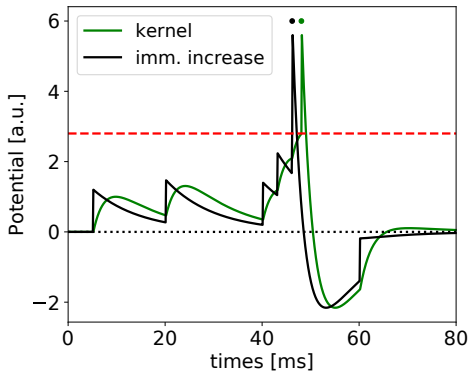


Figure 6.4: **Choice of EPSP.** For the same presynaptic spikes, the postsynaptic voltage behaves similarly in general, but the postsynaptic spike time is different for different shapes for the EPSP. Figure was adapted from Hathway and Goodman (2018).

Table 6.2: Success-relevant implementation details

	works	does not work
simulation time step	continuous, discrete	
time step size	$\leqslant 10^{-4}$ s	$> 10^{-4}$ s
learning rule	RNN	ATA, NN
EPSP shape	kernel	immediate voltage increase

of 64 Hz) and fewer neurons (407 instead of 2000), the threshold for spiking in the output neuron was lowered from 500 [a.u.] to 30 [a.u.]. All other parameters were kept the same as specified in Table 6.1. The simulation time was 900 seconds and each parameter setting was repeated 20 times with randomly initialised initial weights (Gaussian noise around 0.5 with standard deviation of 0.05).

Even though the RNN learning rule proved to be the only successful one for the artificial data, I wanted to be able to compare the three learning rules, since the lower firing rate might work in favour of the ATA rule. Therefore, each input was run through three networks which only differed in the learning rule used (RNN, NN, ATA). To assess whether the network actually potentiates fine temporal patterns, I jittered the spike timings using Gaussian noise with standard deviations of 0 ms, 10 ms and 1000 ms. Additionally, I ran another simulation with the same number of neurons with the same firing rates, in which the spike times were drawn from a Poisson process. This simulation was used as a control since random spiking should not reliably lead to the reliable potentiation of weights from those neurons.

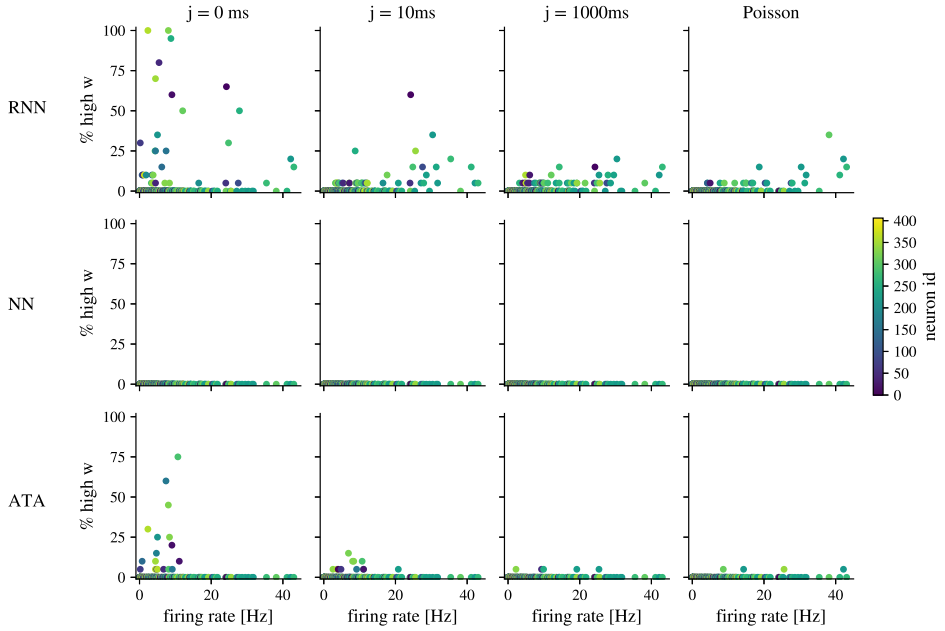


Figure 6.5: **Potentiation of synaptic weights by the STDP spiking neural network in electrophysiological data.** In each subplot, each dot represents one neuron in the electrophysiological recording. The plots show the percentage of runs (out of 20) the weight of that neuron was higher than 0.8 (on a scale from 0 to 1) at the end of the simulation. Three different learning rules are compared (RNN, NN, ATA; rows) for four different noise settings (no noise, jittered with Gaussian noise with standard deviation of 10 ms, jittered with Gaussian noise with standard deviation = 1000 ms, Poisson spikes; columns). The colourbar indicates the neuron ID for a better differentiation of the different neurons.

Regarding the learning rules, it is evident from Figure 6.5 that only the RNN and the ATA rule manage to potentiate weights for some neurons reliably, and the RNN more than the ATA rule. Interestingly, these neurons with potentiated weights seem to have low firing rates, thereby relying on precise timings of their spikes. This is further corroborated by those neurons not being potentiated in the runs where spike times were even slightly jittered (10 ms). In the runs with a lot of jitter (1000 ms) and the Poisson spike runs, higher firing rates become potentiated more often but not reliably (only for around 5-25% of runs). There seems to be some, but no clear overlap between the neurons potentiated under the two learning rules when no jitter iss present.

6.5 DISCUSSION

Masquelier et al. (2008) observed that a single neuron equipped with STDP could reliably detect large multi-neuron repeating spike patterns in continuous noisy spiking activity. This was the case even though no explicit reference to the start of the pattern was given, which was long thought to be essential to the detection of patterns in neural data. Their algorithm showed resilience against additional, unrelated noise and even small variations in spike timing within the pattern.

I could replicate these findings including the resilience against noise in the input. However, I discovered that the network's pattern finding abilities rely on seemingly secondary parameters such as EPSP shape, time step size, and exact learning rule. When applied to the publicly available electrophysiological data sets, I could not identify the large-scale spike patterns I was hoping for, instead only a handful of neurons reliably exhibited a high synaptic weight at the end of a simulation (in more than 75% of runs). Even the much longer simulation time of the neural activity (> 15 min) compared to the artificial data did not result in more potentiated neurons. Weight changes during the simulations do not happen very often because there is a trade-off between lowering the threshold of the output neuron to preserve a high output neuron firing rate (and thereby enable more learning) and the network potentiating random neurons (false positives). With lower thresholds of 20 [a.u.] or 25 [a.u.], neurons with high firing rates are especially prone to reliably appear as false positives in the high jitter and Poisson runs.

6.6 CONCLUSION

This chapter proposed a biologically-plausible spiking neural network which finds repeating spike sequences in large-scale spike trains. The neural network was implemented in easy-to-use Python code and the pattern finding abilities of the network were validated on artificial data sets. It was demonstrated that the neural network is robust towards a number of interventions. The neural network was applied to electrophysiological recordings from mouse brains and was shown to reliably potentiate a number of neurons, indicating that these neurons were part of repeating spike sequences. The code to use the neural network and to create the artificial data is freely available for download and use: https://github.com/ReScience-Archives/Hathway-Goodman-2018.

The advantage of this spiking neural network is that the computational cost does not scale exponentially with the number of neurons, but can be computed more efficiently. Therefore this neural network is suitable for much larger neuronal recordings e. g. calcium imaging datasets with 10 000s of neurons. In addition, since the learning rule used for the pattern detection is a learning rule exhibited by actual neurons in the brain, it is given that any patterns found by the network can also be found by biological neurons. This can give us an insight into how the brain identifies and becomes specific to large spike sequences present in neuronal activity.

Future works could investigate how the network can be made more attuned to the type of patterns typically found in the brain, namely a smaller number of spikes involved in the spike sequences and a lower number of repetitions. In addition, it would be interesting to investigate whether heterogenous initialisations of parameters such as the neuronal and synaptic time constants could assist in pattern detection.

7

CONCLUSION

This chapter summarises the contributions presented in this thesis, discusses some limitations and outlines possible research directions for future work.

7.1 OVERVIEW AND CONTRIBUTIONS OF THE THESIS

Analysing neuronal recordings is becoming more and more challenging as data set size grows due to technological advancements in neuronal recording equipment. This thesis has explored two approaches to the challenge of detecting neuronal assemblies – groups of neurons which show co-modulations of some aspects of their activity. Both approaches make use of biologically-inspired machine learning methods for the analysis of large-scale electrophysiological data sets.

One approach was the use of a spiking neural network with spike timing dependent plasticity (STDP) synapses for the detection of repeated spike sequences in large electrophysiological data sets. Chapter 6 outlined the implementation of the previously published network (Masquelier et al., 2008) in Python. It additionally described the robustness of the network against parameter variations and noise in the data. The choice of specific variation of the STDP learning rule proved to be crucial to the success of the network in identifying large repeating spike sequences. Additionally, the chapter showed the results of applying the spiking neural network to large neuronal recordings. Unfortunately, the network in the implementation presented here did not convincingly identify large repeating spike sequences in the neuronal data. Further adjustments need to be made for the network to perform as well on real life data sets as it does on artificial data.

The second approach was the application of text mining algorithms to neuronal recordings. Chapter 3 presented neural topic modelling, a new data analysis method to find neuronal assemblies in large electrophysiological data sets. Currently, there exist two types of neuronal assembly detec-

tion algorithms: those which look for a set of neurons participating in repeated spike sequences and those which look for groups of neurons with co-modulated firing rates. This limits the search for neuronal assemblies to one of two binary options – a stark contrast to the acknowledgement by researchers that neurons can contribute to assemblies through a variety of different activity patterns. Neural topic modelling allows the combined analysis of timed spikes, firing rate modulations and even other data types such as local field potentials (LFPs). This approach therefore challenges the current binary choice of how to search for neuronal assemblies, each of which exclude all but one aspect of neuronal data. By allowing multiple activity patterns to be extracted from each unit, neural topic modelling accounts for multiplexed coding in neurons (using different strategies to encode different stimuli), something that is not taken into account by any other assembly detection method.

Neural topic modelling was tested and validated on artificial data sets. The results in Chapter 4 show that neural topic modelling often surpasses the performance of the previously developed algorithms on artificial data sets with known assembly structures. The artificial data was created in line with those being used in a previous publication comparing different assembly detection algorithms.

Applying neural topic modelling to large electrophysiological recordings revealed interesting neuronal assemblies as shown in Chapter 5. Some of the assemblies resemble the ones previously characterised by other methods (e. g. firing rate co-modulations in response to visual stimuli or movement of the animal, synchronous spiking in visual areas and hippocampus). In addition to confirming previously found assembly structures, neural topic modelling can extend the current knowledge by also generating assemblies in which different activity patterns interact with each other (e. g. timed spikes and the power in low frequencies of the LFP signal). The analysis of such interactions are unique to neural topic modelling. Furthermore, the found assembly structures can reveal interesting dynamics in cooperation between brain regions across different experiments.

7.2 LIMITATIONS OF APPLYING MACHINE LEARNING TO NEURAL DATA

The scalability and the success in other fields has made machine learning seem like something magical, a tool that can be applied to one's data and will just come up with the correct solution. In reality, this is not the case and

data preprocessing and avoiding overfitting are massive challenges. Additionally, researchers in other fields which rely heavily on machine learning may have different sets of goals to neuroscientists. As an example, computer vision researchers are often not massively interested in how the neural network arrived at the decision that a particular image showed a human, they only care that it did.[1] In neuroscience, the way in which an algorithm arrived at a particular decision is maybe as interesting or even more so than the knowledge that it has. The choice of method can potentially generate insights into what kind of information the brain might encode and how it achieves this.

This thesis has attempted to select methods from machine learning which can be interpreted in biological contexts. These approaches work well in some contexts (neural topic modelling) and not overwhelmingly so in others (spiking neural network with STDP). It is very likely that powerful machine learning algorithms such as Deep Learning would achieve higher accuracy at decoding neural data, but the biological interpretation of any feature used in that kind of model is nearly impossible. Maintaining a link to the underlying biological processes might not always be the easiest path, but it is a worthwhile effort. Carefully selecting models for their transparency and biological interpretability can generate interesting hypotheses about how the brain processes information and provides an opportunity to uncover previously unknown brain communication mechanisms.

7.3 FUTURE DIRECTIONS

As neural topic modelling generates new insights into how neural activity patterns interact with each other in the brain, it also generates hundreds of new questions. Do interactions between assemblies change on separate recording sessions or when different tasks are given to the animal? Are the patterns of interactions between activity patterns consistent across different animals? Will we see other patterns of interactions when recording from

1 And this bias towards solutions at the expense of understanding what is going on is creating its own set of problems. There is a vast amount of evidence that algorithms for face recognition perform poorly for women of colour, recruiting software discriminates against women for tech jobs, and automatic image cropping algorithms remove people of colour from Twitter image previews. Unfortunately the response of some tech companies have been lacklustre – instead of fixing their algorithm, Google banned any image from ever being labelled with the words "gorilla" and "monkey" to prevent its racist image labelling algorithm to continue to add these tags to images of black people. Scientists should do their best to avoid or at least investigate biases in their algorithms.

different brain regions or even when using different recording techniques such as different electrode layouts or calcium imaging data? What is the effect of behaviour on the interactions?

Then there are additional questions about whether neural topic modelling can capture changes in assembly dynamics due to learning or experience. One of the postulations of Hebb for assemblies was that the activity of a subset of neurons in an assembly may activate the whole assembly (Deolindo et al., 2018). With the help of optogenetics, we might confirm this type of behaviour and might also be able to double check the weights of each neuron in the assemblies (Buzsáki, 2004). Another postulation was that continued activity of an assembly should increase the connection strength between neurons in the assembly. This might be a very interesting experiment – to see whether weights of neurons in an assembly may be increased through assembly-based activity. If this is the case, this would support the idea of looking at the weights of each unit in an assembly to better understand assembly dynamics.

In the best case scenario, I envision neural topic modelling as a science accelerator – it can make it easier to find interesting relationships in old and new data sets and can make finding complex relationships between spike patterns faster and easier.

Spatiotemporal spike patterns and firing rate modulations have independently been shown to convey some information about an animal's behaviour and external environment (Allen et al., 2019; Torre, Quaglio, Denker, Brochier, Riehle and Grün, 2016). Understanding the relationships between those aspects of neuronal activity and also other features of neuronal activity will improve our understanding of how the brain encodes information. Such insights might even be used to improve the specificity of neural prostheses (Smith et al., 2018)

7.4 CONCLUDING REMARKS

At the end of the ~~PhD thesis~~ day, we do not know how useful neuronal assemblies are in the grand puzzle of how the brain works. Every year researchers discover new exciting ways in which the brain is more complicated and more sophisticated than we thought – and even single neurons become more and more powerful the longer we look at them, since even individual axons seem to be capable of highly intricate computations (Bicknell, 2019).

Neuronal assemblies might be a major building block or they might just be the currently visible part of the floating iceberg and in a few years we will wonder how we did not see the next part underneath the surface. What is certain though is that finding neural assemblies is one step towards discovering ever more complex interactions between neurons. They might just provide a clearer picture of how these millions of neurons give rise to fascinating abilities (such as these that let me summarise the work of the last four years in this thesis). I can only hope to have contributed a small bit to the overall knowledge leading to a greater understanding of our own brains.

A

AUTHOR'S PUBLICATIONS

This appendix contains a list of the peer-reviewed publications made during the course of this PhD. The reference for each publication is provided as well as a short summary and in which chapters the content of the papers are discussed.

Hathway, P. and Goodman, D.F.M. 2018, [Re] Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains, *in* 'ReScience' 4, 1, #6. doi: 10.5281/zenodo.1327348.

- This article contains the successful reproduction of an algorithm for the detection of repeated spike sequences in neural data. In addition to the open-source implementation in Python, the influence of various parameters on the success of the detection algorithm are discussed.
- Chapter 6 is based on this article.

Hathway, P. and Goodman, D.F.M. 2019, Neural topic modelling, *in* 'Conference on Cognitive Computational Neuroscience, Berlin, Germany', PS-2B.46. doi: 10.32470/CCN.2019.1382-0.

- In this publication, I present neural topic modelling as a new analysis method for large-scale neural data sets. The unsupervised method is shown to recover the known physical and functional relationships of neurons in a electrophysiological recording.
- Chapter 3 is in part based on this article, which included an earlier version of the pipeline presented here.

BIBLIOGRAPHY

Abeles, M. (1982), 'Role of the cortical neuron: Integrator or coincidence detector?', *Israel Journal of Medical Sciences* **18**(1), 83–92. (pages 34, 146).

Abeles, M. and Gerstein, G. L. (1988), 'Detecting spatiotemporal firing patterns among simultaneously recorded single neurons', *Journal of Neurophysiology* **60**(3), 909–924. doi: 10.1152/jn.1988.60.3.909 (page 18).

Abeles, M. and Prut, Y. (1996), 'Spatio-temporal firing patterns in the frontal cortex of behaving monkeys', *Journal of Physiology Paris* **90**(3-4), 249–250. doi: 10.1016/S0928-4257(97)81433-7 (pages 20, 36, 72, 145).

Allen Brain Observatory (2019), 'Neuropixels Visual Coding', *Technical White Paper* **October**, Version 1.0. Retrieved from `https://brainmapportal-live-4cc80a57cd6e400d854-f7fdcae.divio-media.net/filer_public/80/75/8075a100-ca64-429a-b39a-569121b612b2/neuropixels_visual_coding_-_white_paper_v10.pdf` (pages 44, 56, 72, 106, 107, 121).

Allen, W. E., Chen, M. Z., Pichamoorthy, N., Tien, R. H., Pachitariu, M., Luo, L. and Deisseroth, K. (2019), 'Thirst regulates motivated behavior through modulation of brainwide neural population dynamics', *Science* **364**(6437), 0–10. doi: 10.1126/science.aav3932 (pages 26, 28, 42, 53, 162).

Arabzadeh, E., Panzeri, S. and Diamond, M. E. (2006), 'Deciphering the spike train of a sensory neuron: Counts and temporal patterns in the rat whisker pathway', *Journal of Neuroscience* **26**(36), 9216–9226. doi: 10.1523/JNEUROSCI.1491-06.2006 (page 29).

Asuncion, A., Welling, M., Smyth, P. and Teh, Y. W. (2009), On Smoothing and Inference for Topic Models, *in* 'Uncertainty in Artificial Intelligence'. (page 62).

Avitan, L., Pujic, Z., Mölter, J., Van De Poll, M., Sun, B., Teng, H., Amor, R., Scott, E. K. and Goodhill, G. J. (2017), 'Spontaneous Activity in the Zebrafish Tectum Reorganizes over Development and Is Influenced by Visual Experience', *Current Biology* **27**(16), 2407–2419.e4. doi: 10.1016/j.cub.2017.06.056 (pages 40, 50, 73, 102).

Bair, W., Cavanaugh, J. R., Smith, M. A. and Movshon, J. A. (2002), 'The timing of response onset and offset in macaque visual neurons.', *Journal of Neuroscience* **22**(8), 3189–205. doi: 20026257 (page 29).

Bi, G. Q. and Poo, M. M. (1998), 'Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type.', *Journal of Neuroscience* **18**(24), 10464–10472. doi: 10.1038/25665 (page 148).

Bicknell, B. (2019), Learning the principles of single-neuron computation, *in* 'UK Neural Computation'. (page 162).

Bienenstock, E. (1995), 'A model of neocortex', *Network: Computation in Neural Systems* **6**(2), 179–224. doi: 10.1088/0954-898X_6_2_004 (page 34).

Blei, D. M. (2012), 'Probabilistic topic models', *Communications of the ACM* **55**(4), 77. doi: 10.1145/2133806.2133826 (pages 61, 62, 63, 72).

Blei, D. M., Griffiths, T. L. and Jordan, M. I. (2010), 'The nested Chinese restaurant process and Bayesian nonparametric inference of topic hierarchies', *Journal of the American Society for Information Science* **57**(2). doi: 10.1145/1667053.1667056 (page 70).

Blei, D. M., Griffiths, T. L., Jordan, M. I. and Tenenbaum, J. B. (2003), Hierarchical Topic Models and the Nested Chinese Restaurant Process, *in* 'Proceedings of the 16th International Conference on Neural Information Processing Systems', pp. 17–24. (page 140).

Blei, D. M. and Lafferty, J. D. (2006), Dynamic topic models, *in* 'ACM International Conference Proceeding Series', Vol. 148, ACM Press, pp. 113–120. doi: 10.1145/1143844.1143859 (page 140).

Blei, D. M. and Lafferty, J. D. (2007), 'A Correlated Topic Model of Science', *The Annals of Applied Statistics* **1**(1), 17–35. doi: 10.1214/07-AOAS114 (page 140).

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003), 'Latent Dirichlet Allocation', *Journal of Machine Learning Research* **3**(Jan), 993–1022. (pages 43, 59, 60, 62, 111).

Brecht, M., Schneider, M., Sakmann, B. and Margie, T. W. (2004), 'Whisker movements evoked by stimulation of single pyramidal cells in rat motor cortex', *Nature* **427**(6976), 704–710. doi: 10.1038/nature02266 (pages 28, 29).

Buonomano, D. V. and Maass, W. (2009), 'State-dependent computations: spatiotemporal processing in cortical networks', *Nature Reviews Neuroscience* **10**(2), 113–125. doi: 10.1038/nrn2558 (page 145).

Butts, D. A., Weng, C., Jin, J., Yeh, C. I., Lesica, N. A., Alonso, J. M. and Stanley, G. B. (2007), 'Temporal precision in the neural code and the timescales of natural vision', *Nature* **449**(7158), 92–95. doi: 10.1038/nature06105 (pages 29, 52).

Buzsáki, G. (2004), 'Large-scale recording of neuronal ensembles', *Nature Neuroscience* **7**(5), 446–451. doi: 10.1038/nn1233 (pages 29, 35, 126, 162).

Buzsáki, G. (2010), 'Neural Syntax: Cell Assemblies, Synapsembles, and Readers', *Neuron* **68**(3), 362–385. doi: 10.1016/j.neuron.2010.09.023 (pages 29, 31, 33, 38, 42, 44, 72, 138).

Caldas, J., Gehlenborg, N., Faisal, A., Brazma, A. and Kaski, S. (2009), 'Probabilistic retrieval and visualization of biologically relevant microarray experiments', *Bioinformatics* **25**(12). doi: 10.1093/bioinformatics/btp215 (page 59).

Carrillo-Reid, L., Miller, J. E. K., Hamm, J. P., Jackson, J. and Yuste, R. (2015), 'Endogenous sequential cortical activity evoked by visual stimuli', *Journal of Neuroscience* **35**(23), 8813–8828. doi: 10.1523/JNEUROSCI.5214-14.2015 (pages 41, 42, 71).

Carrillo-Reid, L., Yang, W., Bando, Y., Peterka, D. S. and Yuste, R. (2016), 'Imprinting and recalling cortical ensembles', *Science* **353**(6300), 691–694. doi: 10.1126/science.aaf7560 (pages 32, 33, 43, 71, 73, 91).

Castellani, U., Perina, A., Murino, V., Bellani, M., Rambaldelli, G., Tansella, M. and Brambilla, P. (2010), 'Brain morphometry by probabilistic latent semantic analysis', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **6362 LNCS**(PART 2), 177–184. doi: 10.1007/978-3-642-15745-5_22 (page 59).

Chen, X., Hu, X., Shen, X. and Rosen, G. (2010), Probabilistic topic modeling for genomic data interpretation, *in* '2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)', pp. 149–152. doi: 10.1109/BIBM.2010.5706554 (page 59).

Chen, Y., Yin, X., Li, Z., Hu, X. and Huang, J. X. (2012), 'A LDA-based approach to promoting ranking diversity for genomics information retrieval',

BMC Genomics 2012 13:3 **13**(3), S2. doi: 10.1186/1471-2164-13-s3-s2 (page 59).

Cohen, M. R. and Newsome, W. T. (2008), 'Context-Dependent Changes in Functional Circuitry in Visual Area MT', *Neuron* **60**(1), 162–173. doi: 10.1016/j.neuron.2008.08.007 (page 55).

Cunningham, J. P. and Yu, B. M. (2014), 'Dimensionality reduction for large-scale neural recordings', *Nature Neuroscience* **17**(11), 1500–1509. doi: 10.1038/nn.3776 (pages 20, 71, 73).

Deerwester, S., Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. and Harshman, R. (1990), 'Indexing by latent semantic analysis', *Journal of the American Society for Information Science* **41**(6), 391–407. (page 59).

Dejean, C., Courtin, J., Karalis, N., Chaudun, F., Wurtz, H., Bienvenu, T. C. and Herry, C. (2016), 'Prefrontal neuronal assemblies temporally control fear behaviour', *Nature* **535**(7612), 420–424. doi: 10.1038/nature18630 (page 38).

Deolindo, C. S., Kunicki, A. C. B., Brasil, F. L. and Moioli, R. C. (2014), Limitations of principal component analysis as a method to detect neuronal assemblies, *in* '2014 IEEE 16th International Conference on e-Health Networking, Applications and Services, Healthcom 2014', IEEE, pp. 24–30. doi: 10.1109/HealthCom.2014.7001808 (page 38).

Deolindo, C. S., Kunicki, A. C., Da Silva, M. I., Brasil, F. L. and Moioli, R. C. (2018), 'Neuronal assemblies evidence distributed interactions within a tactile discrimination task in rats', *Frontiers in Neural Circuits* **11**(January), 1–15. doi: 10.3389/fncir.2017.00114 (pages 20, 33, 34, 38, 40, 41, 42, 162).

Desimone, R., Albright, T. D., Gross, C. G. and Bruce, C. (1984), 'Stimulus-selective properties of inferior temporal neurons in the macaque', *Journal of Neuroscience* **4**(8), 2051–2062. doi: 10.1523/jneurosci.04-08-02051.1984 (page 28).

Diana, G., Sainsbury, T. T. and Meyer, M. P. (2019), 'Bayesian inference of neuronal assemblies', *PLoS Computational Biology* **15**(10), e1007481. doi: 10.1371/journal.pcbi.1007481 (pages 41, 42, 43, 69, 71, 76, 78, 79, 80, 81, 83, 85, 86, 87, 88, 89, 90, 91, 98, 100, 101, 102).

Doshi-Velez, F. and Kim, B. (2017), 'Towards A Rigorous Science of Interpretable Machine Learning', *arXiv* (Ml), 1–13. arXiv: 1702.08608 (pages 21, 48).

Dragoi, G. (2020), 'Cell assemblies, sequences and temporal coding in the hippocampus', *Current Opinion in Neurobiology* **64**, 111–118. doi: 10.1016/j.conb.2020.03.003 (pages 33, 34, 35, 50, 52, 143).

Dragoi, G. and Buzsáki, G. (2006), 'Temporal Encoding of Place Sequences by Hippocampal Cell Assemblies', *Neuron* **50**(1), 145–157. doi: 10.1016/j.neuron.2006.02.023 (page 33).

Ehret, G. (1975), 'Masked auditory thresholds, critical ratios, and scales of the basilar membrane of the housemouse (Mus musculus)', *Journal of comparative physiology* **103**, 329–341. doi: 10.1007/BF00612025 (page 27).

Farooq, U., Sibille, J., Liu, K. and Dragoi, G. (2019), 'Strengthened Temporal Coordination within Pre-existing Sequential Cell Assemblies Supports Trajectory Replay', *Neuron* **103**(4), 719–733.e7. doi: 10.1016/j.neuron.2019.05.040 (page 35).

Fellous, J.-M., Tiesinga, P. H. E., Thomas, P. J. and Sejnowski, T. J. (2004), 'Discovering spike patterns in neuronal responses.', *Journal of Neuroscience* **24**(12), 2989–3001. doi: 10.1523/JNEUROSCI.4649-03.2004 (page 145).

Friedrich, R. W., Habermann, C. J. and Laurent, G. (2004), 'Multiplexing using synchrony in the zebrafish olfactory bulb', *Nature Neuroscience* **7**(8), 862–871. doi: 10.1038/nn1292 (page 30).

Gansel, K. S. and Singer, W. (2012), 'Detecting Multineuronal Temporal Patterns in Parallel Spike Trains', *Frontiers in Neuroinformatics* **6**(May), 1–16. doi: 10.3389/fninf.2012.00018 (page 36).

Gärtner, M., Duvarci, S., Roeper, J. and Schneider, G. (2017), 'Detecting joint pausiness in parallel spike trains', *Journal of Neuroscience Methods* **285**, 69–81. doi: 10.1016/j.jneumeth.2017.05.008 (page 138).

Gerstein, G. L. and Aertsen, A. M. (1985), 'Representation of cooperative firing activity among simultaneously recorded neurons', *Journal of Neurophysiology* **54**(6), 1513–1528. doi: 10.1152/jn.1985.54.6.1513 (page 26).

Gerstein, G. L., Bedenbaugh, P. and Aertsen, M. H. (1989), 'Neuronal Assemblies', *IEEE Transactions on Biomedical Engineering* **36**(1), 4–14. doi: 10.1109/10.16444 (pages 27, 32, 33).

Gerwinn, S. (2009), 'Bayesian population decoding of spiking neurons', *Frontiers in Computational Neuroscience* **3**(October), 1–14. doi: 10.3389/neuro.10.021.2009 (page 21).

Goldberg, M. K., Hayvanovych, M. and Magdon-Ismail, M. (2010), Measuring similarity between sets of overlapping clusters, *in* 'Proceedings - SocialCom 2010: 2nd IEEE International Conference on Social Computing', pp. 303–308. doi: 10.1109/SocialCom.2010.50 (page 81).

Gollisch, T. and Meister, M. (2008), 'Rapid neural coding in the retina with relative spike latencies', *Science* **319**(5866), 1108–1111. doi: 10.1126/science.1149639 (page 29).

Goodman, D. F. M. and Brette, R. (2009), 'The brian simulator', *Frontiers in Neuroscience* **3**, 192–197. doi: 10.3389/neuro.01.026.2009 (page 146).

Green, S., Şerban, M., Scholl, R., Jones, N., Brigandt, I. and Bechtel, W. (2018), 'Network analyses in systems biology: new strategies for dealing with biological complexity', *Synthese* **195**(4), 1751–1777. doi: 10.1007/s11229-016-1307-6 (page 28).

Gridchyn, I., Schoenenberger, P., O'Neill, J. and Csicsvari, J. (2020), 'Assembly-Specific Disruption of Hippocampal Replay Leads to Selective Memory Deficit', *Neuron* **106**(2), 291–300.e6. doi: 10.1016/j.neuron.2020.01.021 (page 35).

Grossberger, L., Battaglia, F. P. and Vinck, M. (2018), 'Unsupervised clustering of temporal patterns in high-dimensional neuronal ensembles using a novel dissimilarity measure', *PLoS Computational Biology* **14**(7), 1–34. doi: 10.1371/journal.pcbi.1006283 (pages 37, 146).

Grün, S., Abeles, M. and Diesmann, M. (2008), Impact of Higher-Order Correlations on Coincidence Distributions of Massively Parallel Data, *in* 'Dynamic Brain - from Neural Spikes to Behaviors', Springer, Berlin, Heidelberg, pp. 96–114. doi: 10.1007/978-3-540-88853-6_8 (pages 34, 36).

Hafting, T., Fyhn, M., Molden, S., Moser, M. B. and Moser, E. I. (2005), 'Microstructure of a spatial map in the entorhinal cortex', *Nature* **436**(7052), 801–806. doi: 10.1038/nature03721 (pages 28, 31).

Hahnloser, R. H., Kozhevnikov, A. A. and Fee, M. S. (2002), 'An ultra-sparse code underlies the generation of neural sequences in songbird', *Nature* **419**, 65–70. (page 19).

Haller, M., Donoghue, T., Peterson, E., Varma, P., Sebastian, P., Gao, R., Noto, T., Knight, R., Shestyuk, A. and Voytek, B. (2018), 'Parameterizing neural power spectra', *bioRxiv* p. 299859. doi: 10.1101/299859 (page 139).

Hardcastle, K., Maheswaranathan, N., Ganguli, S. and Giocomo, L. M. (2017), 'A Multiplexed, Heterogeneous, and Adaptive Code for Navigation in Medial Entorhinal Cortex', *Neuron* **94**(2), 375–387.e7. doi: 10.1016/j.neuron.2017.03.025 (pages 27, 30, 52, 71, 102).

Harris, K. D. (2005), 'Neural signatures of cell assembly organization', *Nature Reviews Neuroscience* **6**(5), 399–407. doi: 10.1038/nrn1669 (pages 31, 33, 42, 53).

Harvey, M. A., Saal, H. P., Dammann, J. F. and Bensmaia, S. J. (2013), 'Multiplexing Stimulus Information through Rate and Temporal Codes in Primate Somatosensory Cortex', *PLoS Biology* **11**(5), e1001558. doi: 10.1371/journal.pbio.1001558 (page 30).

Hathway, P. and Goodman, D. (2018), '[Re] Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains', *ReScience* **4**(1), #6. doi: 10.1371/journal.pone.0001377 (pages 145, 149, 150, 151, 152, 154).

Hebb, D. (1949), *The Organization of Behavior*, Wiley, New York. (page 47).

Herreras, O. (2016), 'Local field potentials: Myths and misunderstandings', *Frontiers in Neural Circuits* **10**(DEC), 101. doi: 10.3389/fncir.2016.00101 (pages 115, 127, 135).

Hoffman, M. D., Blei, D. M. and Bach, F. (2010), Online learning for Latent Dirichlet Allocation, *in* '24th Annual Conference on Neural Information Processing Systems 2010', pp. 856–864. (page 62).

Hoffman, M., Blei, D. M., Wang, C. and Paisley, J. (2013), 'Stochastic Variational Inference', *Journal of Machine Learning Research* **14**, 1303–1347. arXiv: 1206.7051 (page 58).

Hofmann, T. (2001), 'Unsupervised learning by probabilistic Latent Semantic Analysis', *Machine Learning* **42**(1-2), 177–196. doi: 10.1023/A:1007617005950 (page 59).

Hong, G. and Lieber, C. M. (2019), 'Novel electrode technologies for neural recordings', *Nature Reviews Neuroscience* **20**(6), 330–345. doi: 10.1038/s41583-019-0140-6 (page 25).

Hubel, D. H. and Wiesel, T. N. (1961), 'Integrative action in the cat's lateral geniculate body', *The Journal of Physiology* **155**(2), 385–398. doi: 10.1113/jphysiol.1961.sp006635 (pages 17, 26, 27, 53).

Hughes, M. C., Hope, G., Weiner, L., Mccoy, T. H., Perlis, R. H., Sudderth, E. and Doshi-velez, F. (2018), 'Semi-Supervised Prediction-Constrained Topic Models', *Aistats* **84**(2003). (pages 48, 59, 103, 111, 140).

Hughes, M. C., Weiner, L., Hope, G., McCoy, T. H., Perlis, R. H., Sudderth, E. B. and Doshi-Velez, F. (2017), Prediction-Constrained Training for Semi-Supervised Mixture and Topic Models, *in* 'NIPS Machine Learning for Health workshop', pp. 1–28. arXiv: 1707.07341 (page 103).

Huk, A. C. and Hart, E. (2019), 'Parsing signal and noise in the brain', *Science* **364**(6437), 236–237. doi: 10.1126/science.aax1512 (page 71).

Humphries, M. (2017), 'Dynamical networks: finding, measuring, and tracking neural population activity using network science', *Dynamical networks: Finding, measuring, and tracking neural population activity using network science*. p. 115485. doi: 10.1101/115485 (pages 41, 92).

Insanally, M. N., Carcea, I., Field, R. E., Rodgers, C. C., DePasquale, B., Rajan, K., DeWeese, M. R., Albanna, B. F. and Froemke, R. C. (2018), 'Nominally non-responsive frontal and sensory cortical cells encode task-relevant variables via ensemble consensus-building', *bioRxiv* p. 347617. doi: 10.1101/347617 (pages 19, 29).

Izhikevich, E. M. (2006), 'Polychronization: Computation with Spikes', *Neural Computation* **18**(2), 245–282. doi: 10.1162/089976606775093882 (page 34).

Jones, K. E., Campbell, P. K. and Normann, R. A. (1992), 'A glass/silicon composite intracortical electrode array', *Annals of Biomedical Engineering* **20**(4), 423–437. doi: 10.1007/BF02368134 (page 18).

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S. and Saul, L. K. (1999), 'Introduction to variational methods for graphical models', *Machine Learning* **37**(2), 183–233. doi: 10.1023/A:1007665907178 (page 62).

Juavinett, A. L., Bekheet, G. and Churchland, A. K. (2019), 'Chronically implanted neuropixels probes enable high-yield recordings in freely moving mice', *eLife* **8**. doi: 10.7554/eLife.47188 (pages 19, 33, 106, 119, 126, 135).

Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., Lee, A. K., Anastassiou, C. A., Andrei, A., Aydin, Ç., Barbic, M., Blanche, T. J., Bonin, V., Couto, J., Dutta, B., Gratiy, S. L., Gutnisky, D. A., Häusser, M., Karsh, B., Ledochowitsch, P., Lopez, C. M., Mitelut, C., Musa, S., Okun, M., Pachitariu, M., Putzeys, J., Rich, P. D., Rossant, C., Sun, W. L.,

Svoboda, K., Carandini, M., Harris, K. D., Koch, C., O'Keefe, J. and Harris, T. D. (2017), 'Fully integrated silicon probes for high-density recording of neural activity', *Nature* **551**(7679), 232–236. doi: 10.1038/nature24636 (page 107).

Kandel Eric R., Schwartz James H., Jessell Thomas M., Siegelbaum Steven A., H. A. J. (2013), *Principles of Neural Science Fifth Edition*, McGraw-Hill. doi: 10.1017/CBO9781107415324.004 (page 27).

Kaufman, M. T., Churchland, M. M., Ryu, S. I. and Shenoy, K. V. (2014), 'Cortical activity in the null space: permitting preparation without movement', *Nature Neuroscience* **17**(3), 440–448. doi: 10.1038/nn.3643 (page 73).

Kim, S. S., Rouault, H., Druckmann, S. and Jayaraman, V. (2017), 'Ring attractor dynamics in the Drosophila central brain', *Science* **356**(6340), 849–853. doi: 10.1126/science.aal4835 (page 28).

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C. E., Kepecs, A., Mainen, Z. F., Qi, X.-L., Romo, R., Uchida, N. and Machens, C. K. (2016), 'Demixed principal component analysis of neural population data', *eLife* **5**. doi: 10.7554/eLife.10989 (page 73).

König, P., Engel, A. K. and Singer, W. (1996), 'Integrator or coincidence detector? The role of the cortical neuron revisited', *Trends in Neurosciences* **19**(4), 130–137. doi: 10.1016/S0166-2236(96)80019-1 (pages 30, 37).

Koyejo, O., Patel, P., Ghosh, J. and Poldrack, R. A. (2013), Learning predictive cognitive structure from fMRI using supervised topic models, *in* '2013 3rd International Workshop on Pattern Recognition in Neuroimaging, PRNI 2013', pp. 9–12. doi: 10.1109/PRNI.2013.12 (pages 48, 59, 70).

Kreiman, G., Koch, C. and Fried, I. (2000), 'Category-specific visual responses of single neurons in the human medial temporal lobe', *Nature Neuroscience* **3**(9), 946–953. doi: 10.1038/78868 (pages 26, 28).

Kreuz, T., Mulansky, M. and Bozanic, N. (2015), 'SPIKY: A graphical user interface for monitoring spike train synchrony', *Journal of Neurophysiology* **113**(9), 3432–3445. doi: 10.1152/jn.00848.2014 (page 44).

Kropff, E., Carmichael, J. E., Moser, M. B. and Moser, E. I. (2015), 'Speed cells in the medial entorhinal cortex', *Nature* **523**(7561), 419–424. doi: 10.1038/nature14622 (page 28).

Lara, J. A., Lizcano, D., Pérez, A. and Valente, J. P. (2014), 'A general framework for time series data mining based on event analysis: Application to the medical domains of electroencephalography and stabilometry', *Journal of Biomedical Informatics* **51**, 219–241. doi: 10.1016/j.jbi.2014.06.003 (pages 59, 70).

Leutgeb, S., Leutgeb, J. K., Barnes, C. A., Moser, E. I., McNaughton, B. L. and Moser, M. B. (2005), 'Neuroscience: Independent codes for spatial and episodic memory in hippocampal neuronal ensembles', *Science* **309**(5734), 619–623. doi: 10.1126/science.1114037 (page 31).

Li, W. and Mccallum, A. (2006), Pachinko Allocation: DAG-Structured Mixture Models of Topic Correlations, *in* 'Proceedings of the 23rd International Con ference on Machine Learning'. (page 140).

Lisman, J. and Spruston, N. (2005), 'Postsynaptic depolarization requirements for LTP and LTD: a critique of spike timing-dependent plasticity', *Nature Neuroscience* **8**(7), 839–841. doi: 10.1038/nn0705-839 (page 149).

Liu, K., Sibille, J. and Dragoi, G. (2019), 'Preconfigured patterns are the primary driver of offline multi-neuronal sequence replay', *Hippocampus* **29**(3), 275–283. doi: 10.1002/hipo.23034 (page 35).

Liu, L., Tang, L., Dong, W., Yao, S. and Zhou, W. (2016), 'An overview of topic modeling and its current applications in bioinformatics', *SpringerPlus* **5**(1). doi: 10.1186/s40064-016-3252-8 (pages 21, 48, 62, 70, 72, 73, 140).

Lopes-dos Santos, V., Conde-Ocazionez, S., Nicolelis, M. A., Ribeiro, S. T. and Tort, A. B. (2011), 'Neuronal assembly detection and cell membership specification by principal component analysis', *PLoS ONE* **6**(6). doi: 10.1371/journal.pone.0020996 (pages 20, 28, 34, 39, 40, 50, 53, 69, 71, 102, 103, 124).

Lopes-dos Santos, V., Ribeiro, S. and Tort, A. B. (2013), 'Detecting cell assemblies in large neuronal populations', *Journal of Neuroscience Methods* **220**(2), 149–166. doi: 10.1016/j.jneumeth.2013.04.010 (pages 34, 38, 39, 40, 43, 53, 69, 71, 72, 92, 95, 100, 101, 103).

Lopez, C. M., Mitra, S., Putzeys, J., Raducanu, B., Ballini, M., Andrei, A., Severi, S., Welkenhuysen, M., Van Hoof, C., Musa, S. and Yazicioglu, R. F. (2016), Paper 22.7 A 966-electrode neural probe with 384 configurable channels in 0.13um SOI CMOS, *in* 'IEEE International Solid-State Circuits Conference', IEEE. doi: 10.1109/ISSCC.2016.7418072 (pages 18, 20, 26, 31, 106, 147).

López-Madrona, V. J., Pérez-Montoyo, E., Álvarez-Salvado, E., Moratal, D., Herreras, O., Pereda, E., Mirasso, C. R. and Canals, S. (2020), 'Different theta frameworks coexist in the rat hippocampus and are coordinated during memory-guided and novelty tasks', *eLife* **9**, 1–35. doi: 10.7554/eLife.57313 (page 139).

Luo, T. Z. and Maunsell, J. H. R. (2015), 'Neuronal Modulations in Visual Cortex Are Associated with Only One of Multiple Components of Attention.', *Neuron* **86**(5), 1182–8. doi: 10.1016/j.neuron.2015.05.007 (page 59).

Mackevicius, E. L., Bahle, A. H., Williams, A. H., Gu, S., Denissenko, N. I., Goldman, M. S. and Fee, M. S. (2018), 'Unsupervised discovery of temporal sequences in high-dimensional datasets, with applications to neuroscience', *bioRxiv* p. 273128. doi: 10.1101/273128 (pages 21, 55, 71, 73, 139, 146).

Mackevicius, E. L., Best, M. D., Saal, H. P. and Bensmaia, S. J. (2012), 'Millisecond precision spike timing shapes tactile perception', *Journal of Neuroscience* **32**(44), 15309–15317. doi: 10.1523/JNEUROSCI.2161-12.2012 (pages 29, 36).

Mainen, Z. F. and Seinowski, T. J. (1995), 'Reliability of spike timing in neocortical neurons', *Science* **268**(5216), 1503–1506. doi: 10.1126/science.7770778 (page 29).

Malvache, A., Reichinnek, S., Villette, V., Haimerl, C. and Cossart, R. (2016), 'Awake hippocampal reactivations project onto orthogonal neuronal assemblies', *Science* **353**(6305), 1280–1283. doi: 10.1126/science.aaf3319 (page 38).

Mante, V., Sussillo, D., Shenoy, K. V. and Newsome, W. T. (2013), 'Context-dependent computation by recurrent dynamics in prefrontal cortex', *Nature* **503**(7474), 78–84. doi: 10.1038/nature12742 (page 73).

Marg, E. and Adams, J. E. (1967), 'Indwelling multiple micro-electrodes in the brain', *Electroencephalography and Clinical Neurophysiology* **23**(3), 277–280. doi: 10.1016/0013-4694(67)90126-5 (page 18).

Masquelier, T., Guyonneau, R. and Thorpe, S. J. (2008), 'Spike timing dependent plasticity finds the start of repeating patterns in continuous spike trains', *PLoS ONE* **3**(1). doi: 10.1371/journal.pone.0001377 (pages 22, 37, 55, 71, 145, 146, 148, 149, 153, 156, 159).

Mazor, O. and Laurent, G. (2005), 'Transient dynamics versus fixed points in odor representations by locust antennal lobe projection neurons.', *Neuron* **48**(4), 661–73. doi: 10.1016/j.neuron.2005.09.032 (page 73).

Meliza, C. D. and Dan, Y. (2006), 'Receptive-field modification in rat visual cortex induced by paired visual stimulation and single-cell spiking', *Neuron* **49**(2), 183–189. doi: 10.1016/j.neuron.2005.12.009 (page 149).

Merzenich, M. M., Knight, P. L. and Roth, G. L. (1975), 'Representation of cochlea within primary auditory cortex in the cat', *Journal of Neurophysiology* **38**(2), 231–249. doi: 10.1152/jn.1975.38.2.231 (page 27).

Miller, J. E. K., Ayzenshtat, I., Carrillo-Reid, L. and Yuste, R. (2014), 'Visual stimuli recruit intrinsically generated cortical ensembles', *Proceedings of the National Academy of Sciences of the United States of America* **111**(38), E4053–E4061. doi: 10.1073/pnas.1406077111 (pages 40, 42).

Mishne, G., Talmon, R., Meir, R., Schiller, J., Lavzin, M., Dubin, U. and Coifman, R. R. (2016), 'Hierarchical Coupled-Geometry Analysis for Neuronal Structure and Activity Pattern Discovery', *IEEE Journal on Selected Topics in Signal Processing* **10**(7), 1238–1253. doi: 10.1109/JSTSP.2016.2602061 (page 146).

Mölter, J., Avitan, L. and Goodhill, G. J. (2018), 'Detecting neural assemblies in calcium imaging data', *BMC Biology* **16**(1), 1–20. doi: 10.1186/s12915-018-0606-4 (pages 20, 39, 40, 41, 43, 50, 66, 69, 70, 72, 79, 80, 83, 86, 96, 101, 102, 103, 104, 135, 137).

Montangie, L., Miehl, C. and Gjorgjieva, J. (2020), 'Autonomous emergence of connectivity assemblies via spike triplet interactions', *PLoS Computational Biology* **16**(5), e1007835. doi: 10.1371/journal.pcbi.1007835 (page 38).

Moore, G. P., Perkel, D. H. and Segundo, J. P. (1966), 'Statistical Analysis and Functional Interpretation of Neuronal Spike Data', *Annual Review of Physiology* **28**(1), 493–522. doi: 10.1146/annurev.ph.28.030166.002425 (page 39).

Morrison, A., Diesmann, M. and Gerstner, W. (2008), 'Phenomenological models of synaptic plasticity based on spike timing', *Biological Cybernetics* **98**(6), 459–478. doi: 10.1007/s00422-008-0233-1 (page 149).

Munn, R. G. and Giocomo, L. M. (2020), 'Multiple head direction signals within entorhinal cortex: origin and function', *Current Opinion in Neurobiology* **64**, 32–40. doi: 10.1016/j.conb.2020.01.015 (pages 28, 31).

Musall, S., Kaufman, M. T., Juavinett, A. L., Gluf, S. and Churchland, A. K. (2019), 'Single-trial neural dynamics are dominated by richly varied movements', *Nature Neuroscience* **22**(10), 1677–1686. doi: 10.1038/s41593-019-0502-4 (pages 21, 33, 38, 44, 45, 126).

Naud, R. and Sprekeler, H. (2018), 'Sparse bursts optimize information transmission in a multiplexed neural code', *Proceedings of the National Academy of Sciences of the United States of America* **115**(27), E6329–E6338. doi: 10.1073/pnas.1720995115 (pages 29, 30, 31, 138).

Nawrot, M. P. (2010), Analysis and Interpretation of Interval and Count Variability in Neural Spike Trains, *in* 'Analysis of Parallel Spike Trains', Springer, chapter 3. doi: 10.1007/978-1-4419-5675-0_3 (pages 29, 138).

O'Keefe, J. and Dostrovsky, J. (1971), 'The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat', *Brain Research* **34**(1), 171–175. doi: 10.1016/0006-8993(71)90358-1 (pages 28, 29).

Oostenveld, R., Fries, P., Maris, E. and Schoffelen, J. M. (2011), 'FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data', *Computational Intelligence and Neuroscience* **2011**. doi: 10.1155/2011/156869 (page 44).

Ozden, I., Lee, H. M., Sullivan, M. R. and Wang, S. S. (2008), 'Identification and clustering of event patterns from in vivo multiphoton optical recordings of neuronal ensembles', *Journal of Neurophysiology* **100**(1), 495–503. doi: 10.1152/jn.01310.2007 (pages 52, 139).

Pan, X.-Y., Zhang, Y.-N. and Shen, H.-B. (2010), 'Large-Scale Prediction of Human Protein-Protein Interactions from Amino Acid Sequence Based on Latent Topic Features', *Journal of Proteome Research* **9**(10), 4992–5001. doi: 10.1021/pr100618t (page 59).

Panzeri, S., Brunel, N., Logothetis, N. K. and Kayser, C. (2010), 'Sensory neural codes using multiplexed temporal scales', *Trends in Neurosciences* **33**(3), 111–120. doi: 10.1016/j.tins.2009.12.001 (pages 20, 28, 69, 70, 110).

Panzeri, S., Petersen, R. S., Schultz, S. R., Lebedev, M. and Diamond, M. E. (2001), 'The role of spike timing in the coding of stimulus location in rat somatosensory cortex', *Neuron* **29**(3), 769–777. doi: 10.1016/S0896-6273(01)00251-3 (page 29).

Partha, M. and Bokil, H. (2008), *Observed Brain Dynamics*, Oxford University Press, New York. (page 44).

Pastalkova, E., Itskov, V., Amarasingham, A. and Buzsáki, G. (2008), 'Internally generated cell assembly sequences in the rat hippocampus', *Science* **321**(5894), 1322–1327. doi: 10.1126/science.1159775 (page 35).

Pastore, V. P., Massobrio, P., Godjoski, A. and Martinoia, S. (2018), 'Identification of excitatory-inhibitory links and network topology in large-scale neuronal assemblies from multi-electrode recordings', *PLoS Computational Biology* **14**(8), 1–25. doi: 10.1371/journal.pcbi.1006381 (pages 29, 40, 42).

Paul, M. J. and Dredze, M. (2014), 'Discovering Health Topics in Social Media Using Topic Models', *PLoS ONE* **9**(8), e103408. doi: 10.1371/journal.pone.0103408 (page 59).

Pauli, R., Weidel, P., Kunkel, S. and Morrison, A. (2018), 'Reproducing Polychronization: A Guide to Maximizing the Reproducibility of Spiking Network Models', *Frontiers in Neuroinformatics* **12**, 46. doi: 10.3389/fninf.2018.00046 (page 35).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, É. (2011), 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research* **12**(Oct), 2825–2830. (page 62).

Perez-Orive, J., Bazhenov, M. and Laurent, G. (2004), 'Intrinsic and circuit properties favor coincidence detection for decoding oscillatory input.', *Journal of Neuroscience* **24**(26), 6037–47. doi: 10.1523/JNEUROSCI.1084-04.2004 (page 55).

Perkel, D. H., Gerstein, G. L. and Moore, G. P. (1967), 'Neuronal Spike Trains and Stochastic Point Processes: II. Simultaneous Spike Trains', *Biophysical Journal* **7**(4), 419–440. doi: 10.1016/S0006-3495(67)86597-4 (pages 29, 101).

Peyrache, A., Benchenane, K., Khamassi, M., Wiener, S. I. and Battaglia, F. P. (2010), 'Principal component analysis of ensemble recordings reveals cell assemblies at high temporal resolution', *Journal of Computational Neuroscience* **29**(1-2), 309–325. doi: 10.1007/s10827-009-0154-6 (pages 39, 40, 42, 72, 92, 96, 100, 103, 126).

Pfarr, S., Schaaf, L., Reinert, J. K., Paul, E., Herrmannsdörfer, F., Roßmanith, M., Kuner, T., Hansson, A. C., Spanagel, R., Körber, C. and Sommer, W. H. (2018), 'Choice for drug or natural reward engages largely overlapping

neuronal ensembles in the infralimbic prefrontal cortex', *Journal of Neuroscience* **38**(14), 3507–3519. doi: 10.1523/JNEUROSCI.0026-18.2018 (page 43).

Picado Muiño, D., Borgelt, C., Berger, D., Denker, M., Gerstein, G. L. and Grün, S. (2013), 'Finding neural assemblies with frequenct item set mining', *Frontiers in Neuroinformatics* **7**(9). (pages 40, 100, 101).

Pirschel, F. and Kretzberg, J. (2016), 'Multiplexed population coding of stimulus properties by leech mechanosensory cells', *Journal of Neuroscience* **36**(13), 3636–3647. doi: 10.1523/JNEUROSCI.1753-15.2016 (page 30).

Pons, P. and Latapy, M. (2006), 'Computing communities in large networks using random walks', *Journal of Graph Algorithms and Applications* **10**(2), 191–218. doi: 10.7155/jgaa.00124 (page 141).

Pröpper, R. and Obermayer, K. (2013), 'Spyke Viewer: a flexible and extensible platform for electrophysiological data analysis', *Frontiers in Neuroinformatics* **7**. doi: 10.3389/fninf.2013.00026 (page 44).

Pruszynski, J. A. and Johansson, R. S. (2014), 'Edge-orientation processing in first-order tactile neurons', *Nature Neuroscience* **17**(10), 1404–1409. doi: 10.1038/nn.3804 (page 29).

Pruszynski, J. A. and Zylberberg, J. (2019), 'The language of the brain: real-world neural population codes', *Current Opinion in Neurobiology* **58**, 30–36. doi: 10.1016/j.conb.2019.06.005 (pages 28, 29, 30).

Prut, Y., Vaadia, E., Bergman, H., Haalman, I., Slovin, H. and Abeles, M. (1998), 'Spatiotemporal structure of cortical activity: properties and behavioral relevance.', *Journal of Neurophysiology* **79**(6), 2857–2874. doi: 10.1126/science.1529342 (pages 20, 31, 34, 36, 50, 55, 71, 145, 146).

Quaglio, P., Yegenoglu, A., Torre, E., Endres, D. M. and Grün, S. (2017), 'Detection and Evaluation of Spatio-Temporal Spike Patterns in Massively Parallel Spike Train Data with SPADE', *Frontiers in Computational Neuroscience* **11**, 41. doi: 10.3389/fncom.2017.00041 (pages 20, 26, 31, 33, 34, 35, 36, 40, 42, 50, 55, 71, 139, 146).

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C. and Fried, I. (2005), 'Invariant visual representation by single neurons in the human brain', *Nature* **435**(7045), 1102–1107. doi: 10.1038/nature03687 (page 28).

Reinagel, P. and Reid, R. C. (2000), 'Temporal coding of visual information in the thalamus', *Journal of Neuroscience* **20**(14), 5392–5400. doi: 10.1523/jneurosci.20-14-05392.2000 (pages 29, 30, 42, 138).

Renshaw, B., Forbes, A. and Morison, B. R. (1940), 'Activity of Isocortex and Hippocampus: Electrical Studies with Micro-electrodes', *Journal of Neurophysiology* **3**(1), 74–105. doi: 10.1152/jn.1940.3.1.74 (pages 17, 26, 27).

Riehle, A., Grün, S., Diesmann, M. and Aertsen, A. (1997), 'Spike synchronization and rate modulation differentially involved in motor cortical function', *Science* **278**(5345), 1950–1953. doi: 10.1126/science.278.5345.1950 (pages 34, 55).

Rogers, S., Girolami, M., Campbell, C. and Breitling, R. (2005), 'The latent process decomposition of cDNA microarray data sets', *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **2**(2), 143–156. doi: 10.1109/TCBB.2005.29 (page 59).

Romano, S. A., Pietri, T., Pérez-Schuster, V., Jouary, A., Haudrechy, M. and Sumbre, G. (2015), 'Spontaneous neuronal network dynamics reveal circuit's functional adaptations for behavior', *Neuron* **85**(5), 1070–1085. doi: 10.1016/j.neuron.2015.01.027 (pages 38, 39, 43, 72, 73, 92, 96, 102, 103, 126).

Roudi, Y., Dunn, B. and Hertz, J. (2015), 'Multi-neuronal activity and functional connectivity in cell assemblies', *Current Opinion in Neurobiology* **32**, 38–44. doi: 10.1016/j.conb.2014.10.011 (pages 32, 33).

Roy, S. A. and Alloway, K. D. (2001), 'Coincidence detection or temporal integration? What the neurons in somatosensory cortex are doing', *Journal of Neuroscience* **21**(7), 2462–2473. doi: 21/7/2462 [pii] (pages 20, 34, 37, 45).

Russo, E. and Durstewitz, D. (2017), 'Cell assemblies at multiple time scales with arbitrary lag constellations', *eLife* **6**. doi: 10.7554/eLife.19428 (page 38).

Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B. L., Witter, M. P., Moser, M. B. and Moser, E. I. (2006), 'Conjunctive representation of position, direction, and velocity in entorhinal cortex', *Science* **312**(5774), 758–762. doi: 10.1126/science.1125572 (page 28).

Sasaki, T. (2019), 'Non-structured spike sequences of hippocampal neuronal ensembles in awake animals', *Neuroscience Research* **142**, 1–6. doi: 10.1016/j.neures.2018.05.005 (pages 35, 126, 135).

See, J. Z., Atencio, C. A., Sohal, V. S. and Schreiner, C. E. (2018), 'Coordinated neuronal ensembles in primary auditory cortical columns', *eLife* **7**, 1–33. doi: 10.7554/eLife.35587 (pages 80, 92).

Sheng, M., Lu, D., ming Shen, Z. and ming Poo, M. (2019), 'Emergence of stable striatal D1R and D2R neuronal ensembles with distinct firing sequence during motor learning', *Proceedings of the National Academy of Sciences of the United States of America* **166**(22), 11038–11047. doi: 10.1073/pnas.1901712116 (pages 20, 31, 35, 43, 143).

Shimazaki, H., ichi Amari, S., Brown, E. N. and Grün, S. (2012), 'State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data', *PLoS Computational Biology* **8**(3), 1002385. doi: 10.1371/journal.pcbi.1002385 (page 36).

Shouval, H. Z., Bear, M. F. and Cooper, L. N. (2002), 'A unified model of NMDA receptor-dependent bidirectional synaptic plasticity', *Proceedings of the National Academy of Sciences* **99**(16), 10831–10836. doi: 10.1073/pnas.152343099 (page 149).

Smith, R. J., Soares, A. B., Rouse, A. G., Schieber, M. H. and Thakor, N. V. (2018), 'Modeling Task-specific Neuronal Ensembles Improves Decoding of Grasp', *Journal of Neural Engineering* **15**(3). doi: 10.1088/1741-2552/aaac93.Modeling (pages 31, 35, 77, 101, 162).

Solstad, T., Boccara, C. N., Kropff, E., Moser, M. B. and Moser, E. I. (2008), 'Representation of geometric borders in the entorhinal cortex', *Science* **322**(5909), 1865–1868. doi: 10.1126/science.1166466 (page 28).

Srivastava, K. H., Holmes, C. M., Vellema, M., Pack, A. R., Elemans, C. P., Nemenman, I. and Sober, S. J. (2017), 'Motor control by precisely timed spike patterns', *Proceedings of the National Academy of Sciences of the United States of America* **114**(5), 1171–1176. doi: 10.1073/pnas.1611734114 (page 29).

Staude, B. (2010), 'Higher-order correlations in non-stationary parallel spike trains: statistical modeling and inference', *Frontiers in Computational Neuroscience* . doi: 10.3389/fncom.2010.00016 (page 36).

Stevenson, I. H. (2015), 'github repository: simple script for plotting advances in neural recording using seaborn'. (accessed on: 2020-09-12) (page 25).

Steyvers, M. and Griffiths, T. (2007), Probabilistic Topic Models, *in* 'Latent Semantic Analysis: A Road to Meaning', Laurence Erlbaum. (page 62).

Stimberg, M., Goodman, D. F. M., Benichoux, V. and Brette, R. (2014), 'Equation-oriented specification of neural models for simulations', *Frontiers in Neuroinformatics* **8**, 6. doi: 10.3389/fninf.2014.00006 (page 146).

Strangman, G. (1997), 'Detecting Synchronous Cell Assemblies with Limited Data and Overlapping Assemblies', *Neural Computation* **9**(1), 51–76. doi: https://doi.org/10.1162/neco.1997.9.1.51 (pages 39, 40).

Stringer, C., Pachitariu, M., Steinmetz, N., Reddy, C. B., Carandini, M. and Harris, K. D. (2018), 'Spontaneous behaviors drive multidimensional, brain-wide population activity', *bioRxiv* p. 306019. doi: 10.1101/306019 (pages 20, 21, 26, 30, 44, 45, 52, 69, 71, 72, 90, 99, 100, 107, 119, 126).

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. and Leahy, R. M. (2011), 'Brainstorm: A user-friendly application for MEG/EEG analysis', *Computational Intelligence and Neuroscience* **2011**. doi: 10.1155/2011/879716 (page 44).

Tanaka, K. (1996), 'Inferotemporal Cortex and Object Vision', *Annual Review of Neuroscience* **19**(1), 109–139. doi: 10.1146/annurev.ne.19.030196.000545 (page 28).

Torre, E., Canova, C., Denker, M., Gerstein, G., Helias, M. and Grün, S. (2016), 'ASSET: Analysis of Sequences of Synchronous Events in Massively Parallel Spike Trains', *PLoS Computational Biology* **12**(7). doi: 10.1371/journal.pcbi.1004939 (pages 55, 71, 139).

Torre, E., Picado-Muiño, D., Denker, M., Borgelt, C. and Grün, S. (2013), 'Statistical evaluation of synchronous spike patterns extracted by frequent item set mining', *Frontiers in Computational Neuroscience* **7**(OCT). doi: 10.3389/fncom.2013.00132 (page 36).

Torre, E., Quaglio, P., Denker, M., Brochier, T., Riehle, A. and Grün, S. (2016), 'Synchronous spike patterns in macaque motor cortex during an instructed-delay reach-to-grasp task', *Journal of Neuroscience* **36**(32), 8329–8340. doi: 10.1523/JNEUROSCI.4375-15.2016 (pages 18, 36, 40, 145, 162).

Trouche, S., Koren, V., Doig, N. M., Ellender, T. J., El-Gaby, M., Lopes-dos Santos, V., Reeve, H. M., Perestenko, P. V., Garas, F. N., Magill, P. J., Sharott, A. and Dupret, D. (2019), 'A Hippocampus-Accumbens Tripartite Neuronal Motif Guides Appetitive Memory in Space', *Cell* **176**(6), 1393–1406.e16. doi: 10.1016/j.cell.2018.12.037 (page 32).

Tsitoura, C., Malinowski, S., Mohrhardt, J., Degen, R., DiBenedictis, B. T., Gao, Y., Watznauer, K., Gerhold, K., Nagel, M., Weber, M., Rothermel, M., Hanganu-Opatz, I. L., Ben-Shaul, Y., Davison, I. and Spehr, M. (2020), 'Synchronous infra-slow oscillations organize ensembles of accessory olfactory bulb projection neurons into distinct microcircuits', *Journal of Neuroscience* pp. JN–RM–2925–19. doi: 10.1523/jneurosci.2925-19.2020 (page 38).

van der Meij, R. and Voytek, B. (2018), 'Uncovering Neuronal Networks Defined by Consistent between-Neuron Spike Timing from Neuronal Spike Recordings', *Eneuro* **5**, ENEURO.0379–17.2018. doi: 10.1523/ENEURO.0379-17.2018 (pages 20, 21, 30, 71, 146).

Van Rossum, M. C. W. (2001), 'A novel spike distance', *Neural Computation* **13**(4), 751–763. doi: 10.1162/089976601300014321 (page 30).

Victor, J. D. and Purpura, K. P. (1996), 'Nature and precision of temporal coding in visual cortex: a metric-space analysis', *Journal of Neurophysiology* **76**(2), 1310–1326. doi: 10.1152/jn.1996.76.2.1310 (pages 30, 138).

Von Luxburg, U. (2007), 'A tutorial on spectral clustering', *Statistics and Computing* **17**(4), 395–416. doi: 10.1007/s11222-007-9033-z (page 102).

Wainwright, M. J. and Jordan, M. I. (2008), 'Graphical models, exponential families, and variational inference', *Foundations and Trends in Machine Learning* **1**(1-2), 1–305. doi: 10.1561/2200000001 (page 62).

Wang, Y., Romani, S., Lustig, B., Leonardo, A. and Pastalkova, E. (2015), 'Theta sequences are essential for internally generated hippocampal firing fields', *Nature Neuroscience* **18**(2), 282–288. doi: 10.1038/nn.3904 (page 35).

Watanabe, K., Haga, T., Tatsuno, M., Euston, D. R. and Fukai, T. (2019), 'Unsupervised detection of cell-assembly sequences by similarity-based clustering', *Frontiers in Neuroinformatics* **13**(May), 1–19. doi: 10.3389/fninf.2019.00039 (page 146).

White, R. J. (2017), Using Topic Models to Detect Behaviour Patterns for Healthcare Monitoring, Phd thesis, University of Reading. (page 59).

Wilson, M. A. and McNaughton, B. L. (1993), 'Dynamics of the hippocampal ensemble code for space', *Science* **261**(5124), 1055–1058. doi: 10.1126/science.8351520 (pages 28, 33).

Wilson, R. I. (2013), 'Early olfactory processing in drosophila: Mechanisms and principles', *Annual Review of Neuroscience* **36**, 217–241. doi: 10.1146/annurev-neuro-062111-150533 (page 28).

Worrell, G. A., Jerbi, K., Kobayashi, K., Lina, J. M., Zelmann, R. and Le Van Quyen, M. (2012), 'Recording and analysis techniques for high-frequency oscillations', *Progress in Neurobiology* **98**(3), 265–278. doi: 10.1016/j.pneurobio.2012.02.006 (page 56).

Yamins, D. L., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D. and DiCarlo, J. J. (2014), 'Performance-optimized hierarchical models predict neural responses in higher visual cortex', *Proceedings of the National Academy of Sciences of the United States of America* **111**(23), 8619–8624. doi: 10.1073/pnas.1403112111 (page 30).

Yegenoglu, A., Holstein, D., Phan, L., Denker, M., Davison, A. and Grün, S. (2015), Elephant–open-source tool for the analysis of electrophysiological data sets, *in* 'Bernstein Conference', pp. W–05. (pages 36, 44).

Young, J. M., Waleszczyk, W. J., Wang, C., Calford, M. B., Dreher, B. and Obermayer, K. (2007), 'Cortical reorganization consistent with spike timing–but not correlation-dependent plasticity', *Nature Neuroscience* **10**(7), 887–895. doi: 10.1038/nn1913 (page 149).

Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S. I., Shenoy, K. V. and Sahani, M. (2009), 'Gaussian-Process Factor Analysis for Low-Dimensional Single-Trial Analysis of Neural Population Activity', *Journal of Neurophysiology* **102**(1), 614–635. doi: 10.1152/jn.90941.2008 (page 73).

Yuste, R. (2015), 'From the neuron doctrine to neural networks', *Nature Reviews Neuroscience* **16**(8), 487–497. doi: 10.1038/nrn3962 (pages 19, 20, 26, 27, 42, 74).

Zeldenrust, F., Wadman, W. J. and Englitz, B. (2018), 'Neural Coding With Bursts — Current State and Future Perspectives', **12**(July), 1–14. doi: 10.3389/fncom.2018.00048 (page 19).

Zenke, F., Gerstner, W. and Ganguli, S. (2017), 'The temporal paradox of Hebbian learning and homeostatic plasticity', *Current Opinion in Neurobiology* **43**, 166–176. doi: 10.1016/j.conb.2017.03.015 (page 149).

Zhang, B., Dai, J. and Zhang, T. (2017), 'NeoAnalysis: a Python-based toolbox for quick electrophysiological data processing and analysis', *BioMedical Engineering OnLine* **16**(1), 129. doi: 10.1186/s12938-017-0419-7 (page 44).

Zhang, L. I., Tao, H. W., Holt, C. E., Harris, W. A. and Poo, M.-m. (1998), 'A critical window for cooperation and competition among developing

retinotectal synapses.', *Nature* **395**(6697), 37–44. doi: 10.1038/25665 (page 149).

Zylberberg, J. and Strowbridge, B. W. (2017), 'Mechanisms of Persistent Activity in Cortical Circuits: Possible Neural Substrates for Working Memory', *Annual Review of Neuroscience* **40**(1), 603–627. doi: 10.1146/annurev-neuro-070815-014006 (page 30).