Modeling population dynamics of neural circuits: from *in vitro* neural systems to structured cortical networks

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften

der Mathematisch-Naturwissenschaftlichen Fakultät
und
der Medizinischen Fakultät
Eberhard-Karls-Universität Tübingen

vorgelegt
von

Oleg Vinogradov

aus Moskau, Russland

2024

Tag der mündlichen Prüfung:	20.03.2024
Dekan der Math.-Nat. Fakultät:	Prof. Dr. Thilo Stehle
Dekan der Medizinischen Fakultät:	Prof. Dr. Bernd Pichler
1. Berichterstatter:	Jun. Prof. Dr. Anna Levina
2. Berichterstatter:	Prof. Dr. Jan Benda
Prüfungskommission:	Jun. Prof. Dr. Anna Levina
	Prof. Dr. Jan Benda
	Prof. Dr. Jakob Macke
	Prof. Dr. Martin Giese

Erklärung / Declaration:
Ich erkläre, dass ich die zur Promotion eingereichte Arbeit mit dem Titel:

" Modeling population dynamics of neural circuits:
from *in vitro* neural systems to structured cortical networks"

selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

I hereby declare that I have produced the work entitled

" Modeling population dynamics of neural circuits:
from *in vitro* neural systems to structured cortical networks"

submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

Tübingen, den ....
Datum / Date Unterschrift / Signature

Statement of contributions

This thesis was written and edited by **Oleg Vinogradov**. The first chapter of the thesis is based on collaboration with Moses and Segal labs from the Weizmann Institute of Science (WIS). The results of this project were previously published (Sukenik* , Vinogradov* et al., 2021) with the shared first author contribution of Nirit Sukenik and **Oleg Vinogradov**. The chapter at least partially recapitulates the findings in the paper. The research project was designed by **Oleg Vinogradov**, Nirit Sukenik, Elisha Moses, Anna Levina, and Menahem Segal. Anna Levina and Elisha Moses supervised the project. Experimental work is done by Nirit Sukenik, Elisha Moses, and Menahem Segal. Network modeling, fitting, and analysis of the results are done by **Oleg Vinogradov**. Data analysis is performed by me and Nirit Sukenik, Eyal Weinreb, Menahem Segal, Elisha Moses, and Anna Levina.

The second chapter is based on the collaboration with Lerche lab at the Hertie Institute for Clinical Brain Research (HIH), and the Weizmann Institute of Science (Moses lab). **Oleg Vinogradov** designed the project, analyzed the publicly available data, model fitting, analytical derivations, and designed experiments. Culture preparations were done in HIH by Betül Uysal. MEA recordings at HIH were done by **Oleg Vinogradov** and Betül Uysal. Experimental work at HIH was supervised by Holger Lerche. Culture preparation work at WIS was done by **Oleg Vinogradov**, Shlomo Ron, Eyal Weinreb, and Elisha Moses. All imaging experiments are done by **Oleg Vinogradov**. Experimental work at WIS was supervised by Elisha Moses. Victor Buendia contributed to the analytical derivation of the reduced model. Emmanouil Giannakakis contributed to the classification analysis. Anna Levina supervised all parts of the study.

The third chapter is based on the collaboration with Emmanouil Giannakakis and Victor Buendia. The results of this project were previously published as a preprint (Giannakakis et al., 2023) and will be submitted to a journal. The results in the second chapter at least partially recapitulate the findings in the preprint. Emmanouil Giannakakis and Anna Levina conceptualized the project. Emmanouil Giannakakis designed the network and performed data analysis. **Oleg Vinogradov** performed network fitting and data analysis. Victor Buendia performed analytical derivations and data analysis. Anna Levina supervised the project.

The fourth chapter is based on the collaboration with the University of Helsinki (Totah lab). The study was conceptualized by Nelson Totah. Ryo Iwai collected behavioral data, single unit, and LFP data; Dmitrii Vasilev collected behavioral data, single unit data, LFP data, and EEG data, and performed data analysis; **Oleg Vinogradov** performed data analyses, fitting of GLM and dPCA. The work is done under the guidance of Nelson Totah and Anna Levina.

Contents

Glossary

ABC Approximate Bayesian Computation. 18, 19, 44, 45, 47, 100–102, 104, 105

ACC Anterior Cingulate Cortex. 24, 26, 110, 112

adEx (also adExp) adaptive exponential integrate-and-fire neuron (Brette & Gerstner, 2005). 16

AHP Afterhyperpolarization. 39, 55, 63

AP Action Potential. 15

CR Correct Rejection trials in the Go-NoGo task. 113–115

CV Coefficient of Variation. 36, 66, 67, 73

DIV Days in vitro; usually refers to cell culture development from the day of seeding. 43, 136

FA False Alarm trials in the Go-NoGo task. 113, 115

FACS Fluorescence Activated Cell Sorting. 34

GMM Gaussian Mixture Model. 19, 121

Hopf bifurcation occurs when a stable equilibrium point of the system undergoes a transition, leading to the emergence of sustained oscillations or limit cycles. 17, 152

hPSC Human pluripotent stem cells, including embryonic and induced pluripotent stem cells. 25, 60, 62–64, 72, 74, 75, 77, 79, 80, 83, 85, 130

HT Hit trials in the Go-NoGo task. 113, 115, 116

IBI Inter-Burst Interval. 36, 38, 54, 66, 67, 69, 71, 145

iPSC Induced pluripotent stem cell. 133, 134

KNN K-Nearest Neighbors. 67, 68

LIF Leaky Integrate-And-Fire neuron. 17, 41–43, 48, 83

MAP Maximum a Posteriori. 74, 103

MEA Microelectrode array. 22, 72, 137

mEPSC miniature Excitatory Post-Synaptic Current. 54

mPSC miniature Post-Synaptic Current. 41

PFC Prefrontal Cortex. 24

PSC spontanoeus Post-Synaptic Current – measured without TTX. 37, 41, 53, 54, 77, 80

RNN Recurrent Neuronal Network. 24, 132

Saddle-node bifurcation two equilibria in dynamical systems collide and disappear (typically one stable and one unstable); after the bifurcation, only one equilibrium point remains in the system. 17, 152

SBI Simulation-Based Inference. 14, 18, 44, 69, 132

STDP Spike-Timing-Dependent Plasticity. 6, 23, 89, 93, 94, 98, 99

"Snakes", Theodor Kondra

Acknowledgements

When I began my PhD, I never anticipated that we would have to endure a global pandemic or that my home country would become a totalitarian state and start a war. People around me were often the only source of support and I could barely express all my gratitude here.

I would like to start by acknowledging my colleagues at Levina Lab. When I first joined the group, Roxana Zeraati was the only student, who worked on her lab rotation project. Over these last five years, it was inspiring to see how hardworking and brilliant she is. Next, I am also extremely lucky that I got to know and work with Tanguy Fardet. Tanguy was the first person in our group who had a lot of experience in modeling *in vitro* devices and he really helped me to shape my own research on this topic. Thanks to him I never forget to check if the light and heaters are off. But on a more serious note, I am extremely grateful for his profound influence on my own ethical values and for raising awareness regarding climate change. I am thankful to Sina Khajehabdollahi for all the stimulating discussions that we had. They often boosted my curiosity and made me very motivated to keep going. Not only my research projects but also my mental health owes a lot to Manos Giannakakis. Manos was the first person who really took the idea of collaboration in our lab very seriously. We worked on many projects together, two of which are featured in this thesis (Chapter 3 and Chapter 4). It was fun and many fun things have happened. He is also a person, who voluntarily read this thesis and this goes beyond any possible gratefulness. Victor Buendía is one of a few real theoreticians, whom I could work with without being intimidated. I never was afraid to ask stupid questions and it pushed my understanding of how the theory is done. I often envy the master students, who are supervised by Victor. I also thank Tim Schäfer for all the discussions we've had since he joined the group. Tim's kindness and ability to stay organized. I am also very grateful to all the students that I got to supervise over the years in Levina lab, I learned a lot from them. Finally, I thank the members of the Lisbon secret society for all the coffee breaks we have had during the COVID pandemic.

I am grateful to my supervisor Anna Levina. It was a great fortune to get into Anna's lab and to be one of the first PhD students there. This felt like a gamble at the time but within the very first months realized how lucky I got. Anna invested a lot of her time and resources into my scientific (and personal) development. Her personal skills helped me to manage through even the darkest years of my doctoral work. She has always been extremely supportive and this support applied to many areas of my life. I was also lucky to have an advisor who generally strives toward excellent research and inspired me to do the same, and I absolutely loved this. And I was lucky to have an advisor, who reads the same media and understands your problems.

During my PhD, I learned a lot from Elisha Moses at the Weizmann Institute of Science. I was privileged to visit his lab three times in the past five years. Thanks to Elisha, I know how to talk to physicists of all kinds, grow cell cultures, mix chemicals, degrease disk breaks, and respond to difficult questions during presentations among many other things. I am also very grateful that Elisha supported my own experimental work and gave me a lot of freedom in the lab very early on. I am beyond grateful for Elisha's academic and personal generosity.

I was extremely lucky to share a whole project with Nirit Sukenik, who over the year became a good friend. I believe we made a great research team. Nirit always had my back even when the project did not look too great.

Next, I am very grateful to other members of Elisha's lab. I would like to thank Shlomo Ron for teaching me all the basics of lab work and being always around. Eyal Weinreb, for his help with experiments and all the awesome conversations we have had. I am also happy that I got to know Nik Freche, who was always happy to share his time and discuss the project with me. I was also privileged to have worked with Menahem Segal, whose energy and passion set a remarkable example.

I would like to thank my dear friend Betül Uysal, who also ended up being an efficient collaborator. I am also grateful to Holger Lerche for supporting our work with Betül.

Next, I would like to thank Nelson Totah for being a fantastic collaborator, inspiring mentor, and one of the most passionate scientists I met. I am also grateful to my colleague and friend Dmitrii Vasiliev.

I was fortunate to have one of the best possible advisory boards. So I like to thank Jan Benda and Fabian Sinz for their excellent support and availability.

I am grateful for all the influence that Viola Priesemann had on the research that I've done in my PhD, especially in the first years. I also owe a lot to members of the Priesemann lab. First of all to Paul Spitzner, who luckily for me was also doing research on *in vitro* systems. We even organized a workshop on *in vitro* systems at the Berstein conference together, which was just an amazing experience (and would not have happened without Paul's organizational skills). Johannes Zierenberg helped me a lot over these five years and encouraged my theoretical research. I would never be able to do integration on a napkin in a bar as well as he does. Speaking of theory, I was also lucky to get to know Farzada Farhooi and Tilo Schwalger, with whom we had occasional conversations about my work and who directly and indirectly shaped my theoretical work. I am also grateful to Alexander Ecker, who managed to be an amazing mentor through short conversations at conferences.

I was lucky to get to know Jakob Macke and to be just a floor away from his lab. In this

thesis, I used a lot of methods and software, developed in his group. I am grateful to Richard Gao, Auguste Schultz, Michael Deistler, and other members of Macke Lab. I would like to separately thank Pedro Gonçalves, who advised me a lot on how to choose a PhD position. Next, I would to thank Florian Dehlmet, who advised me a lot on many things throughout my PhD. I was lucky to tutor a course that Florian taught and thanks to that I have very high teaching standards. I am also grateful to Aristides Arrenberg, who supported this tutoring and was always very supportive of me personally.

Next, shot out to Konstantin, Lara, Max, Jen, Julia, Kate, (Betül was mentioned above), and all my master's classmates, without whom I probably wouldn't be doing a PhD in computational neuroscience in the first place. I thank Ivan Ustyuzhaninov for being an awesome person and a fantastic flatmate. Finally, I received a lot of support from the Graduate Training Centre for Neuroscience since my first day in Tuebingen. I would like to thank Horst Herbert, Katja Thieltges, Marc Himmelbach, and Monika Lam for that. I was very lucky to be a part of the vibrant neuroscience and I am grateful to researchers here, who make Tuebingen such a special place. I am grateful to all my friends here and there. I thank my father who helped me to find my way around in Germany.

I am privileged to have a large and very supportive family. I am grateful to my mom, who always supported me and whatever I came up with. I thank my sisters Anna and Liza, my uncle and aunt Alexey and Larisa, my cousins Anastasia and Ivan, and their partners and families. I would like to thank my partner's parents Julia Anatolievna and Andrey Ivanovich (aka Andrey). Finally, I am thankful to my wife Anna, and our son Theodor, whose painting is featured in this thesis. They made days brighter, and the future more certain, and were a constant source of love and inspiration. This is dedicated to them.

Summary

The collective dynamics of neuronal populations form the basis of complex physiological processes and adaptive behavior in animals. Computational modeling is one of the key tools that facilitate the understanding of population dynamics and its functions. Recent progress in machine learning and brain recording techniques allowed a close integration of experimental recordings and computational modeling. This, on the one hand, enables detailed quantitative fits of experimental data that improve our understanding of basic physiology, network organization, and variability. On the other hand, modern methods help to better characterize neuronal activity and learn the details of the computations these systems perform.

In the first part of the thesis, I demonstrate how modeling approaches and simulation-based parameter inference, integrated closely with experiments, can enhance our understanding of network organization principles. In the second part, I apply statistical modeling methods to characterize the neuronal dynamics underlying complex self-correction behavior.

Chapter 2 leverages simulation-based inference (SBI) to integrate single-cell properties, network structure, and population dynamics in a model of networks of dissociated neurons *in vitro*. This approach allowed us to discover that networks adjust the inhibitory connectivity and maintain excitation/inhibition balance under chronic changes in the cellular excitatory/inhibitory composition. **Chapter 3** identifies the dynamical states underlying the self-organization of networks of cultured neurons towards collective bursting activity. I show how a reduced model of network activity can explain the difference between networks of rodent and human pluripotent stem cell-derived neurons *in vitro* that exhibit seemingly identical dynamics. **Chapter 4** focuses on the functional implications of the excitation/inhibition connectivity in the spiking networks model and examines how recurrent network structure allows the development of excitation and inhibition balance and stimulus tuning. **Chapter 5** turns to the more complex dynamics of cortical networks underlying self-correction behavior in rats. There, statistical modeling and data analysis techniques helped us to uncover the complex organization of neural responses in the Anterior Cingulate Cortex. In the final **Chapter 6**, I discuss the main limitations of this work and the main future direction.

Chapter 1
Introduction

Neuroscience has a long history of tight integration of experiments and computational modeling (Bialek, 2018; O'Leary et al., 2015). Modeling is used to describe and understand neuronal phenomena across measurement scales from molecular and single neuron to cognitive levels. One of the most famous examples of such a tight integration is a description of action potential (AP) generation (Hodgkin & Huxley, 1952). By measuring and modeling ionic conductances in the giant squid axon[1], Hodgkin and Huxley discovered that the action potential is governed by a complex interplay of sodium and potassium and the properties of ionic channels.

To this day, the Hodgkin-Huxley-like models remain the standard way to link properties of single neurons and ion channels (McCormick et al., 2007; Ori et al., 2020) and build basic predictions on the single neuron levels. Furthermore, it inspired multiple modeling directions from simplified theories of neuronal excitability (Izhikevich, 2007; LeMasson et al., 1993) to detailed compartmental modeling (Mainen & Sejnowski, 1996; Sterratt et al., 2011). Research questions in neuroscience, however, go beyond explaining single neuron properties and span across multiple scales. Therefore, modeling approaches also vary, ranging from detailed single-neuron models to abstract models of cognitive functions (O'Leary et al., 2015).

Generally, the models can be categorized into descriptive, mechanistic, and normative (Levenstein et al., 2023; Sterratt et al., 2011). This distinction is based on what empirical phenomena are used as an entry point, and on the modeling goals[2].

Descriptive models aim at characterizing and summarizing the observed phenomena. One famous example is Weber-Fechner law (Fechner, 1860) that relates the intensity of physical

[1] As pointed out by William Bialek (Bialek, 2018), the axon's diameter is close to the diameter of a small straw, which makes it very easy to record from

[2] Note here that the word theory can also be used in this context. In this thesis, I often use both interchangeably. See, however, Levenstein et al., 2023 for a more rigorous discussion

stimuli and the perception threshold. Another illustrative example is the approximation of orientation tuning of neurons with the von Mises distribution (Berens et al., 2008).

Mechanistic models explain how exactly a phenomenon arises. Typically, these models are built on some known physical principles of the system studied. The Hodgkin-Huxley model (Hodgkin and Huxley, 1952) is an example of a mechanistic model. Sometimes mechanistic models help identify the most relevant building blocks involved in a neural process (O'Leary et al., 2015). These building blocks can be on completely different levels and are not always measurable directly. Thus, many simplified single-neuron models could be used to find a mechanistic explanation, yet include some effective parameters that can be measured only indirectly (Benda & Herz, 2003; Brette & Gerstner, 2005).

Normative models or interpretative models try to explain why a phenomenon exists at all. They are, loosely speaking, at least in some cases a way to find an Aristotelean final cause for a phenomenon (Aristotle, n.d.). For instance, theories of sampling in sensory systems (Fiser et al., 2010; Haefner et al., 2016) or theories of efficient sensory representations (Bialek et al., 2006).

Whether a certain model falls within a particular category often depends on context (Levenstein et al., 2023). For instance, an adaptive leaky integrate-and-fire neuron model (adEx, Gerstner et al., 2014) can be considered descriptive or mechanistic depending on the research questions.

The main focus of this thesis is mechanistic and descriptive modeling of neural population activity. In this chapter, I first introduce the basic concepts behind modeling network activity and parameter inference. Next, I review the systems that I focus on across this thesis: *in vitro* cultures of neurons and cortical networks *in vivo*.

1.1 Building a model of network activity

Collective dynamics of neurons emerge from single-neuron properties and connectivity between the neurons. It also can reflect the computations that a network performs. Recent progress in experimental neuroscience allows monitoring the dynamics of large populations of neurons in alive and behaving animals (Kim et al., 2017). Making sense of huge amounts of complex multivariate data (yet with a lot of missing details) and relating it to behavior, computations, and physiology is often difficult. In such cases, modeling-based approaches serve as essential tools for interpreting these intricate datasets (O'Leary et al., 2015).

Building good mechanistic models of collective dynamics remains, however, nontrivial. One of the main approaches is to use *dynamical systems* models. It is, however, not always

clear which biological details should be considered. Usually, modelers seek some compromise between the complexity and expressiveness of the model (O'Leary et al., 2015). Thus, network models based on simple single-neuron models are often used.

A classical example of a network model used to explain neural activity (N. Brunel, 2000) is a sparse random recurrent neuronal network of leaky integrate-and-fire neurons (LIF). This model takes one of the most basic single-neuron descriptions (LIF) of a neuron and assumes that neurons are connected randomly and sparsely. These simplifications allowed Brunel (2000) to analytically compute bifurcations in the model parameter space and study the dynamical repertoire of the model (Fig. 1.1). Specifically, he computed the transitions from a single fixed point to two stable and one unstable (Saddle-node bifurcation), or a transition to oscillations (Hopf bifurcation).

Dynamical states of the Brunel network have been linked to different network phenomena observed experimentally ranging from gamma-oscillations (Buzsáki & Wang, 2012) to asynchronous irregular activity (Renart et al., 2010) and up-down dynamics (Tartaglia & Brunel, 2017). Despite its limitations (Ostojic, 2014), this model is often used as a starting point in describing physiological principles of cortical dynamics and allows us to test how computation may arise in such semi-realistic settings (Litwin-Kumar & Doiron, 2012; Ostojic, 2014; T. Vogels & Abbott, 2009).

1.1.1 Data-consistent models

Mechanistic modeling of network dynamics often relies on qualitative comparisons (N. Brunel, 2000; Mongillo et al., 2018; Renart et al., 2010; T. Vogels & Abbott, 2009) and quantitative fits are rather rare (Prinz et al., 2004). Thanks to advances in machine learning techniques, we can now infer the parameters of mechanistic models and achieve good quantitative fits (Gonçalves et al., 2020). Furthermore, it is now possible to infer parameters even for rather detailed models (Tolley et al., 2023). However, exploring data-consistent model parameters remains challenging (O'Leary et al., 2015).

There are crudely two approaches to parameter inference for computational models: point estimation (i.e. maximum likelihood estimation) and Bayesian inference of the parameter posteriors[3].

Progress in automatic differentiation (Paszke et al., 2017) allows us to find the maximum likelihood for a wide range of dynamical systems. Thus, there are now gradient-based techniques

[3]There are also other methods of parameter estimation, such as evolutionary algorithms, which are beyond the scope of this thesis.

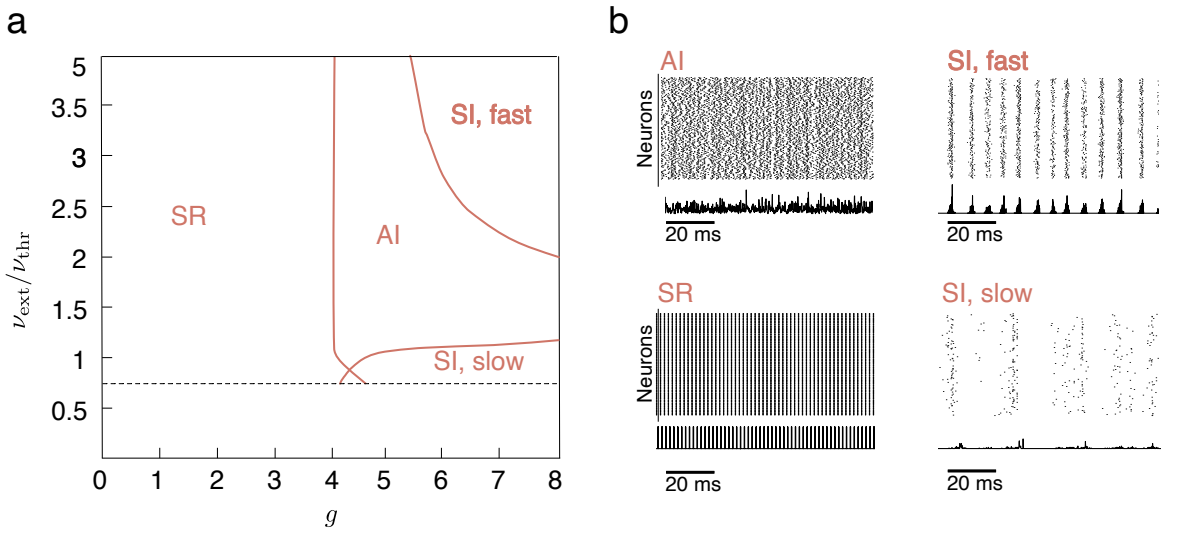


Figure 1.1: Dynamical states in the sparse random spiking network of excitatory and inhibitory neurons (N. Brunel, 2000). (**a**) Phase diagram of the network model for the relative inhibitory strength (g) and the background noise intensity relative to the firing threshold ($\nu_{\text{ext}}/\nu_{\text{thr}}$). Hopf bifurcations are colored in coral, below the dashed line the network has three types of solutions: stable high, stable low activity, or can be bistable (transitions not shown). The small triangular region between SI and SR is quasi-periodic, where both slow and fast oscillations overlap. (**b**) The model can produce asynchronous irregular spiking, synchronous irregular, and different types of synchronous irregular activity (N. Brunel, 2000).

for very complex non-linear problems even with discontinuity (Li et al., 2020; Oesterle et al., 2022; Schmiester et al., 2021). Furthermore, computing gradients through solvers of differential equations became very efficient (Chen et al., 2018; Kidger et al., 2021). There are also approximate methods like surrogate gradient descent that have been shown to efficiently find parameters for spiking networks (Huh & Sejnowski, 2017; Zenke & Vogels, 2021).

The second approach is to find the posterior distributions of parameters consistent with the data in the context of Bayesian inference. Although Bayesian inference is typically considered difficult and in a closed form only possible for rather simple models, there are powerful simulation-based approximate methods that can tackle this problem. *Simulation-based inference* (SBI) is often model agnostic and could be used for models of arbitrary complexity as long as one can simulate them (Cranmer et al., 2020; Gonçalves et al., 2020).

One of the most famous approaches is the Approximate Bayesian Computation (ABC, Sunnåker et al., 2013), which uses rejection sampling to approximate the posterior (see Chapter 2 and 4 for more details). This approach relies on a few ingredients: a prior distribution of the model parameters, a fitness function that compares the model output to the experimental data (or typically summary statistics of the data), and a threshold for this fitness function, below

which the model fits the data. Using those, one can sample parameters, run simulations, evaluate the fitness function, and keep the parameters for which the fitness is below a threshold. From the resulting samples, a posterior distribution can be approximated (Beaumont et al., 2009). Parameters sampled from this posterior should lead to data-consistent fits. The approach can be used for different models, however, in practice, requires a lot of simulations and may be limited to fit only a few parameters at a time (Lueckmann et al., 2021).

A closely related solution to the posterior inference problem is to build a surrogate model for the posterior (Cranmer et al., 2020)[4]. This approach requires a probabilistic model that can represent the conditional distribution of $p(\text{paramters}|\text{data})$ well enough. Then one needs to sample parameters for the model and run a simulation to collect enough data points to fit this probabilistic model (Fig. 1.2).

One classical solution for this problem is to use the Gaussian mixture model (GMM) linked with a feedforward neuronal network (Bishop, 1994). The feedforward network learns the mapping between the data point and the parameters of the mixture of Gaussians (the mean, covariance, and mixing coefficients). This way rather complex conditional distributions can be learned. This approach allows us to flexibly learn the model posterior and outperforms ABC (Lueckmann et al., 2017). Recently more flexible deep probabilistic models were introduced (Durkan et al., 2019; Germain et al., 2015; Papamakarios et al., 2017) as well as novel approaches to the sequential inference (Greenberg et al., 2019) that significantly boost the performance of these algorithms (Lueckmann et al., 2021). One particular advantage of the surrogate models is that they can generalize the observations such that, one can evaluate $p(\text{paramters}|\text{data})$ given novel recently unobserved data (Cranmer et al., 2020; Gonçalves et al., 2020; see also an example in Chapter 3).

1.1.2 Model sloppiness

Mechanistic models of biological systems often have degenerate parameter combinations (Gutenkunst et al., 2007). Thus, in some cases, the same model behavior can be generated by vastly different parameters. This degeneracy is called parameter *sloppiness* (Gutenkunst et al., 2007; O'Leary et al., 2015; Transtrum et al., 2015). The *sloppiness* of model parameters can both be a challenge for parameter inference and also in some cases can provide valuable insights into the system's organization.

[4]This also works for approximating a likelihood function (Papamakarios, 2019)

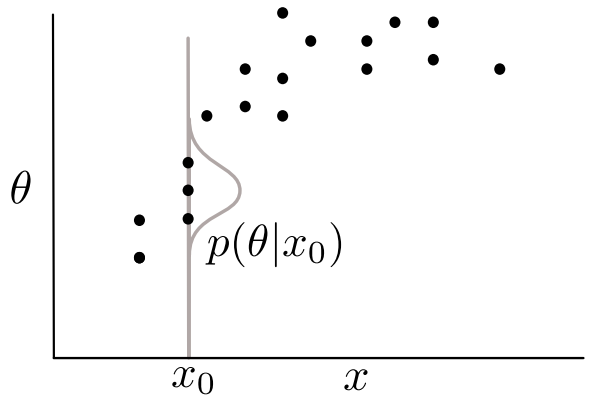


Figure 1.2: Estimating conditional probability for single data summary statistics x and a single model parameters θ. Black dots represent the results of simulating the model for with randomly sampled parameters. The approximate posterior probability $p(\theta|x)$ can be found by using probabilistic surrogate models (density estimators) such as the Mixture Density Network (Bishop, 1994). Amortized inference with the density estimators allows to evaluate $p(\theta|x_0)$ for x_0 that were not a part of the training set.

Typically if a model has many parameters there are only a few sensitive directions in the parameters space that affect the model behavior (Gutenkunst et al., 2007). These directions are not always aligned with model parameter axes. The sensitive directions are called *stiff* or non-slopy (Fig.1.3). Inferring distributions of parameters consistent with the data (e.g. posterior distributions) is one way to assess the stiffness and sloppiness in the parameter space (Gonçalves et al., 2020).

Careful examination of sensitive directions also allows us to improve the model and to find interesting directions for experimental perturbations (Fisher et al., 2013). Beyond that, finding sloppy subspaces of parameters might be informative of biological variability (O'Leary et al., 2015). A classical example of this is a model of the stomatogastric ganglion (Prinz et al., 2004). The stomatogastric ganglion (STG) is a simple circuit with a few neurons and highly stereotypical connectivity (Bal et al., 1988). The pyloric central pattern generator within STG produces a series of bursts that control the constriction of the pyloric part of the crab stomach (Harris-Warrick et al., 1992). A detailed realistic model of the pyloric network is well established, yet it includes a lot of parameters of the synaptic conductances, that are difficult to estimate directly. Prinz et al. (2004) simulated about 20 million network models of STG with various combinations of the parameters. They found that the bursting pattern could arise from a widely disparate set of parameters. Next, they discovered that this degeneracy in the model is linked to the real variability of STG between animals. Furthermore, this redundancy then could account for the differences in responses to perturbations. Interstingly, with modern methods of simulation-based inference, redundancy in parameters can be identified faster and much more precisely (Deistler et al., 2022).

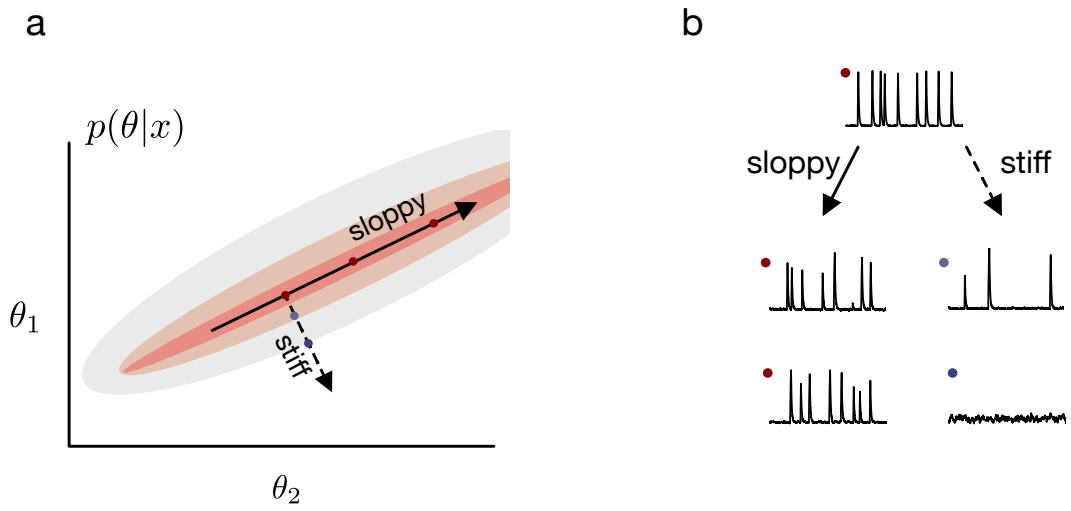


Figure 1.3: A sketch of the posterior distribution of the parameters shows the sloppy and stiff direction in the model parameter space. In a sloppy direction, a change in the parameters does not affect the resulting dynamics, whereas even a small change in the stiff direction dramatically changes the model output. The example dynamics is from a slow-fast model of the network bursting introduced in Chapter 3.

1.2 Modeling "simple" networks

The main contribution of this thesis is related to modeling and interpreting the population activity of living neuronal network *in vitro*. Neuronal systems *in vitro* are one of the most simple experimental preparations, in which neurons display collective activity. Such systems offer a high degree of experimental control and are well-suited to study the relationship between neurons, connectivities, and network activity (Soriano, 2023).

The results of Prinz et al. 2004 show that simple experimental model systems sometimes allow drawing very general conclusions about the biology of neuronal systems (see Marder and Goaillard, 2006 for a review). Thus, cultures of dissociated neurons *in vitro* have been a major vehicle for discoveries in neuroscience starting from single neuron properties, connectivity (Barral & D Reyes, 2016; Breskin et al., 2006; Okujeni et al., 2017; Soriano et al., 2008), plasticity (Abbott & Nelson, 2000; Bi & Poo, 1999; Segal, 2005), network homeostasis (O'Leary et al., 2010), excitation/inhibition balance (Barral & D Reyes, 2016; G. Liu, 2004), effects of stimulation (Wagenaar et al., 2005; Weinreb & Moses, 2022).

In vitro cultures of neurons also can be used to directly test theoretical predictions about neuronal networks. A prominent example is the scaling of synaptic strength by $1/\sqrt{\mathrm{K}}$ (where K is the number of connections) predicted by the theory of balanced networks. Barral and Reyes

(2016) used *in vitro* cultures of neurons and showed that networks of different sizes follow this rule within an experimental error.

Cultures are typically easy to manipulate and perturb (Soriano, 2023) and modern MEA and imaging techniques allow us to visualize and record large populations of neurons simultaneously (Soriano, 2023). Furthermore, advances in stem cell technologies, enable us to grow cultures of human-derived neurons, which is one of a few ways to generate human-specific conclusions about neurons and networks (Choudhary et al., 2022; Hoffmann et al., 2019b; Lotila et al., 2022; Stern et al., 2018; Uysal et al., 2019).

Progress in experimental techniques also allows us to directly engineer connectivity *in vitro* (Montalà-Flaquer et al., 2022) and build 3D cultures (Frega et al., 2014). In addition, there are multiple novel methods of growing complex brain organoids (Quadrato & Arlotta, 2017). These systems serve as an excellent middle ground between 2D cell cultures and *in vivo* networks (Trujillo et al., 2019).

By default, in many preparations, neuronal cultures produce collective activity – network bursting (see Fig. 1.5a). Network bursting is usually characterized by synchronous events that span the whole network and long irregular periods of silence in between (Maeda et al., 1995). This type of activity robustly emerges in healthy cultures. A normative theory for network bursting has not yet been developed. There are some indications that network bursting might be important for normal network development (Gu & Spitzer, 1995). But it also could be an epiphenomenon related to the absence of stimulation (Zierenberg et al., 2018).

Networks of spiking neurons are often used to model population bursting *in vitro*. Thus, coupled neurons that have a slow firing-rate adaptation mechanism can very well reproduce the statistics of bursting activity and burst shapes (Fardet et al., 2018; Giugliano et al., 2004; Orlandi et al., 2013; Segev et al., 2001; Sukenik* , Vinogradov* et al., 2021; Yamamoto et al., 2022). Note, however, although the fact that although adequate minimal models are established, there is no standardized model of bursting activity *in vitro* (Fardet, 2018 and also see discussion in Chapter 3). Nonetheless, models of network bursting allow us to combine single-neuron properties and connectivity to identify the principles of network functioning and derive useful predictions (J.-P. Eckmann et al., 2008; Giugliano et al., 2004; Orlandi et al., 2013; Sukenik* , Vinogradov* et al., 2021; Yamamoto et al., 2022).

In Chapter 2, I use population bursting to identify a model that predicts how the networks of neurons adapt to chronic change in ratios of inhibitory and excitatory cells. In Chapter 3, we show how statistics of network bursting can be used to infer network excitability and map dynamical states underlying network bursting. In both cases, we used simulation-based inferences in a close

loop with experiments and analytical techniques.

1.3 Adding functions to networks

1.3.1 Synaptic plasticity and neural tuning

Although basic features of neuronal functioning can be discovered *in vitro*, there has been only limited success in training *in vitro* networks to encode stimuli and control the external world [5] (Kagan et al., 2022; Shahaf & Marom, 2001; Wagenaar et al., 2005). Therefore, it is critical to focus on structured networks that encode information.

One of the open questions in the field is how input selectivity emerges in the output of a balanced network of excitatory and inhibitory neurons that is assumed to be a steady-state model of the cortex (Renart et al., 2010).

Synaptic plasticity is one of the key mechanisms of learning in biological networks (Abbott & Nelson, 2000; Humeau & Choquet, 2019). Modeling studies have suggested a variety of possible plasticity mechanisms ranging from basic Hebbian plasticity to more various types of spike-time dependent plasticity (STDP). The exact implementation of synaptic plasticity as a basis of learning is still a topic of active research. Inhibitory plasticity has recently received a lot of attention (Froemke, 2015; Sprekeler, 2017; Y. K. Wu et al., 2022). Since, it has been suggested that it may play a very special role in learning and memory formation (Hennequin et al., 2017; T. P. Vogels et al., 2011). Simultaneously, it can help to maintain balanced dynamics and network homeostasis (Sprekeler, 2017).

Recent studies showed that synaptic plasticity can promote neural tuning (T. P. Vogels et al., 2011). Some forms of plasticity allow hierarchical networks to learn complex tasks and reach performance close to state-of-the-art deep learning methods (Payeur et al., 2021). In Chapter 3, we show how to leverage simulation-based inference to infer recurrent network structures that a realistic read-out neuron can use to efficiently learn to encode stimuli under biological constraints (Giannakakis et al., 2023).

1.3.2 Dynamics in higher brain networks

One of the final questions that this thesis touches upon is how dynamics in higher brain areas are implicated in the control of behavior. Studying this question is often complicated because it

[5] although some of the basic synaptic plasticity mechanisms were established *in vitro* (Abbott & Nelson, 2000; Bi & Poo, 1999)

is difficult to establish a direct relationship between population dynamics, computations, and behavior. Thus, in higher brain areas such as the prefrontal cortex (PFC), neurons show very complex features and potentially implement rather sophisticated computations (Mante et al., 2013).

One approach to this problem is to use converging evidence from the mapping of neural activity and studying responses of a network model trained to perform a behavioral task (Williamson et al., 2019). This approach became possible thanks to progress in neural data (Kobak et al., 2016; Mante et al., 2013; S. Schneider et al., 2023; Sussillo et al., 2016 just to name a few examples) and progress in training RNN (Barak, 2017).

A classical example of this approach is work by Mante et al. (2013). First, using the targeted dimensionality reduction they showed how exactly neurons change their coding dimensions depending on the context. Then, by training an RNN of rate units, they showed that similar responses would emerge when a network is trained to perform a task with rules switching (Fig. 1.4).

Typically, one of the first data-driven steps in this approach is a detailed description of the variability of neural responses and their relationship to the animal's behavior. In Chapter 5, we start building a map of responses for the ACC that is involved in the complex control of animal behavior.

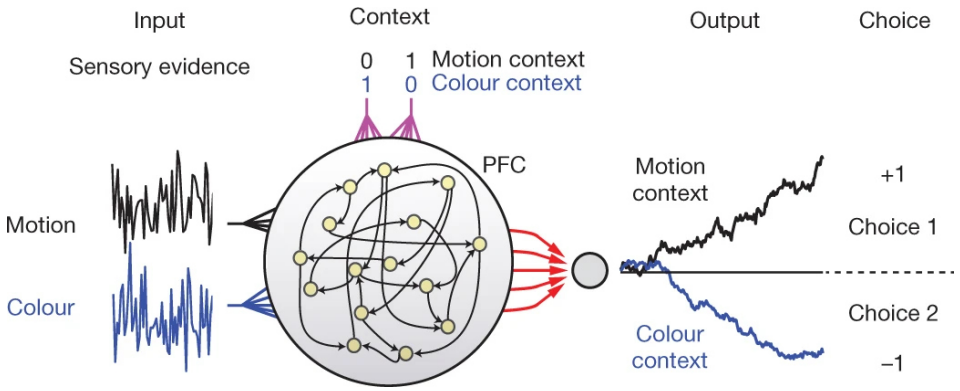


Figure 1.4: Recurrent neural network from trained to distinguish motion and color depending on the context. The model combines dynamics and computation and explains how mixed selectivity of neurons in the prefrontal cortex. Reproduced from Mante et al., 2013 with the permission of Nature Publishing Group.

1.4 Thesis outline

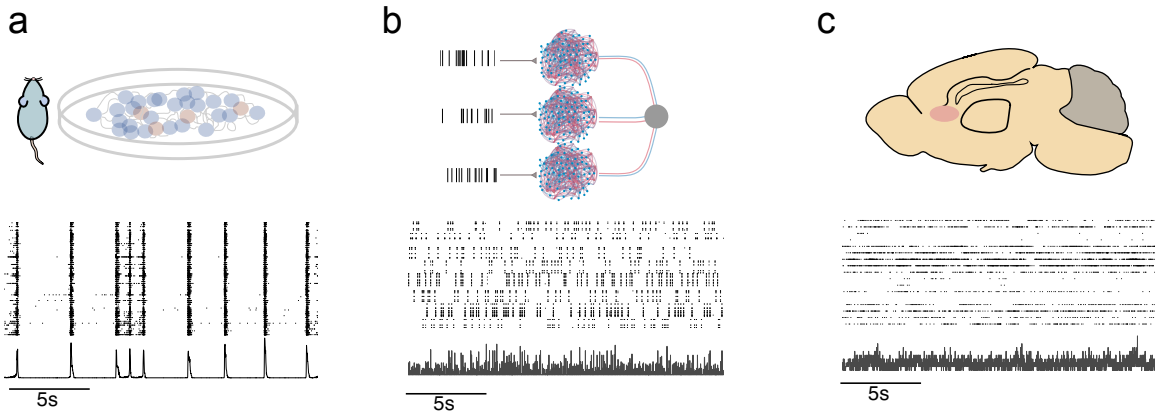


Figure 1.5: Spiking dynamics across different levels of system complexity **(a)** Bursting network activity in a network of neurons dissociated from the rat brain cortex at 21 days *in vitro* (DIV). The network exhibits collective bursting activity, a characteristic type of dynamics observed in *in vitro* systems (see 2 and 3). **(b)** Network dynamics in a recurrent spiking neuronal network model that preserved the input structure and allows for neuronal tuning on a postsynaptic neuron (see Chapter 4). **(c)** The asynchronous population activity of neurons in the medial frontal cortex in rate performing Go-NoGo task (see Chapter 5).

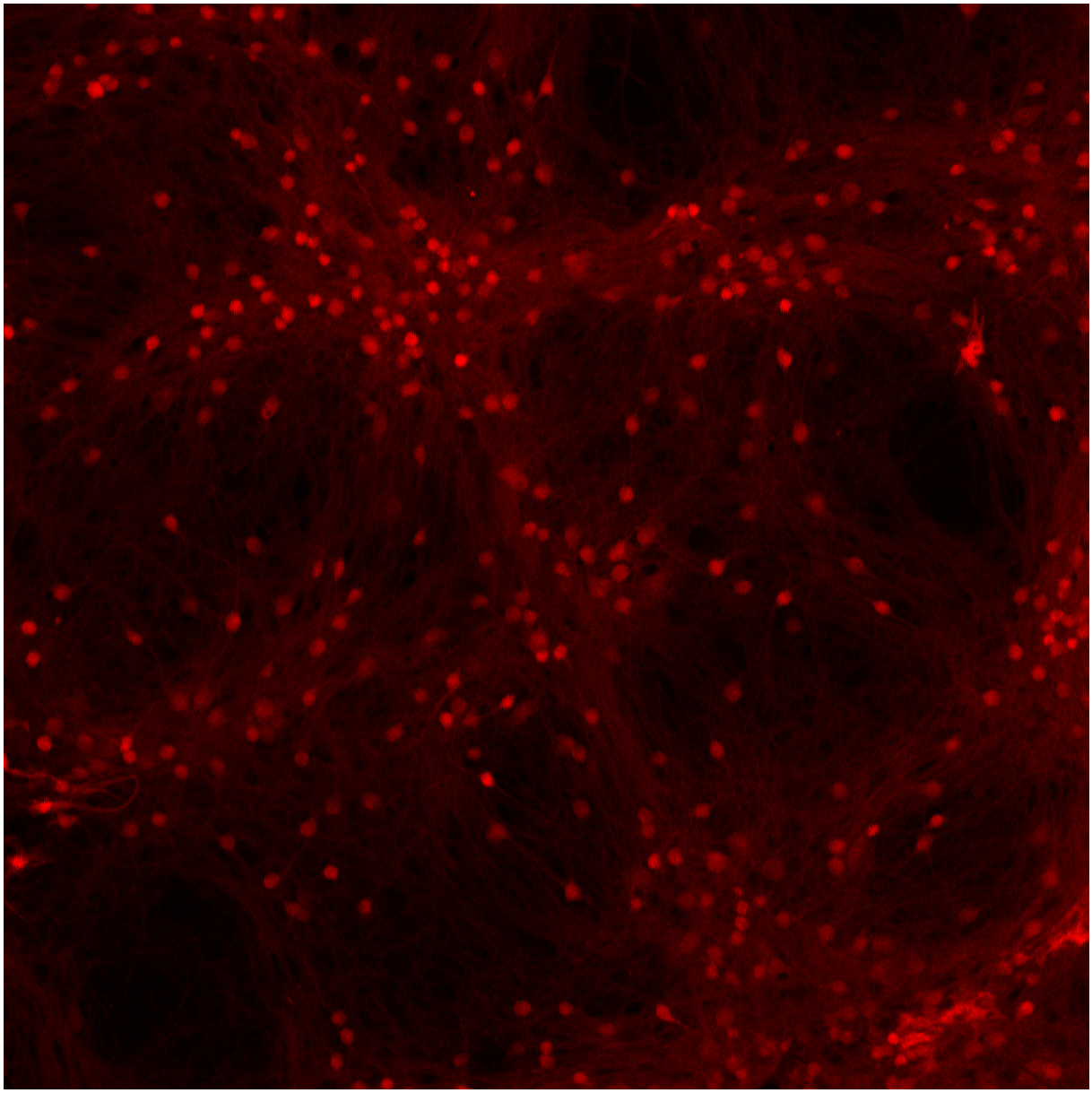
The main aim of my work was to infer mechanistic models of population neural activity and to discover the basic principles of network organization and function. I first focused on data-driven modeling of *in vitro* neural systems. In Chapter 2, we used a mechanistic model to uncover mechanisms of network adaptation to chronic perturbation in cellular composition. To give a broader context, we used a network model to fit collective dynamics in neuronal networks *in vitro* with different cellular inhibitory compositions. The model identified that keeping these dynamics should require maintenance of E/I balance and thus guided patch-clamp experiments. We used these results, in turn, to further improve the model (Sukenik* , Vinogradov* et al., 2021).

Chapter 3 identifies the general principles behind the collective dynamics of networks in vitro. In this case, we use temporal statistics of a network to fit a low-dimensional model of network activity. We map the collective dynamics of different *in vitro* preparation onto a few dynamical states. This helped us to identify the differences between primary hippocampal, and cortical as well as hPSC-derived neuronal cultures and stage their development based on the phase transitions. One of the exciting parts of this project was identifying the invariant manifolds

of this simplified model. Thus, simulation-based inference guided the discovery of the analytical description of the model.

In the following chapter, I focus on the functional implications of excitation and inhibition in recurrent neuronal networks. In Chapter 4, we explore how biologically realistic details of E/I connectivity and neuronal plasticity constrain the dynamics and structure of networks. Here we used simulation-based inference to link the low-dimensional correlation structure of the network activity to its ability to produce E/I balance and stimulus tunning.

Finally, in Chapter 4, I move toward the frontal cortex, where collective activity appears to have mixed selectivity. There, we use an array of descriptive models in an attempt to start parsing the code of ACC, while it is mediating conflict signals. Thus, we provide the first building blocks for a mechanistic understanding of conflict monitoring in ACC.



Inhibitory neurons labeled with tdTomato in a network with 80% of inhibitory cells, (Sukenik* , Vinogradov* et al., 2021)

Chapter 2
Neuronal networks self-organize to maintain the balance of excitation and inhibition

Summary

In this chapter, I focus on one of the most basic components of neuronal circuits, excitatory and inhibitory neurons. Inhibitory cell fractions in the neocortex and hippocampus are found to be around 15-30%. It is assumed that a specific inhibitory percentage plays an important role in the stability of neuronal dynamics. Using cultures of dissociated neurons with precisely controlled rations of excitatory and inhibitory cells we found that networks are highly robust to unusual cellular composition. Thus, they maintain their characteristic bursting dynamics. Using network modeling and recordings from individual cells we identified that the main mechanism behind this stability is the maintenance of excitatory and inhibitory balance.

Contributions disclaimer: The results discussed in this chapter have been published (Sukenik* , Vinogradov* et al., 2021). The chapter at least partially recapitulates the findings in the paper. The research project was designed by Oleg Vinogradov, Nirit Sukenik, Elisha Moses, Anna Levina, and Menahem Segal. Anna Levina and Elisha Moses supervised the project. Experimental work is done by Nirit Sukenik, Elisha Moses, and Menahem Segal. Network modeling, fitting, and analysis of the results are done by Oleg Vinogradov. Data analysis is performed by Oleg Vinogradov and Nirit Sukenik, Eyal Weinreb, Menahem Segal, Elisha Moses, and Anna Levina.

Contents

2.1 Introduction

The neuronal circuits of many living organisms are comprised of excitatory and inhibitory neurons. The role of excitatory cells is traditionally associated with the spreading of network activity. The inhibitory neurons provide recurrent feedback regulation of activity (Sprekeler, 2017). Thus it is assumed that inhibitory cells implement one of the basic forms of negative feedback that coordinate the activity of excitatory neurons. The balance (E/I balance) of these two opposing forces is the focus of most network models comprised of both neuron types.

The E/I ratio has been shown to control many aspects of the activity of large-scale neural networks. For instance, experimental studies show that precise coordination of excitatory and inhibitory inputs shapes the activity of populations of sensory neurons in the cortex (Brunel, 2000; Isaacson & Scanziani, 2011; Okun & Lampl, 2008). At the same time, the interplay of excitation and inhibition is often proposed as a fundamental mechanism for generating oscillations in the brain (Isaacson & Scanziani, 2011). Theoretical work has shown that changing the overall E/I ratio plays a major role in controlling dynamic states, stability, and coding capabilities of neuronal networks, with the resulting network activity ranging from asynchronous, irregular firing to synchronized network bursting (Brunel, 2000; Gigante et al., 2015).

Inhibition in the cortical areas is implemented by GABAergic neurons, which comprise about 20-30% of all cortical neurons. This proportion is conserved across mammalian species and during the lifespan of an animal (Sahara et al., 2012). The importance of keeping a specific fixed inhibitory percentage may be crucial for efficient storage capacity (Mongillo et al., 2018) and has been linked to multi-task learning (Capano et al., 2015) as well as efficient coding under the sparsity constraints (Alreja et al., 2022). However, the importance of having this particular fraction of inhibitory neurons for the general control of network dynamics remains unclear.

Another well-studied aspect of cortical organization is that excitation and inhibition are balanced both structurally and dynamically. Dynamically, excitatory and inhibitory inputs strongly correlate and synchronize in both spontaneous and evoked activity (G. Liu, 2004; Okun & Lampl, 2008). Structurally, the ratio of excitatory and inhibitory synapses converging to one cell is approximately constant (G. Liu, 2004). The role of inhibition is further complicated by the fact that there are several types of inhibitory neurons that can be clustered by their locus of action on the excitatory neurons as well as the formation of inhibitory synapses on interneurons (Kullmann, 2011). In this study, we do not discriminate the different sub-types of inhibitory cells.

Perturbations of the E/I ratio that move the network away from its balance have been

reported recently and can be applied both acutely and chronically. Acute blockade of inhibitory synapses *in vitro* by application of pharmacological agents causes the dynamics to be excitatory dominated, more uniform, and synchronized (Eytan & Marom, 2006; Feinerman et al., 2005). Blocking inhibition acutely *in vivo* has been found to create epileptic seizures (Dichter & Ayala, 1987). In contrast, chronic blockade (about 48h) or over-activation of inhibition, causes neuronal networks to adjust their activity (G. Liu, 2004). Changes in the E/I balance have been linked to different brain states like deep anesthesia (Taub et al., 2013). They also can be achieved by repetitive sensory stimulation (Yizhar et al., 2011). Finally, shifts in E/I balance were found to have far-reaching behavioral effects on freely moving mice (Berg et al., 2007).

Given that the conservation of E/I balance is a basic property of large-scale neuronal networks, we asked what mechanisms contribute to the creation of this balance, and where their limitations become apparent. E/I balance was extensively studied in brain areas with 20-40% inhibitory cells (Berg et al., 2007; Okun & Lampl, 2008), however, there is no systematic view of how the inhibitory cell fraction and balance are related. To address this, we have engineered cultures of hippocampal neurons with precisely controlled numbers of excitatory and inhibitory cells over a wide range of E/I ratios, from 0% inhibitory neurons to 100%. Our design imposes a global and chronic change, to which the network must respond. We asked whether, given enough time to adapt and rewire, a neuronal network can compensate for the perturbation and reach balanced and stable dynamics.

In the present study, we focus on the ability of neuronal networks with artificially obtained E/I ratios to adapt during their development *in vitro* by monitoring both the whole network dynamics and the single-cell behavior. In parallel, we employ finite network models to directly relate the network properties with collective dynamics. This enables the emergence of a unified picture of the accommodation to changing E/I ratios in ensembles of active neurons.

2.2 Results

2.2.1 Experimental platform to study the effect of cellular excitatory/inhibitory ratio on network dynamics

When dissociated neurons grow in culture, they evolve into a highly connected network that generates synchronized bursting activity separated by quiet periods (Maeda et al., 1995; Orlandi et al., 2013; Penn et al., 2016; Wagenaar et al., 2006). This activity emerges during the development of a network and is controlled by the architecture of its excitatory and inhibitory

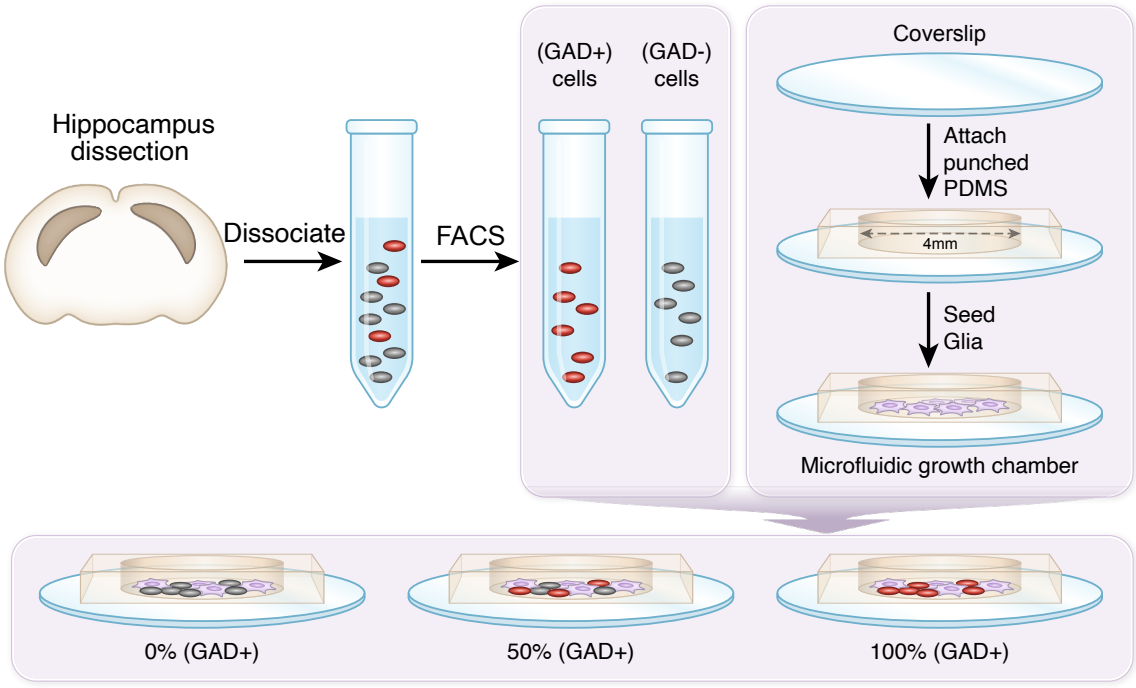


Figure 2.1: Experimental platform to study the effect of cellular E/I ratios on network dynamics. Excitatory and Inhibitory neurons were dissociated from transgenic mice hippocampi at E19. Inhibitory cells express fluorescent proteins. Neurons are consequently sorted with FACS, mixed with different ratios of excitatory and inhibitory cells, and plated onto a microfluidics growth chamber.

cells (Soriano et al., 2008) and by the balance between them. To study the role of inhibitory cells in balancing and regulating network dynamics and connectivity, we change the fraction of inhibitory cells between 0% and 100% while keeping identical seeding densities (Fig. 2.1), and measure whole network spontaneous activity with Ca-imaging.

We used transgenic mice, in which the GABAergic neurons express tdTomato. The hippocampi of mice embryos were dissected and dissociated at E19, and excitatory and inhibitory cells were then sorted with Fluorescence-activated cell sorting (FACS, see Box 2.2.1). After the sorting, neurons were seeded on a microfluidics growth chamber along with glial cells. The ratios of inhibition to excitatory cells were preserved after the growths for 14-30 days *in vitro* (Fig. 2.2). A detailed description of the experimental methods can be found in (Appendix A.1 and Sukenik* , Vinogradov* et al., 2021)

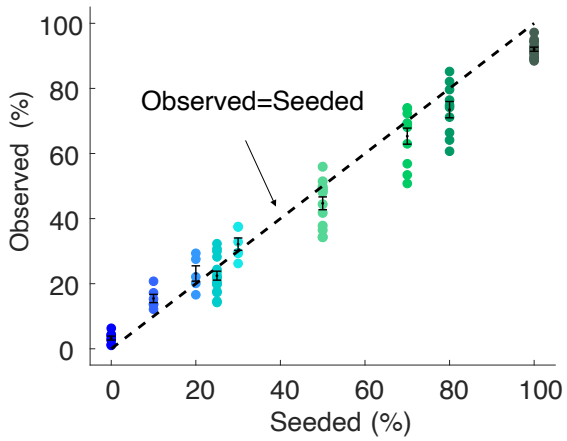
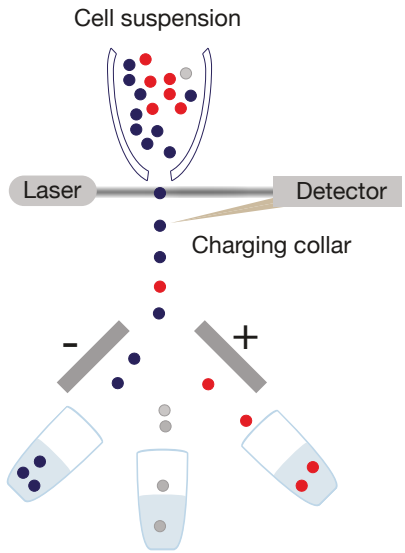


Figure 2.2: Results of manual cell counting based on fluorescent images show that the fractions of excitatory and inhibitory cells after the network is fully developed (14-30 DIV) are preserved.



Fluorescent Activated Cell Sorter (FACS) allows separating cells based on their fluorescent properties. A cell suspension passes through an ultrasonic nozzle vibrator that forms droplets that typically are comprised of a single cell. Cells flow through a laser beam with a certain preset wavelength that excites a fluorescent marker. The light then passes to the detector that identifies the presence of fluorescence. When the fluorescent light is detected a positive or negative charge is applied to a droplet. The charged droplets then pass through an electric field that separates fluorescent and non-fluorescent cells.

2.2.2 Development of spontaneous population activity in networks with various fractions of inhibitory cells

Cultures with varying amounts of inhibitory cells were spontaneously active and developed network bursting (Fig. 2.3). Including the cultures with very high inhibitory percentages, which indicated that the hippocampal cultures are different from almost fully inhibitory striatal culture (about 95% inhibitory cells), which typically do not show network bursting (Segal et al., 2003).

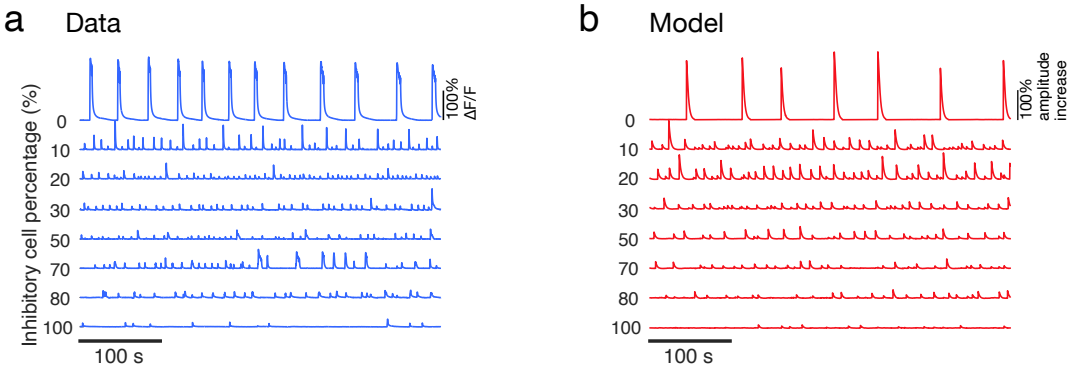


Figure 2.3: Examples of calcium traces from the experimental data and fitted model. For the data, traces are averages of the whole recording field with about 100-300 neurons recorded simultaneously. All cultures exhibit spontaneous activity, even when the inhibitory percentages are high. The control cultures (no FACS) are shown under 25% inhibition.

Almost purely excitatory cultures (Fig. 2.3, top trace), had higher burst amplitudes compared to the cultures with 20-100% inhibition. This effect was larger than expected by the acute blockade of inhibition, which typically leads to a 2-5 fold increase in amplitude (Feinerman et al., 2005).

To characterize network bursting we computed inter-burst intervals (IBIs), burst amplitudes, and burst durations and their variability. IBIs were defined as the time between two consecutive bursts. The mean of the IBIs as a function of inhibitory cell percentage follows a U-shaped trend (Fig. 2.4). The IBIs remained similar for most E/I ratios, except in cultures with extreme inhibitory percentages. In cultures with 10-80% inhibitory cells, the mean IBIs values range between 10s to 21s, whereas cultures with 0% and 100% seeded inhibitory percentages are characterized by significantly higher IBI values of 65s and 70s respectively ($p<0.0001$, permutation ANOVA, please refer to Appendix E). A small positive trend of the IBIs' mean is observed in 10-80% inhibitory cultures ($p=0.039$). In contrast, the variability of the IBIs, measured by the Coefficient of Variation (CV), grew linearly with the inhibitory percentage from 0.29 to 0.75 (Fig. 3b). Thus, bursting in networks with larger inhibitory fractions appears more random and closer to a Poisson process (CV=1).

A change of 10% in the number of inhibitory cells led to a nearly 5-fold decrease in the burst amplitude (Fig. 2.4, permutation ANOVA, $p<0.001$). This behavior has a remarkable similarity to the effect of blocking inhibition (Feinerman et al., 2005). Burst durations (Fig. 2.4e) were the largest in 0% inhibitory cultures (8.5 ± 0.6 s, mean$\pm$s.e.m.), an effect that is comparable to that of blockade of inhibition (Feinerman et al., 2005), that also indicates larger network excitability (Giugliano et al., 2004). Burst durations for 0%, 80%, and 100% cultures were significantly

longer than in control cultures ($p<0.05$). We also observe a positive trend in burst durations between 10% and 100% of inhibitory neurons ($p=0.001$). This increase can be related to the number of input connections. The CV of burst durations grew from 0.15 ± 0.01 to 0.46 ± 0.04 between 0% and 10% and stays approximately constant for 10% to 100% of inhibitory neurons (Fig. 2.4).

Generally, despite small differences, the network bursting features of cultures with varying E/I ratios did not change considerably. The only exception is the changes in extreme E/I presentation (0% and 100%).

2.2.3 Patch clamp experiments reveal that the number of connections decreases as the fractions of inhibitory cells increase

Recordings of the population activity showed that the network bursting dynamics is very robust even to a large change in the fraction of inhibitory cells. Using patch clamp recording, we identified that the spontaneous synaptic release changes as the fraction of inhibitory neurons decrease. This indicates that networks likely adjust the connectivity to accommodate changes in the cellular E/I ratios.

First, we checked if neurons in our culture had the typical electrophysiological properties of excitatory and inhibitory neurons. Fig. 2.5 shows the properties of inhibitory and excitatory cells recorded in control cultures (20% inhibitory neurons, no FACS). The spike durations were significantly longer in excitatory cells, whereas the after-hyperpolarization current was larger in inhibitory cells. We did not see a significant difference in the resting potential of input resistance. Thus, we observed a typical electrophysiological portrait of excitatory and inhibitory cells in the hippocampus (Kullmann, 2011).

Next, we asked if the adaptation to a different fraction of inhibitory cells is due to changes in the synaptic strength or the number of synaptic connections between neurons. We used spontaneous post-synaptic currents (PSC) as a proxy for the synaptic strength and measured it in three E/I ratios. We did not observe any statistically significant differences in the PSC sizes ($p=0.98$). This rules out the adjustment of synaptic strength as the mechanism of network

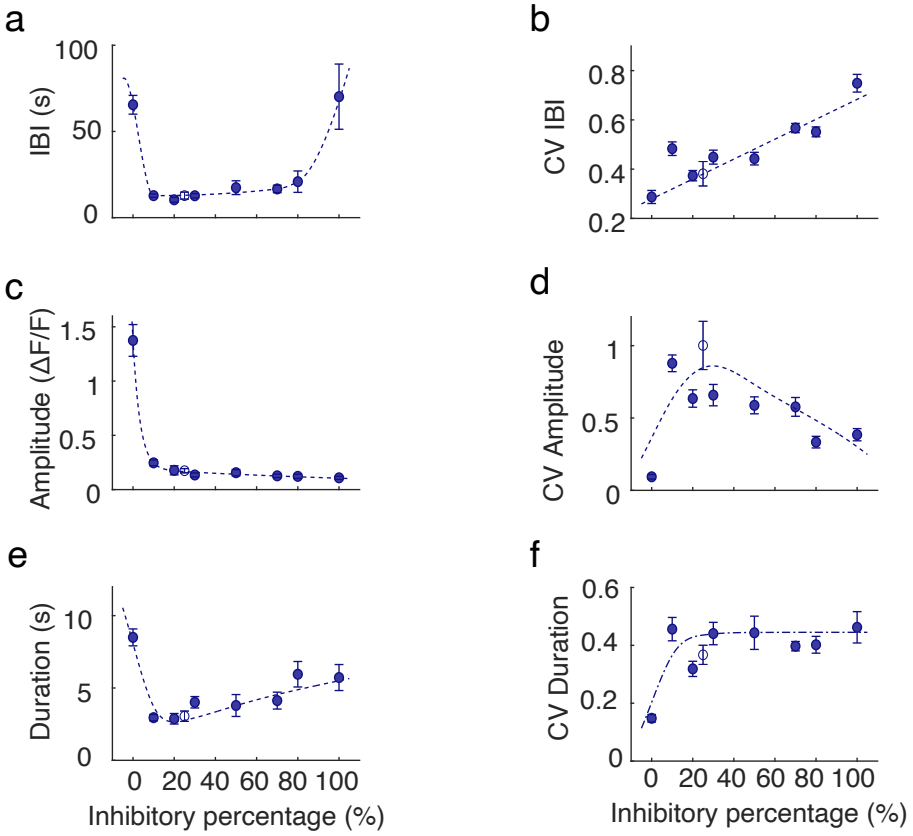


Figure 2.4: Features of population bursting dynamics in cultures with various cellular E/I ratios **(a)** Interburst intervals (IBIs) display a U-shaped trend as a function of inhibitory percentage, with higher values at extreme cases (0 and 100%), while remaining relatively constant in the midrange (10 to 80%). **(b)** The coefficient of variation (CV) for IBIs increases linearly with the inhibitory percentage. **(c)** Burst amplitudes decrease from 0 to 10% inhibition and continue to decline until 100% inhibition. **(d)** The CV of burst amplitude is minimal for 0% inhibition (0.09±0.02). Raising the inhibitory percentage to 10% results in a sharp increase in amplitude variance, which then slowly decreases at higher inhibition fractions. **(e)** Burst duration is maximal for 0% inhibition and rapidly decreases at 10% inhibition. The duration progressively increases with a higher inhibitory percentage. **(f)** Durations for 0% inhibitory cultures exhibit a relatively low CV of 0.15±0.01. Other fractions in the range of 10 to 100% inhibition display a relatively constant value of around 0.45. In all panels, the dashed lines serve only as a visual guide, the error bars represent the standard error of the mean (s.e.m.), and the control (no FACS) is depicted as an empty circle.

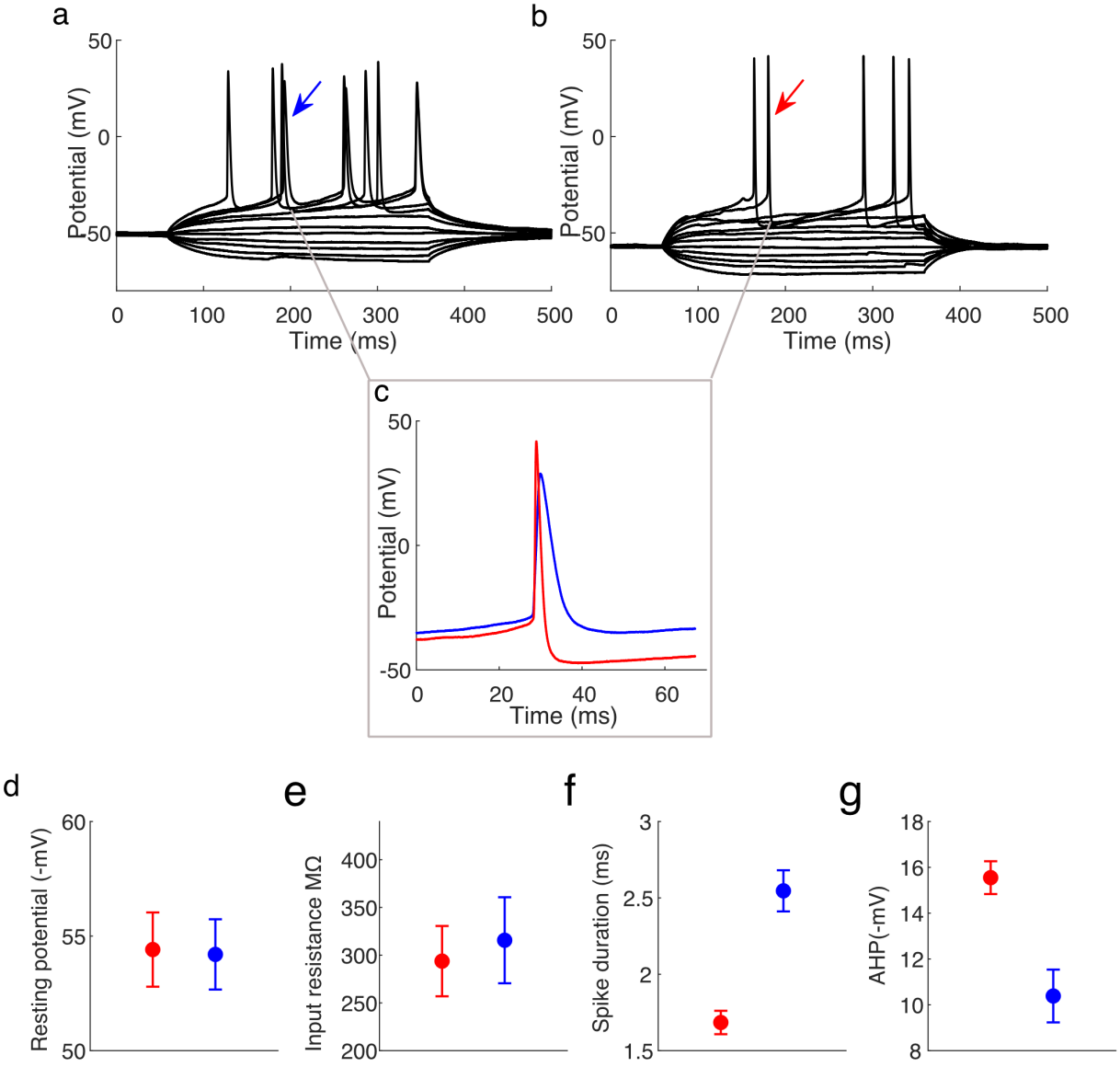


Figure 2.5: Electrophysiological properties of inhibitory (red) and excitatory (blue) cells in cell-sorted cultures. (**a**) The firing of individual cells when stimulated by currents of between around -45 to +70pA in excitatory neurons and stimulated by currents around -45 to +80pA in inhibitory (**b**) neurons. (**c**) Spike shapes for individual excitatory and inhibitory neurons. (**d-e**) We did not see statistically significant differences between the resting potential (d, N(inhibitory)=17, N(excitatory)=10, p=0.931), input resistance (N(inhibitory)=8, N(excitatory)=8, p=0.712). (**f**) Spikes of excitatory neurons were longer (f, N(inhibitory)=21, N(excitatory)=13,p<0.001). (**g**) The AHP was more negative in inhibitory cells (N(inhibitory)=20, N(excitatory)=13, p=0.001). The results show that the inhibitory (fluorescent) cells and excitatory (non-fluorescent) cells in FACS-sorted cultures have typical features of excitatory and inhibitory neurons in the hippocampus. Error bars – standard error of the mean.

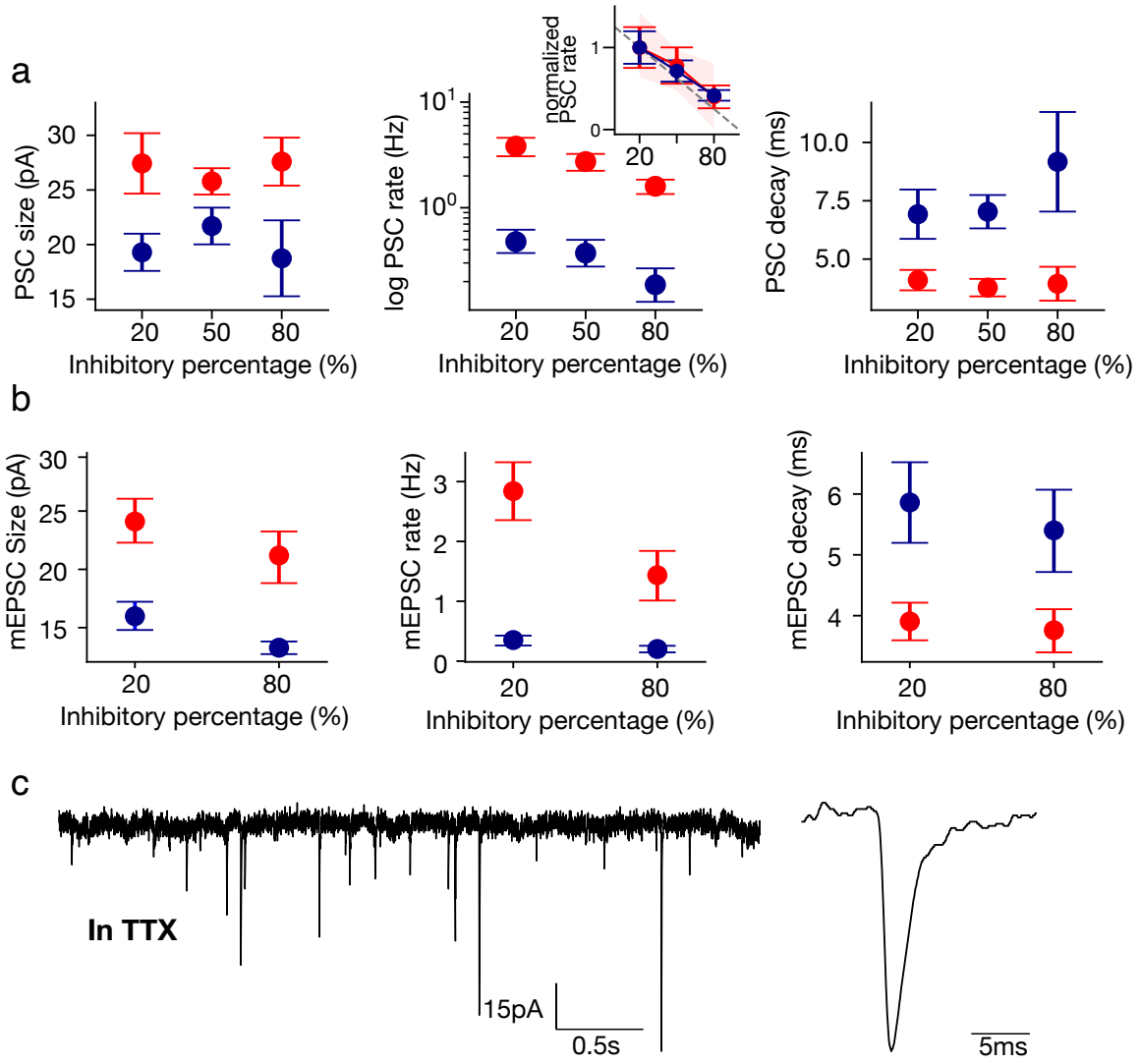


Figure 2.6: Patch clamp results suggest that neurons adjust the number of connections to adjust to different factions of inhibitory neurons. **(a)** The size of PSC (left) did not change as we increased the fraction of inhibitory cells ($F(2,64)=0.0156$, $p=0.98$); (middle) The rate of PSC decreased linearly with increasing fractions of inhibitory neurons ($F(2,64)=6.348$, $p=0.003$); the inset shows the normalized decrease for both excitatory and inhibitory cells (dashed line - linear fit, 95% CI-red area). We also did not see a significant difference in the PSC decay time ($F(2,64)=0.24$ $p=0.786$). **(b)** To confirm that the effects are due to connections and not network activity, we recorded miniature PSC (mEPSC) in under 0.5μM TTX and 10μM bicuculline. (left) mEPSC size did not significantly change with a higher fraction of inhibitory cells; (middle) mEPSC rate significantly decreased with a higher fraction of inhibitory cells (similar to the PSC, $F=12.9$, $p<0.0001$); (right) we did not observe differences in the mESCP decay times. **(c)** Example traces of the currents and mESCP. Error bars – s.e.m.

adaptation to different E/I ratios. The decay time of PSCs also did not show any significant difference as a function of inhibitory percentages (Fig. 2.6a, left, p=0.786).

Next, we asked if the spontaneous synaptic release changed as a function of inhibitory percentage. We observed that PSC (as well as miniature post-synaptic currents mPSC) rates dramatically decrease as a function of the E/I ratio (Fig. 2.6, p=0.003). We also saw a difference in the PSC frequency of inhibitory versus excitatory cells. Given that the spontaneous release rate is typically a function of the number of connections, these results suggest that the number of inputs to both inhibitory and excitatory neurons is directly proportional to the number of excitatory neurons (see also Fig. 2.14).

2.2.4 Modeling population activity in networks with various fractions of inhibitory neurons

To relate the dynamics of neuronal networks with network and single neuron properties, we build a model of randomly connected excitatory and inhibitory leaky integrate-and-fire (LIF) neurons with spike-frequency adaptation. We change the number of inhibitory neurons in the network to follow the experimental preparations and infer model parameters. The results show that in order to maintain stable bursting while changing the percentage of inhibitory neurons, the networks should adjust the number of inhibitory connections and keep the structural E/I balance.

We also compare the result to a baseline model without spike-frequency adaptation (N. Brunel & Hakim, 2008; Tartaglia & Brunel, 2017). Despite the fact that the model without adaptation can show burst-like network activity it can only partially describe the temporal bursting dynamics observed *in vitro*. Next, we block inhibitory synapses both in the model and in experiments and compare the changes in bursting activity. The network with spike-frequency adaptation close to the E/I balance can closely follow experimental results. At the same time, the model without adaptation predicts an opposite change and thus fails to replicate the experiments. Finally, we analyzed the network bursting in both models using mean-field approximation a show that realistic network bursting emerges as an interplay between the network excitability and negative feedback mechanisms.

Network of leaky integrate-and-fire (LIF) neurons

The network is comprised of N^I inhibitory and N^E excitatory adaptive leaky integrate-and-fire neurons (Giugliano et al., 2004; Tartaglia & Brunel, 2017). The dynamics of neuron $i (i = 1, \ldots, N = N^E + N^I = 1000)$ follows:

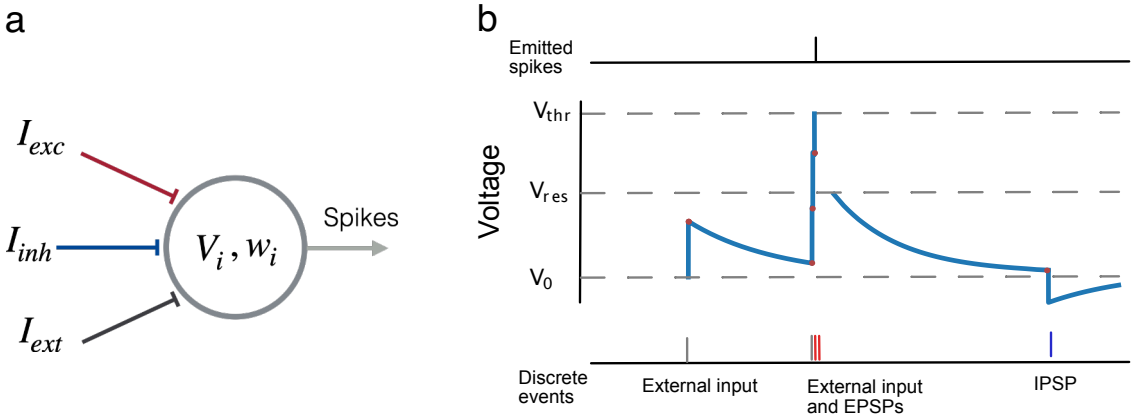


Figure 2.7: To relate the population activity and single neuron properties we built a network model of excitatory and inhibitory adaptive LIF neurons. **(a)** A single neuron integrated excitatory, inhibitory, and external inputs.**(b)** A single neuron receives the inputs and emits a spike only when its voltage reaches a firing threshold, after which the neuron's potential is reset and adaptation current (w) increases (not shown).

$$\dot{V}_i = -\frac{V_i}{\tau_{\text{m}}} + C_m^{-1}(I_i^{\text{syn}} - w_i)$$
$$\dot{w}_i = -\frac{w_i}{\tau_w} + b\sum_k (t - t_i^k). \qquad \text{If } V_i > V_{\text{thr}}, \quad \text{then} \quad V_i \mapsto V_{\text{reset}}$$

where w_i is the strength of adaptation, I^{syn} is a synaptic input, τ_m and τ_w are the membrane and adaptation timescales. When V_i reaches a firing threshold V_{thr}, the neuron emits a spike and the strength of adaptation is increased by b (Fig. 2.7). In the model without adaptation, b is set to zero, which results in a standard LIF neuron (N. Brunel & Hakim, 2008).

The synaptic input to a neuron is the sum of excitatory, inhibitory, and external post-synaptic potentials

$$I_i^{\text{syn}}(t) = \underbrace{J\sum_k \sum_{j\in S_i^E} \delta(t - t_j^k - D)}_{\text{Excitatory}} \underbrace{-gJ\sum_k \sum_{j\in S_i^I} \delta(t - t_j^k - D)}_{\text{Inhibitory}} + \underbrace{J^{\text{ext}}\sum_k \delta(t - t_k^{\text{ext}})}_{\text{External}}.$$

Here, J is the synaptic strength, $t_j^{k,a}$ is the kth spike emitted from a set of excitatory S_i^E or inhibitory S_i^I presynaptic neurons through jth connection, and D is a synaptic delay. Each neuron receives a fixed number of random excitatory $K^E = pN^E$ and inhibitory connections $K^I = pN^I$ as well as external input events $t^{k,\text{ext}}$ from a Poisson distribution with a rate ν_{ext}.

Each random input depolarizes the neuron by J^{ext}. The external input models various sources of spontaneous activations of neurons.

We set the number of incoming excitatory K^E connections to be proportional to the number of excitatory neurons, with the proportionality constant inferred from the data (see all parameters in and the comparison with physiological values in Appendix C). The number of inhibitory connections K^I is a free parameter (see below). Inhibitory synaptic weights were 4 times larger than excitatory ($J^I = 4J^E$) and both were kept constant for all cellular compositions (G. Liu, 2004). The size of the excitatory postsynaptic potential (PSP) J^E is kept constant at 2mV across networks with varying fractions of inhibitory neurons. The synaptic strength is meant to approximate a "functional synapse" with multiple axon collaterals. The choice of 2mV for the strength of the connection is based on cultured neuronal networks with similar densities (Penn et al., 2016). A connection strength of 2mV means that a neuron needs about 10 simultaneous incoming events to generate a spike. Network dynamics are simulated numerically with a 0.5 ms integration time step using the community-supported network simulator NEST 2.20. The burst detection algorithm was based on the inter-spike interval threshold and is described in detail in Appendex B.

The model exhibits diverse network bursting that can be controlled by the synaptic strength, network connectivity, rate of external inputs, and the strength of adaptation (Giugliano et al., 2004; Tartaglia & Brunel, 2017). The network bursting activity produced by the model very closely matches the bursting in real recordings (Fig. 2.8). Moreover, the model activity also matches the basic features of the burst initiation and burst termination when compared to the classical result from (Eytan & Marom, 2006). Thus, we found a good agreement between the initiation and termination timescale for an average burst (Fig 2.9a). The model, however, is limited to spiking activity, and, in fact, single neuron voltages during bursts show strong negative deflection that could be attributed as a limitation to the LIF neuron model(Fig 2.9b).

Feature	Model (20% inh.)	Data
Burst duration	0.14 ± 0.8 s mean $\pm$ s.t.d)	$0.1-0.2; 0.4-1.750$ s*(Eytan & Marom, 2006)
Firing rate inside the bursts	51 ± 11 Hz	$12 - 67$ Hz** (Charlesworth et al., 2015; Eytan & Marom, 2006)
% of spikes inside the burst	96%	$80 - 99\%$**(Charlesworth et al., 2015)

Table S5 Features of the bursts in the network model and in the data
*Patch-clamp recordings of our data
**18DIV

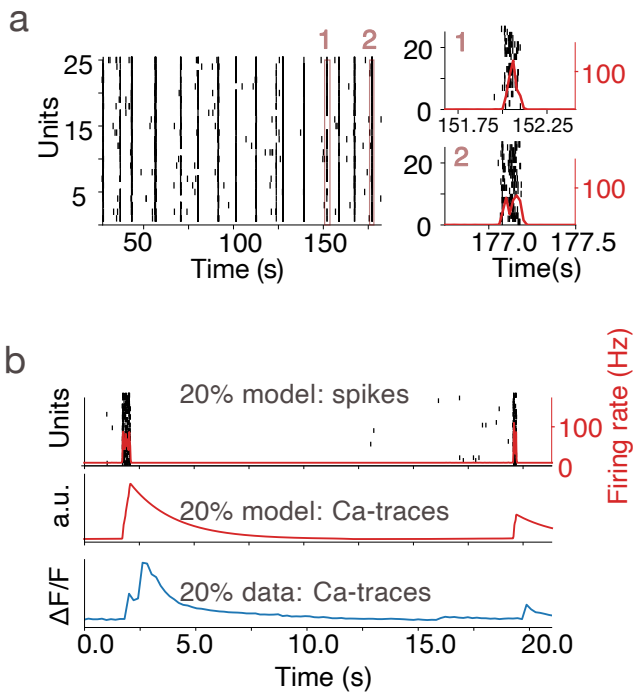


Figure 2.8: Qualitative comparison of the model dynamics and Ca-activity of network *in vitro*. **(a)** Population activity (on spiking level) of the network model with 20% inhibitory neurons. Side panels show two example bursts **(b)** An example of a network burst in the network model and in a real culture with 20% inhibition. Blue tace - Ca-trace of the data, red - spiking activity convolved with 2s exponential kernel to imitate the Fluo-4 dynamics.

2.2.5 Inference of model parameters

The model simulations confirm that the model can qualitatively fit the neuronal dynamics of cultured neurons (Fig. 2.8). To obtain a quantitative fit that allows us to interpret differences and similarities between the cultures with different E/I ratios, we applied simulation-based inference (SBI). We use one of the classic approaches called approximate Bayesian computation (ABC, Beaumont et al., 2009).

The ABC framework allows us to estimate the full posterior of the parameters of interest and to visualize dependencies between different parameters. The method is widely used for Bayesian inference for models where a likelihood function is not available and allows estimating the likelihood by comparing the outcomes of simulation with observed data. To fit the bursting network, we define a distance function:

$$E(\Theta) = \frac{1}{2}[(\mu_{\text{IBI}} - \hat{\mu}_{\text{IBI}})^2 + (\text{CV} - \hat{\text{CV}})^2] \tag{2.1}$$

where $\hat{\mu}_{\text{IBI}}$ and $\hat{\text{CV}}$ are the mean and CV of IBIs estimated from the model simulations.

Here we fit only the number of inhibitory connections per neuron and the properties of the external drive. The fixed parameters of the model were constrained to lie within the corresponding intervals from the experiments where available as described in the parameters table (Appendix C).

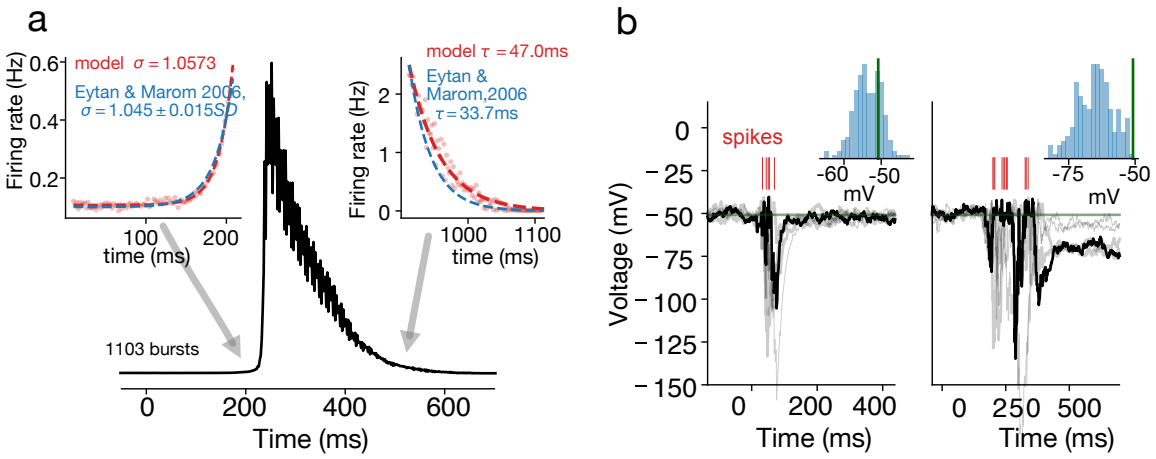


Figure 2.9: Single population bursts in the model and in the data. **(a)** Burst shape (average over 1103 bursts) in the control 20% inhibitory cells. The burst propagation throughout the network is characterized by a fast exponential rise. The average exponential rise (σ) in the model is 1.05, which well matches the burst initiation in experimental MEA recordings (Eytan & Marom, 2006). The average burst was about 47ms compared to 33ms for the data. **(b)** The single neuron voltage in the model reflects the integration of excitatory and inhibitory currents. After high-intensity bursts (right), the voltage is decreased compared to the burst baseline (green line), which indicates a dynamic decrease in the single neuron excitability (Gal et al., 2010). The inset shows the distribution of voltages after the burst termination.

We choose a set of wide uniform priors for each parameter (from 100 to 2,000Hz and from 2 to 50 connections). Then we sample a combination of network parameters from the prior, run the network simulation for 500 s., estimate the statistics of the bursting dynamics, and compute the distance function. If the distance function is smaller than a threshold value, we accept the combination of parameters (Fig. 2.10). The sampling is repeated until 50 combinations of parameters are accepted. Then we compute a Gaussian probability density function for the accepted parameters, rescale it with the original prior, and use it as the new prior for the next ABC step (population Monte Carlo method, Beaumont et al., 2009). In every step, the acceptance threshold (ϵ) decreases and as a result, we obtain a set of parameters that estimates the posterior. Each network with its different fraction of inhibitory neurons is fitted independently. Typically, the networks converge to $\epsilon < 0.05$ within 10-20 steps. After the threshold ϵ has dropped below 0.05, we run a final ABC step with a 1500 s.-long simulation in order to match the number of bursts in experimental data. Implementation of the ABC algorithm is based on https://github.com/rcmorehead/simpleabc/.

We separately fit networks with E/I ratios corresponding to the various seeded fractions of inhibitory neurons. For extreme 0% and 100% inhibitory networks, we assumed 5% and 92%

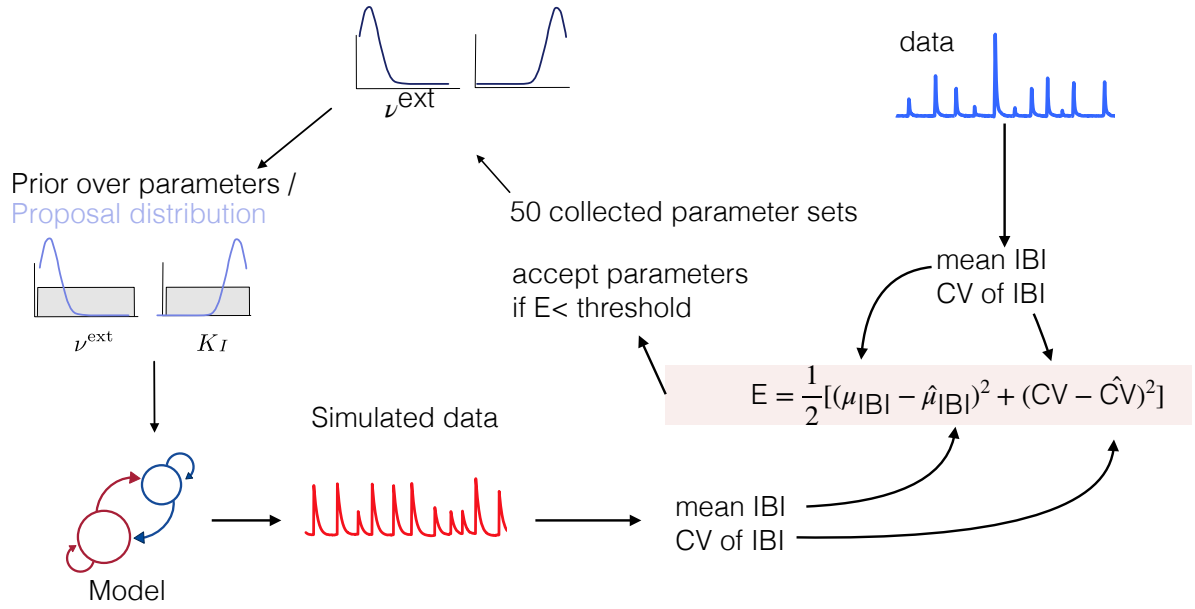


Figure 2.10: Schematics of the Approximate Bayesian Computation (Beaumont et al., 2009) . We start by choosing a prior over the model parameters. The parameter values are then sampled from the prior, and the model with these parameter sets is simulated. We extract the summary statistics from the simulated data and compare them with the summary from the real data using a distance function. In this project, we used a mean squared difference between the IBIs and the CV of IBI. We accept a parameter set if its loss is below a prespecified threshold. This part is repeated until at least 50 parameter pairs are accepted. The accepted samples are smoothed and weighted with the prior and used a new proposal distribution. The proposal is instead of the prior in the next iteration of the ABC.

of inhibitory neurons. The model without adaptation ($b = 0$) is matched to the mean IBI of the experimental data manually by adjusting the model parameters. The reason for that is that network-bursting-like activity lies within a very narrow part of the parameter space.

2.2.6 Networks adjust E/I balance to maintain stable dynamics

The fitted network model well matches the bursting activity in cultured networks with different fractions of inhibitory neurons (Fig. 2.3). With just two individually adjusted parameters, our network accurately fits the important features of the recordings, including nearly constant mean IBIs in networks with 10-80% inhibitory percentages and an increase of the mean IBI in networks with extreme inhibitory fractions (Fig. 2.11). The model also reproduced the linear increase of the CV of IBIs and the burst amplitudes (Fig. 2.11c,d).

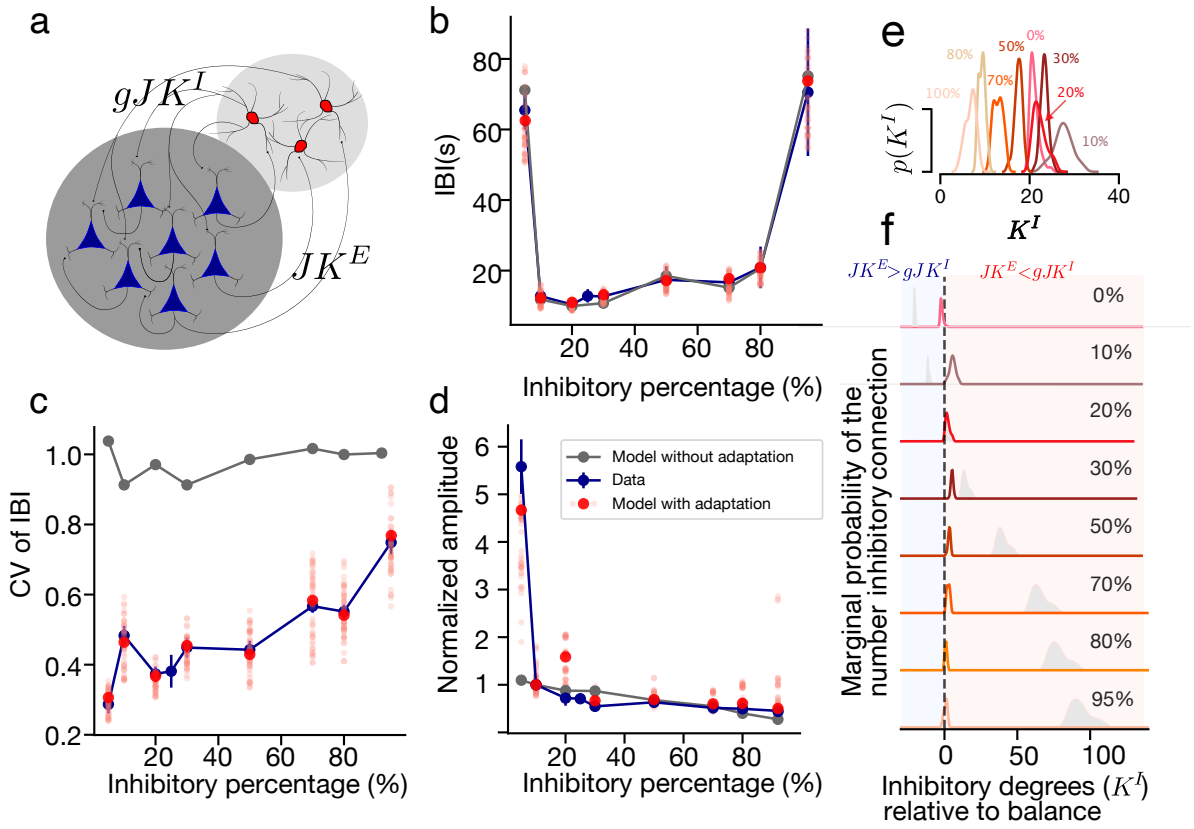


Figure 2.11: Results of parameter inference show that in order to maintain stable population bursting the network should maintain a balanced E/I connection. **(a)** A schematic of the model. **(b)** The fitted models both with (red) and without (gray) the spike-frequency adaptation can match the experimental data. **(c)** Only the model with the spike-frequency adaptation can reproduce experimentally found CVs of IBI. **(d)** The mean amplitudes are also better fitted in the model with adaptation (note that we did not include the burst amplitudes to the fitness function). The plotted amplitudes are normalized to 10% network for both the data and model. **(e, f)** The number of inhibitory connections stays proportional to the (decreasing) number of excitatory neurons. **(e)** The marginal distributions of the inhibitory connection found with ABC are shifted leftwards. **(f)** Same distributions with the x-axis shifted such that the balance for every network is at zero. Gray shadows show the distribution of inhibitory connections expected when the numbers stay proportional to the number of inhibitory neurons.

Next, we analyzed the marginal posterior distributions of inhibitory connections (Fig. 2.11e). In networks with 10-100% inhibitory neurons, the distribution of inhibitory connections stays proportional to the number of excitatory connections. This way, excitatory and inhibitory connections can balance each other, in comparison to the naively expected distribution, where the number of connections is proportional to the number of inhibitory neurons (Fig. 2.11f, grey lines). In the 0% inhibitory neurons model, the distribution was shifted towards the excitation-dominated region (Fig. 2.11f, blue region). This likely indicates that these networks fail to adapt to their cellular compositions, which results in the difference in IBI, CV of IBI, and burst amplitudes.

We also compared the results to a more simple version of the network that does not include the spike-triggered adaptation. This network can exhibit noise-induced population bursts that recapitulate some of the features of bursting *in vitro* (Tartaglia & Brunel, 2017). However, the CV of IBIs in the model without adaptation always stays close to one (Fig. 2.11c, grey). The reason for this is the random nature of burst initiation and the absence of refractoriness given the spike-triggered adaptation (see also Appendix D).

2.2.7 Interaction of adpatation and E/I balance drive the bursting dynamics

The leaky integrate-and-fire network with slow spike-triggered adaptation allows us to effectively describe the network bursting as bistable dynamics in the timescale separation limit. These dynamics are then driven by the interaction of excitation, inhibition, and adaptation (Giugliano et al., 2004). The adaptation current in our model sets the slowest timescale of the IBIs. When the network does not have inhibition (the relative inhibitory strength g is set to 0), each burst is terminated only by a strong adaptation that counteracts excitation. After that, the adaptation current decays exponentially, and the probability of initiating the next burst gradually increases (Fig. 2.12a). In this regime, the variability of IBIs is small. Adding inhibition to the network allows for shorter IBIs. In this case, both inhibition and adaptation counteract excitation, which stops the burst. That results in a much smaller adaptation current at the point of burst termination, increasing the probability of starting the next burst earlier (Fig. 2.12a).

To explore this observation, we analyzed the bursting dynamics, assuming that the spike-triggered adaptation timescale is much slower than the membrane timescale ($\tau_m \ll \tau_w$). We estimate the stationary firing rate at different fixed levels of adaptation using the current-to-rate transfer function of the white noise-driven LIF neuron and mean-field approximations of the synaptic input (N. Brunel & Hakim, 2008).

This approach assumes that every neuron receives a large number of small inputs such that the total input can be approximated by the Gaussian white noise (N. Brunel, 2000; Gerstner et al., 2014). Dynamics at a short temporal scale are considered so that the influence of the adaptation current is approximately constant. We find the stationary firing rate of the network by solving a self-consistent system of equations:

$$\nu_0\left(\mu_0, \sigma_0^2\right) = \left[\tau_{ref} + \sqrt{\pi}\tau_m \int_{\frac{v_r-\mu_0}{\sigma_0}}^{\frac{\theta-\mu_0}{\sigma_0}} e^{u^2}(1+\mathrm{erf}(u))du\right]^{-1} \tag{2.2}$$

$$\mu_0 = J\left(K^E - gK^I\right)v_0\tau_m + J_{\text{ext}}\, v_{\text{ext}}\, \tau_m - \langle w\rangle \tag{2.3}$$

$$\sigma_0^2 = J^2\left(K^E + g^2K^I\right)v_0\tau_m + J_{\text{ext}}^2\, v_{ext}\tau_m \tag{2.4}$$

$$\tag{2.5}$$

where $\nu_0\left(\mu, \sigma^2\right)$ is the stationary firing rate, μ_0 is the mean synaptic input, σ_0 is the stationary standard deviation of the synaptic input, and $\langle w\rangle$ is the average adaptation. The details of derivation can be found in (N. Brunel, 2000; Gerstner et al., 2014)

This system of equations can be solved numerically for different values of adaptation. There are 3 types of solutions: a fixed point at a high firing rate with a small adaptation current, a low firing rate fixed point for large adaptation, and a bistable firing rate for intermediate adaptation. Levels of adaptation at the beginning and end of a burst in the network simulation approximately correspond to analytically computed bifurcations (Fig. 2.12b). Decreasing the strength of inhibitory synapses gradually increased the size of the bistable region, which corresponds to an increase of the adaptation current at the end of the burst (Fig. 2.12b, inset) and leads to longer IBIs and larger burst amplitudes. For the comparison with full network simulations, we used a 20% inhibitory neurons network with N=10^4 and recorded bursting activity for 500s (Fig. 2.12b).

The idea that network bursting can be reduced to the interaction of average noisy input and slow adaptation current is further explored in Chapter 3.

2.2.8 Blocking inhibition *in vitro* and *in silico* probes the limitations of the model

Our findings, so far, are that networks with all but extreme fractions of inhibitory neurons adapt during development and behave similarly to control cultures. To test what further changes occur

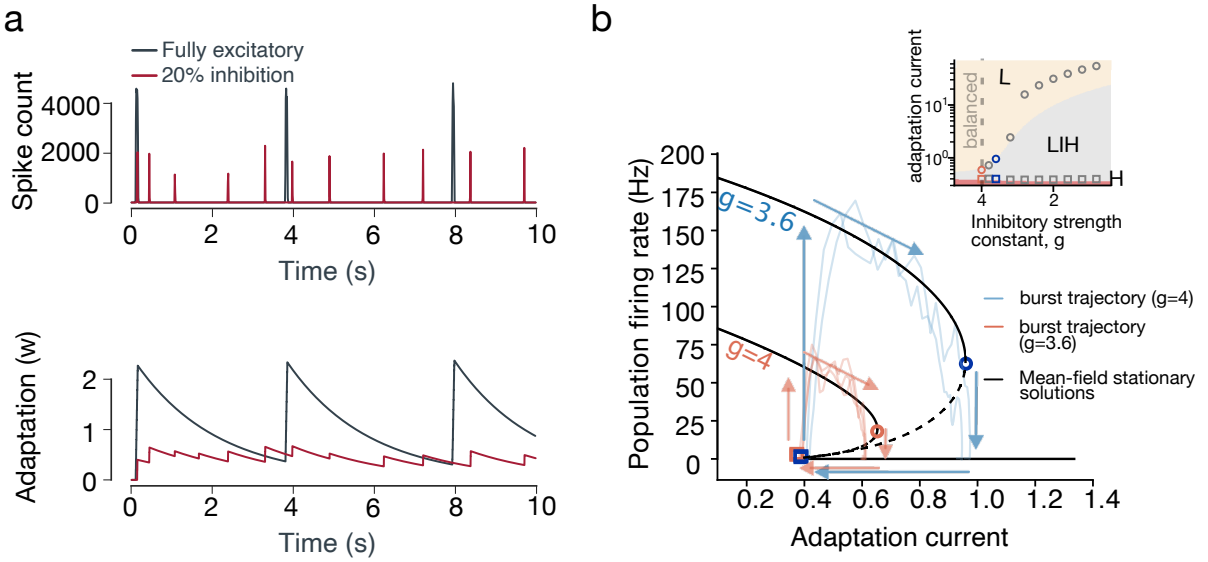


Figure 2.12: Network bursting as an interplay of recurrence, inhibition, and spike-frequency adaptation. **(a)** An adaptation current with and without inhibition. The figure illustrates the coupling between the average network firing rate and adaptation current. In the model without inhibition (dark grey), the bursts are terminated by the adaptation alone and the adaptation current is typically large after the burst termination. The new burst can be then initiated by the external input noise (that represents spontaneous activation of neurons) only when the adaptation current reaches a baseline. Inhibition in this system can terminate a burst earlier while the adaptation current is not yet very strong. This in turn allows the next burst to be initiated earlier. **(b)** Phase space of the population firing rate and spike-frequency adaptation for inhibitory strengths g=3.6 and g=4.0 (solid black lines, stable solutions; dashed lines, unstable solutions). At g=4, the network is in the balance condition; at g=3.6, it is in the excitation-dominated condition. The pale lines show examples of individual burst trajectories (pink, balanced network; blue, excitation dominated). Larger values of the adaptation at the burst end lead to longer IBIs. Inset shows how the size of the bistable region expands with decreasing inhibitory strength. The squares and circles indicate the average adaptation at the beginning and end of simulated bursts. Decreasing the inhibitory strength leads to higher burst amplitudes and longer IBIs.

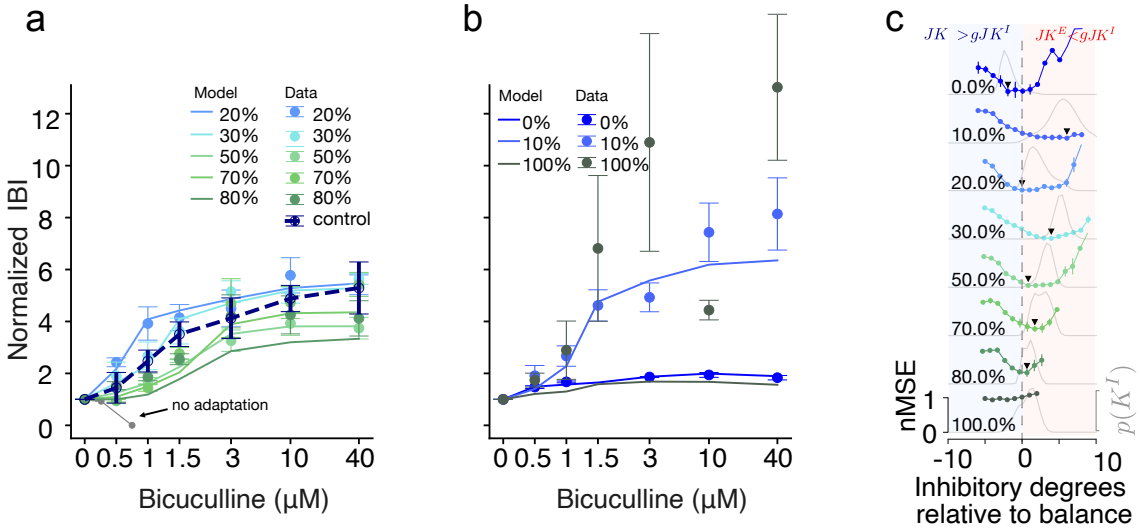


Figure 2.13: Blocking of inhibitory synapses reveals the bursting mechanisms. **(a)** In networks with 20 to 80% inhibitory neurons, the increase in bicuculline con-centration (logarithmic scale) results in longer IBIs—that is, within experimental error, similar to the control cultures. The model (lines) reproduces the experimental results (dots). In contrast, no adaptation model (gray) decreases the IBIs and transitions to a non bursting dynamic. The IBIs are normalized by the mean at 0μM bicuculline.((**b**)) Bicuculline application to networks with extreme inhibitory percentages and the corresponding responses of the model. ((**c**) Normalized mean squared errors between the experiment and model (nMSE, dots) are shown together with the posterior probability of inhibitory connections (grey lines). The number of inhibitory degrees on the x-axis is shown relative to balance (0 corresponds to $JK = 4gJK$). Red area – inhibition dominated the region, blue area – excitation dominated the region. External input rates for the models are samples from posterior probability conditioned on the number of inhibitory connections (10 samples, error bar – s.e.m.); The maximum of the posterior distribution coincides with optimal or nearly optimal nMSE (black arrows).

under an acute change of synaptic strength, we block inhibitory connections using the synaptic inhibitory blocker, Bicuculline (Fig. 2.13). All cultures, with different E/I ratios, show a gradual increase in the mean IBI values as a function of inhibitory receptor blockade. The change of IBI in 20-80% cultures with saturating amounts of Bicuculline ([Bicuculline] = 40μM) is between 3.7±0.8 and 5.6±0.3 and they behave approximately like control cultures (Fig. 2.13).

Cultures with almost no inhibition are only slightly sensitive to inhibition blockade in comparison to control (Fig. 2.13) and show an increase of 1.8±0.1 in IBI values (p=0.0181). Whereas, cultures that are mainly inhibitory are highly sensitive to Bicuculline (11±6 times increase of IBIs already at [Bicuculline] = 3 μM) and show high variability of responses. Some of the 10% inhibitory cultures exhibit a unique behavior, with a sharp increase in IBI values ranging between 7 to 16 times at [Bicuculline]=40μM.

To repeat these experiments in silico, we gradually decrease the strength of inhibitory connections in models fitted to the experimental data without the application of Bicuculline. We approximate the decrease of the inhibitory strength by the fraction of inhibitory receptors blocked with $1/\left(1+\frac{[\text{bicuculline}]}{K_d}\right)$, where $K_d = 3\mu M$. Networks with adaptation responded by increasing the mean IBI closely matching the experimental results (Fig. 2.13a,b note that we did not refit the networks to match the Bicuculline responses). In contrast, networks without adaptation decreased the mean IBIs (Fig. 2.13a grey).

Next, we quantify how deviations from balance affect the response of the model network to bicuculline, where deviation 0 means that the excitatory and inhibitory connections are at balance (Fig. 2.13c). To this end, we investigate the set of networks parametrized by the deviation from balance. We measure the mean squared error (MSE) between the bicuculline responses recorded *in vitro* and the responses of networks. In 20-80% networks, the best matching responses came from the models with nearly balanced E/I connections. In 0% and 10%, the best parameters are found to be in the excitation-dominated region and in the inhibition-dominated region respectively. The network model with 100% inhibitory neurons fails to show a response to bicuculline with the magnitude observed *in vitro*. Altogether, for all but extreme fractions, the networks faithfully reproduce the responses to increasing bicuculline concentrations, with the best matching parameters either at the maximum a posteriori estimate or situated slightly closer to the balance point.

2.3 Discussion

Neuronal networks *in vitro* often maintain their bursting dynamics even in the presence of prolonged perturbations (Kaufman et al., 2014; O'Leary et al., 2010). One of the main findings of this study is that a neuronal network of hippocampal neurons self-organizes towards a state attractor characterized by an E/I balance. This tendency is observed in different cellular E/I ratios, covering a range of approximately 10% to 80% of inhibitory neurons. Thus, these networks adapt and develop stable spontaneous network activity that is similar to the control cultures. We suggest that the main mechanism behind this stability is the development of balanced E/I connectivity.

2.3.1 Adjusting the connectivity to maintain E/I balance

The theoretical role of E/I balance in network dynamics has been established in earlier works (N. Brunel, 2000; Renart et al., 2010; Vreeswijk & Sompolinsky, 1998) and it is believed to be one of the important features of highly recurrent networks in the brain (Renart et al., 2010 and see more on that in Chapter 4). Experimental studies demonstrated dynamic E/I balance (e.g. closer correlation of excitatory and inhibitory inputs to a neuron) in different brain circuits such as the neocortex, hippocampus, and spinal cord (Berg et al., 2007; Isaacson & Scanziani, 2011; G. Liu, 2004; Okun & Lampl, 2008). Furthermore, it has been shown that the dynamic balance of excitatory and inhibitory currents relies on balanced E/I connectivity (Barral & D Reyes, 2016; G. Liu, 2004). Most of the networks that have balanced E/I dynamics also have a tightly preserved ratio of excitatory and inhibitory neurons (Sahara et al., 2012).

Our results are aligned with earlier studies on the development of input connectivity (G. Liu, 2004). Liu (2004) reported that the ratio of E/I connections on a dendrite of a single neuron is fixed and that E/I currents balance each other already at the single dendrite level. Liu also proposed that the balance is controlled by a compensatory push-pull mechanism. That was established by the chemical perturbation of E/I balance with blocking or activation inhibition (with bicuculline and flunitrazepam) or blocking of excitation (with CNQX). Our study indicates that cultured networks tend to maintain E/I balance even under the challenge of extreme and long-term changes in cellular composition.

We found that this structural E/I balance is regulated by changes in network connectivity. The patch clamp and modeling results show that structural E/I balance in the dissociated hippocampal cultures is achieved by adjusting the number of connections in networks that have an increased number of inhibitory cells. First, we observed that the frequency of PSCs decreases with the percentage of inhibitory neurons. Simultaneously, the PSCs size does not significantly change. This suggests that the number of connections decreases in proportionality to the E/I ratio (Fig. 2.14).

We also explored this possibility by fitting a network model to reproduce the collective dynamics in cultures with varying fractions of inhibitory neurons. The model fit indicated that keeping the structural E/I balance by changing the number of connections is sufficient to account for the experimentally observed differences in the network dynamics.

Our results complement earlier studies of systematic manipulation of the network structure (Barral & D Reyes, 2016; Ivenshitz & Segal, 2010; Wilson et al., 2007). Ivenshitz and Segal (Ivenshitz & Segal, 2010) showed how changing the density of hippocampal cultures affects spontaneous activity and linked these changes to the differences in synaptic organization. In

their study, the average duration of IBIs, as well as the burst amplitudes, increases in sparser networks. This change is governed by an increase in the amplitudes of the PSCs, along with a decrease in the number of connections, which is opposite to the synaptic changes that we found.

Wilson et al. (Wilson et al., 2007) studied the effects of the network size on connectivity and dynamics. They show that the firing rates of neurons were preserved across networks of different sizes even though the number of neurons and synapses increased. The result was due to the fact that the number of excitatory inputs was compensated by synaptic strength. Interestingly, they showed that amplitudes of mEPSCs are larger in the smaller cultures while, in our case, we do not see substantial differences in mEPSC amplitudes. What is most striking in our networks is the changes in mEPSC frequencies, which do not change in their study.

Barral and Reyes (Barral & D Reyes, 2016) extended this result to cortical cultures of various densities. They showed that the synaptic strength changes as the inverse of the square root of the number of connections while preserving the balance between incoming excitatory and inhibitory connections as predicted by theory (Vreeswijk & Sompolinsky, 1998). Alternatively, homeostatic plasticity theory (Turrigiano, 2011, 2012) predicts that neurons adjust their excitability in order to maintain firing under prolonged perturbation (O'Leary et al., 2010). In contrast, our results indicate that long-term changes in the cellular E/I ratio radically change the input connectivity, rather than the synaptic strength.

2.3.2 Inhibition controls variability

We have found that while the number of inhibitory neurons does not affect the average spontaneous activity in the cultured network, it strongly affects its variability. The CV of IBIs grows linearly with the number of inhibitory neurons in the culture. Thus, networks with the highest numbers of inhibitory cells and only a small fraction of excitatory cells have the CV that approaches one. The bursting dynamics in these networks thus resemble a Poisson process. This is possibly a consequence of a smaller number of connections in networks with high numbers of inhibitory neurons, which decorrelates the input into neurons and randomizes bursting events (Lonardoni et al., 2017). This could obviously have an impact on the reliability of neuronal circuitry.

Inhibition can also directly influence the bursting variability, as indicated by several theoretical and experimental studies (Gigante et al., 2015; Giugliano et al., 2004; Sanchez-Vives et al., 2010). An increase in the inhibitory synaptic strength or number of connections makes inputs less synchronized, which consequently decreases the burst amplitude and increases the variability of IBIs.

2.3.3 Bursting is an interplay of adaptation and inhibition

Our theoretical analysis shows that the interaction of slow spike-frequency adaptation and inhibition can efficiently describe the main temporal aspects of bursting dynamics. We use a simplified network model to link network properties with the features of its collective dynamics. Note, however, that we ignore the complex interplay of dynamics and topology in burst initiation that might account for additional variability (Okujeni et al., 2017; Orlandi et al., 2013; Yamamoto et al., 2022). This model has two major components: a very slow adaptation current coupled with the mean-field recurrent activity. Bursting in this context is described by bistable dynamics that are controlled by the adaption current (Giugliano et al., 2004). Generally, when the adaptation current is small, the noise inevitably pushes the network into a synchronous high firing rate state and the firing rate adaptation eventually terminates the burst. The introduction of inhibition decreases the network synchrony and balances the recurrent excitation, which in turn reduces the size of the adaption current needed to terminate the burst and increases the chance that a burst would be terminated earlier. The presence of this interplay allows us to precisely predict an increase of the IBI as the response of the network activity when the inhibition is blocked pharmacologically.

We have implemented a very general form of spike-triggered adaptation with the timescale on the order of seconds. In any real network, spike-frequency adaptation with such a timescale could be implemented by a wide variety of physiological mechanisms. Examples are slow AHP (Sanchez-Vives et al., 2000), persistent leakage Na-currents (Penn et al., 2016) or depletion of synaptic vesicles (Orlandi et al., 2013). We also cannot exclude the possibility that multiple feedback mechanisms contribute to bursting dynamics (Fardet et al., 2018; Pozzorini et al., 2013). However, all these mechanisms would have a very similar effect on spike-frequency adaptation, making it practically impossible to discriminate them within the scope of the current study.

2.3.4 The functional benefits of maintaining the excitation/inhibition balance

The specificity of the cellular excitation/inhibition ratio remains unclear. There are several recent attempts to provide theoretical frameworks that would explain the functional implication of various E/I ratios (Alreja et al., 2022; Capano et al., 2015). For instance, Capano et al. (2015) analyzed how different fractions of inhibitory cells affect multitask learning in a network of binary neurons. The authors found that the performance is optimized in a network with about 30% inhibitory units. The network, however, does not include the differences between synaptic

strengths of excitatory and inhibitory cells (Barral & D Reyes, 2016). More recently, Alreja et al. (2022) analyzed the performance of the efficient coding model. They show that the representation of sensory stimuli is optimal in terms of the performance and metabolic costs at the physiological E/I ratios given a volume constraint. This work makes predictions about the changes in optimal E/I composition given the neuronal densities, which could indeed be further tested using *in vitro* models that allow us to flexibly control neuronal densities and examine the excitatory/inhibitory connectivity. Also recently, Mongillo et al. (2018) theoretically showed that inhibitory connectivity is disproportionally important for memory storage capacity. Our new experimental platform allows us to precisely control the E/I network architecture and opens new possibilities for detailed studies of E/I circuitry. I will return to the discussion of the functional benefits of the balance E/I in Chapter 4

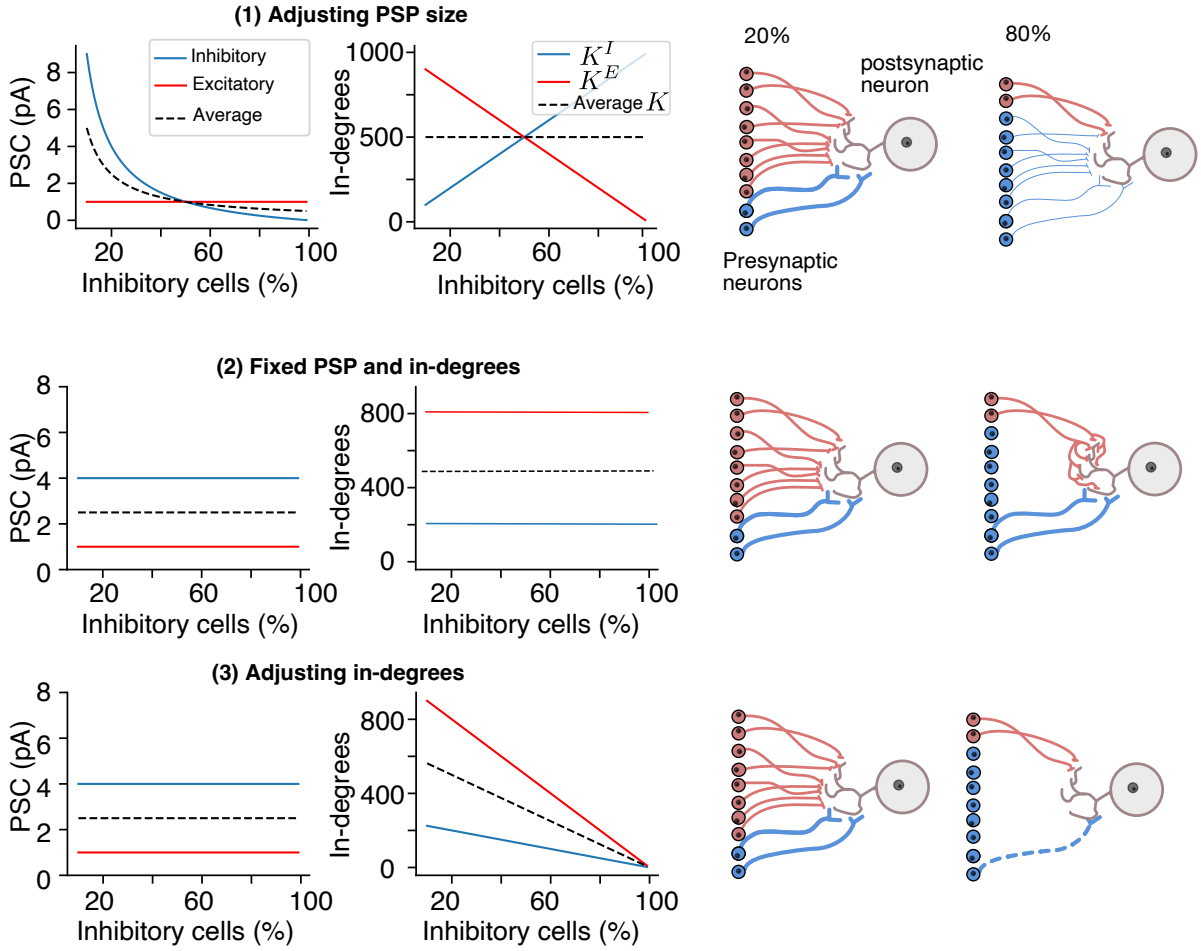


Figure 2.14: Theoretically, the balance of excitation and inhibition could be achieved in different ways. **(1)** Synaptic strength balance: the in-degree per neuron is always proportional to the number of neurons and the strength of inhibitory synapses changes. **(2)** Fixed connections balance: synaptic strengths and a number of in-degrees are always the same. Neurons maintain the same number of inputs and the ratio between E/I inputs. **(3)** Connections balance: the number of in-degrees is proportional to the number of excitatory neurons, inhibitory connections follow the excitatory ones. The synaptic strength does not change. The results of the patch clamp experiments show that the average input per neuron decreases as we increase the inhibitory percentage, while the PSC amplitudes do not change. Modeling shows that to maintain the population burst the inhibitory connections should balance excitatory connections. Together, this shows that our results are most consistent the option (3).

Chapter 3
Principles of network bursting dynamics

Yes, these invariances are worth exploring, but they are features of the model, not the real networks.

- Victor Buendía

...Thus, it is not a model of anything in particular, it is a theory for a class of things that can happen, and within that theory there are questions such as how one should set the (many) parameter values. Stated this way, the question is internal to the theory, but then we can jump to suggest that this is a problem that neurons themselves actually need to solve.

- William Bialek, 2018

Summary

Network bursting is a form of collective dynamics frequently observed in cultures of dissociated neurons. Such dynamics are considered a marker of successful network maturation *in vitro*. Bursting dynamics emerges in cultures of neurons dissociated from a mature or developing brain as well as in cultures of human pluripotent stem cells-derived neurons (hPSC).

In this chapter, we use a macroscopic rate model with a neural adaptation mechanism to examine the dynamical states underlying network bursting. We found that network bursting can occur in bistable, excitable, and oscillatory states. Using simulation-based inference, we fit the model to match the activity of a wide range of cultures and characterize the differences between cortical, hippocampal, and hPSC cultures. We found that cortical cultures have consistently higher network excitability compared to the hippocampal cultures and it is possible that cortical cultures develop oscillatory network dynamics, whereas hippocampal exhibit only excitable dynamics. Cultures of hPSC neurons, in contrast, exhibit oscillatory dynamics early on in their development and later transition towards an excitable state.

Overall, we found that despite the discrepancies between preparations mature networks *in vitro* often exhibit either noisy oscillatory or excitable dynamics.

Contributions disclaimer: In this project I designed the study, performed the mathematical analysis of the model, and the simulation-based fitting. Victor Buendía helped with the mathematical analysis of the model. Emmanouil Giannakakis performed a feature-based burst classification analysis. I recorded the data at the Weizmann Institute of Science with help from Shlomo Ron, Eyal Weinreb, and Elisha Moses. Cortical cultures from HIH were recorded by me and Betül Uysal, under the supervision of Holger Lerche.

Contents

3.1 Introduction

Cultures of dissociated neurons are one of the major vehicles for basic research in single-cell and neuronal network physiology and pathophysiology. Neuronal networks *in vitro* help to examine basic properties of network dynamics (Beggs & Plenz, 2003; Penn et al., 2016), plasticity mechanisms (Bi & Poo, 1999; O'Leary et al., 2010), mechanisms of network formation (Breskin et al., 2006; Okujeni et al., 2017), and the development of E/I balance (Barral & D Reyes, 2016; G. Liu, 2004; Soriano et al., 2008; Sukenik* , Vinogradov* et al., 2021). Cultures of embryonic and neonatal rodent neurons are a classical tool to study the interaction between single-cell properties and network activity (Gal et al., 2010). Recent progress in stem cell research helped to establish a number of experimental protocols for growing and maintaining functional networks of hPSCs both as 2D cultures as well as brain organoids (Hoffmann et al., 2019a; Hyvärinen et al., 2019; Trujillo et al., 2019). These new platforms allow us to examine how specific human genotypes affect single-cell and network properties.

Cultures of neurons develop network activity shortly after the neurons in a dish start forming connections (Soriano et al., 2008; vanPelt et al., 2004; Wagenaar et al., 2006). The activity is usually characterized by large synchronous network events (e.g. network bursts) that propagate through the whole network followed by long periods of quiescence. This type of collective dynamics *in vitro* has been found across different preparation and cultured neuron types (Charlesworth et al., 2015; Hyvärinen et al., 2019; Soriano et al., 2008; Trujillo et al., 2019). Networks of cultured hPSC also appear to exhibit network bursting at least in some protocols (Hyvärinen et al., 2019) and often bursting activity is considered to be a marker of the successful network formation (Cabrera-Garcia et al., 2021).

The variability of network bursting is reported to be large both within and between experiments (Wagenaar et al., 2005). Network activity appears to be affected by both experimental protocols for growth and maintenance (Eytan & Marom, 2006; Maeda et al., 1995; Okujeni et al., 2017; Soriano et al., 2008) as well as the individual features of single cells that form the cultures (Ahtiainen et al., 2021; Charlesworth et al., 2015; Hyvärinen et al., 2019). Thus, network bursts sometimes can behave like Poisson-like network spikes and sometimes can appear as organized quasi-oscillatory bursts (Giugliano et al., 2004). Bursting also dramatically changes in the course of development. Thus, collective dynamics typically manifest with small and sparse bursts and then transform into more frequent and sometimes complex network events (Cabrera-Garcia et al., 2021; Charlesworth et al., 2015; Wagenaar et al., 2005).

The exact mechanism for the bursting activity as well as whether there is a unique network

mechanism across preparations remains unclear (Fardet, 2018). Thus, it has been suggested that bursting is governed by random synaptic transmission and terminated by neuronal adaptation governed by slow AHP currents (Giugliano et al., 2004; Maeda et al., 1995; Robinson et al., 1993). For small *in vitro* cultures, it has been shown that the main neural mechanism that controls bursting is the depletion of synaptic resources (D. Cohen & Segal, 2011). Also, some authors indicate a critical role of a persistent sodium current and pacemaking activity of individual neurons (Penn et al., 2016; Suresh et al., 2016).

Models of population bursting *in vitro* typically account for the fast recurrent activity of neurons as well as slow adaptation mechanisms. A variety of these slow adaptation mechanisms has been proposed, including (Segev et al., 2001), dynamics of synaptic resources (Gigante et al., 2015; Orlandi et al., 2013; Yamamoto et al., 2022), homeostatic plasticity (Zierenberg et al., 2018), slow firing rate adaptation (Fardet et al., 2018; Ferguson et al., 2015; Giugliano et al., 2004; Masquelier & Deco, 2013; Sukenik* , Vinogradov* et al., 2021). Most of these models have been shown to reproduce various aspects of bursting dynamics such as the distribution of network event times as well as burst sizes and burst shapes. Thus, the presence of a slow mechanism that controls the activity propagation is necessary to model the temporal distributions of network events (Giugliano et al., 2004; Masquelier & Deco, 2013; Segev et al., 2001; Sukenik* , Vinogradov* et al., 2021).

Simple models that include noise-driven slow-fast dynamics can exhibit transitions from low to high firing states that can resemble network bursts (Jercog et al., 2017; Levenstein et al., 2019; Mattia & Sanchez-Vives, 2012). Assuming that the network bursting can be well modeled using models of slow-fast dynamics it remains unclear whether there is an associated dynamical state. Furthermore, it is unclear if differences between different preparations, culture types, and development can be attributed to the changes in dynamical states.

In this project, we model network events in different preparations of cultures of dissociated neurons *in vitro* using a mean field model of the recurrent neuronal population with slow firing rate adaptation. The model allows us to directly examine the dynamical states associated with parameter combinations. We fit the model parameters to reproduce the key temporal statistics of network bursts recorded *in vitro* and analyze the differences in dynamical states in cortical, hippocampal, and one type of hPSC cultures from openly available datasets (Charlesworth et al., 2015; Hyvärinen et al., 2019; Kapucu et al., 2022) as well as from novel recordings (Appendix A.2). We find that rodent cultures exhibit either oscillatory or excitable dynamics. Mature cultures of hippocampal neurons are most consistent with excitable dynamics whereas the cultures of cortical neurons can show oscillatory dynamics. Over the course of development,

cultures of primary rodent neurons typically increase their excitability which drives the changes in dynamics. The cultures of hPSC, in contrast, appear to decrease their excitability and transition from constant firing to population bursting.

3.2 Results

3.2.1 Network Bursting driven by adaptation and network noise

To analyze the network bursting dynamics, we focus on a simplified model of the recurrent rate (x) and neuronal adaptation (w) that can exhibit burst-like activity. The firing rate of the model follows:

$$\tau \dot{x}(t) = -x(t) + \phi[-a(cx - w + \theta)] + \sigma\eta(t) \tag{3.1}$$

$$\tau_w \dot{w}(t) = -w(t) + bx(t) \tag{3.2}$$

where ϕ is a sigmoid activation function

$$\phi(z) = \frac{1}{1 + e^z}. \tag{3.3}$$

and $\eta(t)$ is the Gaussian white noise. The model has a number of parameters: τ is the firing rate evolution timescale, a defines the gain of input-output relationship, c is the strength of recurrence, θ sets the model excitability, σ is the noise strength, τ_w defines the relative timescale of the adaptation variable, and b is the adaptation strength. We typically set the recurrence parameter $c = 1$ and the gain $a = 5$ unless specified otherwise. The timescale of the rate variable τ was typically set to 20ms. The model that we study here is similar to the noise-driven Fitz-Hugh-Nagumo model for single-neuron activity (Lindner & Schimansky-Geier, 1999).

We assume that the spiking activity is given by an inhomogeneous Poisson process with a rate $x(t)$, which is defined as follows:

$$y_i(t) \sim \text{Poiss}\left[\exp(C_i x(t) + r_i)\right] \tag{3.4}$$

where the the C_i is the weight for a neuron i and r_i is the constant offset. Across the chapter, we typically find the C_i and r_i using a Poisson regression or set them manually. Note, however, that the underlying rate does not depend on these parameters. We analyzed the deterministic behavior of the model using linear stability analysis (Strogatz, 2018, see details in Appendix F.1) and used simulation-based techniques to study the interaction between deterministic states and the noise-driven output of the model.

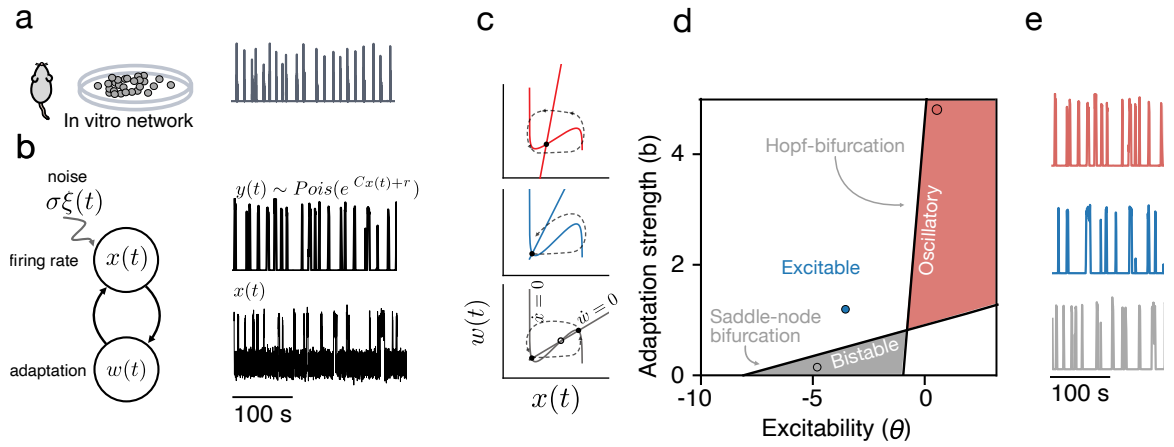


Figure 3.1: Reduced model of bursting activity and redundancy of dynamical states that can govern *in-vitro*-like bursting. **(a)** Cultures of dissociated neurons *in vitro* often exhibit network bursting. The gray trace shows the total spike count of activity recorded in hippocampal cultures at 28DIV (Charlesworth et al., 2015).**(b)** Schematics of the noise-driven slow-fast model. The modeled population activity that drives the network spiking is denoted by $x(t)$, and $y(t)$ are the generated spike counts ($C = 0.9, r = 9$). **c,d,e** The model can generate network bursting that closely resembles the real data in oscillatory, excitable, and bistable states. **c** Nullclines of the model in the three state **(d)** Bifurcation diagram of the model in θ-b space. **(e)** Examples of the generated spike counts for the model in 3 dynamical states. Parameters in the simulations are bistable state $b = 0.073, \theta = -4.7, \tau_w = 4s, \sigma = 1.6, J = 9, a = 5$, excitable state:$b = 1.2, \theta = -3.4, \tau_w = 32s, \sigma = 1.56, J = 9, a = 5$, oscillatory state:$b = 4.9, \theta = -0.56, \tau_w = 146s, \sigma = 1.57, J = 9, a = 5$. Statics of the bursting in the model mean IBI=7s, burst duration=1s, CV(IBI)=0.7.

3.2.2 Population bursting can be governed by bistable, excitable, or oscillatory dynamics

The noise-driven model with recurrence and slow adaptation can exhibit bursting-like dynamics in three different dynamical states: bistable state, excitable state, and oscillatory state. In all three states, the model can qualitatively match the network bursting recorded *in vitro* (Fig. 3.1).

In the case of *bistability*, noise forces the system to jump between the states. The frequency of jumps is then determined by the noise intensity and modified by the slow adaptation. Note that, when $b = 0$, the system exhibits classical bistable dynamics and the transitions between up and down states can be seen as the escape times in a double well potential, where the relative depth of each well is controlled by θ.

In the *oscillatory state*, the noise can force the transition from down to up state or vice versa earlier than the full period of the oscillation, which leads to irregular oscillatory dynamics.

Summary	Definition
Mean IBI (μ_{IBI})	$\frac{1}{N}\sum_{i=1}^{N-1}(t_i^{\text{start}} - t_{i-1}^{\text{end}})$
CV of IBI	$\frac{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(t_i^{\text{start}} - t_{i-1}^{\text{end}} - \mu_{\text{IBI}})^2}}{\mu_{\text{IBI}}}$
Mean burst duration (μ_{dur})	$\frac{1}{N}\sum_{i=1}^{N} t_i^{end} - t_i^{start}$

Table 3.1: We denote the time when burst i begins as t_i^{start} and the end of a burst by t_i^{end}

Finally, when the system has only a single low-activity fixed point (Fig. 3.1c), noise can perturb the dynamics such that it first has to transition towards the upstate (e.g. the left branch of the nullcline depicted in Fig. 3.1c) and then it relaxes back to the fixed point. An analogous *excitable state* exists also for the high fixed point.

3.2.3 Temporal statistics of population bursting discriminates different dynamical states

We next quantify the temporal bursting activity in the model and use it to classify different states. Using summary statics of bursting is one of the most simple ways to characterize phenotypes of the culture. We use three summary statistics: mean inter-burst intervals (IBI), burst duration, CV of IBI (see in Table 3.1).

To check if the dynamical states are separable based on the summary statistics, we broadly sample the model parameters. We sample from 4-dimensional uniform prior for b, θ, τ_w, σ (Table 3.2).

Parameter	Lower Bound	Upper Bound
b	0.05	20.0
θ	-10.0	15.0
τ_w	$\log(200.0)$	$\log(200000.0)$
σ	0.01	2.0

Table 3.2: Prior values for parameters.

Then we ran the simulations and computed the summary statistics (Fig. 3.2a). We obtained a $3 \times N$ table of summaries, where N is the number of samples, and a corresponding N vector

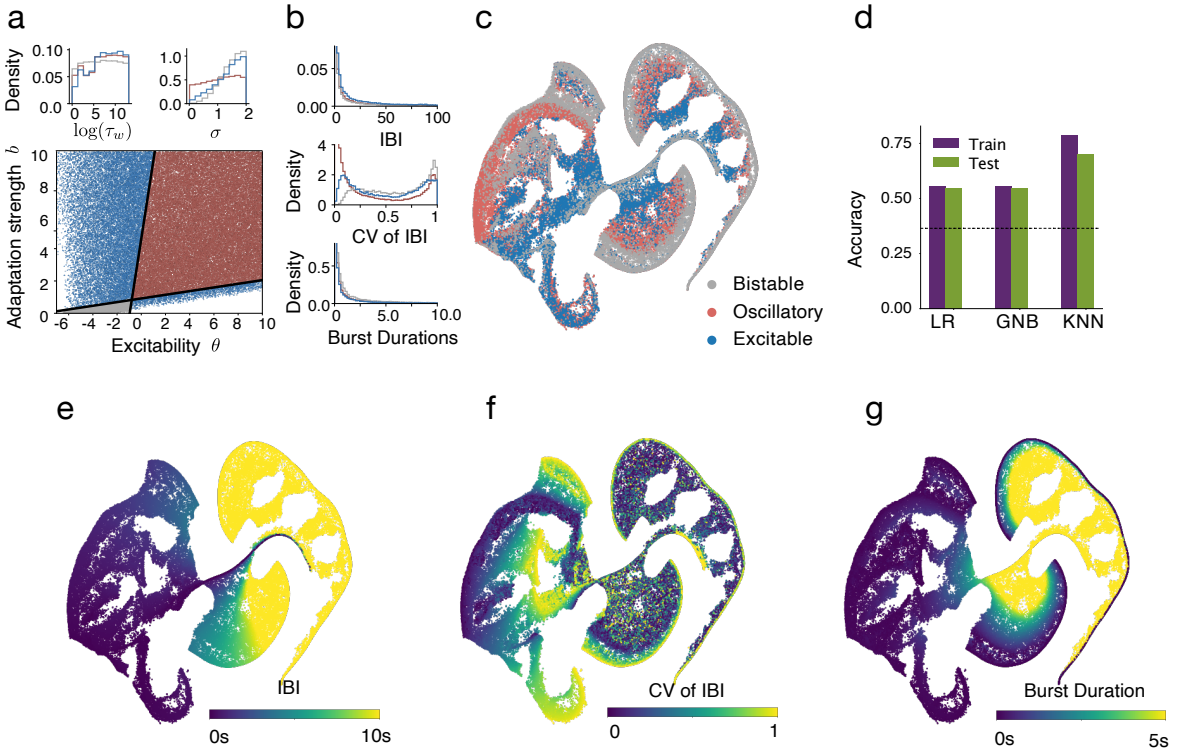


Figure 3.2: Classification of the dynamical states that govern bursting activity based on the bursting summary statistics. **(a)** Samples of the parameters that give rise to bursting activity (red – oscillatory, blue – excitable, grey – bistable). Overall, we took 20000 valid samples per state for the classification analysis. The top panel shows the marginal distributions of sampled parameters for the timescale of adaptation in log scale (τ_w) and the noise intensity (σ). Note bursting activity emerges in excitable and bistable states only when the noise is sufficiently strong. The oscillatory state can exhibit up-down transitions even without any noise **(b)** Distribution of bursting summary statistics for the three dynamical states. There were no apparent differences in the distributions apart from the distribution CV of IBI. Thus, bursting with CV close to zero appears predominantly in the oscillatory state. **(c)** 2D tSNE embedding of the bursting summary statistics indicates that the states are not linearly separable and do not form isolated classes. Each point is a 2D embedding for 3 summary statistics with colors indicating different dynamical states. **(d)**. Accuracy of three classifiers trained to discriminate the dynamical states based on the bursting statistics. The best performance is achieved by the KNN classifier. Logistic Regression (LR) and Gaussian Naive Bayes (GNB) are worse and only slightly higher than chance, indicating that linear separation between the states based on the bursting summary statistics is not reliable. **(e ,f ,g)** show the distribution of the bursting summary statistics on the tSNE embedding.

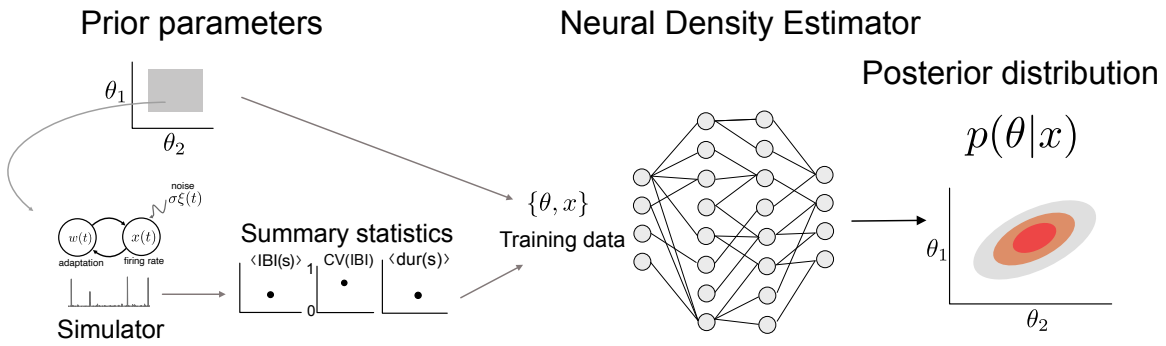


Figure 3.3: Simulation-based inference with a neural density estimator

of labels, indicating whether the summary corresponds to the excitable, oscillatory, or bistable dynamics. We then fit a logistic regression (LR), Gaussian naive Bayes (GNB), and KNN classification to the summaries.

We found that the KNN classifier can distinguish all states with a classification accuracy of around 0.7 on the test set (Fig. 3.2d). Linear classifiers perform only slightly above chance. This suggests that there is no clear linear boundary that separates the dynamical states based on the burst summary statistics. However, the states can be separated using the information about a neighborhood of the burst summary. We further visualized this using t-SNE (Fig. 3.2c).

3.2.4 Simulation-based inference of the model parameters

In order to directly determine how the bursting statistics map onto the model parameters, we apply a simulation-based inference approach. In contrast to the Approximate Bayesian Computation (ABC) used in the previous Chapter 2, we use the surrogate model approach to approximate the posterior distribution of parameters (Fig.3.3 and see also Chapter 1). The main difference is that we use the simulations to build a conditional density estimator p(paramters|summary) that can be evaluated for other summary statistics without additional sampling (Cranmer et al., 2020; Gonçalves et al., 2020).

Such conditional density estimators can be built, for example, using classical Mixture Density Networks (Bishop, 1994), more recently developed Masked autoregressive density estimators (Germain et al., 2015) or normalizing flows models (Papamakarios et al., 2017). Here, we use a neuronal spline flow (Durkan et al., 2019) within the SBI-toolbox (Tejero-Cantero et al., 2020). The model included 5 transformations and 50 features per layer.

We fit the conditional density estimator using 300,000 samples from a wide prior of parameters(Table 3.2). The estimator allows us to obtain an approximate posterior probability of the model parameters given a set of summary statistics ($\mathrm{p}(\mathrm{parameters}|\mathrm{ibi}, \mathrm{CV\ of\ IBI}, \mathrm{duration})$). In contrast to rejection-based multi-round inference used in the previous chapters, fitting a simple density estimator us to efficiently use all 300,000 samples and fit a density estimator that can be evaluated for different summary statistics, this approach is also called amortized inference (Cranmer et al., 2020).

We evaluated the predictive performance of the model and found that it achieves excellent results for a wide range of summary statistics (Fig. 3.4). We run the predictive checks analysis by randomly sampling parameters from a uniform prior, simulating the model with these parameters to get a set of summary statistics. Then we obtained parameter samples from the inferred posterior distribution, ran the simulation with these parameters, and compared the results with the ground truth summary statistics. Summary statistics of the bursting activity for the inferred parameters almost perfectly matched the ground truth (Fig. 3.4).

Next, we evaluated if the ground truth parameters were included in the high-density region of the fitted posterior (Fig. 3.4b). We found that the ground truth parameter was typically located within the high density of the posterior probability. However, the inferred approximate posterior distribution also clearly indicates a number of invariances of the bursting summary with respect to the model parameters. Thus, the excitability, adaptation strength, and adaptation timescale appear to linearly depend on each other (Fig. 3.4b). In the next section, we use the insight of the SBI and analytical methods to define the invariant manifolds of model parameters where the summary statistics stay constant.

3.2.5 Invariant manifolds of bursting dynamics

Using the results of SBI, we found that the model can produce invariant bursting activity for a wide range of values of the excitability (θ), adaptation strength (b), and the timescale of adaptation (τ_w). An increase in the excitability can be counteracted by a stronger (b) and slower (τ_w) rate adaptation. Thus, the invariance emerges in a 3D space of the model parameters. We explore this dependency analytically, by considering the dependencies pair-wise.

The first observation that we made was that the θ and b strongly linearly depend on each other. The slope of $b - \theta$ line depends only on the ratio between IBI and burst durations (Fig. 3.5a).

We observed a similar change in dependency between the excitability θ and the adaptation timescale τ_w. We found that for large values of τ_w they are also almost linear (Fig. 3.5b).

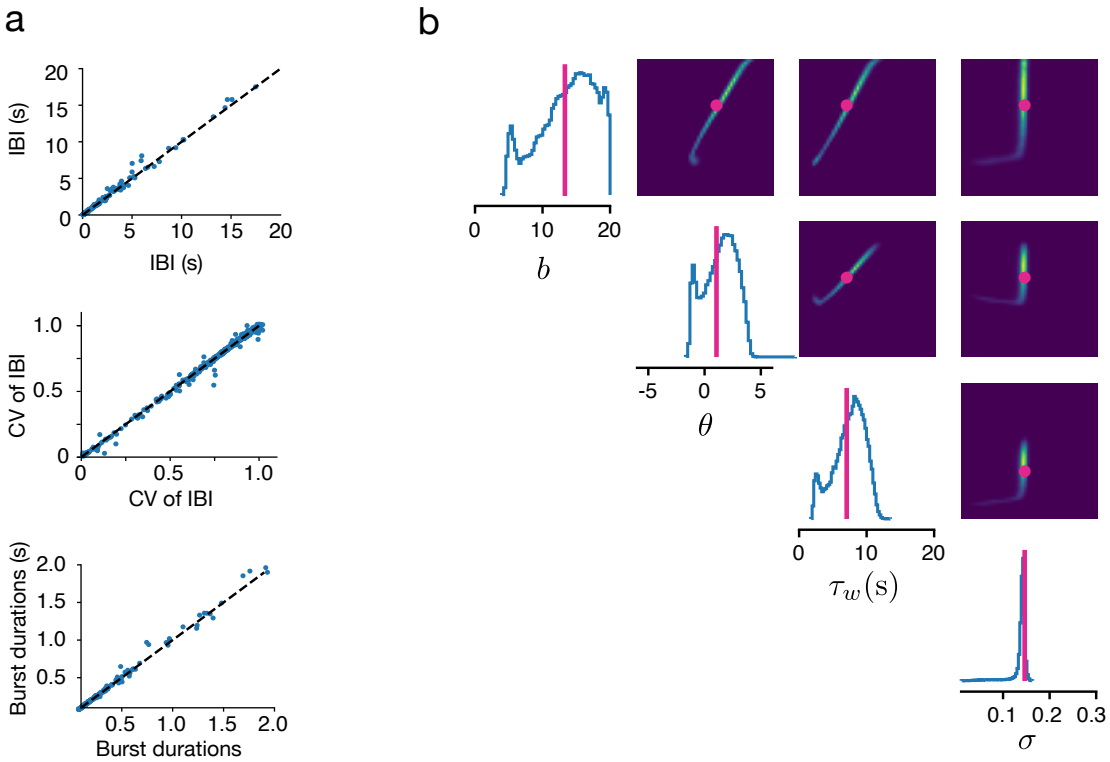


Figure 3.4: Predictive performance of the amortized posterior fitted with 300000 sets of parameters and bursting summaries. **(a)** Shows the ground truth vs the predicted bursting summary statistics for 800 pairs of parameters. To perform the predictive checks analysis, we first randomly sample parameters from a uniform prior $\hat{\Theta}$ and simulate the model using these parameters, which results in a set of summary statistics $\hat{x}$ that follow a distribution given by model($\hat{\Theta}$). Next, we obtain parameter samples from the inferred posterior distribution $p(\Theta|\hat{x})$ and run the simulation with these parameters. Finally, we compare the results of the simulation with the ground truth summary statistics $\hat{x}$. **(b)** Examples of the inferred posterior distribution of the model parameters given a set of summary statistics and the ground truth parameters that were used to generate these summary statistics (purple dots and lines). The histograms on the diagonal – are marginal distributions of the parameters. The diagonals of the plot show marginal distributions of pairs of parameters.

In order to identify the origin of these dependencies, we approximated the sigmoid non-linearity of our model with a piece-wise linear function (see Appendix F.2 for more details). We then focused on a deterministic case and only on the slow oscillatory regime, where up and down transitions happen without noise (e.g. CV of IBI = 0). Using this simplification we found an explicit relationship between the model parameters, IBI, and burst durations. Given, the excitability levels and the required summary statdefistics, we can express the values of b and τ_w

$$b = \frac{w^+(\theta) + w^-(\theta)e^{T^{\text{dur}}/\tau_w}}{x(e^{-T^{\text{dur}}/\tau_w} - 1)} \tag{3.5}$$

$$\tag{3.6}$$

$$\tau_w = -\frac{T^{\text{IBI}}}{\log(\frac{w^-(\theta)-bx}{w^+(\theta)-bx})} \tag{3.7}$$

where w^+ and w^- are the values of the adaptation variable at the end and the beginning of the burst, T^{dur} denotes the burst duration and T^{IBI} is the IBI. Thus, in the case of a deterministic limit cycle, one can always adjust the values of b and τ_w to get the same burst duration and IBIs.

As θ grows to large values, the equation for the $b - \theta$ above approaches a line (Fig. 3.5a and c) given by

$$b \approx \frac{\theta}{\langle x \rangle} + \frac{J}{2\langle x \rangle} \tag{3.8}$$

where $\langle x \rangle$ is an average activity, that can be obtained either numerically or by noting that the x is bounded between its two fixed points, and the amount of time it spends close to each fixed point is given by the ratio of IBI and burst durations. Thus, $\langle x \rangle \approx JT^{\text{dur}}/(T^{\text{IBI}} + T^{\text{dur}})$, if the low fixed point is zero and the upper fixed point is close to J. The ratio of the mean burst duration and the sum of the burst duration of the IBIs $T^{\text{dur}}/(T^{\text{IBI}} + T^{\text{dur}})$ here simply describes the probability of being in the upstate and J typically sets the value of x fixed point in the upstate. Fig. 3.5c shows that both the full piece-wise linear approximation as well as the samples from the posterior approach the line 3.8 for increasing values of τ_w.

The relationship between $\theta - \tau_w$ can also be expanded and the slope of the dependency $\tau_w \approx \frac{aT^{\text{IBI}}\theta}{aJ-8}$. Interestingly, the slope of this invariance depends only on the IBI, but not on the burst duration (note that in Fig. 3.5b the ratio of IBI and duration change by increasing the IBI, while the durations stay constant).

Gaussian white noise makes the transition irregular, and forces irregular burst events in the excitable state as well as the state transitions in the bistable state. Interestingly, this does not destroy the invariance between model parameters but rather modifies the lines (Fig. 3.5d,e). This is probably due to the fact that the values of adaptation at the end and at the beginning of the

burst become stochastic, which changes the positions of the invariances. The $b - \theta$ dependency shifts towards the excitable state and eventually towards bistable states (Fig. 3.5d), but the slope stays the same and still can be found from the analytics.

The slope of $b - \tau_w$ invariance increases with the variability of transitions (Fig. 3.5e) and starts to deviate from the analytical expression in the deterministic case. The increase of the slope depends on the noise intensity (Fig. 3.5f). Obtaining the distribution of the adaptation values at which transitions are happening can help to further resolve the invariance in the deterministic case (Appendix F.3).

Thus, simulation-based inference allowed us not only to reliably fit the model to reproduce bursting summary statistics but also to infer some of the basic parameter invariances in the model that we further explored using analytical techniques. When the adaptation timescale can be arbitrarily set, the model for some combination of summary statistics can exhibit identical dynamics in excitable and oscillatory states. In the following sections, we bound the timescales of the slow adaptation to 15s. to match the experimentally found timescale of network excitability (Gal et al., 2010). This boundary allows us to probabilistically discriminate the excitable and oscillatory dynamics.

3.2.6 Excitable and oscillatory dynamics of neuronal network *in vitro*

To identify the dynamical states in cultures of neurons *in vitro*, we fit bursting dynamics from recordings of cortical, hippocampal, and hPSC cultures (Table 3.2.6). The data were recorded either with multielectrode arrays (MEA) or with Ca-imaging[1]. For datasets **1,2,4,5** the culturing methods and spike detection methods can be found in the corresponding publications (Charlesworth et al., 2015; Hyvärinen et al., 2019; Kapucu et al., 2022). For experimental methods for datasets **3** and **6**, please refer to Appendix A.2

First, we focused on the average bursting statistics for mature neuronal cultures (between 14 and 28 DIV) of cortical, hippocampal, and hPSC neurons. Despite a large inter-dataset variability, we find that most cultures are excitable or oscillatory (Fig. 3.6a). The model with parameters sampled from the posterior well matched the average bursting statistics (Fig. 3.6b).

[1]This is possible because we fit the model based on the summary statistics of IBI and burst durations that can be extracted from both Ca and MEA recordings

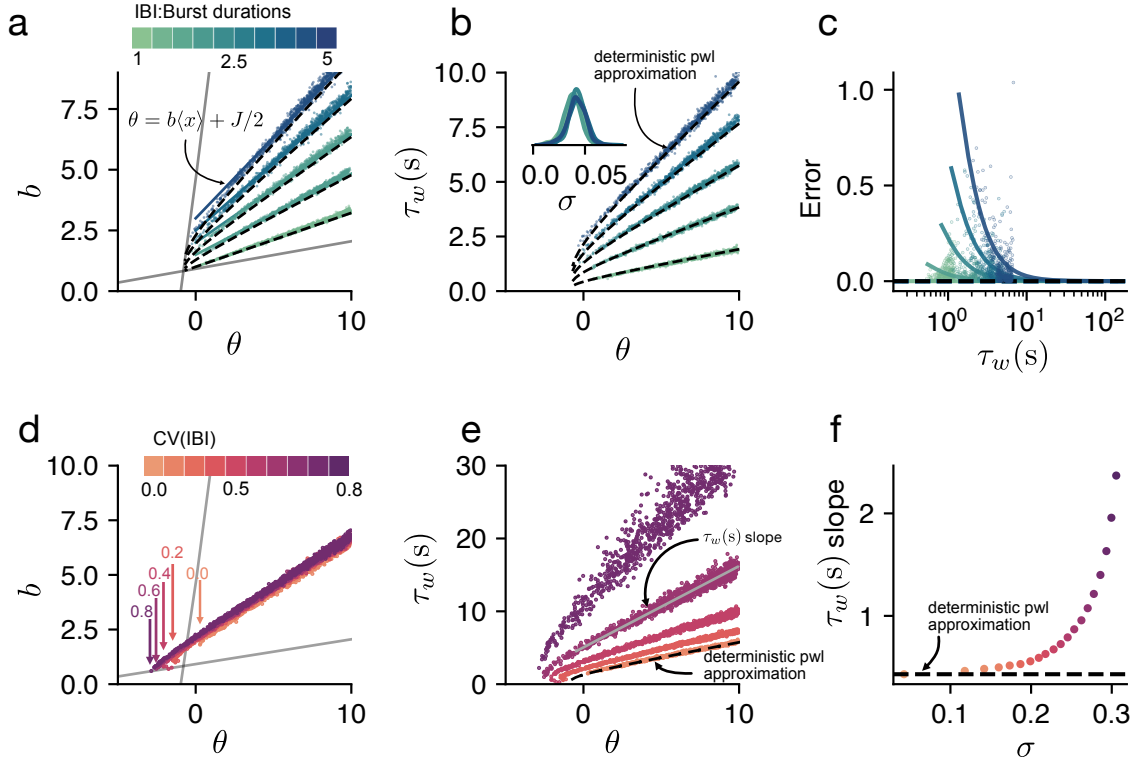


Figure 3.5: Invariant manifolds in the model with slow firing rate adaptation. Increasing overall excitability θ can be balanced by higher adaptation strength and a longer adaptation timescale. (**a**) shows the samples from the approximated posterior for an increasing IBI, while burst duration stays constant and CV is close to 0. The piece-wise approximation of the model dependencies (dashed line) very accurately predicts the growth of the b with τ_w. As excitability increases the dependency approaches the line $\frac{\theta}{\langle x\rangle} + \frac{J}{2\langle x\rangle}$ (see main text), which is also illustrated in **c**. **c** Shows how the adaptation timescale τ_w changes along θ to maintain the invariant bursting statistics. The dependency again can be well approximated with analytics. (**c, e, f**) Illustrate the changes in the dependencies as we keep the IBI and burst duration fixed (3s and 1s accordingly) and only change the variability of IBIs measured with the CV. The slope of the $b - \theta$ dependency stays the same (**c**), although the posterior protrudes deeper towards the excitable state. (**e**) To keep the IBI and burst durations the same, while increasing excitability, the model also has to have longer adaptation timescales for the same values of θ. Thus, the slope of $\theta - \tau_w$ dependency increases and exponentially deviates from the deterministic line (**f**).

Dataset	Source	Type	Recordings	DIV
1	Charlesworth et al., 2015	mice cortex	MEA	8-28
2	Charlesworth et al., 2015	mice hippocampus	MEA	8-28
3	HIH	rat cortex	MEA	17,18
4	Hyvärinen et al., 2019; Kapucu et al., 2022	rat cortex	MEA	1-35
5	Hyvärinen et al., 2019; Kapucu et al., 2022	hPSC	MEA	7-63
6	WIS	rat hippocampus	Ca-imaging	14-27

Table 3.3: Datasets used to fit the model

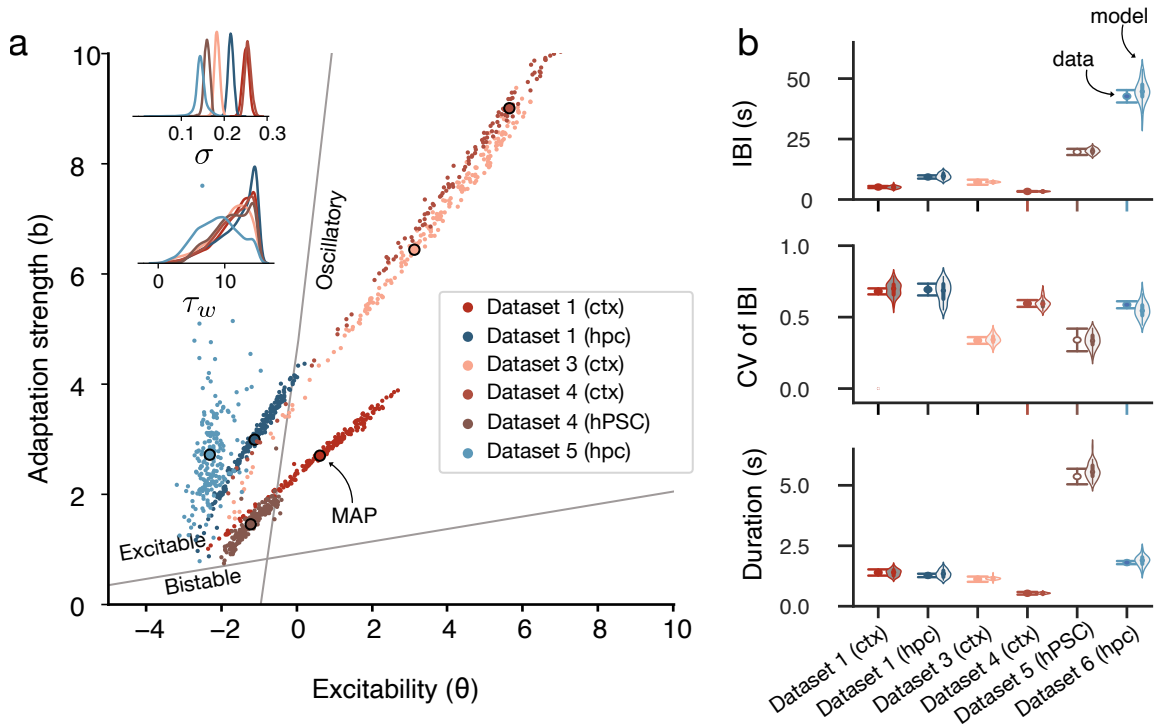


Figure 3.6: Model in excitable and oscillatory states fits the statistics of living neuronal cultures *in vitro*. **(a)** Shows the samples from the approximated posterior for the average bursting summaries across 6 datasets. Points with black edges show the maximum a posteriori (MAP). note that the prior distribution of the adaptation timescales τ_w is bound between 1-15s. MAP for all cortical cultures that we considered were found in an oscillatory state. The MAP for two datasets of hippocampal cultures was found in the excitable state. The most probable solution for a single dataset of hPSC cultures was also found in the excitable state. **(b)** Bursting statistics from the model simulated with the parameters sampled from the priors match the means of IBI, CV of IBI, and burst duration in the data. The main effects were significant between groups for all summary statistics (one-way ANOVA; IBI $F(5,402)= 128.26$, $p<0.0001$, CV of IBI $F(5,402)= 5.8$, $p<0.0001$; Durations $F(5,402)= 87.3$, $p<0.0001$)

Cortical cultures are oscillating and hippocampal are excitable

Next, we specifically analyzed the differences between two standardly used cortical and hippocampal cultures. To reduce the inter-experimental variability, we consider the recordings from a single laboratory (Charlesworth et al., 2015). Here, we found that the model most consistent with cortical cultures was more excitable than the model for hippocampal cultures (Fig.3.7). The maximum a posteriori solution suggests that bursting dynamics in cortical cultures are governed by oscillatory dynamics while bursting in hippocampal cultures is most consistent with the excitable state (Fig.3.7a). The model suggests that this difference is caused by the larger excitability of the cortical cultures. Additionally, the timescales of the adaptation variable were similar but the noise intensity was lower in hippocampal cultures. The observed changes might be consistent with previously reported differences in the input connectivity (Soriano et al., 2008) that would effectively change the overall network excitability.

Previously it has been shown that both culture types exhibit very similar network bursting, which could be, however, distinguished by considering a number of features of bursting statistics (Charlesworth et al., 2015). The reduced model allows us to find differences between the two types of cultures using a very limited number of basic summary statistics.

3.2.7 Development of network excitability

In rodent cultures, network-wide events typically emerge at around 10-14 DIV (Charlesworth et al., 2015; Wagenaar et al., 2005). The hPSC cultures usually start showing stable network events at later stages, typically around 20DIV (although it might strongly depend on the protocol (Estévez-Priego et al., 2023; Hyvärinen et al., 2019). In this section, I analyze how the dynamical states unfold over the course of culture development.

Cortical vs hippocampal cultures

First, we look at the development of the network activity in cultures of dissociated hippocampal and cortical neurons (datasets 1 and 2, Fig.3.8 and 3.6). In both cortical and hippocampal cultures the mean IBIs typically decrease over the course of development, although more prominently in the cortical cultures (Fig.3.6c data, the interaction term). The CV of IBI in both cortical and hippocampal cultures fluctuates between 0.5 and 1. The largest difference was observed around 10-14 DIV. Burst durations showed a small increase throughout development (Fig.3.6c, data).

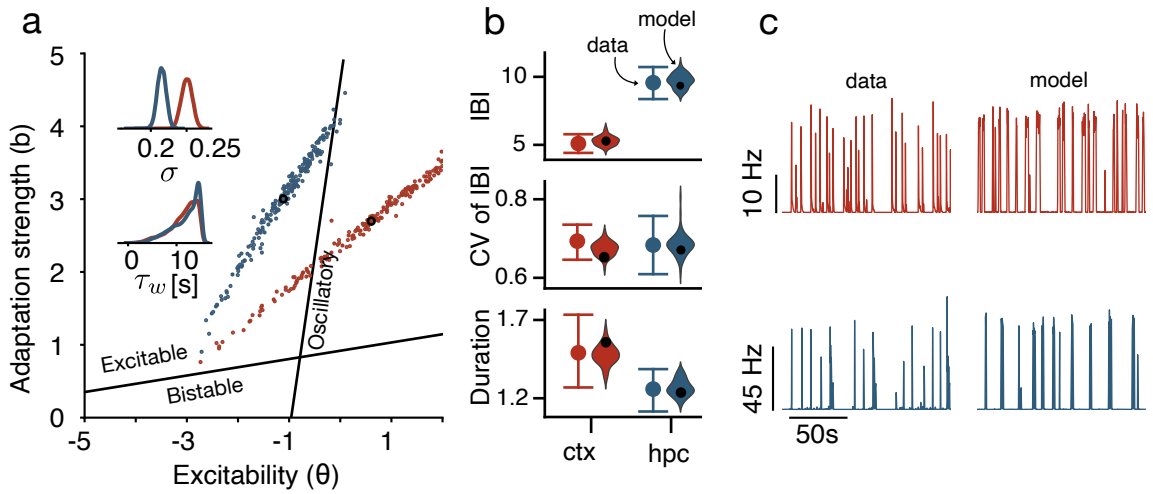


Figure 3.7: Most probable parameters of the fitted model to the average bursting statistics of hippocampal cultures lie in the excitable state, and most probable parameters for the cortical cultures are found in the oscillatory state (averaged from 17 to 28 DIV). **(a)** Infered distribution of parameters of the reduces model. Main figure - samples from the posterior probability plotted jointly for b and θ. Insets show the marginal distributions of the noise intensity and adaptation timescale; Posteriors are conditioned on the average statistics of the hippocampal cultures (blue) or cortical cultures (red). **(b)** Mean bursting statistics of the cortical and hippocampal cultures (error bars - s.e.m.) used to condition the posterior distribution of the model parameters. Mean IBIs were significantly different between cortical and hippocampal cultures (posthoc Tukey HSD, FWER=0.05, p=0.04), we did not see observe significant differences between CV of IBI and burst durations. Violin plots show distributions of bursting statistics for the model simulated with parameters sampled from the approximate posterior distribution. **(c)** Example traces of the dynamics of the model in 200ms bins (model parameters correspond to the maximum a posteriori in **(a)** and Poisson process coefficients are $k1 = 0.7, k2 = -9$ and $k1 = 0.8, k2 = -9$).

We fit the reduced model and analyze the changes in the parameters conditioned on the fixed adaptation strength $b = 2.5$. This way we slice through the invariance discussed in the previous section and consider only one possible combination of θ and τ_w.

The network excitability linearly grew over the course of development in cortical cultures (3.6a) and stayed almost the same with only a small positive trend in hippocampal cultures. The timescale of adaptation fluctuated between 5 and 15 s. Finally, the noise intensity grew in both cortical and hippocampal cultures. Throughout development cortical cultures showed a larger change in the network excitability and possibly transition from the excitable state towards the oscillatory state. Hippocampal cultures displayed excitable dynamics across development.

These results are consistent with earlier percolation experiments that suggested that cortical cultures develop stronger connectivity, which would be associated with higher susceptibility to

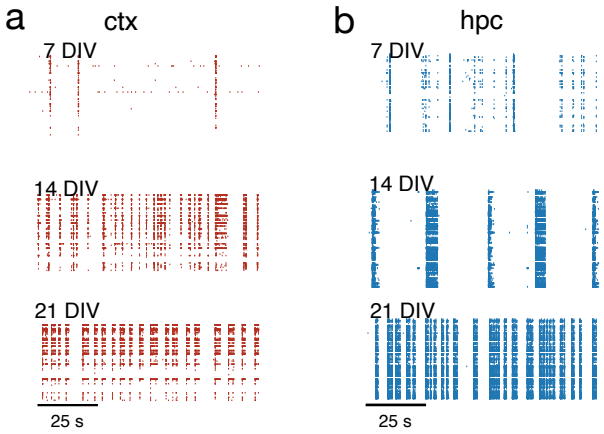


Figure 3.8: Raster plots of the network activity in cortical and hippocampal cultures at 7, 14, and 21 DIV recorded with MEA (Charlesworth et al., 2015)

stimulation and therefore higher network excitability (Soriano et al., 2008).

Rodent primary cultures vs human PSC cultures

Next, we compared the development pattern of rodent cortical cultures and hPSC cultures (dataset 4, 5). The development profile of the cortical cultures was very similar to the one in dataset 1 analyzed in the previous section, although bursting activity emerged slightly later, at around 10 DIV. The cortical cultures were monitored until 35 DIV (Fig.3.10). Cortical cultures in this dataset showed an early decrease in the IBI and an associated decrease of the CV of IBI, which was larger than in dataset 1. Bursts showed a small increase (Fig.3.11c).

Excitability was the main parameter that explained the changes in the network development. It grew initially between 14-25 DIV and then decayed for later stages of development, which might be associated with the decay of cultures. The timescale of adaptation shows a decrease from 10s at early developmental stages and later converges to values around 2s. In contrast to dataset 1, the noise intensity was slowly decaying as cultures matured (Fig. 3.11a,b).

Here, cortical cultures showed a transition to the oscillatory state and later decay back towards an excitable state after which they likely lose the ability to generate network events.

Network bursting of hPSC cultures appeared later during the development and had a different developmental trajectory. hPSC cultures first showed network-wide events at around 20 DIV. Before this point, single channels appeared more active than rodent cortical cultures (Hyvärinen et

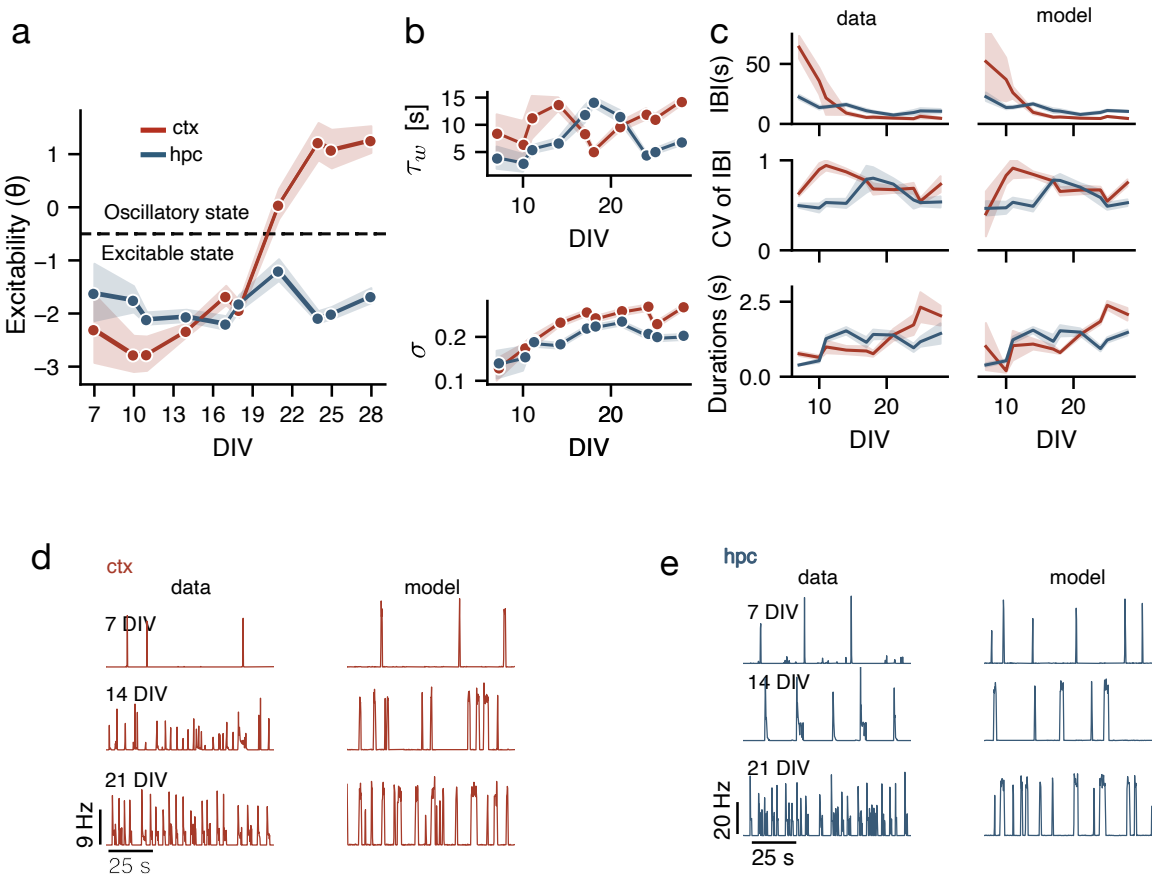


Figure 3.9: The model captures the changes in bursting statistics over the course of development and magnifies the differences between cortical and hippocampal cultures. (**a**) Cortical cultures show a large change in the network excitability and possibly transition to the oscillatory states, whereas hippocampal cultures stay inside the excitable region. The parameters here are conditioned on $b = 2.5$. (**b**) Adaptation timescales τ_w fluctuate between 2 and 15s over the course of development. Noise intensity grows as cultures mature, which is possibly related to the increasing intrinsic noise in the networks as more connections appear. (**c**) Changes in the bursting summary statistics over the course of development in the data(left) and in the model with parameters from **a, b**. For mean IBIs, the effect of DIV was significant, 2-way ANOVA ($F(9,494)=19.05$, $p<0.0001$) as well as the main effect of the culture type ($F(1,494)=4.06$, $p=0.04$), we also saw a significant interaction ($F(9,494)=12.8$, $p<0.0001$). Hosthoc comparisons (Bonferroni corrected) revealed in the case of hippocampal cultures only IBI at early stages were different from the rest. The CV of IBIs were significantly different both over the development and between culture types (DIV $F(9,494)=2.5$, $p<0.007$; Type DIV $F(1,494)=19.3$, $p=0.000013$; Interaction $F(9,494)=3.7$, $p=0.000149$). Average durations showed the effect of the development but we did not see significant differences between the culture types (DIV $F(9,494)=9.2$, $p<0.0001$; Type $F(1,494)=0.26$, $p=0.6$; Interaction $F(9,494)=4.9$, $p<0.0001$). (**d, e**) Comparison of the network spike rate for both the model and data. The model well captures the main changes in the bursting activity, however, it fails to closely reproduce the burst shapes.

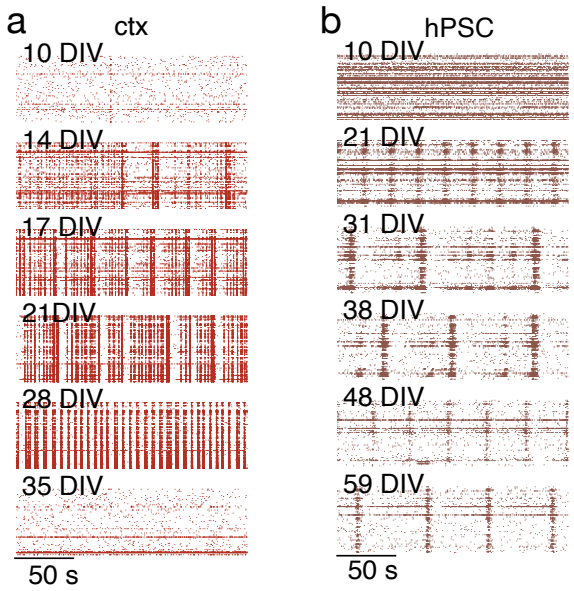


Figure 3.10: Raster plots of the network activity in cortical and hPSC cultures over the course of development recorded with MEA (Hyvärinen et al., 2019; Kapucu et al., 2022)

al., 2019). The mean inter-burst intervals moderately increased between 20 to 60 DIV (Fig. 3.11c). Variability was typically quite low and did not have any development-related trends. The CVs of IBIs fluctuated below 0.5. The burst durations decreased throughout development.

When we analyzed the model parameters conditioned on b, we saw the main change associated with network excitability. However, the change was the opposite compared to the rodent cultures (Fig. 3.11a). Network activity appeared to emerge in the oscillatory state and then slowly decayed toward the excitable state where it stabilized. We hypothesize that desynchronized activity with a high firing rate at the early stages of development in these cultures may be associated with a single hight-activity fixed point state that appears in the model with high excitability and low adaptation strength (Fig. 3.11d). In Fig. 3.11f we show the activity in this state. Adaptation timescale as well as noise intensity showed only a small decrease over the course of development in hPSC cultures.

Our results suggest that hPSC culture has an inverted developmental trajectory. However, at mature states, their dynamics appear to converge to an excitable state and thus might be consistent with standard network bursting observed in rodent cultures.

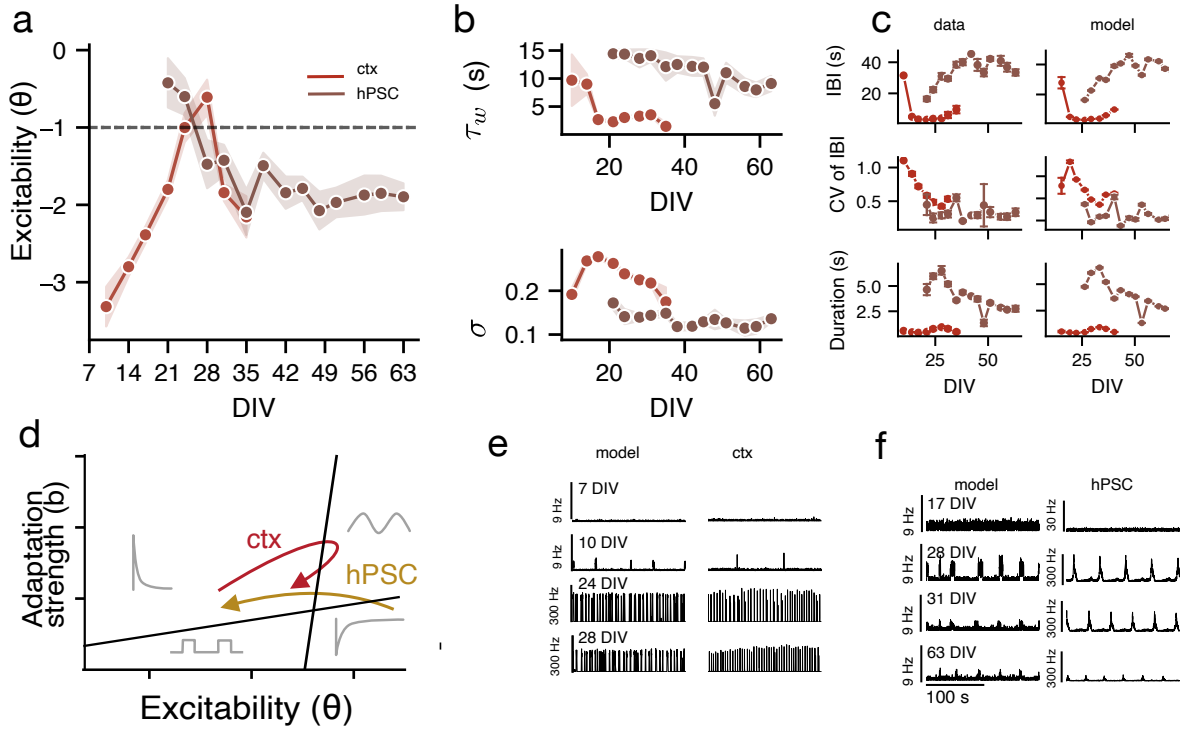


Figure 3.11: Human PSC cultures show an opposite pattern of development compared to primary rodent cortical cultures and appear to decrease their excitability and transition to the excitable state **(a)** The rodent cortical cultures in this dataset consistently show a large change in the network excitability and transition to the oscillatory states. The human hPSC cultures, in contrast, first show oscillatory-like bursting and then slowly transition to the excitable state. The parameters here are conditioned on $b = 2.5$. **(b)** Adaptation timescales τ_w fluctuate between decreased over the course of developments. Noise intensity grows as cultures mature, which is possibly related to the increasing intrinsic noise in the networks as more connections appear. (**c**) Changes in the bursting summary statistics throughout development in the data (left) and in the model with parameters from a,b. IBIs were significantly different both over the development and between culture types, yet the directions of change were different (DIV $F(4,103)=22.5$, $p<0.0001$; Type $F(1,103)=529.7$, $p<0.0001$; Interaction $F(4,103)=8.3$, $p<0.0001$, note that we fit 2-way ANOVA only for overlapping days 17-35 DIV). The CV of IBIs was significantly lower for hPSC cultures ($F(1,103)=16.1$, $p=0.00012$) and decreased with the development ($F(4,103)=4.17$, $p=0.003$), we did not see a significant effect of the interaction ($F(4,103)=1.04$, $p=0.38$). Average durations were also significantly longer for hPSC cultures ($F(4,103)=827.5$, $p<0.0001$), there was also a significant effect of the development ($F(1,103)=13.1$, $p<0.0001$) and interaction between the culture type and development ($F(1,103)=6.8$, $p<0.0001$). **(d, e)** Comparison of the network spike rate for both the model and data. The model well captures the main changes in the bursting activity, however, it fails to closely reproduce the burst shapes.

Probing network excitability experimentally

To further relate the excitability in the model to the changes associated with an increasing network excitability in cultures, we recorded network dynamics of hippocampal cultures in various concentrations of potassium chloride (KCl) and show that increasing excitability is associated with changes in the overall model excitability and decreasing $b - \theta$ dependency slope (Fig. 3.12).

In previous sections, we showed that the major effects that we observed during development might be attributed to an increase in overall excitability. This effect might be associated with an increasing number of connections in the network as well as an increase in single-cell excitability. To further relate the changes in the model parameters with the changes in the network dynamics, we recorded network activity of cultures of dissociated hippocampal neurons in 5 different concentrations of KCl ranging from 1.5mM to 7mM.

Changing potassium chloride is one of the standard ways to control neuronal excitability (He et al., 2020; O'Leary et al., 2010; Penn et al., 2016; Rhoades & Gross, 1994; Rienecker et al., 2020; Sancristóbal et al., 2016). The main effect of potassium is associated with the changes in the resting potential of the membrane and according to the Goldman-Hodgkin-Katz-equation increasing the KCl concentration leads to a logarithmic change in the resting potential (Sterratt et al., 2011).

The hippocampal neurons were obtained from the E19 rat embryos using standard dissection protocols (see details in Appendix A.2). Neurons were seeded onto 13mm radius coverslips and cultured for 14-28 days *in vitro*. We recorded the population activity with Ca-indicator Fluo-4 in different concentrations of KCl.

Cultures recorded in 1.5-7mM of KCl showed bursting activity. We did not observe network bursting in 10mM of KCl, which might be either associated with depolarization block or the effects of inhibition on the network activity that we do not explicitly model here (He et al., 2020). On average, hippocampal cultures decreased their IBIs as the concentration of KCl increased from 1 to 5.5mM ($F(4,86)=5.28$ $p=0.0007$). We did not see a significant difference in IBI between 5.5mM and 7mM (post hoc t-test, $p=0.94$). We did not see significant differences in the burst durations ($F(4,86)=1.9$,$p=0.1$) and the CV of IBIs ($F(4,86)=0.7$,$p=0.55$).

We fit the model to the network bursting activity to validate how model parameters are related to the increasing or decreasing excitability *in vitro*. We sample from the approximated posterior given the experimental summary statistics and visualize the model parameters. The main effect of changing KCl concentration can be related to increasing overall excitability in the network, assuming that the adaptation parameters do not change. As discussed in the previous sections,

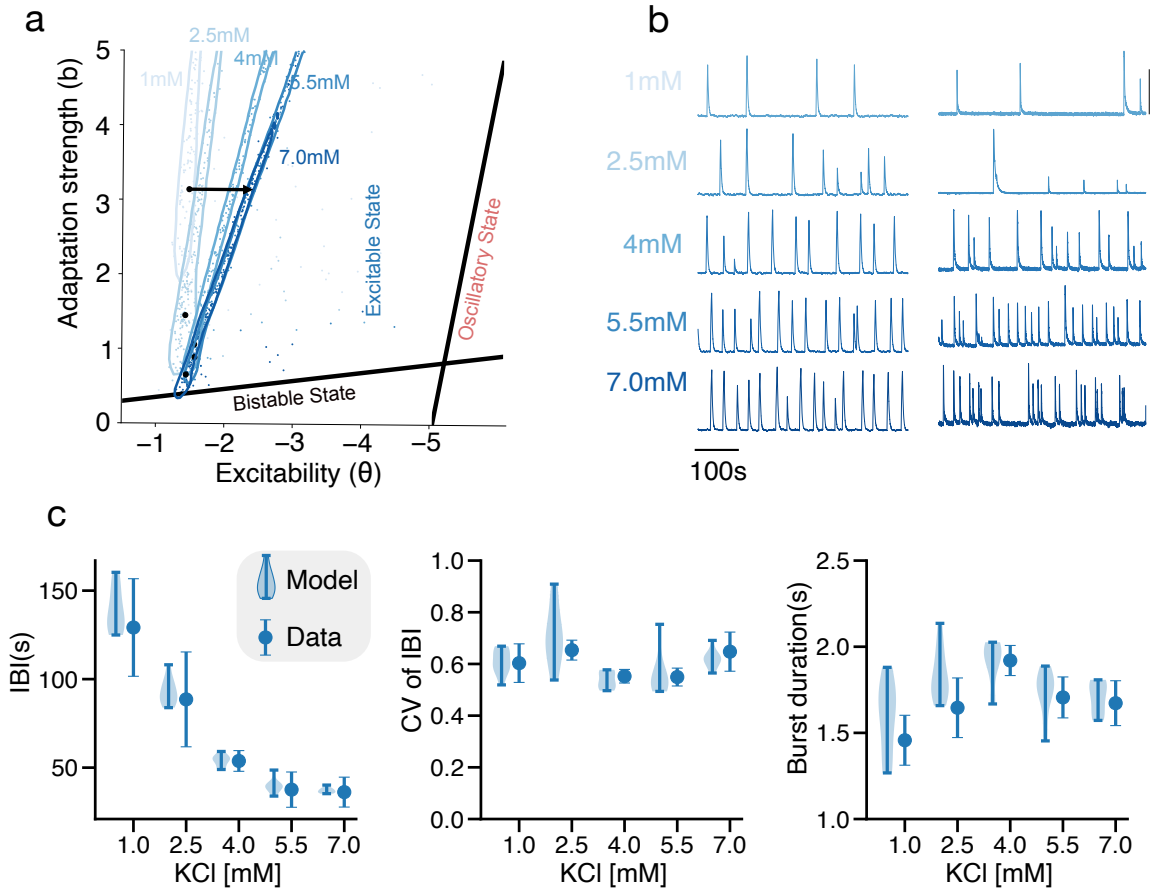


Figure 3.12: Increasing the excitability of the network *in vitro* is associated with the decrease of adaptaiton-excitability $b - \theta$ slopes. **(a)** The marginal posterior distributions of the model excitability and adaptation fitted to match the bursting activity in networks with an increasing concentration of potassium (KCl). **(b)** Example traces for the Ca-recordings over the whole field of view and the fitted model (rate is convolved with 2s exp kernel). **(c)** Summary statistics in the data (errorbars show mean and s.e.n.) and in the model with parameters sampled from the posterior (n=50, violin plots)

because of the invariance in $b - \theta - \tau_w$ space, the changes in excitability can also be matched by decreasing the slope of the $\theta - b$ dependency.

Overall, the main effect of the increasing excitability experimentally is captured by the changes in θ and a corresponding change in the slope of the $\theta - b$ dependency.

3.3 Discussion

Network bursting emerges in networks *in vitro* comprised of different neuron types, ranging from rodent cortical or hippocampal neurons to human hPSC. The temporal statistics of network bursting can be well captured by a model that includes noise recurrent activity and a slow adaptation/recovery mechanism. Network bursting may then be governed by three different types of dynamics: oscillation, excitable dynamics, and bistability. We build a simplified model that allows us to probabilistically map the statistics of network bursting onto different dynamical states and a small number of corresponding model parameters. Fitting a large variety of experimental data suggests that the network dynamics of cultures are consistent with excitable and in some cases with oscillatory dynamics. We find that cortical cultures, on average, have larger network excitability, which also may lead to noisy oscillatory dynamics. The activity in mature hippocampal cultures is, at the same time, governed by excitable dynamics.

We map the development of collective activity *in vitro* and find that the network activity transitions between different dynamical states. Thus, cortical cultures transitioned from the excitable state to the oscillation as cultures matured. Hippocampal cultures, in contrast, showed only a small change in overall excitability and stayed within the excitable state. Additionally, we analyzed the developmental pattern of one type of hPSC culture. We found that these cultures had a completely different maturation pattern. Their immature activity was governed by oscillations and then the networks slowly converged toward excitable dynamics.

The simplified model that we used is closely related to the models used to approximate cortical network dynamics *in vivo* and *ex vivo* (Jercog et al., 2017; Levenstein et al., 2019; Mattia & Sanchez-Vives, 2012). Thus, a similar model that uses a more realistic nonlinearity (e.g Siegert function that is derived directly from a network of LIF neurons, see also Chapter 2) has been used to map the activity of cortical slices (Mattia & Sanchez-Vives, 2012; Sanchez-Vives et al., 2000). Slices of the cortex exhibit slow-wave oscillations that strongly resemble network bursts, although they appear to be less variable (Sanchez-Vives et al., 2000). Mattia et al.(2012) showed that the slow waves in slices of the cortex are likely governed by oscillatory activity. Our results in this context might suggest that cortical cultures of dissociated neurons may be developing to

exhibit a similar type of dynamics.

Levenstein et al. (2019) recently proposed that the cortical and hippocampal activity in vivo during sleep might be governed by excitable dynamics, either in up or down states. Their modeling approach is very similar to ours, apart from a sigmoid nonlinearly in the adaption variable and the autocorrelation in the noise (in contrast to our white noise). To find the model parameters, they also use a set of summary statistics that include the average event duration and the duration inter-event intervals as well as their variability. Finally, Jecorg et al.(2017) used a simplified model to show that the up-down transitions in the steady-state cortical activity might be consistent with bistable dynamics.

We fit the model parameters, using amortized simulation-based inference (Gonçalves et al., 2020). This approach allows us to fit a single flow-based probabilistic network to predict the posterior distribution of parameters given the data (Cranmer et al., 2020; Lueckmann et al., 2017). Then we evaluate the approximated posteriors for a wide range of parameters. This allowed us to explore the dependencies between parameters and identify invariant manifolds. Furthermore, by exploring the posteriors we could derive some equations for the dependencies that guided our analytical exploration of the model. This illustrates that using simulation-based inference for simple models can help to gain useful insights into the analysis of models.

Fitting the model to the summary statistics of bursting activity *in vitro*, allowed us to detect the differences between cortical and hippocampal cultures. The results of fitting showed that these differences are driven by the changes in the network excitability. Physiologically this should indicate either higher single-neuron excitability or more connections between neurons with a higher rate of spontaneous synaptic release. Differences in excitability directly predict that cortical networks are more susceptible to stimulation, which can be further tested experimentally. There are studies showing these differences between cortical and hippocampal networks. Thus, cortical cultures appear overall to grow more connections and have proportionally fewer inhibitory connections as identified in percolation experiments (Soriano et al., 2008). Furthermore, there is also recent direct evidence for higher responsiveness of cortical network in vitro (Callegari et al., 2022).

We explored the difference between the network states over development and found that hippocampal and cortical cultures increase their overall excitability. Cortical cultures, however, show a much stronger increase and transition toward oscillatory states, which explains the differences when only average statistics in the mature phase are considered. The changes in overall excitability, in this case, are likely due to a growing number of connections (Maeda et al., 1995; Soriano et al., 2008; Wagenaar et al., 2005). Here, we again expect that mature culture

should show a higher level of responses to external stimulation, which has been previously observed (Wagenaar et al., 2005). In one dataset (Hyvärinen et al., 2019; Kapucu et al., 2022), cortical cultures only briefly transitioned to the oscillatory state and then returned to the excitable states.

In contrast to the rodent cultures, the dataset of hPSC culture recording (Hyvärinen et al., 2019; Kapucu et al., 2022), showed the opposite developmental trajectory. At early developmental stages, hPSC neurons displayed asynchronous firing, then at around 20 DIV, the networks started to show slow oscillatory activity and finally slowly converged to excitable activity. The statistics of the network events were very different from the rodent cultures. However, the model suggests that the dynamical states might be consistent, although the underlying adaptation timescale for hPSC was much longer and the intrinsic noise was smaller. Here we analyzed only one type of hPSC culture, which displayed a very specific pattern of network activity, which might be quite different from other preparations (Estévez-Priego et al., 2023).

The development of collective activity indeed could be the hallmark of the successful development of hPSC cultures and may help to establish hPSC as a platform for human-specific disease modeling (see Chapter 6). The number of protocols in which the collective activity has been established is rapidly growing. However, staging of human hPSC based on the displayed dynamics remains challenging, because of the high variability between protocols (Estévez-Priego et al., 2023). Using computational models for studying network dynamics might be a plausible way to relate different preparations to each other and compare the resulting activity to the activity of cultures of dissociated human neurons and rodent cultures.

Finally, to relate the changes in living networks and associated differences in the model, we experimentally changed the excitability in the network of neurons by changing the level of KCl. This helped us identify that higher network excitability for our network could be matched either by higher model excitability or weaker adaptation strength. The changes in network dynamics that we observed for higher or lower levels of KCl are consistent with earlier observation (Rhoades & Gross, 1994; Sancristóbal et al., 2016). The main effect was related to an increase in the rate of network bursting. Interestingly, early models Giugliano et al., 2004 as well as a simple change of excitability in our simplified model predict a decrease in the variability of IBIs. That is not what we found experimentally. Finally, when KCl is increased to 10mM and above, collective network events disappear. This result could be consistent with a phase transition from excitable dynamics to a single high-fixed point in our model. However, the model failed to resolve this quantitatively. Thus, we would expect to see a more gradual transition and some intermediate activity with longer burst durations. Secondly, this could be related to the

discoordination of collective dynamics mediated by inhibition or depolarization block (He et al., 2020; Rienecker et al., 2020) and might require a different modeling approach to fully capture this transition.

Chapter 4
Network topology leading to the formation of balanced E/I inputs

Summary

Experimental studies show that the balance of excitation and inhibition in cortical circuits emerges not only structurally but also functionally (Okun & Lampl, 2008). This means that excitatory and inhibitory inputs to a neuron are fine-tuned and strongly correlated during both spontaneous and evoked activity. Theoretical work suggests that such a balance might be essential for efficient neural coding (Deneve & Machens, 2016) as well as signal propagation (T. Vogels & Abbott, 2009). One of the suggested mechanisms that help to develop a tight balance of excitation and inhibition (E/I balance) is inhibitory plasticity (Sprekeler, 2017; T. Vogels & Abbott, 2009). This type of long-term plasticity dynamically adjusts the weights of incoming inhibitory connections. It has been shown that this type of plasticity both creates the tight E/I balance and helps to develop neuronal tunning when a post-synaptic neuron receives structured feedforward sensory inputs (T. P. Vogels et al., 2011).

The ability of inhibitory plasticity to develop E/I balance and neuronal tuning rapidly deteriorates when inputs are noisy or read out from a recurrent network. In this project, we show that introducing a specific topology in the presynaptic network restores both the neural tuning and E/I balance on the post-synaptic neuron. Using simulation-based inference, we found that optimal connectivity of the presynaptic networks should include strong assemblies of excitatory neurons that receive shared sensory inputs along with more diffuse inhibitory connectivity that controls the signal propagation.

Contributions. This work is published as a preprint (Giannakakis et al., 2023). The project is led by Emmanouil Giannakakis, who designed the network, selected the plasticity rules, and obtained the initial results for neuronal assemblies. Victor Buendía performed analytical derivation for the equivalent linear model, details of which can be found in the original publication (Giannakakis et al., 2023). My main contribution is the application of simulation-based inference and analysis of the dynamics of the network with different assembly strengths.

Contents

4.1 Introduction

4.1.1 Excitation/inhibition balance and input selectivity

Input selectivity is a ubiquitous mechanism for encoding information in neuronal networks. Neurons display preferential sensitivity to stimuli, beginning with selectivity for basic features of environmental stimuli in lower sensory areas (such as orientation and direction) and moving to more complicated spatiotemporal patterns in higher areas (Riesenhuber & Poggio, 1999). Such selectivity is usually called neuronal tuning (Kriegeskorte & Wei, 2021). The mechanisms by which this input selectivity arises in the nervous system have been extensively investigated in experimental and modeling studies (Bienenstock et al., 1982; Blais et al., 1998; Brito & Gerstner, 2016; Clopath et al., 2016).

Input selectivity is attributed to excitatory and inhibitory neurons (Isaacson & Scanziani, 2011). Some experimental studies found that excitatory and inhibitory neurons show very similar tunning properties and it has been suggested that excitation and inhibition are co-tuned (Wehr & Zador, 2003). Other studies found evidence for broader tunning of inhibitory neurons (tip of the iceberg model, Isaacson and Scanziani (2011) and G. K. Wu et al. (2008)). The complete picture of whether different types of tuning coexist or should be attributed to different coding schemes is still missing (Deneve & Machens, 2016; Sprekeler, 2017; G. K. Wu et al., 2008).

Simultaneously, there is evidence for a strong temporal correlation of excitatory and inhibitory currents that arrive on a single post-synaptic neuron. This has been shown both during spontaneous and evoked activity in cortical circuits (Barral & D Reyes, 2016; Okun & Lampl, 2008). Theoretical studies clearly show how such a balance of excitation and inhibition can govern asynchronous irregular activity, which is assumed to be the default dynamical state of cortical networks (Renart et al., 2010). However, it is not immediately obvious how balancing E/I currents that cancel each other out would promote coding, memory, or signal prorogation. Early studies showed that balanced E/I allows for faster signal processing (Vreeswijk & Sompolinsky, 1998). More recently, it also has been shown that under some assumptions balanced E/I is beneficial for efficient coding and computations (Deneve & Machens, 2016).

The balance of excitation and inhibition can emerge by design or can be promoted by synaptic plasticity. Early studies analyzed the activity of random networks (Vreeswijk & Sompolinsky, 1998), where tight E/I balance can emerge simply by the choice of the connectivity matrix. More recently, there has been a lot of effort in trying to resolve the emergence of E/I balance for inhomogeneous structured networks (Sprekeler, 2017). One of the dominant views in the

field is that the tight balance of excitation and inhibition can be promoted by synaptic plasticity mechanisms (T. P. Vogels et al., 2011; Y. K. Wu et al., 2022). Thus, inhibitory plasticity has been suggested as one of the main candidates for balancing inhibitory weights to match a given excitatory input structure (Hellyer et al., 2016; Khajehabdollahi et al., 2021; Luz & Shamir, 2012; T. P. Vogels et al., 2011). It also has been shown that plasticity together with a normalization mechanism can endorse simultaneous development of matching excitatory and inhibitory connectivities (Effenberger et al., 2015; Giannakakis et al., 2020; Larisch et al., 2021; Litwin-Kumar & Doiron, 2014; Mackwood et al., 2021; Miehl & Gjorgjieva, 2022; Zenke et al., 2015).

Most of the studies on the development of E/I balance and stimulus tuning relied on a simple setting where a postsynaptic neuron receives feedforward projections that carry inputs and adjusts its weights using various plasticity mechanisms. Such settings allow a neuron to learn statistical dependencies present in the inputs (S. Eckmann & Gjorgjieva, 2022; T. P. Vogels et al., 2011). It remains, however, unclear if inputs that propagate through a recurrent network can still maintain the necessary statistical structures for E/I balance and input selectivity to emerge. Similarly, it is unclear if the balance and tuning emerge when inputs are corrupted by noise. Here we analyze the E/I balance and input tunning in a single post-synaptic neuron that receives inputs from either a feedforward or recurrent network of neurons. The E/I weights in these post-synaptic neurons develop according to a combination of excitatory and inhibitory plasticity mechanisms with a divisive normalization mechanism (S. Eckmann & Gjorgjieva, 2022; Gjorgjieva et al., 2011; J.-P. Pfister & Gerstner, 2006; T. P. Vogels et al., 2011). In a standard feedforward setting such a post-synaptic neuron develops E/I balance and stimulus tunning. Next, we show that input noise or propagation of inputs through a recurrent presynaptic network strongly deteriorates the E/I balance and tuning. To counteract these detrimental effects of recurrence and noise we introduce stimulus-specific assemblies of neurons in the presynaptic network. The near-optimal connectivity configurations involve strong excitatory assemblies and weaker inhibitory assemblies. Finally, we demonstrate that assemblies allow co-tuning to emerge even in sparsely connected networks if their relative strengths are adjusted for the sparsity level.

4.2 Results

4.2.1 Self-orgnization of E/I balance and neuronal tuning in a feedforward network

Network model

We assume that cortical neurons tuned to a stimulus receive balanced excitatory and inhibitory inputs. Plasticity mechanisms can easily promote the emergence of these features. We model a single read-out postsynaptic neuron stimulated by a population of $N = 1000$ pre-synaptic neurons. The pre-synaptic neurons are divided into M subpopulations G_i, $i \in \{1, \dots, M\}$. Each group receives identical group-specific inputs (Fig. 4.1). The groups are comprised of excitatory (80%) and inhibitory (20%) neurons. Additionally, every neuron is driven by independent external noise (T. P. Vogels et al., 2011). In this setting, each presynaptic group has a highly correlated pattern of activity.

Neurons are modeled as leaky integrate-and-fire (LIF) neurons with dynamic synapses:

$$C_m \cdot \frac{dV(t)}{dt} = g_{\text{leak}} \cdot (V_{\text{rest}} - V(t)) + g_I(t) \cdot (V_I - V(t)) - g_E(t) \cdot (V_E - V(t)), \tag{4.1}$$

where V_{rest} is the resting potential, V_E, V_I are the excitatory and inhibitory reversal potentials and g_{leak} the leak conductance. Additionally, the excitatory and inhibitory conductances g_E, g_I decay exponentially over time and get boosts upon excitatory or inhibitory pre-synaptic spiking respectively, as

$$\begin{aligned} \tau_E \cdot \frac{dg_E(t)}{dt} &= -g_E(t) + \overline{g_E} \cdot \sum_j W_j^E \cdot \sum_f \delta(t - t_j^f), \\ \tau_I \cdot \frac{dg_I(t)}{dt} &= -g_I(t) + \overline{g_I} \cdot \sum_j W_j^I \cdot \sum_f \delta(t - t_j^f). \end{aligned} \tag{4.2}$$

Here t_j^f denotes the time of a f-th spike of the $j-$th neuron. When the membrane potential reaches the spiking threshold V_{th}, a spike is emitted, and the potential is changed to a reset potential V_{reset}. Thus, the model is very similar to the one used in the previous chapter 2. The biggest difference is that single neurons do not have spike-frequency adaptation but have more realistic synapses.

Connections to a post-synaptic neuron in our model also evolve in time according to the inhibitory and excitatory plasticity rules. The weights are also additionally normalized with

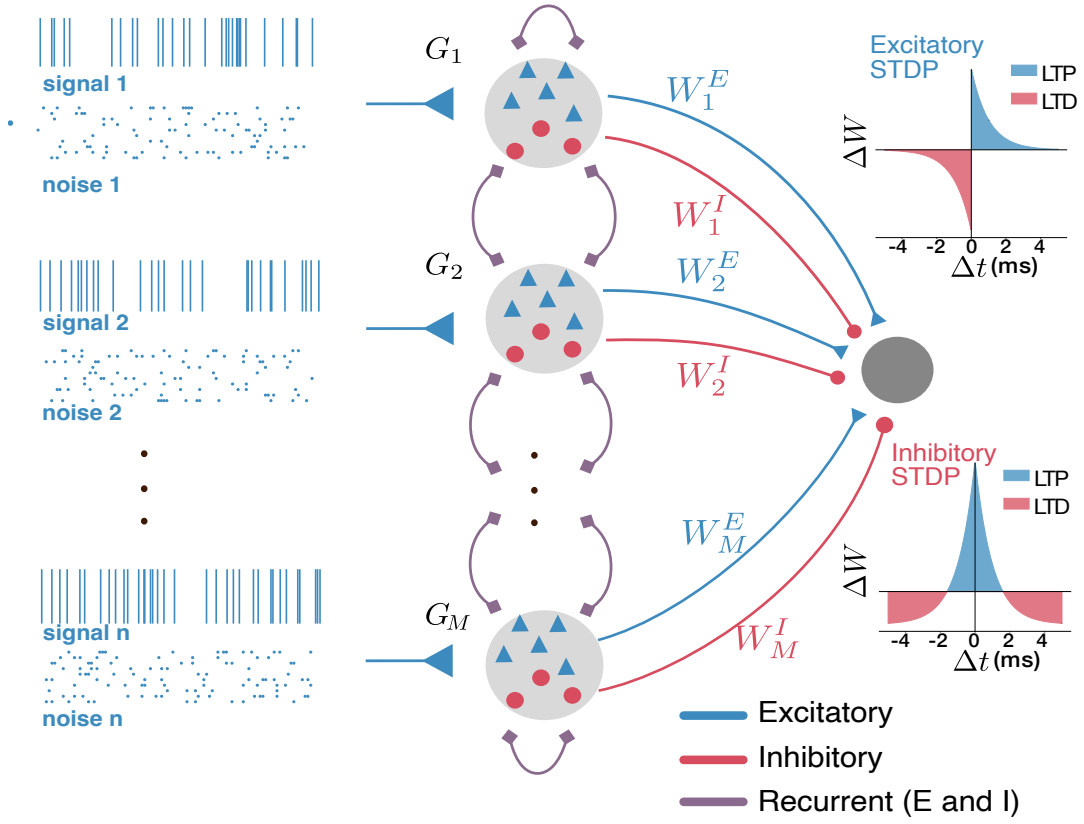


Figure 4.1: A diagram of the network. A post-synaptic neuron receives connections from M groups, each of which processes an independent input sequence. The purple recurrent connections are absent in the feedforward version of the model. The inputs to post-synaptic neurons develop according to excitatory and inhibitory plasticity rules.

a competitive normalization method (S. Eckmann & Gjorgjieva, 2022), which prevents the runaway weight's growth.

Excitatory connections follow the classical triplet STDP rule (J.-P. Pfister & Gerstner, 2006). This rule has been shown to approximate the synaptic weight changes found under the triples of stimulation in experimental studies (Wang et al., 2005). The firing rates of the pre-synaptic excitatory neurons and the post-synaptic neuron are approximated by traces with two different

timescales,

$$\begin{aligned}
\tau_1^{estdp} \cdot \frac{dy_k^E}{dt} &= -y_k^E + \sum_f \delta(t - t_k^f), \\
\tau_2^{estdp} \cdot \frac{dz_k^E}{dt} &= -z_k^E + \sum_f \delta(t - t_k^f), \\
\tau_1^{estdp} \cdot \frac{dx_1}{dt} &= -x_1 + \sum_f \delta(t - t_x^f), \\
\tau_2^{estdp} \cdot \frac{dx_2}{dt} &= -x_2 + \sum_f \delta(t - t_x^f),
\end{aligned} \tag{4.3}$$

where $\tau_1^{estdp} < \tau_2^{estdp}$ are the two timescales of the plasticity rule, y_k^E, z_k^E and x_1, x_2 represent the slow and fast traces of the k-th excitatory pre-synaptic and the single post-synaptic neuron respectively. t_k^f and t_x^f are the firing times of the pre and post-synaptic neurons. The connection weights are updated upon pre and post-synaptic spiking according to

$$\Delta W_k^E = \eta_E \cdot A_{LTP} \cdot x_1 \cdot z_k^E \cdot \sum_f \delta(t - t_k^f) - \eta_E \cdot A_{LTD} \cdot x_2 \cdot y_k^E \cdot \sum_f \delta(t - t_x^f), \tag{4.4}$$

where η_E is the excitatory learning rate and A_{LTP}, A_{LTD} the amplitudes of long term depression and potentiation respectively. The detailed description of the triplet STDP can be found in (Gerstner et al., 2014; Gjorgjieva et al., 2011; J.-P. Pfister & Gerstner, 2006)

Inhibitory connections follow the homeostatic STDP rule (T. P. Vogels et al., 2011). Approximations of the firing rates are kept via a trace for each of the pre-synaptic inhibitory neurons as well as the post-synaptic neuron,

$$\begin{aligned}
\tau^{istdp} \cdot \frac{dy_k^I}{dt} &= -y_k^I + \sum_f \delta(t - t_k^f), \\
\tau^{istdp} \cdot \frac{dx}{dt} &= -x + \sum_f \delta(t - t_x^f),
\end{aligned} \tag{4.5}$$

where τ^{istdp} is the single timesclae of the plasticity rule, y_k^I and x are the traces of the the k_{th} inhibitory pre-synaptic and the single post-synaptic neuron, and $t_k^f,\ t_x^f$ are the spike times of the k_{th} inhibitory pre-synaptic and the post-synaptic neuron respectively. The connection weights are updated upon pre and post-synaptic spiking as

$$\Delta W_k^I = \eta_I \cdot (x - 2\rho_0 \tau^{stdp}) \cdot \sum_f \delta(t - t_k^f) + \eta_I \cdot y_k^I \cdot \sum_f \delta(t - t_x^f). \tag{4.6}$$

Here, η_I is the inhibitory learning rate, and ρ_0 is the target rate of the post-synaptic neuron.

Finally, since the triplet STDP rule is unstable and can lead to uncontrollable weight growth (Gerstner et al., 2014), we use a recently proposed normalization mechanism (S. Eckmann & Gjorgjieva, 2022).

$$W_k^A \leftarrow W_k^A \left(1 - \eta_N + \eta_N \cdot \frac{W_{target}^A}{\sum_{i=1}^{N_A} W_i^A}\right), \quad A \in \{E, I\}. \tag{4.7}$$

Where W_{target}^A is the target total weight of each connection type and η_N is the normalization learning rate. The normalization maintains the sum of the excitatory and the sum of the inhibitory feedforward connections weights close to the set target total weights W_{target}^E and W_{target}^I. Note that Eckmann et al. (2022) used this mechanism for the rate model, here we for the first time introduce it for a spiking neuron.

The external input to each of the pre-synaptic neurons is comprised of two Poisson spike trains. The first Poisson spike train is shared with all the other neurons of the same group, while the second Poisson spike train is the individual noise of the neuron,

$$C_{\text{total}} = C_{\text{signal}} + C_{\text{noise}}, \tag{4.8}$$

where $C_{\text{signal}} \sim \text{Poisson}((1-c) \cdot r)$ and $C_{\text{noise}} \sim \text{Poisson}(c \cdot r)$. Here, r is the total firing rate of the input, and c is the strength of the noise. C_{signal} is the same for all neurons of the same input group, while C_{noise} is individual to each neuron

Development of E/I balance on the post-synaptic neuron

Excitatory and inhibitory plasticity balances the excitatory and inhibitory synapses, which promotes a tight balance of incoming excitatory and inhibitory currents. We simulate the postsynaptic neuron that receives feedforward inputs (recurrent connections in each group are turned off) and analyze the resulting dynamics. Usually, after about 50 seconds of weights development, the incoming weights from each group converge to individual fixed points (Fig. 4.2b). The average excitatory and inhibitory currents appear to be tightly balanced in time, with inhibition slightly lagging behind excitation, because of the fixed difference in synaptic timescales. This allows for brief changes in the firing rate induced by an incoming stimulus. On average excitation and inhibition, currents cancel each other out (Fig. 4.2a). This type of balance is similar to a tight E/I balance, which has been shown to promote efficient coding and computations (Deneve & Machens, 2016). Note that for our model, the necessary prerequisite for this type of E/I balance is a strong correlation between incoming excitatory and inhibitory weights. We thus can quantify

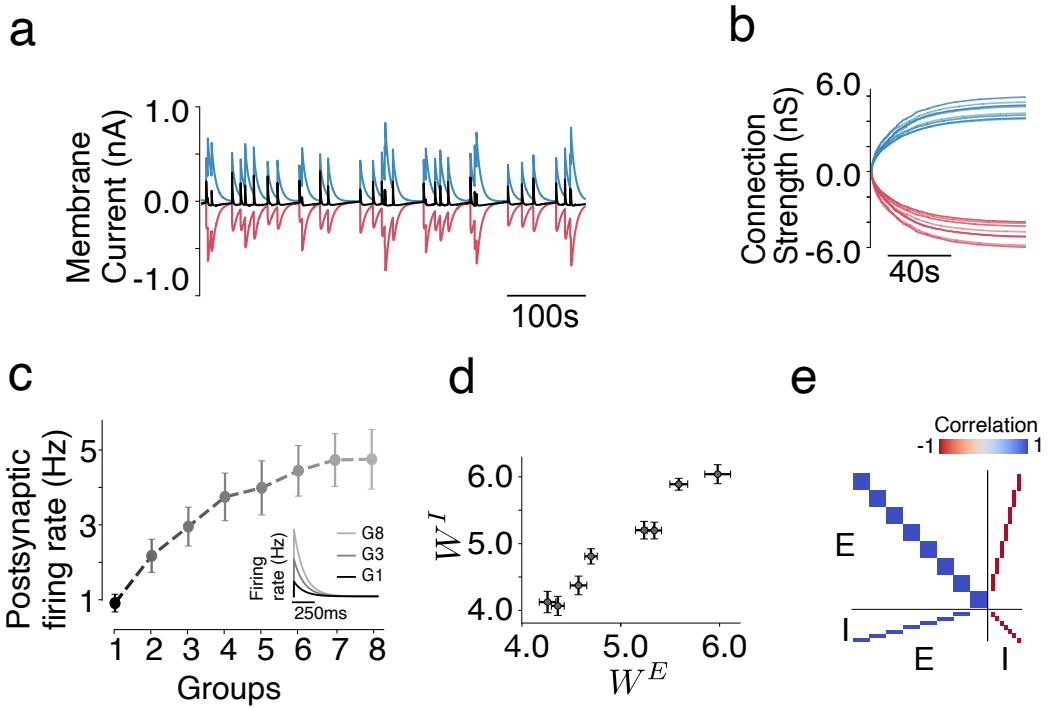


Figure 4.2: Balance and tuning in a feedforward network. **(a)** Excitatory and inhibitory currents at the post-synaptic neuron (after plasticity has converged). The black trace is the average of two. Transients occur because inhibition slightly lags behind the excitation. **(b)** Convergence of the weights under plasticity. The modified (the sign of inhibitory activities is inverted) covariance matrix determines the convergence point of the plasticity protocol. Here we see a near-optimal matrix that leads to very clear co-tuning. **(c)** The post-synaptic neuron's response to inputs given to different groups. The resulting firing rates depend on which group was stimulated, which leads to the input selectivity. **(d)** E/I weights projecting onto a postdynaptic neuron demonstrate co-tuning: most groups have distinct weights and the E and I weights of each group match each other. **(e)** Correlation in the feedforward connectivity that allows for perfect balance and tuning at a postsynaptic neuron

the detailed balance with the Pearson correlation coefficient between the mean excitatory and inhibitory weights of each group,

$$B = \frac{\mathrm{Cov}\left(\langle W_G^E\rangle, \langle W_G^I\rangle,\right)}{\mathrm{Std}\left(\langle W_G^E\rangle\right) \cdot \mathrm{Std}\left(\langle W_G^I\rangle\right)}, \tag{4.9}$$

where $\langle W_G^k\rangle = \left(\langle W_{G_1}^k\rangle, \langle W_{G_2}^k\rangle, \ldots, \langle W_{G_m}^k\rangle\right)$, $k \in \{I, E\}$ and $\langle W_{G_i}^k\rangle$ is the average projection weight from the excitatory ($k = E$) or inhibitory ($k = I$) neurons in group i. In networks with high balance B, the strength of incoming E and I currents are highly correlated.

Balanced post-synaptic neurons can discriminate incoming inputs

When we let the network run and develop weights converging on the postsynaptic neuron according to the plasticity rule, the responses of a readout neuron to stimulation of different groups become diverse and discriminable (Fig. 4.2c). Thus, neuronal tuning emerges. The post-synaptic neuron discriminates the inputs from different groups when projections from each group are diverse (Fig. 4.2d). If the firing rate in response to stimulation of each group is sufficiently different the stimulus identity can be read out by the relative values of the output firing rate. This type of coding is easily achieved if the connections from each group are well separated by their synaptic strength. In this case, groups with stronger projections to the postsynaptic neuron would elicit stronger responses. Groups with weak projections would accordingly elicit weaker responses. Note that the outgoing connections from neurons inside one group are highly correlated (Fig. 4.2e). We quantify the neuronal tuning using a weight diversity metric:

$$D = 1 - \frac{1}{M \cdot \mathrm{Std}\left(W^E\right)} \sum_{i=1}^{M} \mathrm{Std}(W^E_{G_i}), \tag{4.10}$$

where $W^k,\ k \in \{E, I\}$ is the set of (E or I) feedforward connection weights and $W^k_{G_i},\ k \in \{E, I\}$ is the subset of (E or I) feedforward connection weights from input group i. Diversity $D \in [0, 1]$ equals unity when the feedforward connections from each group are similar within a group, but different across groups; D is close to zero when there is no difference between groups.

The point at which the feedforward weights converge is determined by the covariance matrix of the activity of presynaptic neurons. For a rate network, the fixed points to which the weights converge are fully determined by the eigenvectors of the covariance matrix (Clopath et al., 2016; S. Eckmann & Gjorgjieva, 2022; T. P. Vogels et al., 2011).

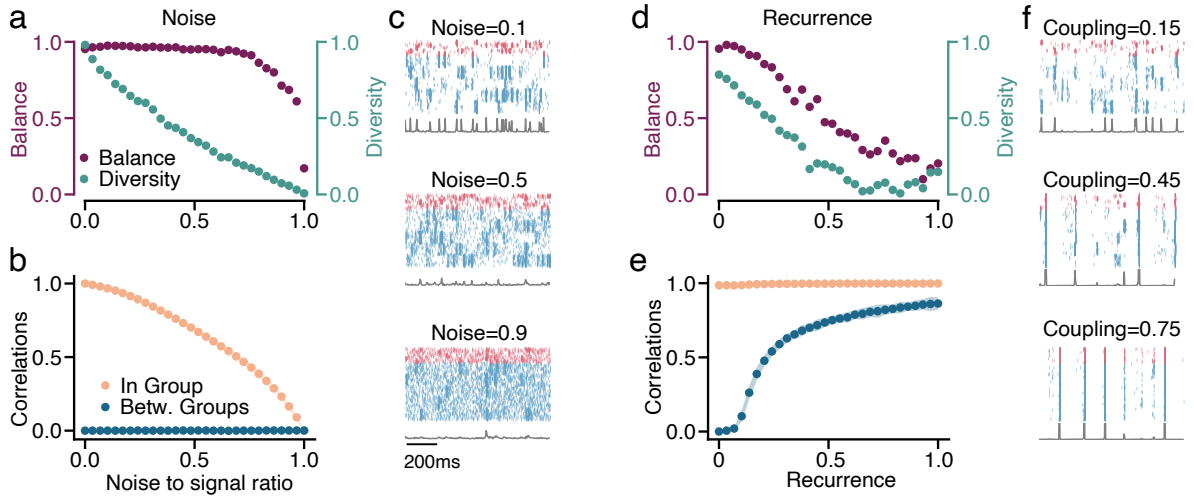


Figure 4.3: Noise and Recurrence Destroy E/I Co-Tuning. **(a)** An increase in the noise the network receives leads to a reduction in diversity (green) and after some point also balance (purple). **b.** This decrease is caused by an increase in in-group correlation. **(c)** As the noise increases (indicated above the panel) spiking activity becomes more asynchronous. **(d)** An increase in the recurrent coupling strength leads to a rapid decrease in balance and diversity, which is caused by **(e)** an increase in the between-group correlation. **(f)** The spiking activity becomes more synchronous as the coupling strength increases.

4.2.2 Noise and recurrent connectivity destroys the ability of STDP to produce E/I balance and neuronal tuning

While it is reasonable to assume that in some early sensory areas, the connections that convey the information are sufficiently feedforward (Antolík et al., 2016; Klindt et al., 2017), most areas of the brain receive inputs from highly recurrent networks. Additionally, incoming stimuli might be noisy (Aponte et al., 2021; Ecker & Tolias, 2014). We use our network model to probe the effect of noise and recurrence on the E/I balance and neuronal tuning. To control the noise level we change the fraction of signal spikes that presynaptic neurons of the same group receive and the noise spikes from an independent Poisson process. Thus, we can flexibly set the noise-to-signal ratio.

Input noise decorrelates the activity of neurons from the same input group. This leads to the reduction of the weight diversity (Fig. 4.3a). The E/I balance is much less affected by the noise and starts to decrease only when stimuli contain almost only noise (Fig. 4.3b,c).

Next, we start adding recurrent connections between the presynaptic neurons and observe that it jeopardizes both the diversity and balance (Fig. 4.3d). Recurrence introduces cross-correlations in the presynaptic neurons (Fig. 4.3e). Stimulus-evoked activity in this case propagates through-

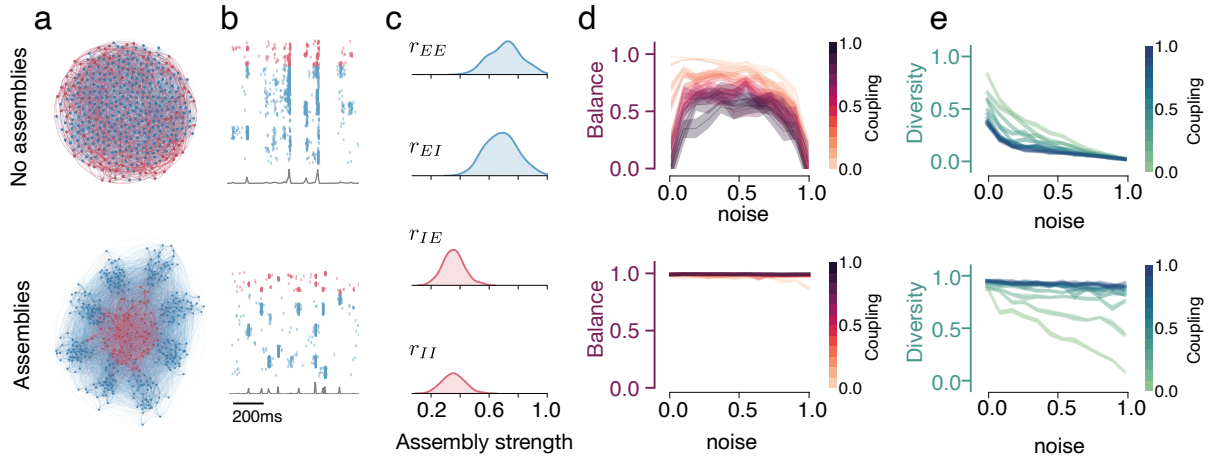


Figure 4.4: Optimized assemblies of neurons restore the co-tuning in recurrent noisy networks. Top row — networks with uniform connectivity, bottom — networks with inferred optimal assembly strengths. **(a)** Diagram of the network, with uniform connectivity and inferred optimal assembly strengths. **(b)** Stimulus-driven activity of the networks with (bottom) and without (top) assemblies. **(c)** Approximate posterior distributions of excitatory and inhibitory assemblies strength. **(d)** Balance in homogeneous networks (top) and optimized assemblies (bottom) networks with different strengths of noise and coupling. **(e)** The same for diversity. Assemblies restore co-tuning even with strong noise if given sufficient coupling strength. In the homogeneous networks, the strong coupling is detrimental to the development of co-tuning.

out the whole presynaptic network, which naturally compromises the ability to discriminate stimuli (Fig. 4.3f)

4.2.3 Assemblies of neurons restore the ability of STDP to produce E/I balance neuronal tuning

Increasing connectivity between neurons that receive a shared input helps to restore the necessary covariance structure that allows for the E/I balance and neuronal tuning. Homogeneous recurrent connectivity in the presynaptic population destroys both E/I balance and tuning on the post-synaptic read-out neuron. One possible way to counteract the effects of recurrence and, as we will see later, noise is to introduce inhomogeneous connectivity.

We introduce stronger recurrent connectivity between neurons of the same input group. This forms assemblies of neurons that share input and helps to maintain the beneficial correlations in the activity of the presynaptic network (Fig. 4.4a bottom). To investigate the effects of assemblies, we define an assembly strength metric as a ratio of the average input from the same

group and type (E/I) neurons to the total average input from the given type of neurons:

$$r_{ab} = \frac{C_{\text{in}}^{ab}}{C_{\text{in}}^{ab} + C_{\text{out}}^{ab}} = \frac{W_{\text{in}}^{ab}}{W_{\text{in}}^{ab} + (M-1) \cdot W_{\text{out}}^{ab}}, \quad a, b \in \{E, I\}, \tag{4.11}$$

where C_{in}^{ab} in the total input a neuron of type b receives from neurons of type a, for $a, b \in \{E, I\}$, of its own input group and C_{out}^{ab} is the total input a neuron of type b receives from neurons of type a, for $a, b \in \{E, I\}$, from all other input groups.

Since we keep all connections of the same type equal $C_{\text{in}}^{ab} = \frac{p \cdot N}{M} \cdot W_{\text{in}}^{ab}$ and $C_{\text{out}}^{ab} = \frac{p \cdot (M-1) \cdot N}{M} \cdot W_{\text{out}}^{ab}$, where p is the probability of connection between two neurons, W_{in}^{ab} the connection strength between neurons of the same group and W_{out}^{ab} the connection strength between neurons of different groups. We vary assembly strengths for each type of connection r_{EE}, r_{EI}, r_{IE}, and r_{II} while keeping the total input to a neuron $C_{\text{in}}^{ab} + C_{\text{out}}^{ab} = \frac{p \cdot N}{M} \cdot (W_{\text{in}}^{ab} + (M-1) \cdot W_{\text{out}}^{ab}) =: p \cdot N \cdot W$ constant. Here, W is a coupling strength, the same as in the network without assemblies. Thus, we can vary the fraction of input coming to a neuron from its input group without changing the average recurrent excitatory and inhibitory inputs it receives.

Simulation-based inference of assembly strengths

Overall we have four different assembly strengths r_{EE}, r_{EI}, r_{IE}, r_{II} for the presynaptic population. We observed that introducing strong assemblies for all types of connections restores the diversity and balance at the postsynaptic neuron. However, to systematically find the optimal combination of assembly strengths, we use simulation-based inference (Cranmer et al., 2020; Gonçalves et al., 2020). With this approach, we find the combination of parameters that produce the detailed E/I balance ($B \approx 1$) and maximum weight diversity ($D \approx 1$).

Similarly to the previous chapter (2), we apply the population Monte-Carlo Approximate Bayesian Computation (PMC-ABC) method (Beaumont et al., 2009). One challenge that we faced in this project is that to obtain the balance and diversity metrics, we need to simulate networks until the plasticity converges. Given that ABC requires a substantial number of simulations (Lueckmann et al., 2021), we had to look for surrogate metrics that we would optimize. Luckily, as discussed above, the ability to produce E/I balance and neuronal tuning (e.g. optimize balance (B) and weights diversity (D)) is determined by the covariance of the presynaptic network activity. In an ideal case that allows for perfect balance and diversity, the correlation between neurons from different groups is close to zero, whereas the correlation between neurons from one group is close to one. We define the loss function using the in-group

and between-group correlations of the presynaptic network:

$$\mathcal{L} = -\alpha C_{in}^2 - \beta \left[(1 - C_{out}^{E\to E})^2 + (1 - C_{out}^{E\to I})^2 + (1 - C_{out}^{E\to I})^2 \right].$$

. The loss is then zero when the in-group correlations are equal to one and all between-group correlations vanish.

We define a uniform prior $p(\theta)$. A set of parameters $\theta = [r_{ee}, r_{ei}, r_{ie}, r_{ii}]$ is sampled from it and used to run the simulations for 3 seconds. We accept a parameter set if the loss is below the error ϵ, and keep sampling until the number of accepted samples is 60. We use the kernel density estimate on the accepted samples to obtain an approximate posterior. Next, we rescale this approximate posterior with the original before obtaining a proposal distribution that we use as a prior in the next step of the ABC. In each step, we reduce ϵ by setting it to the 75th percentile of the losses for the accepted samples. As a rule, we run 20 to 30 steps of the sequential ABC until the loss converges. The fitting was done using a modified version of the simple-abc toolbox https://github.com/rcmorehead/simpleabc/.

4.2.4 Networks with optimized assemblies of neurons regain the ability to develop E/I balance and neuronal tunning

We first start by exploring the optimal assembly strengths in a fully connected presynaptic network. We found that the optimized assembly strengths result in a covariance structure that is very similar to the one observed in a feedforward/low-noise network. This allows the plasticity to produce near-optimal co-tuning of the feedforward connections (Fig. 4.4c). The approximated posterior distribution of the assembly strengths indicates that very strong $E \to E$ and $E \to I$ assemblies and medium-strength $I \to E$ and $I \to I$, (Fig. 4.4c) are the most probable solutions. Fig. 4.4d shows the difference in balance and weight diversity when the postsynaptic neuron receives projection from homogeneous and assembled networks. We checked how well assemblies perform for different levels of coupling and noise. Given sufficient coupling strengths, assemblies efficiently ameliorate the effects of noise and recurrence.

4.2.5 Perturbing the optimal assembly strengths indicates that the networks are more robust to changes in the inhibitory but not in the excitatory assemblies

Modifications in the connections of inhibitory and excitatory neurons have distinct impacts on the dynamics of the network (Mongillo et al., 2018). Here we asked if tuning and diversity are

also more sensitive to changes in individual types of assemblies.

We perturb various assemblies away from the optimal solutions to investigate the relative importance of the different connection types. We take individual assemblies and one by one check how fast the balance and diversity deteriorate as the parameters are perturbed away from the optimal solution inferred with ABC (Fig. 4.5a). We find that $E \to I$ and $I \to E$ assemblies have a strong impact on the E/I balance and neuronal tuning (Fig. 4.5b,e). In contrast, the perturbations of $I \to I$ and $E \to I$ assemblies have a weaker effect on the balance and weight diversity.

We also explored how network dynamics change when we perturb different assembly types and how it relates to the balance and diversity. A network with optimal assemblies exhibits asynchronous irregular activity with occasional stimulus-driven bursts of activity inside individual groups (Fig. 4.5d and Fig. 4.4b).

When $E \to E$ assemblies are reduced, the weight diversity rapidly drops, while the balance metric first decreases and then increases. This behavior is related to two different bifurcations in the whole-network dynamics. First, synchronous full-network bursts emerge on top of the stimulus-induced firing. Then, as $E \to E$ are reduced further, the network enters an asynchronous irregular state (Fig. 4.5b) without any stimulus-driven events. The asynchronous irregular state is beneficial for keeping the E/I balance but in the current settings is suboptimal for discriminating stimuli.

Reduction of the $I \to E$ assemblies is related to the simultaneous decay of both balance and diversity metrics. When $I \to E$ assemblies are too strong, the network exhibits synchronous behavior that does not allow the emergence of co-tuned connectivity (Fig. 4.5e). Interestingly, when $I \to E$ assemblies are weaker than optimal, one group of neurons is active. This eventually leads to perfect discrimination of only a single input. Interestingly, the optimal solution is found at a transition between these two types of activity.

The balance and diversity are more robust to changes in both $E \to I$ and $I \to I$ assemblies. Yet, perturbing the strength of these assemblies also leads to changes in the network dynamics (although less detrimental). When $E \to I$ assemblies are reduced, the weight diversity slowly decreases, specifically because of the decrease in the diversity of inhibitory weights. Thus, the activity of the inhibitory population becomes more asynchronous (Fig. 4.5c). On the other hand, a network with uniform $E \to I$ connection (i.e., without $E \to I$ assemblies), shows a dramatic reduction of the balance metric because of the very strong activation of a single neuron group accompanied by the simultaneous activity of all inhibitory neurons.

When $I \to I$ assemblies are weakened the network displays synchronous bursts on top of the

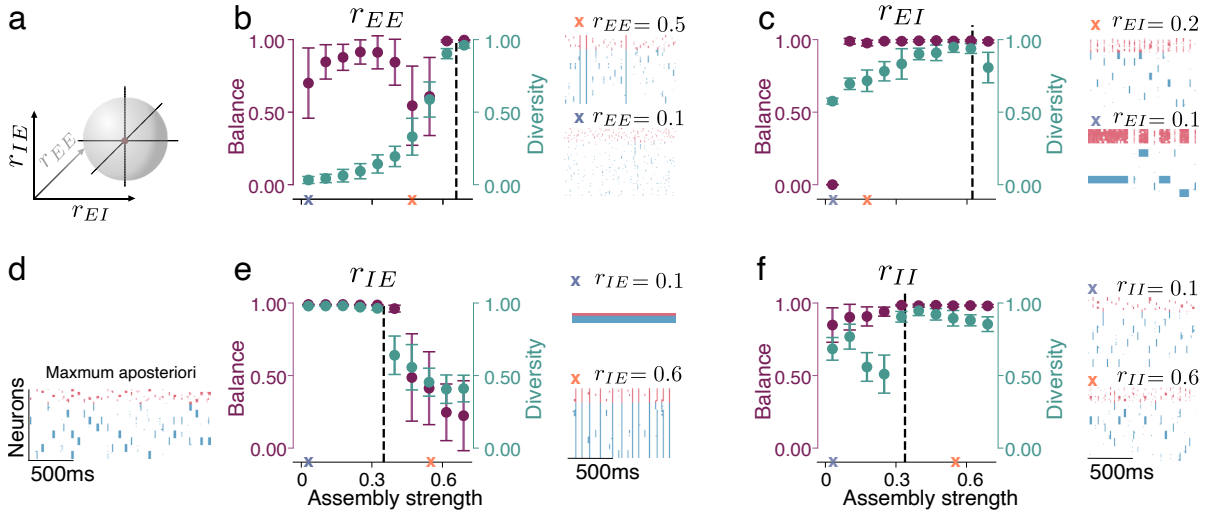


Figure 4.5: Changes of the various assemblies strengths differently affect balance and diversity. **(a)** We sequentially vary the value of each assembly strength, while keeping the rest of the parameters fixed at the maximum a posteriori (MAP) solution (**d**). **b.** A decrease in the $E \to E$ assembly strength (R_{EE}) introduces synchronous burst-like events that jeopardize co-tuning and weight diversity. Further reduction of the $E \to E$ assembly strength results in sparse, asynchronous spiking which significantly reduces tuning quality. **(c)** Reduction of $E \to I$ assembly strength first leads to synchronous inhibitory firing across groups and further reduction leads to the persistent activity of the whole inhibitory population combined with bursts of excitatory activity that prevents the development of diversity. **(e)** Decreasing the $I \to E$ assembly strength leads to persistent activation of a single group of neurons (which affects the tuning only marginally). While the increase of the $I \to E$ assembly strength leads to oscillatory behavior of the whole network. **(f)** Weakening the $I \to I$ assemblies decreases the weight diversity by introducing occasional synchronous bursts in the network while strengthening them leads to asynchronous inhibitory activity.

stimulus-driven activity that only minimally reduces the balance and weight diversity (Fig. 4.5f).

The sparsity of a network's recurrent connectivity shifts the optimal assembly structure

Finally, we investigate the effects of the sparsity of the presynaptic populations on the formation of E/I balance and neuronal tuning. Biological neural networks are usually very sparsely connected (Barral & D Reyes, 2016; Seeman et al., 2018; Wildenberg et al., 2021), and the sparsity of connections is associated with distinct dynamics (Brunel, 2000).

We observed that the impact of noise and recurrence on the deterioration of the balance and weight diversity in sparse networks without assemblies is qualitatively similar to fully connected networks. Thus, we examined the ability of neuronal assemblies to produce activity that restores

balance and weight diversity in sparsely connected recurrent networks that receive noisy input.

We ran ABC to approximate the optimal distributions of assembly strengths for 5 different levels of sparsity with the probability of connection $p = 1.0$, $p = 0.75$, $p = 0.5$, $p = 0.25$, and $p = 0.1$. We preserved the total input per neuron across different sparsities by scaling the coupling strength inversely proportional to p.

The first thing that we found is that the higher sparsity levels decrease the ability of assemblies to improve the tuning and E/I balance on the postsynaptic neuron (Fig. 4.6a). The overall loss after 21 ABC steps is larger for the sparse networks than for fully-connected networks and increases with sparseness, (Fig. 4.6b). It was, however, still beneficial to have assemblies (Fig. 4.6a). Thus, both the balance and diversity were significantly higher in the network with assemblies, although the absolute values were low compared to a fully connected network.

Secondly, we find that the optimal assembly strength values are shifted for different sparsity levels. The optimal strength of the $E \to I$, $I \to E$, and $I \to I$ are weaker for the sparser networks. The strongest effect was observed for the $I \to I$ assemblies (Fig. 4.6c), where the optimal solution included no $I \to I$ assemblies.

Overall, despite an improvement in the balance and diversity metrics for most sparse networks, the diversity was strongly affected by sparseness and could not be fully recovered by assemblies (Fig. 4.6a). This reduced effectiveness is expected given the smaller number of connections and the greater variance in the network's connectivity.

4.3 Discussion

Input selectivity is a common characteristic of brain networks across different levels of hierarchy. The emergence of such selectivity in higher brain areas depends on a neuron's ability to differentiate downstream inputs that arrive from highly recurrent networks. Nevertheless, most studies of the emergence of E/I co-tuning via plasticity (Clopath et al., 2016; S. Eckmann & Gjorgjieva, 2022; Luz & Shamir, 2012; T. P. Vogels et al., 2011) assume a near-optimal input structure of distinct, minimally correlated signals.

While there is a lot of evidence that in lower areas that receive direct sensory that strongly preserves the original inputs (Antolík et al., 2016; Klindt et al., 2017), in higher cortical areas, the topology of the recurrent networks places strong constraints on the statistics of these inputs.

Here, we demonstrate that two ubiquitous features of biological networks, internal noise, and recurrent connectivity, can impact the statistics of inputs and prevent the plasticity-mediated development of input selectivity on a postsynaptic neuron. Furthermore, both recurrence and

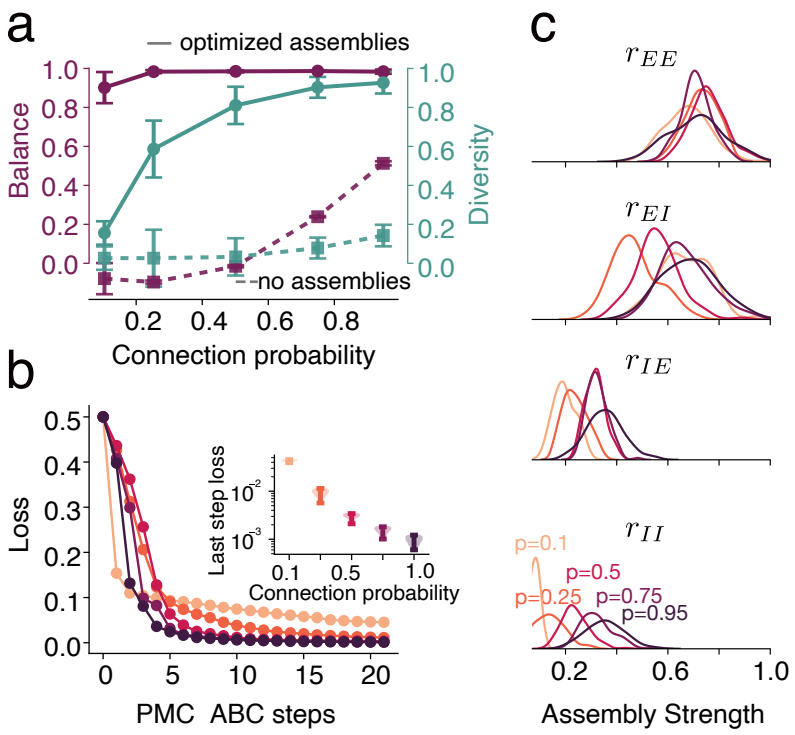


Figure 4.6: Assemblies improve co-tuning and allow for co-tuning in sparse networks. **(a)** E/I balance (purple) and weights diversity (teal) in the networks with assemblies compared to non-structured networks (dashed lines), error bars — standard deviation. The noise level is 0.1 for all sparsities and the coupling is 0.85 (scaled by $1/p$). **(b)** Loss for the sparser networks is higher, which results in the overall worse performance, inset shows the loss for 50 accepted samples at the last ABC step. **(c)** Posterior distributions of all assembly strengths change with sparsity. Sparser networks require weaker inhibitory assemblies (more uniform connections) to produce co-tuning.

noise strongly influence the ability of neurons to maintain E/I balance. We hypothesized that the negative effects can be overcome by introducing inhomogeneous connectivity in the presynaptic network. We examined whether the connectivity of the pre-synaptic network can be modified to create a more beneficial statistical structure for E/I balance and input selectivity to emerge.

We investigated the idea that diverse neuron types may create functional assemblies. Introducing neuron-specific assemblies indeed helped to overcome the negative effects of recurrence and noise. Plasticity mechanisms acting on a post-synaptic neuron driven by assemblies in a presynaptic network could produce balanced E/I weights and input selectivity.

Cortical networks are known to be highly clustered (D'Souza et al., 2019; Hilgetag & Kaiser, 2004) with different kinds of neurons being more likely to follow different connectivity patterns. Moreover, there is strong evidence for the presence of neuronal assemblies, characterized by

strong interconnectivity and correlated functions (Badin et al., 2017; Miehl & Gjorgjieva, 2022; Umbach et al., 2021). Recent experimental results that track the connectivity in the cortex with high precision also suggest that functionally grouped neurons form assemblies and that such assemblies can even span more than one cortical column (Ding et al., 2023).

The optimal assembly structure that we found is to some extent similar to classical lateral inhibition (Blakemore et al., 1970). The optimal correlation structure in the presynaptic network is promoted by the ability of each group to localize the activation by inhibiting the other groups. Simultaneously, the full picture appears to be more complex as the optimal structure that we found includes a significant degree of inhibitory assemblies. Moreover, there is a similarity between the presynaptic network structure that we found and the recently proposed efficient balanced network model (Deneve & Machens, 2016). In the network proposed by Denéve and Machens (Deneve & Machens, 2016) inhibition acts laterally and cancels the input signal. This allows the network to automatically represent an incoming signal as an error and leads to efficient coding of input signals. The degree to which the structure discovered here can be used in the efficient coding scheme remains to be studied.

To examine the functional implications of synaptic type-specific neuronal assemblies, we parameterized the network topology using a quantitative metric for the strength of different assembly types. This enabled efficient simulation-based inference of optimal assembly strengths and analytical treatment of a linear model with similar connectivity constraints (Giannakakis et al., 2023). However, our approach was not intended to identify the optimal recurrent connectivity for the emergence of E/I co-tuning, but rather to demonstrate the functional significance of overlapping functional assemblies.

Direct optimization of individual recurrent weights is likely to yield a better solution than our simplified parametrization. Several gradient-based methods for training spiking networks or an evolutionary algorithm could potentially accomplish this optimization. Whether this optimal recurrent connectivity would maintain any features of the topology we identified through our parameter inference method remains an open question (Bellec et al., 2020; Zenke & Vogels, 2021).

Finally, here we focus on one specific setting in which plasticity mechanisms form E/I balance and tuning at the postsynaptic neuron. However, the general picture of how tunning and E/I balance emerge and are maintained in larger networks without initial structure is still missing. Theoretical work shows how signals might propagate throughout such a network, which may allow for fast and target signal propagation governed by local changes in E/I gain (T. Vogels & Abbott, 2009). Simultaneously, there are studies showing how network structure can emerge

via plasticity rules (Effenberger et al., 2015). It is still not clear how and whether plasticity mechanisms alone can support the emergence of tuning in an initially unstructured network. *In vitro* networks of neurons can be proposed as an optimal experimental platform that can further guide this research. Although, it is worth mentioning that inducing tuning in non-structured networks experimentally in cultured neuronal networks was attempted multiple times by different groups with only limited success (Renault et al., 2015; Schottdorf, 2017; Shahaf & Marom, 2001; Wagenaar et al., 2005).

In summary, we show that a combination of plasticity mechanisms (S. Eckmann & Gjorgjieva, 2022; Gjorgjieva et al., 2011; T. P. Vogels et al., 2011) allows the development of co-tuned excitatory and inhibitory connectivity in a feedforward, low noise spiking neuronal network. The introduction of high noise levels or recurrent connectivity in the pre-synaptic network leads to cross-correlations between groups that support input selectivity. Strong excitatory assemblies together with weaker inhibitory assemblies restore the input covariance and allow for the emergence of co-tuning.

Chapter 5
Conflict monitoring and change of mind mediated by population of Anterior Cingulate Cortex neurons

Summary

In this chapter, we analyze population responses in Anterior Cingulate Cortex (ACC) in rats during near-mistakes using statistical modeling of single neuron and population responses. ACC is one of the key brain regions implicated in behavior monitoring, and detection of conflict between two action plans. In this study, we measured responses in ACC neurons during the conflict in the Go-NoGo task. We recorded single-unit activity in 574 ACC neurons during near-mistakes, a type of behavior when the animals have initiated a wrong response but rapidly self-correct. Unlike classically used stop signal tasks, near-mistakes allow us to have a clear behavioral readout of the conflict with precise timing. By modeling single-cell and population responses we showed that ACC neurons encode behavioral and task-specific information and display conflict-related responses that scale with the conflict intensity. We find that neural responses that scale with conflict start as early as the initiation of the incorrect action and persist even when the conflict is already resolved. Our research expands the understanding of the ACC's role by demonstrating that neurons in this region may also drive change of mind behavior.

Contributions. This work has been done in collaboration with Nelson Totah (NT), Dmitii Vasiliev (DV), and Ryo Iwai (RI) NT, DV, and RI performed all the experimental work. DV performed the spike sorting and implemented single-unit ANOVA analysis and the analysis of rat EEG data (not discussed here). My contribution included data analysis, designing and fitting the Linear Model, and implementation and fitting of the demixed PCA.

Contents

5.1 Introduction

Engaging in a task with alternative choices that have different values often leads to errors that animals need to adapt to (Rabbitt, 1966). Furthermore, the ability to perform a task accurately may be disrupted by other prepotent responses (Eriksen & Eriksen, 1974; Stroop, 1935) or internal states (Benoit et al., 2019; Boksem et al., 2006; Lorist et al., 2005). Conflict monitoring is one of the mechanisms that help overcome these limitations and engage cognitive control when it is needed to reduce the number of errors (Botvinick et al., 2001; Fu et al., 2023; Yeung, 2013).

Conflict monitoring has been linked to the activation of the Anterior Cingulate Cortex (ACC), located on the medial surface of the frontal lobe. ACC is connected with the prefrontal and parietal cortex, as well as with subcortical areas, which makes it anatomically well-suited for action monitoring (Monosov et al., 2020).

ACC activation has been shown in a variety of tasks since the early days of brain imaging (see for example Taylor et al., 1997 and Botvinick et al., 2001 for an overview of early studies). Later, conflict responses in ACC neurons were also found in human and primate electrophysiological studies (Ebitz & Platt, 2015; Fu et al., 2022; Fu et al., 2023; Monosov et al., 2020). The main conflict monitoring theory states that the ACC increases its activity when the opposite representations (of action or stimuli) are co-activated (Botvinick et al., 2001). Thus, the activity of ACC increases in tasks where subjects encounter conflicting stimuli or need continuous monitoring to determine which action to select. Among others, a simple Go-NoGo task, in which a subject has to respond to one stimulus and ignore a different stimulus, drives the activation of ACC as has been shown in human fMRI studies (Hester et al., 2004).

Conflict monitoring allows for the dynamical adjustment of incorrectly selected actions. This may result in the correction of behavior and it is a mechanism through which animals, including humans, mitigate undesired outcomes resulting from incorrect choices (M. X. Cohen & van Gaal, 2014). Some electrophysiological studies in humans and monkeys showed conflict responses in ACC neurons even during action preparation (Fu et al., 2023). Others, however, did not find evidence for early conflict signaling in ACC (Ito et al., 2003). Most of these studies focused on conflict responses in different types of stop signal tasks, where a cue indicates that a subject should not engage in an action. Thus, in these tasks, there is no clear behavioral readout of the conflict.

The theory of conflict monitoring (Botvinick et al., 2001) predicts that conflict arises when the correct action (NoGo) and chosen action (Go) are in conflict and interfere with one another.

Thus, the magnitude of the conflict should scale with a behavioral mismatch between the correct and chosen actions. Responses in ACC should scale with the conflict magnitude, which has not yet been shown on a single neuron level.

In this study, we focus on conflict responses in the ACC during self-correcting (near-mistake behavior) in the Go-NoGo task. We establish a platform to study the near mistakes in a rodent model and record the activity of a single neuron in rat ACC. With single neuron and population analysis, we discovered that ACC neurons closely monitor both the animals' behavior and the task. We also show that conflict can be related to ACC population activity. These responses scaled with the magnitude of the conflict and emerged both during the conflict and after it had already been resolved. In summary, we show diverse responses in the ACC that allow tracking the behavior and strongly indicate that the ACC might be mediating the Change of Mind (CoM).

5.2 Results

5.2.1 Conflict in a Go-NoGo task

We quantify the conflict in ACC by considering the response conflict in trials where animals nearly made a mistake in a Go-NoGo task. We trained the head-fixed rats (N=43) to discriminate visual grating stimuli (Fig.5.1a). The rats were trained to run on a treadmill past a distance threshold for a Go stimulus or to remain immobile for a NoGo stimulus. When animals committed an error, e.g. ran past the distance threshold for NoGo stimuli, they received the brown noise as negative feedback. Correct responses for Go stimuli were rewarded with water. Finally, in some trials animals incorrectly started running to NoGo stimuli but stopped before crossing the distance threshold (Fig. 5.1b). These Correct Rejection trials(CR) allow us to study *near-mistake* behavior. Overall, in this study we focus on the following conditions:

- Hit (HT) - running past a threshold for a Go stimulus (correct)
- False Alarm (FA) - running past a threshold for a NoGo stimulus (incorrect)
- Near mistakes in CR trials - suppression of initial running for a NoGO stimulus before the threshold is crossed (correct)

From a cognitive control theory (Botvinick et al., 2001) we expect that conflict should emerge in CR trials. Running is a prepotent response for rats and in CR trials during near mistakes, they should suppress the ongoing actions. Thus, competing populations of motor neurons are

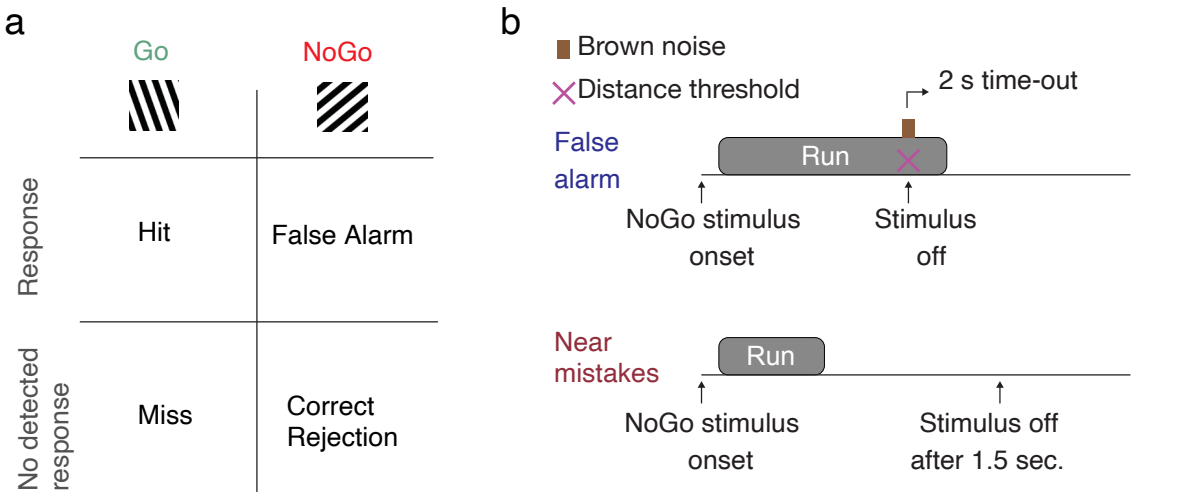


Figure 5.1: Near-mistakes in a Go-NoGo task. **(a)** Trials types in the Go-NoGo task. **(b)** In a subset of (CR) trials, animals started running but self-corrected the wrongly initiated behavior, which led to Near-Mistakes. In FA trials, rats run until they receive negative feedback – brown noise.

activated, which generates response conflict. To quantify the conflict strength, we calculated the running velocity of rats on the treadmill. We expect the conflict to scale with the running velocity in CR trials and peak, as the activations of running and stopping motor commands would be proportional to the animal's running speed in CR trials. The average velocity of the 49 rats in CR trials showed a clear bell-shaped profile. The peak of the velocity, thus, indicated the change from running to stopping and can be used as a behavioral proxy for the conflict (Fig. Fig. fig:population:velocitya). We further split the peak velocity into three tertiles with respect to the maximum running speed on each trial type, which allows us to average over 3 levels of conflict.

5.2.2 Single unit responses

Statistical analysis of single-neuron activity demonstrates scaling with conflict. We analyzed the activity of 603 single units in the ACC of 4 rats during the Go-NoGo task. To examine conflict responses, we aligned spiking activity with peak velocity and assessed the difference between mean responses in CR trials. We found that 283 neurons (out of 603 considered) showed a significant change in the firing rate prior to, during, and after the peak velocity (one-way ANOVA, $p<0.05$, not corrected). The responses showed both positive and negative scaling (Fig.5.3a)

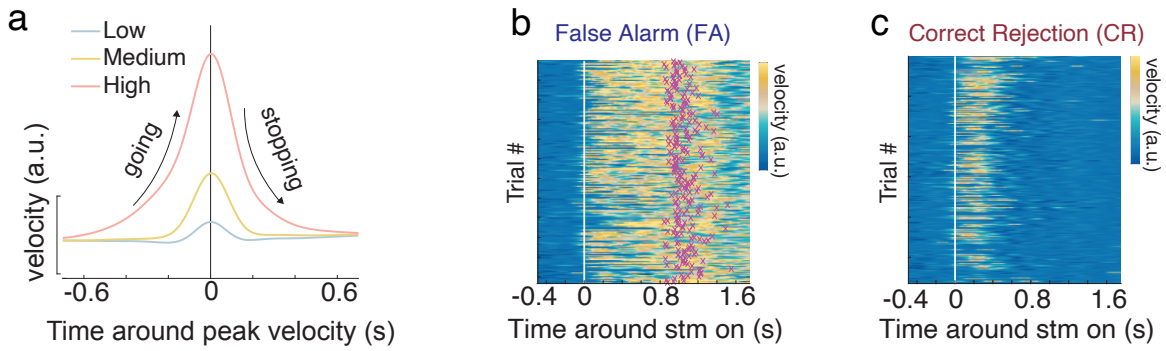


Figure 5.2: Peak running velocity as a proxy for conflict. Stopping required the initiation of a NoGo response. We hypothesize that a Go response requires a matching NoGo response to counterbalance and, therefore, conflict is proportional to the running velocity. **(a)** Average velocities in tertiles averaged over $\sim$ 24 thousand trials in 43 rats show clear bell-shaped patterns. We expect the conflict to be the largest at the peak of it. **(b)** Rats keep running until the distance threshold in the False Alarm (FA) trials and stop early in the Correct Rejection (CR, **c**) trials.

To rule out the velocity-related responses, we ran the analysis on HT where no conflict was present. We found that indeed a part of the ACC population modulated its firing rate with speed (105 units) (Fig.5.3b). However, only a fraction of them overlaps with the conflict-related modulation of firing rate (70 cells). Thus, approximately 196 single units scaled their firing rate in CR trials but did not show a scaling with speed in CR trials (Fig.5.3a).

5.2.3 Linear model reveals mixed and isolated selectivity of single neuron responses

To relate the neural responses to a broader set of task-related and behavioral variables in the Go-NoGo task, we developed a linear encoding model for ACC neurons. Testing neuronal responses with ANOVA showed that neurons in ACC may be monitoring speed and possibly other task-related variables. To account for a more extensive set of continuous and discrete variables, we designed a GLM-based encoding model. Inspired by Engelhard et al. (2019), our model utilized an expressive spline basis to fit neuronal responses (Silverman & Ramsay, 2002). The model had 3 types of predictors: discrete events such as stimulus or feedback onset, and time of velocity peak, continuous variables, and trial-wide variables. Discrete events were convolved with a set of splines extending backward and forward (refer to methods G for additional details). Running velocity was a continuous variable in the model. We also added trial-type regressors (1 or 0) for HT, FA, and CR trials. The model also incorporated a set of session-wide splines

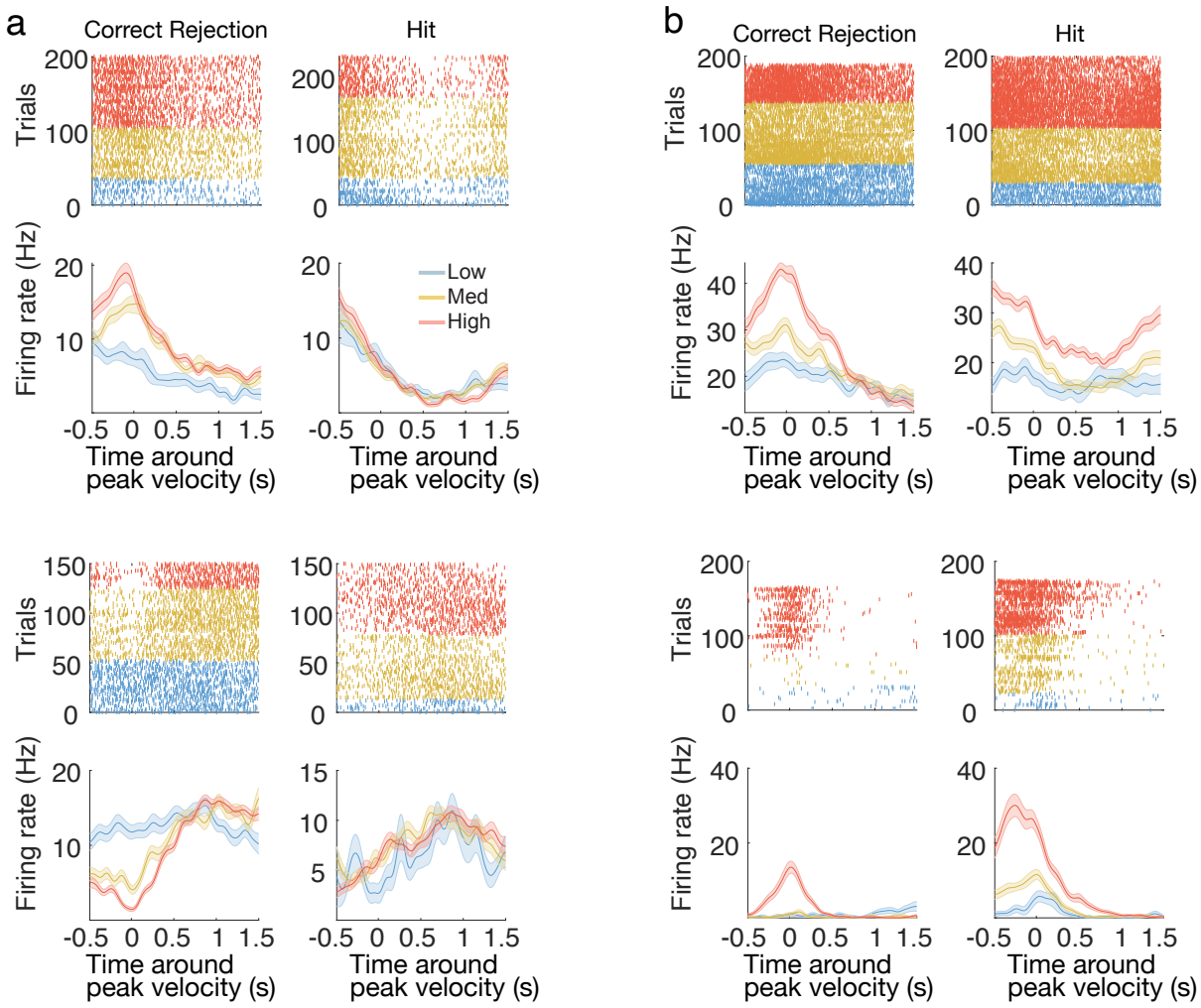


Figure 5.3: Examples of isolated conflict responses and scaling of the activity with velocity in firing rates of single units in rat ACC. **(a)** Examples of units that increase (upper) and decrease (lower) their firing rate with different levels of conflict, while not modulating their firing rates for Hit (HT) trials. **(b)** Velocity units with high and low background firing rates, the firing rate changes proportionally to the running velocity in both CR and HT trials.

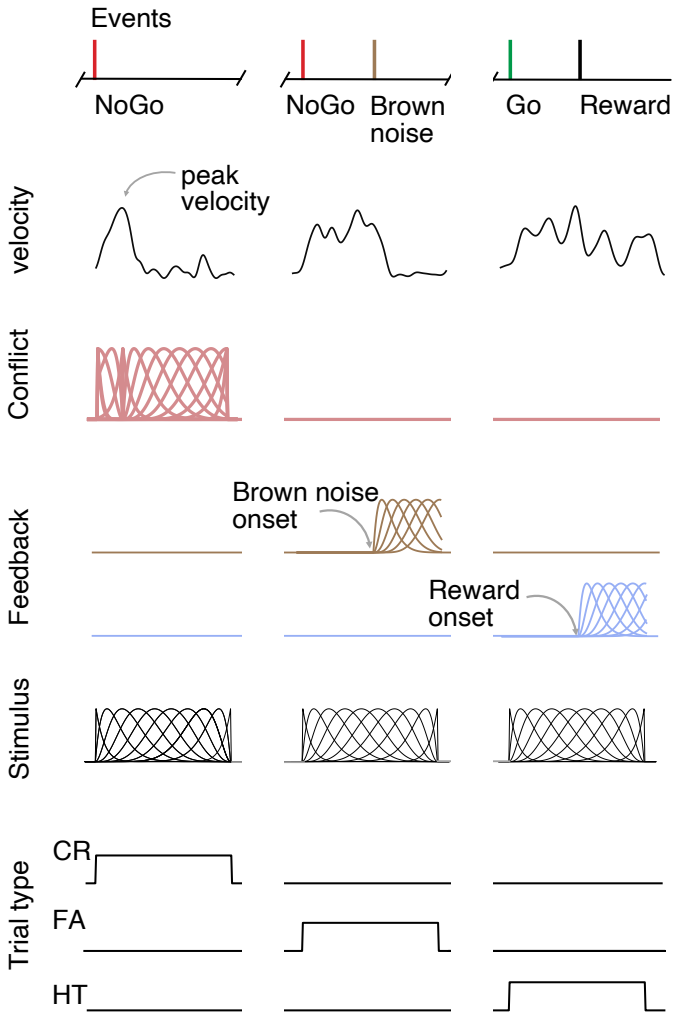


Figure 5.4: Regressor groups in the linear encoding model. We used running velocity as a continuous regressor for neural activity. Events such as conflict, stimulus, feedback, and peak velocity (not shown) onsets were convolved with a set of splines. The amplitude of conflict and velocity splines was scaled by the peak velocity for each trial. Trial types are the whole-trial regressors (0 or 1 within the trial time)

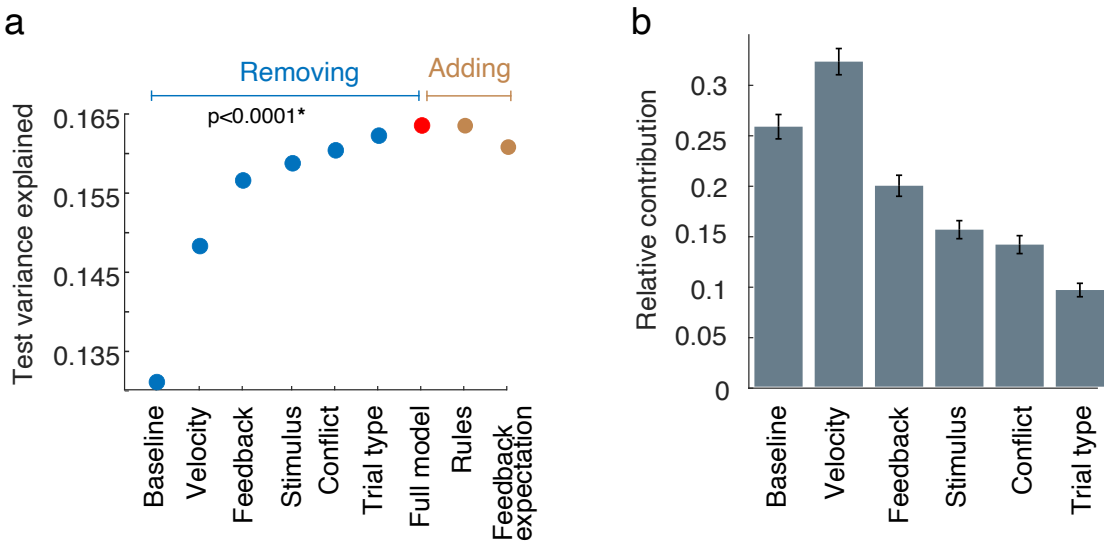


Figure 5.5: Model selection and the average contribution of each regressor in the population of ACC neurons **(a)** The model was selected based on the improvement of the 5-fold cross-validated variance explained. The red dot shows the final model (includes the regressor groups to the left of it). Excluding any of the regressors from the full model leads to a statistically significant decrease in performance (blue dots). Components that were tested and not included in the final model (orange) do not change or decrease the performance when they are added. **(b)** Shows the average relative contribution of each regressor group. Contributions are calculated as the relative decrease in the test variance explained when the regressor group is excluded.

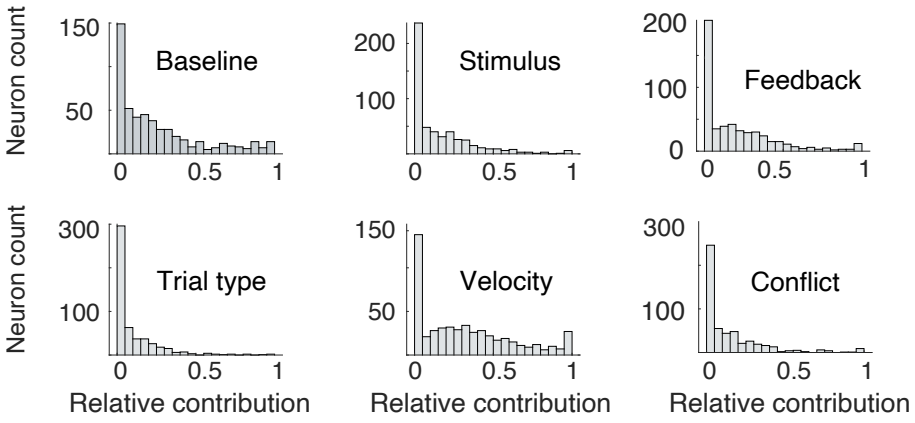


Figure 5.6: Contributions of the encoding variables across the whole recorded population of ACC neurons.

to account for drift in single-unit recordings The regressors were unique for each recording session (Fig. 5.4).

The recorded spike trains for each neuron were convolved with a 50ms Gaussian kernel to approximate an instantaneous firing rate. We cut all the data outside of the experimental trials and fit the encoding model (see G). Overall, we fit the responses of 522 neurons recorded in 26 sessions. To determine the composition of regressors in the final model, we compared the test variance explained (R^2) in 5-fold cross-validation, averaged over all neurons between models with different regressors (Fig.5.5). We grouped the regressors according to their different behavior or task relevance. Thus, we considered 6 groups (Fig. 5.5a). The final model incorporated regressors that significantly enhanced test set performance ($p<0.05$, Bonferroni-Holm correction). We excluded two groups of regressors: the interaction of the stimulus spline with a trial type (Go or NoGo, e.g. task rule regressors) and reward expectation regressors. The final model explained about 16.3% of the overall variance (Fig. 5.5a).

We analyzed the relative contribution of regressors to neuronal responses to build the response profiles of individual neurons. To calculate the relative contribution, we removed each type of regressor group from the final model and compared the residual variance to the full model's variance (Fig. 5.5b). On average, the velocity regressors had the largest contribution, and the contribution of the conflict regressor was about 15%. On the individual neuron level, we found that for some neurons the relative contribution was large for isolated regressors. For others, the contributions were less specialized. The encoding model allows us to build a profile of contributions for each neuron. Fig 5.7 displays examples of predicted average responses from the encoding model and relative contributions of the regressors.

ACC neurons respond to conflicts with varying latencies, both during and after the peak of the running speed. The first group mediates conflict before behavioral change, potentially driving a change of mind. For each neuron with a non-zero contribution from the conflict regressors, we calculated the peak conflict response by projecting the neuron's GLM coefficient onto the conflict splines and identifying the maximum value. The distribution of peak conflict timescales ranged broadly, spanning from -400 ms to 1200 ms around the peak velocity (Fig. 5.8). Therefore, the regression analysis strongly indicates that conflict units are heterogeneous and may both mediate the change of mind and assist in propagating information about conflict magnitude.

5.2.4 Clustering the neurons based on the dominant contribution

Neurons in the ACC can be clustered according to their dominant role, yet many show mixed encoding properties. The distribution of contributions-regressors over the whole population

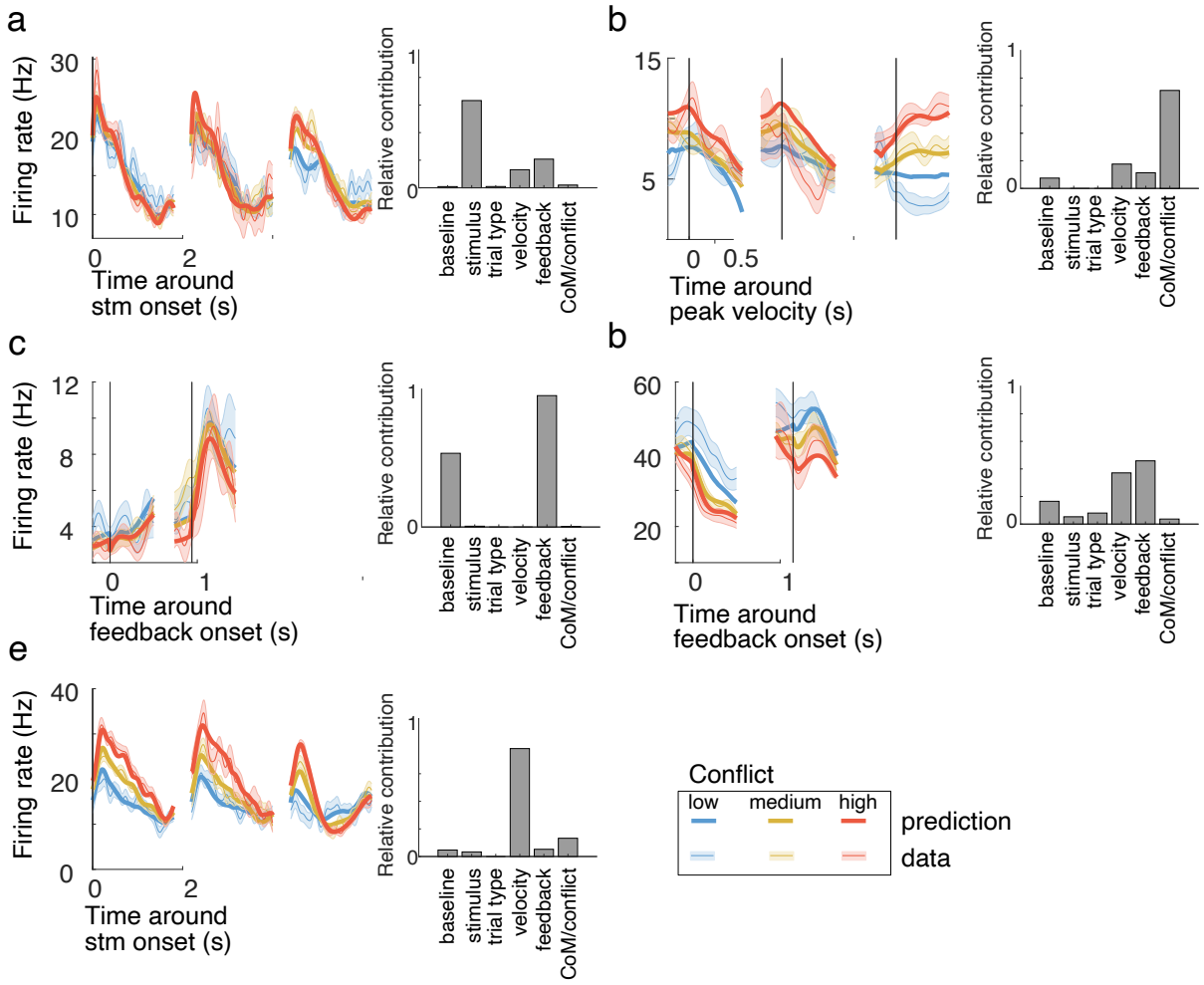


Figure 5.7: Examples of units with different encoding profiles based on the regressor contributions. Mean firing rate ± s.e.m – pale colors, model prediction – solid line; **(a)** is an example of a neuron primarily driven by stimulus presentation (from the stimulus cluster, see the main text), **(b)** is a late conflict unit, **(c)** and **(d)** are pure feedback encoding and a mixed feedback encoding neurons, **(e)** is a velocity neuron.

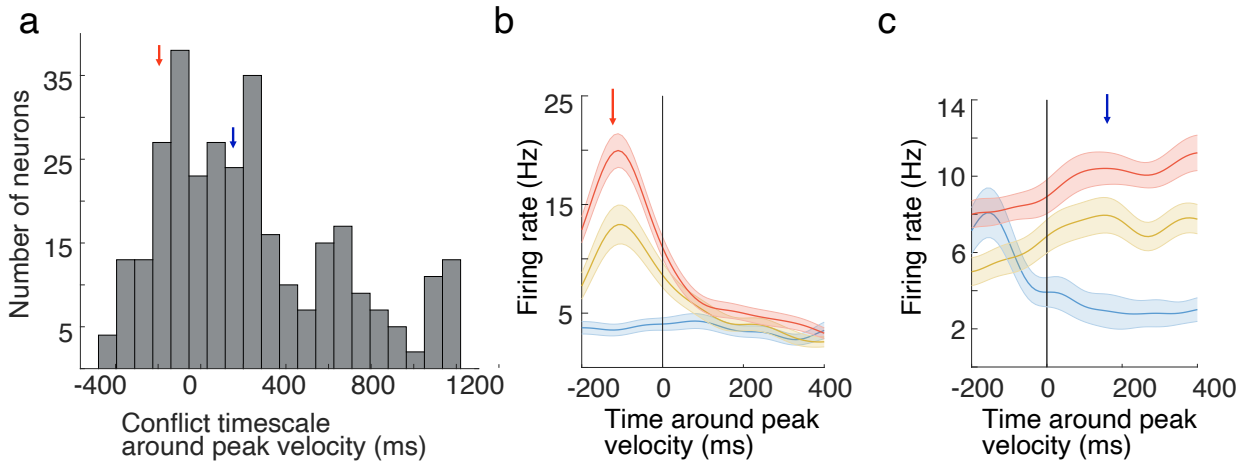


Figure 5.8: ACC neurons mediate behavioral conflict with a wide range of response timescales. **(a)** The distribution of the peaks of the conflict responses was calculated by projecting the GLM coefficient to the conflict spline basis. Some units scale their firing rate before the peak velocity **(b)**, others afterward **(c)**.

of ACC neurons is not uniform and suggests that there might be several specialized groups of neurons (Fig. 5.6). We asked if neurons can be clustered according to their contributions. We combined the relative contributions from all considered single units and fitted a Gaussian Mixture Model (GMM). To determine the number of clusters, we employed the Bayesian Information Criterion (BIC), which balanced the trade-off between complexity and goodness of fit and allowed us to select a model. It penalizes the overfitting through a term proportional to the number of parameters in the model and the logarithm of the sample size (Fig. 5.9b).

We detected four clusters of neurons whose activity could be mostly predicted by: stimulus, velocity, conflict, and feedback (Fig. 5.9). The clustering explained the contributions better than chance (Fig. 5.9c, $p<0.0001$). Interestingly, even though neurons could be clustered according to the dominant contributions, many still carried information from the other group. Some single units still have a significant contribution from other regressor types and show mixed selectivity (see examples in Fig. 5.3).

5.2.5 Demixed Principle Components Analysis extract low-dimensional trajectories of conflict responses in ACC

To further elucidate the variability of conflict and mixed responses, we mapped the activity of the entire population of neurons pooled across recording sessions onto a small number of task-relevant principal components. To this end, we used demixed Principal Component Analysis

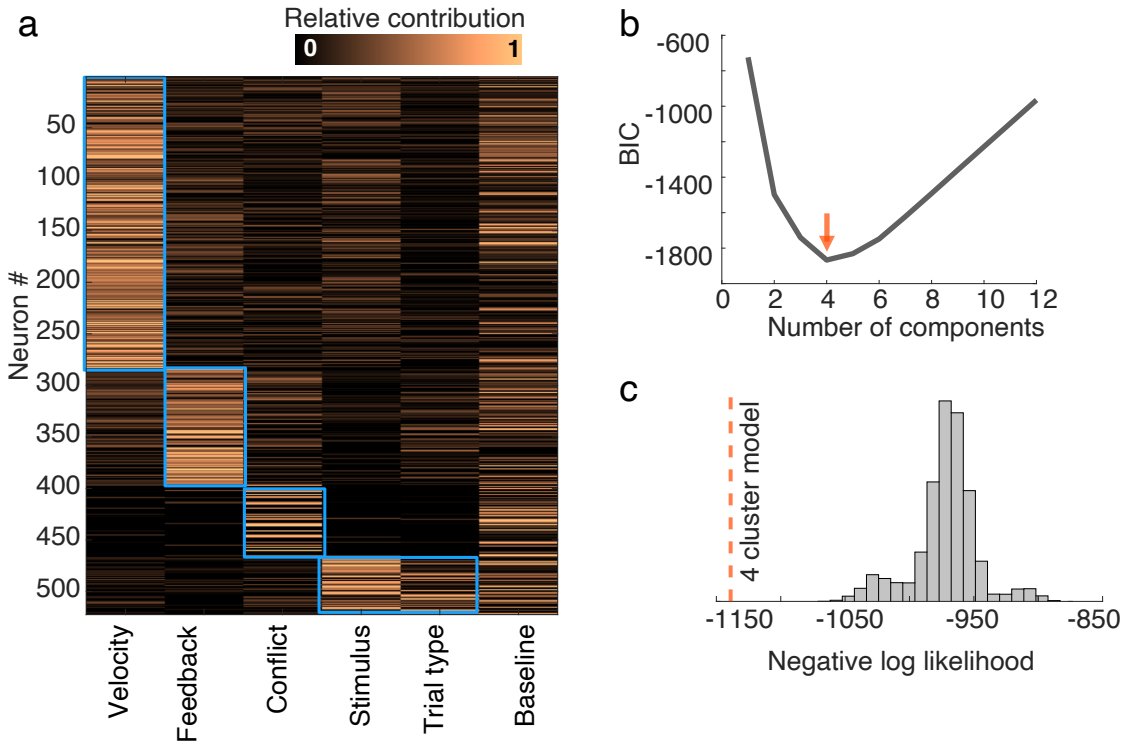


Figure 5.9: Neurons clustering based on the relative contribution of model regressors to their neural responses. **(a)** Heatmap of the relative contributions per neuron sorted according to the four clusters: velocity, feedback, conflict, and stimulus (that also has a large contribution from the trial type regressors). Baseline contributions were not used for the clustering. (**b**) Bayesian information criterion (BIC) is minimal in the GMM with 4 clusters (average over 10 runs). **(c)** The model with four clusters explained the contributions better than chance ($p<0.0001$). The histogram shows the distribution of negative log-likelihoods for the 4 clusters GMM trained on randomly reshuffled contributions per neuron (n=10000).

(dPCA, Kobak et al., 2016). This method enables the identification of principal directions in the tensor of average neural responses while maximizing behavior or task-relevant information in each component (Fig. 5.10 and see also H for further methodological description). We employed a 3x2 ANOVA-like factorization with three velocity tertiles in CR and HT trials. Thus, we extract the components that can decode trial type, speed, and interaction between them, as well as the independent evolution of the firing rate that is not related to any of these factors.

The evolution of the firing rate in the space of the top components indicates a clear separation of the responses depending on the condition (Fig. 5.10b). The firing rate in CR trials projected onto the interaction components starts diverging prior to the peak, which indicates the early presence of the change-of-mind signal along with conflict monitoring (Fig. 5.10b).

Here, the components are projected such that, on correct rejection trials, the interaction component is positive and has a clear separation between the velocities, whereas, in the hit trials, the velocities are hard to separate. Note that the interaction components were symmetric (Fig. 5.11a) and there is a different projection in 3D in which speed could be decoded only during HIT trials but not in CR. The main reason for that is that in HIT trials, the animals keep running until the trial ends and the velocity can be detected, at the same time rats quickly terminate the movement in CR trials (Fig. 5.12).

The first 8 components of the model accounted for approximately 80% of the explained variance. Most of the variance was attributed to the trial type, followed by two components explaining the shared time evolution of the firing rate independent of the factors, then speed, and approximately 7% of the variance was attributed to the first two interaction components (Fig. 5.11a,b). The interaction components account for the variance associated with units that scale their responses with speed differently between CR and HIT trials. We examined whether the trial type could be decoded from the top two components for each condition (Fig. 5.11c). For all of these components, we could decode the trial types of the left-out trials above chance (G). The speed could be decoded with a test accuracy of 0.7, the trial type with an accuracy close to one, and the interaction with a test accuracy of 0.5.

Finally, the decoder weights distributions highlight the isolated and mixed responses of individual units to conflict, trial types, and velocity (Fig. 5.12). We looked at the decoder weights pairwise for each combination of condition components, considering only the first components for each condition. As a result, units with pure responses were more likely to have orthogonal weights; for example, pure speed neurons possessed a large speed weight and near-zero weight for the other components (Fig. 5.12a). Nonetheless, this analysis revealed that most cells exhibited mixed selectivity, displaying responses across all conditions.

5.3 Discussion

According to a classical theory, conflict can arise mechanistically when two opposing actions are present (Botvinick et al., 2001; Botvinick & Cohen, 2014; Yeung, 2013). The medial frontal cortex, particularly the ACC, has been proposed as the brain region responsible for implementing conflict monitoring. This idea was further supported by imaging studies (Carter et al., 1998; Hester et al., 2004; Iannaccone et al., 2015; Taylor et al., 1997) and several electrophysiological studies in monkeys and humans (Ebitz & Platt, 2015; Fu et al., 2023; Monosov et al., 2020). Here, we develop a novel paradigm to study the conflict responses that naturally arise in moving

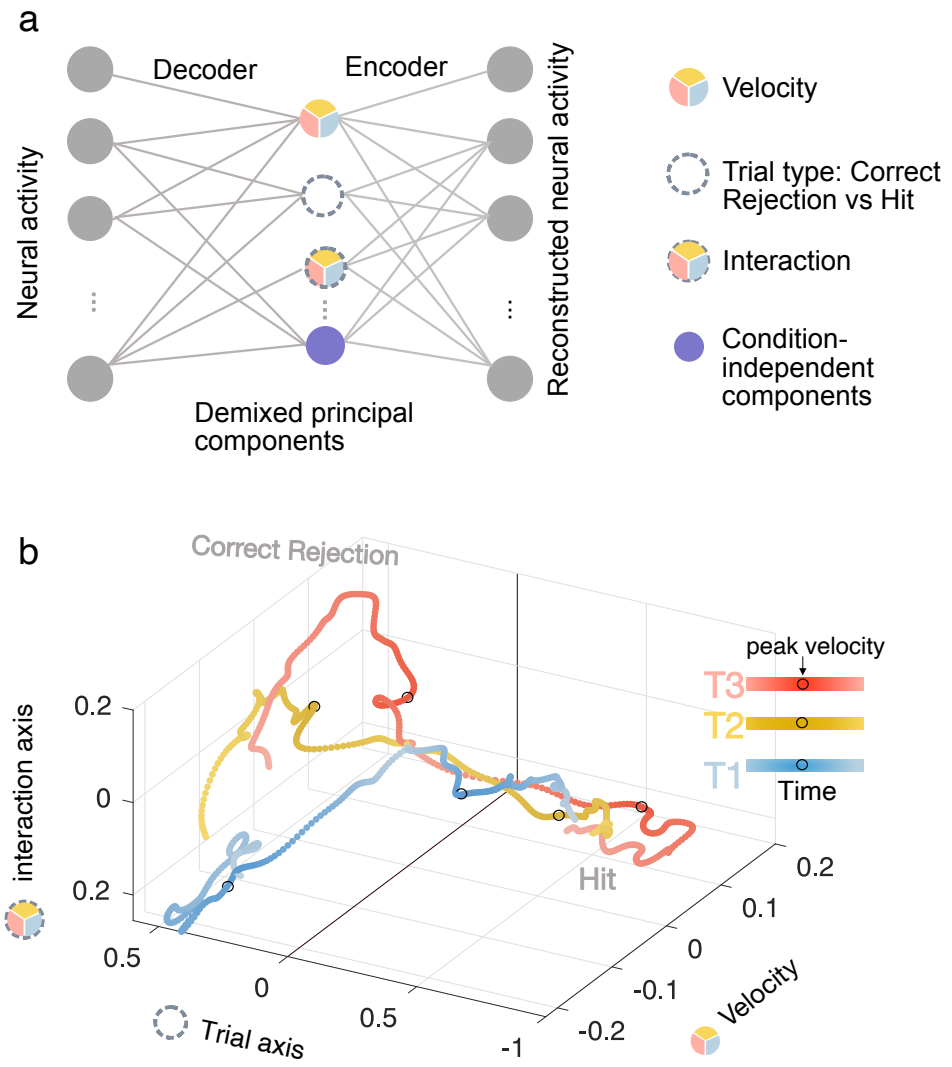


Figure 5.10: Demixed principal component analysis of population responses in Go-NoGo task during near mistakes **(a)** Demixed PCA uses marginalizations of the average responses with behavior/task-related parameters: running velocity tertiles (T1, T2, T3), hit and correct rejection trials, the interaction of the two, as well as task-independent marginalizations (averages over all trials). **(b)** Population conflict responses in the dPC space. Average neural responses were projected onto the 3 top demixed PCs for running velocity, trial type, and interaction of the two.

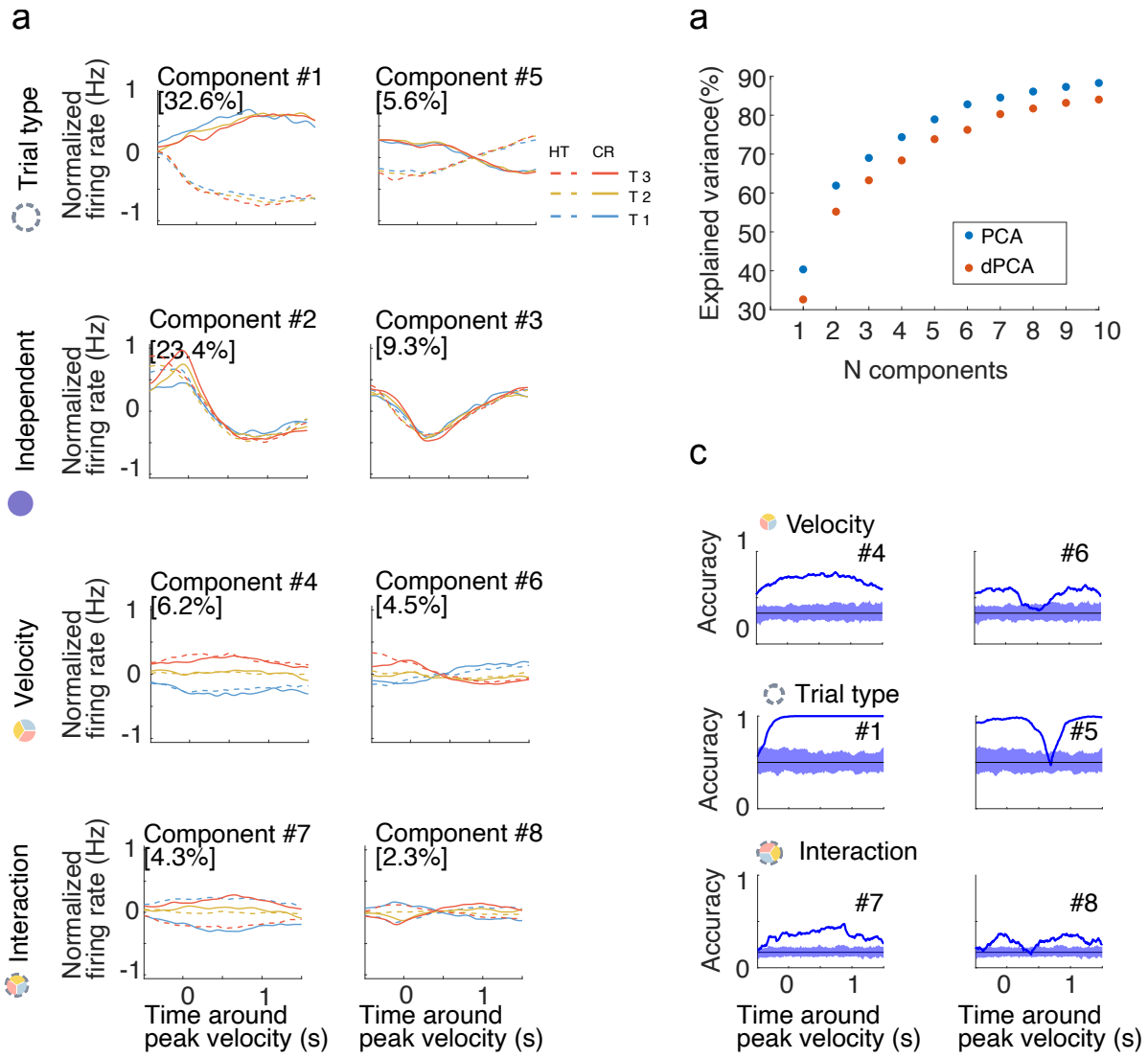


Figure 5.11: Demixed principal component analysis of population responses in Go-NoGo task during near mistakes **(a)** Two top demixed components for each marginalization: trial-type, condition-independent (time), speed, the interaction of trial type and speed. **(b)** Variance explained by first 10 PC vs first 10 demixed components. PCA was applied to trial average responses **(c)** Decoding of individual trial labels (trial type, velocity in this trial, etc.) by projecting them onto the decoding axes for the top 6 components. The accuracy is calculated for every time bin.

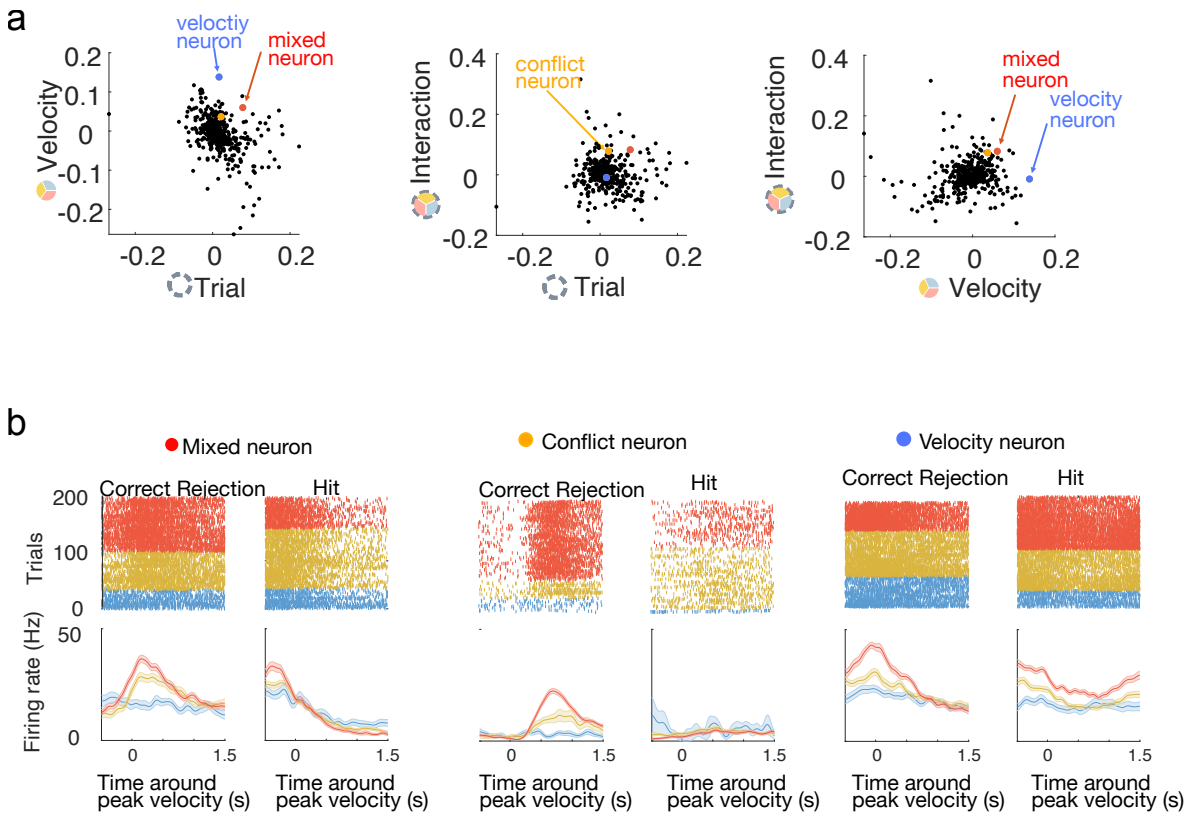


Figure 5.12: Examples of pure speed, and conflict neurons as well as the neurons with mixed responses. The encoder weighs for three top dPCs and examples of single-unit responses. **(a)** Encoder weighs for three top dPCs **(b)** Example conflict, speed, and mixed neurons.

rats during near mistakes. We directly show that conflict responses arise in a large population of ACC neurons when rats change their behavior from running to stopping in an attempt to correct the erroneous response. The responses scale with conflict magnitude and vary in terms of timescales and arise both before, during, and after the conflict.

The fact that ACC neurons scale their firing rates with the conflicting strength is in direct agreement with the conflict monitoring theory (Botvinick et al., 2001). Using velocity magnitude as a proxy for conflict between running and stopping, we show that ACC neurons scale their firing rate with conflict. This type of scaling, as well as the precise timing of conflict, is hard to obtain in standard tasks used in electrophysiological experiments that typically use stop-signals to generate errors and conflict (Fu et al., 2023). The reason for this is the absence of a clear behavioral readout for the conflict. It remains uncertain whether the responses scale proportionally to the activation for Go and NoGo action in the motor cortex. We can expect

that this is the case, given that scaling of the motor cortex activation with the velocity has been shown in the past (Zimnik & Churchland, 2021). Future studies with simultaneous recordings from both the motor cortex and ACC should help to provide further evidence.

The presence of conflict responses during the action selection as well as responses after an action is initiated were shown in recordings from human ACC (Fu et al., 2022; Fu et al., 2023; Sheth et al., 2012; Smith et al., 2019). In non-human primates, only responses following the action have been shown in various stop-signal tasks (Ito et al., 2003; Nakamura et al., 2005; Sajad et al., 2022). Therefore, the role of ACC in resolving the conflict is debated (Fu et al., 2023).

By examining near-mistakes, we demonstrate that neurons in the ACC exhibit a wide range of conflict-response timescales. We find that ACC neurons scale their activity up to 400ms prior to the change in behavior that resolves the conflict. This indicates that ACC neurons might indeed have a causal role in resolving conflict and might be driving the Change of Mind. ACC neurons might mediate the necessity to change the behavior to downstream areas. However, this prediction requires further validation through optogenetic experiments.

To link the responses in ACC to the behavior, task, and conflict, we used a linear encoding model (Engelhard et al., 2019) with a rich set of regressors. The results establish that ACC encodes different aspects of the behavior. The model also helped to cluster the cells based on their dominant encoding features and indicated that neurons may be involved in monitoring and conflict propagation simultaneously. Although there is a clear indication of cell encoding, the single-neuron GLM accounts for only about 20% of the variance. This could be improved by considering a more suitable neural noise model as Poisson or negative binomial. Additionally, this model does not account for network effects, e.g. components shared across cells that could be potentially generated recurrently or come as an unobserved input. Thus, in the fit of a linear encoding model to responses of VTA neurons (Engelhard et al., 2019) adding the 1st PC of the network activity improved the variance explained up to 80%.

Finally, we directly study the conflict signals at the population level by learning the demixed principal components of the activity. We found that the first 6 PCs were sufficient to account for about 60% of the variance in trial average responses. Interaction components that explain the differences in neural responses between correct rejection and hit trials show that the conflict response robustly emerges already the stimulus presentation and before the peak of the conflict. Demixed PCA is one of the methods used to extract the latent condition-specific components from population recordings (see Hurwitz et al., 2021 for an overview). The resulting projection is based on trial-average activity, which in our case allows us to consider a large population of

neurons that were not recorded simultaneously. That is in contrast to other trial-based methods, such as GPFA (Yu et al., 2009), which are typically used to fit neurons recorded simultaneously.

One famous alternative to dPCA is targeted dimensionality reduction (tDR,Mante et al., 2013). The initial version of tDR underperformed dPCA in terms of demixing quality, yet a more recent version (Aoi & Pillow, 2018) might perform better than dPCA. Given that we see a good convergence between conflict signals detected by three different methods, we expect that tDR should learn a similar latent structure of conflict in ACC responses.

While dPCA enables the direct use of task and behavioral information, it is limited to ANOVA-like factors. Therefore, we could not use continuous behavioral signals like in GLM analysis. Using advanced non-linear methods like LFADS (Sussillo et al., 2016), piVAE (Zhou & Wei, 2020), and recently introduced CEBRA (S. Schneider et al., 2023) can help detect latent task/behavior-related components.

We observed that condition-related variance was distributed across different neurons (Fig. 5.12). This implies that trial-related variance is distributed across average neuron responses. Ebitz et al. 2015 using dTR recently showed a distributed representation of the task-specific information in human ACC. However, in contrast to their results, we observed a separate axis for conflict in the population responses, and we found a population of neurons that is more responsive to conflict.

One of the standing questions is whether conflict detection is a single neuron or a network-level computation. Electrophysiological in humans and monkeys also support that conflict might be mediated by specialized neurons. However, more generally, ACC would need to monitor rules and behavior in completely different contexts (Ebitz et al., 2020; Mante et al., 2013), which might require complex population coding schemes (Mante et al., 2013). Future recordings from an even larger population of ACC as well as RNN modeling should further shed light on this problem and help to establish the computations that ACC is performing whether on a single neuron or population level. However, even from a theoretical perspective, it is not always trivial to determine the necessity of having a specialized population (Dubreuil et al., 2022). Addressing this issue presents both methodological and conceptual challenges (Barack & Krakauer, 2021).

Chapter 6
Concluding remarks and outlook

In this thesis, I demonstrate how model-based approaches can be used to discover the basic mechanisms of network organization and simple dynamical principles of spontaneous collective activity *in vitro*. Next, I show how to efficiently leverage simulation-based techniques to identify the network structure that allows for the emergence of E/I balance and neuronal tuning. Finally, I start untangling complex responses of ACC underly conflict monitoring using statistical modeling. In this last chapter, I will discuss some of the limitations of the modeling work introduced in this thesis and provide perspectives for future research.

Data-consistent models of *in vitro* systems

One of the main contributions of this thesis is related to understanding the dynamics and organization of *in vitro* neural systems. I took two alternative but complementary approaches using a spiking network model and a reduced model.

In Chapter 2, we used a spiking network model and fit a few well-selected parameters of this model to reproduce network bursting dynamics. This approach helped us to identify that networks of hippocampal neurons balance excitatory and inhibitory connectivity to maintain stable dynamics upon a chronic perturbation of cellular E/I ratio.

One of the important limitations of this study is that we do not differentiate between different types of inhibitory neurons. Hippocampal circuits have at least 16 morphologically different subtypes of inhibitory cells (Parra et al., 1998) and they are involved in the regulation of various aspects of cellular and circuit development (Pelkey et al., 2017). The contribution of particular inhibitory neuron subtypes to the development of structural E/I balance in the developing brain remains unclear.

In experiments, we considered only hippocampal neurons. Given that E/I balance both structurally and functionally has been shown for cultures of cortical neurons (Barral & D Reyes,

2016), we can expect a similar adaptation property in these networks (although this remains to be checked explicitly). One of the standing questions, is if differentiated excitatory and inhibitory iPSC neurons with cortical phenotypes, would also exhibit similar network adjustment towards balance or if they are more sensitive to the exact cellular ratios.

From the modeling perspective, one of the main limitations of this study is that we only explore a small part of the model parameter space. Thanks to a very close collaboration with experimentalists we could constrain many parameters based on the data. However, in many cases, this is not possible. Thus, based on modeling alone we would not be able to precisely identify the exact mechanism of E/I balance and could only suggest a few alternatives.

Another challenge is a thorough model selection. We selected the model based on its simplicity and biological plausibility (Giugliano et al., 2004; Segev et al., 2001). However, there is a large number of physiologically plausible alternatives (see discussion in Chapter 2 and 3). One particular example is the precise implementation of the slow adaptation mechanism. Thus, slow spike-frequency adaptation is often suggested (Fardet et al., 2018; Giugliano et al., 2004; Sukenik* , Vinogradov* et al., 2021), although synaptic depression is also commonly used (Orlandi et al., 2013; Yamamoto et al., 2022). As discussed in Chapter 3 different mechanisms might be at play depending on the culture type. This possibility needs to be further explored both in theory and experimentally.

Towards a general model of network bursting

In Chapter 3, we took the key mechanistic ingredients of the network model from Chapter 2 and started building a more general theory of network bursting. Here, we again fit the model to network bursting statistics from real recordings and find that bursting can come from a few dynamical states: bistable, oscillatory, and excitable. Interestingly, various types of networks such as hippocampal, cortical, and hPSC appear to converge to slightly different parts of the phase space. This allowed us to identify the differences between experimental preparations and profile the culture development in terms of its population dynamics.

One of the big open questions is whether these differences in the dynamical states explain some of the experimental variability. One example is the difference between cortical and hippocampal cultures (Charlesworth et al., 2015). Our simplified model appears to capture the changes in the network excitability and reflect the developmental differences (Soriano et al., 2008). However, it remains to be validated if the differences in the excitability over the course of development directly lead to the differences in response to stimulation or to pharmacological perturbations.

One of the findings in Chpater 3 is the invariance of excitability and adaptation parameters. This suggests that the model may (and probably should) be reduced further, which would allow us to build an even more abstract (but hopefully better) model of collective bursting dynamics. Simultaneously, the invariances might be resolved when the model is more constrained by the data, either by considering additional summary statistics or by setting some parameters to physiologically plausible values. However, most of the parameters in the reduced model are not measured directly and require further experimental validation.

Finally, in Chapter 3 we specifically focus on the spatial average bursting, thus ignoring all the details of the spatial structure of the network. We assume that the networks that we study are sufficiently coupled and random that the network effects would be captured by the white noise (Breskin et al., 2006). One good way to overcome this limitation would be to consider a proper mean-field theory for different types of networks (Georges et al., 1996; Gerstner et al., 2014), which can be done in future research.

Network dynamics and plasticity

In Chapter 4 we discover the connectivity structure that allows for E/I and receptive field formation. In this work, we assume a rather simple presynaptic network and try to optimize its connectivity based on surrogate metrics of neuronal assemblies.

This approach allowed us to very efficiently use simulation-based inference. Thus, the assembly strengths were well defined and we only needed to compute the correlation of network activity as summary statistics. We did not optimize the network structure directly, which is a harder task and probably requires a slightly different approach. Yet this should be explored in the future.

In the model we considered, only the projections onto the postsynaptic neuron develop with plasticity rules. The connections in the presynaptic network remain, however, static. It remains to be shown if these presynaptic assemblies could arise from the plasticity rules alone.

Finally, in Chapter 4 we started probing the dynamics that allow balance and tunning to be formed (see Fig.4.5). In the optimized network, the groups of neurons that share connections have strong transient responses, that are seen on top of the asynchronous irregular activity. At the same time, when the assembly structure is perturbed, synchronous events appear. Network dynamics here might be a good starting point to disentangle how learning is linked to inhibitory assembly structure (Mongillo et al., 2018) and oscillatory dynamics (Hellyer et al., 2016; Jutras & Buffalo, 2010).

Beyond non-linear responses of ACC during the conflict

In Chapter 5, we show how statistical models help uncover complex responses in higher brain areas. This work sets the foundation for the repertoire of responses that yet remains to be mechanistically explained. Interstingly, our data-based approach here helped to show that the repertoire of responses is rather large and depends on task and behavior.

One of the important limitations of statistical modeling is that it relies on linear models such as GLM e and dPCA. It has been shown that even early sensory areas implement non-linear computations (Cadena et al., 2019; Klindt et al., 2017). Using a non-linear model (Laboratory et al., 2022) might indeed capture more complex interactions between behavior and neuronal responses and further boost the predictive performance.

Finally, mechanistic modeling of computations in the ACC remains a topic for future research. There are at least two possible further directions. First of all, it appears reasonable to directly test the prediction of a simple cognitive model that assumes that ACC directly reads from action representations (Botvinick et al., 2001). This could indeed explain the scaling responses with conflict. Secondly, a more elaborated RNN-based modeling is likely needed to explain all the variability that we observed. In this case, the main challenge is to identify an appropriate task to train the RNN on that would require a separate monitoring subsystem to emerge.

The art of choosing summary statistics

In this thesis, I relied on choosing the right summary statistics for every problem of the simulation-based inference (SBI). This process typically requires domain knowledge and a good grasp of the problem. Thus, in the case of fitting models of population bursting activity, we relied on the summary statistics used in earlier work that dealt with similar population phenomena (Giugliano et al., 2004; Levenstein et al., 2019; Mattia & Sanchez-Vives, 2012). Perhaps even more satisfying was a choice of summary statistics to infer the network structure in Chapter 4 (Giannakakis et al., 2023). There, it has been theoretically shown that in simplified cases neuronal tuning and balance at postsynaptic neurons depends on the correlation structure (Clopath et al., 2016), which we then used as a summary statistics. However, for more complex problems, the question of whether the features or summary statistics are adequate for model fitting is hard to resolve.

When the summary statistics are hard to define, building a lower dimensional representation of the phenomena of interest might be a reasonable approach (Williamson et al., 2019). In that case, one could try to match the model output to the data in this lower-dimensional space.

Luckily in the past years, there has been a lot of effort in extracting interpretable low-dimensional manifolds of neuronal activity (see Hurwitz et al., 2021 for review).

A closely related approach is extracting features from raw data using neuronal networks. In this case, the output of such a network can be directly linked to the density estimators used to train surrogate models of the posterior (Greenberg et al., 2019 and see also the implementation in Tejero-Cantero et al., 2020). A skeptic may contend that such an approach would hinder the relationship between data and model. However, based on our experience in mapping the invariances identified by the deep network model to analytical descriptions, I am inclined to be optimistic that machine learning will not hinder but rather facilitate the description of neural phenomena.

Finally, a decision on the informative summary can be made based on exploring the posterior distributions or by using the approximated likelihood function (Beck et al., 2022).

6.1 Towards the model-driven discovery of network pathology

In Chapter 2 and 3, we used the model-based approach paired with Bayesian parameter inference helped to uncover mechanisms and generate testable predictions using extremely limited observations. This gives a good example of how a model for *in vitro* system can generate insights into the network structure. This approach could be directly transferred to study the pathophysiology of neuropsychatric disorders with *in vitro* experimental models.

Developing understanding and novel therapies for treating neurological diseases is a major challenge for basic research and medicine. Progress in stem cell research established a whole field of using (iPSC) derived neurons to study pathology. This method allows us to keep the exact patient genotype and opens possibilities to study the pathological processes that uncover on different levels from genes to networks (Choudhary et al., 2022; Hoffmann et al., 2019b; Lotila et al., 2022; Stern et al., 2018; Uysal et al., 2019).

So far, most of the focus has been directed to genetic, molecular, and cellular properties that could be affected (Choudhary et al., 2022; Hoffmann et al., 2019b; Karagiannis et al., 2019; Lotila et al., 2022; Stern et al., 2018; Uysal et al., 2019). However, in some cases, even when the exact etiology of a disease is known (Uysal et al., 2019), it gives little insight into how the pathological process unfolds and where it could be mitigated. Progress, in this case, depends on studying systems' effects on different levels. Network properties and collective dynamics are

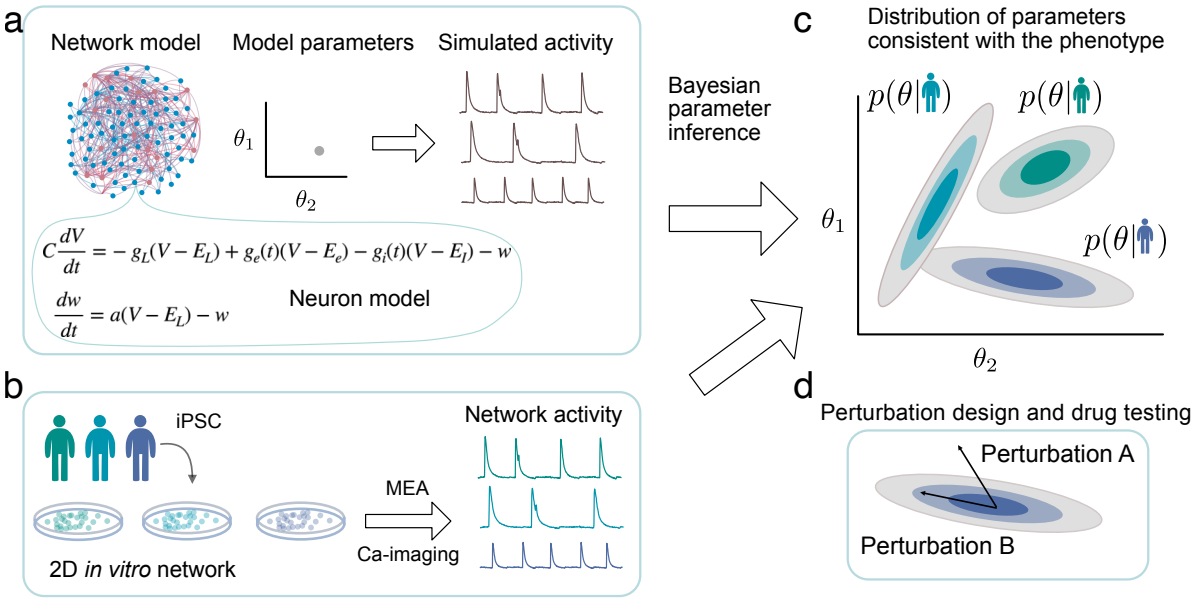


Figure 6.1: Mapping network activity of human iPSC cultures to mechanistic models. **(a)** A network model of spiking adaptive integrate-and-fire neurons produces network bursting for various combinations of parameters. **(b)** *In vitro* networks of human iPSC neurons derived from patients with known phenotypes and their activity. **(c)** Inferred distributions of parameters for different patient phenotypes reveal the differences between subjects. **(d)** Model-guided perturbation design for model testing and drug discovery.

one of these levels and there is a growing body of evidence that understanding network effects is a key to some pathologies (Rosh et al., 2023).

Data-constrained models of *in vitro* systems can be used to capture the collective dynamics of networks of iPSCs neurons and to relate the model parameters with network pathologies (Fig. 6.1). In the context of polygenetic disorders, modeling-based phenotyping could supply researchers with additional insights into what could be affected on the level of single-cell physiology or networks.

Appendix A
Experimental details

A.1 Experimental methods for Chapter 2

Culturing and recording from networks with different ratios of excitatory and inhibitory cells

We control the E/I ratio of neuronal cultures using the procedure illustrated in 2.1. Inhibitory and excitatory cell populations are separated by fluorescence-activated cell sorting (FACS, Fig. 2.2.1) and then mixed them in the desired proportions. We used genetically modified mice, in which inhibitory neurons express a tdTomato fluorescent marker under the promoter of Glutamic acid decarboxylase II (GAD II) (Taniguchi et al., 2011). FACS is then used to sort dissociated hippocampi into two populations, GAD+ and GAD-. After sorting, GAD+ (mostly inhibitory) and GAD- (mostly excitatory) populations are mixed and seeded with prescribed E/I ratios (e.g. 0%, 50% and 100% GAD+ cultures shown in 2.1).

Culturing neurons Both glia and neuron dissections are performed according to standard protocols (Renault et al., 2015). The neurons and glia were cultured on a microfluidics growth chamber. Glia cells are dissected from the cortices of Wistar Hanover/Ola rat pups at postnatal day 0 (P0) or embryonic day 19 (E19). The glial cells were seeded directly into the 4 mm diameter chamber in a density of 2,000 cells$/$mm^2. After 4 days *in vitro* (DIV 4), the medium is replaced by glia changing medium which consists of 1mM L-glutamic acid (Sigma G1626) in addition to the original glia plating medium, to trigger neuronal excitotoxicity that eliminates any remaining neurons. At DIV 7 the medium is replaced by glia plating medium to rinse L-glutamic acid remnants one day before neurons are seeded. At DIV 8 of the glia dissection, labeled neurons are dissected and dissociated following the protocol mentioned above from mice hippocampi at E17 (Taniguchi et al., 2011). The sorted excitatory and inhibitory neurons are added at a density of 4500 cells/mm^2 on top of the glial sheet that covers the surface. After 1.5 hours of incubation, serum-free medium (SFM) is slowly added to avoid detachment of cells (Neurobasal (Gibco 21103-049), 2mM GlutaMAX, 20μg/m gentamicin, 4% B27, and 1% Fetal

Table A.1: Seeded and observed counts of excitatory and inhibitory neurons

Seeded	0%	10%	20%	control	30%	50%	70%	80%	100%
Observed	4 ± 1 %	15 ± 1 %	23 ± 2 %	23 ± 1 %	32 ± 2 %	46 ± 2	65 ± 2 %	73 ± 3 %	92 ± 1%
N	10	6	5	17	6	16	11	10	15

Bovine Serum). We find that the final cellular density, after growing *in vitro*, is independent of inhibitory percentage with a mean ± standard deviation value of $950 \pm 130 \text{cell}/\text{mm}^2$.

Recordings of network activity We recorded network activity in cultures after 14-30 days *in vitro* using calcium imaging. The experiments are done on cultures 14-30 days *in vitro* (DIV) after the GABA switch had already occurred (Soriano et al., 2008). This ensures that GABAergic cells have already acquired their inhibitory behavior. The activity was recorded in the recording medium (5mM KCl, 130 mM NaCl, 2mM $CaCl_2$, 1mM MgCl2 , 10mM HEPES, 45mM Sucrose, and 10mM Glucose). Before the imaging, the cultures were incubated in darkness for 1 hour with 4µl of 1µg/µl Fluo-4 A dissolved in DSMO and 1ml of the recorded medium. Then the medium was changed to 2ml of the fresh recording medium. The experiments were done with a confocal microscope Leica SP5 with an argon laser. Frames of 256×256 pixels were taken every 200ms. We recorded each culture only once.

Seeded and observed percentages of inhibitory neurons

We compared the seeding fraction of inhibitory cells (seeded) to the number of inhibitory cells observed in culture after two to three weeks *in vitro*. For that, we manually counted the number of inhibitory cells labeled with GAD-tdTomato and static images taken with 1024x1024 resolution with the confocal microscope. In each culture, we took both transmitted light and fluorescent z-stack images. We obtained the fluorescent maximal z-stack projection using ImageJ (C. A. Schneider et al., 2012) and overlaid it with the most focused transmitted light image. First, cells were manually marked and counted (ImageJ cell counter plugin) at the transmitted light image alone in an area of around 1mm^2. Then, we checked whether these marked cells were tdTomato positive in the corresponding fluorescent image. The results indicate that the seeded and observed percentages of inhibitory cells strongly correlate. Overall we estimated the deviation from the seeded value is about 5±1% 2.2. For the extreme seeded percentages (0% and 100% of inhibitory cells) we get 3.3% and 92% of observed inhibitory neurons.

Electrophysiology To quantify single neuron properties, we used patch-clamp recordings (done in the Segal lab with cultures prepared as described above). Cells are recorded in current clamp and voltage clamp modes, and responses to current steps or ramps are recorded and analyzed for passive and active properties. Miniature excitatory postsynaptic currents (mEPSCs) are recorded in the presence of TTX (0.5µM) and bicuculline (10μM). The features of spike and PSPs are analyzed using Minianalysis and Origin software by Menahem Segal.

Extended details of the experimental methods and data analysis can be found in (Sukenik* , Vinogradov* et al., 2021).

A.2 Expeimental methods for Chapter 3

HIH dataset

The neurons of rat hippocampi (E18) were dissected and cultured according to the previously published protocols (Y. Liu et al., 2019). Animal protocols for primary cell culture were approved by the local Animal Care and Use Committee. Cortical cultures were seeded on 24-well MEA plates (Multi Channel Systems MCS GmbH, Reutlingen, Germany) at the seeding density of around 5000 neurons/mm^2. Each well contained 12 electrodes (Fig. A.1). The medium was changed every 3 days.

The spontaneous network activity was recorded at 17 and 18 DIV in the recording medium (see Table A.2). After the MEAs were transferred to the amplifier, we let it adjust for 10 minutes. The recordings were done while keeping the temperature at 37°C. The signal was recorded for 20 minutes and sampled at 10 kHz. Spike detection was performed via the noise threshold method provided by the Multi-well Analyzer Software (Multi Channel Systems MCS GmbH, Reutlingen, Germany). For each recording, the default algorithm was set to calculate the standard deviation from 50 different 100 ms long segments of raw data, and the threshold for spike detection was set at $\pm$5 times the standard deviation level. For a subset of wells, we also blocked the inhibitory activity by adding 40μM of bicuculline, these data were not included in this project (Fig.A.1).

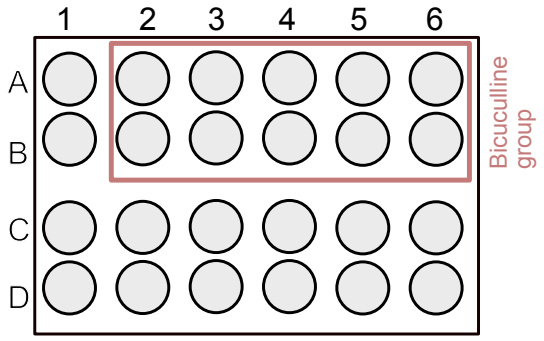


Figure A.1: Layout of the 8-channel multiwell MEA

EM-2	
NaCl	140mM
KCl	4.2mM
$CaCl_2 \cdot 2\,H_2O$	2mM
$MgSO_3 \cdot 7\,H_2O$	1mM
Na_2HPO_4	0.5mM
NaH_2PO_4	0.45mM
HEPES	5mM
Glucose	10mM

Table A.2: External medium-2, the solution was prepared with DDW, and pH was controlled by adding NaOH until pH was within the 7.35-7.45 range

WIS dataset

Primiary cultures preparation

All procedures were approved by the Weizmann Institutes Animal Care and Use Committee. The dissections and cell dissociations were done according to established protocols (Feinerman et al., 2005; Renault et al., 2015). Hippocampal neurons were obtained from Winstar rat embryos at E19.

Brains from embryos were dissected on ice in L-15 medium with 0.6mg/ml D-glucose and 20 μg/ml gentamycin. Hippocampi were dissociated in papain solution (papain 100 units, DNAse 1000 units, L-Cystein 2 mg, NaOH 1M 15 μL, EDTA 50 mM 100 μL, $CaCl_2$ 100 mM 10 μL, dissection solution 10 mL) at 37°C. After 20 minutes of incubation at 37°C with papain, the solution was replaced with 10 mL of plating medium (MEM without glutamine supplemented with 0.6% glucose, 1% GlutaMAX, 5% Horse Serum, 5% Fetal Calf Serum and 0.1% B27) with 25 mg of trypsin inhibitor and 25 mg of Bovine Serum Albumin, that effectively stops tissue

dissection. The resulting tissues were triturated with glass pipettes after which the cells were counted with an automatic cell counter.

The suspension was then seeded with a density of around 5000 neurons/mm^2 on glass coverslips coated with poly-L-lysine placed in 24-well plates. The plates were incubated for half an hour in a humidified, $37C$ and 5% CO_2 incubator to allow the attachment of neurons. Finally, 2 mL of serum-free medium (Neurobasal, B27 4%, GlutaMAX 1%, and FCS 1%) were added and the cultures were placed back into the incubator. At 4 DIV the glial proliferation was stopped by adding in the medium 20 μg/mL 5-fluoro-2-deoxyuridine and 50 μg/mL uridine (Sigma, Israel). Cultures were fed every day, replacing 0.5 mL of the old medium with a new feeding medium containing 90% MEM and 10% of inactivated Horse serum.

Ca-imaging

Recordings were done between 14 and 28 DIV. The recordings were done in an external medium (EM) A.3. On the recording day, a coverslip with culture was transferred to a petri dish with 1mM of EM A.3 with 4μL Fluo-4. The culture was then incubated in darkness for 1 hour, after which the medium was changed to 2ml of EM1 (Table A.3). Before imaging the Perti dish was mounted in Zeiss Axiovert 135TV. The imaging time resolution was 30Hz. To analyze population activity, we averaged the fluorescence from the whole field of view (which typically contained about 30-200 neurons).

EM-1	
NaCl	130mM
KCl	4mM
CaCl2	2mM
MgCl2	1mM
HEPES	10mM
Glucose	10mM
Surcose	45mM

Table A.3: External Medium-1 (EM1), the solution was prepared with DDW, and pH was controlled by adding NaOH until pH was within the 7.35 - 7.45 range

KCl changes

In addition to standard EM-1 (KCl 4mM), we prepared a set of EM with modified concentrations of KCl: 1,2.5,5.5,7,10mM. After the initial 10-minute recording with the baseline concentration, we changed the medium to a medium with a modified concentration. Then we let the networks adjust to 10 minutes and record another 10 minutes of spontaneous activity in the new concentration of KCl. After that, we change the medium again to the baseline condition. We typically start with a baseline of 4mM, but for a small number of validation experiments, we started with higher or lower concentrations. For the final averages, all these data were pooled together.

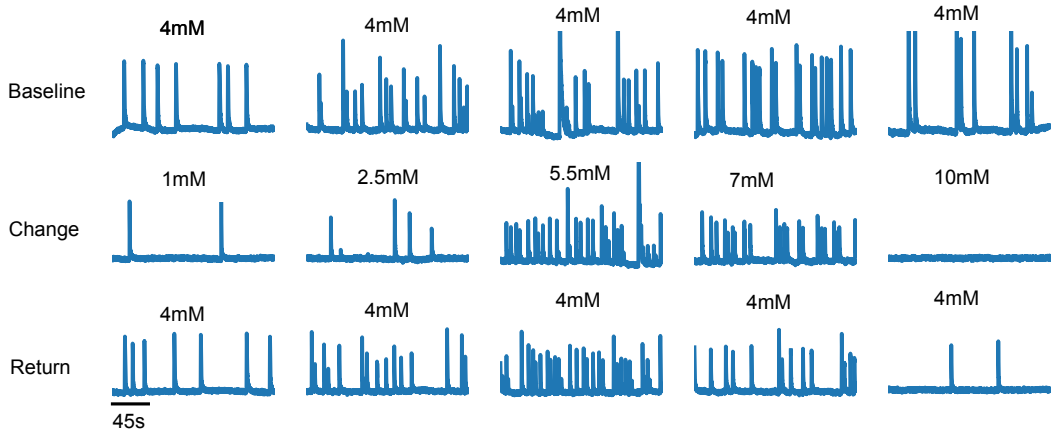


Figure A.2: Examples of average Ca-traces in experiments with different concentrations of KCl and return condition.

Appendix B
Burst detection methods

Population bursts in spiking data

In Chapter 2 the network burst events in the model simulations were identified using a modified version of the max interval algorithm (Cotterill et al., 2016; Kirillov, n.d.). We applied the algorithm to a vector of sorted spike times from the whole network. The algorithm detected burst start and end using a set of parameters: interspike interval (ISI) threshold (4.5ms), the minimum number of spikes in the burst (50spikes), minimal burst duration (40ms), and minimal inter-burst interval (40ms). Burst amplitudes were computed as the highest peak in the 20ms population spike count within a burst at the detected burst onset.

To detect bursts for MEA data in Chapter 3, we first excluded the recordings that did not have a clear bimodal distribution of the events. To this end, we check if the median of the spike counts (in 200ms bins) divided by the number of channels is below 3 and check if the bimodality coefficient (R. Pfister et al., 2013) of the spike count is above 0.5.

Next, we identify the bursts in single recording channels, combine the results of detection, and identify the population bursts. The bursts in single channels were identified by thresholding the interspike intervals using the algorithm described above. We used the following parameters: The minimum detection threshold was 100ms, the minimum of spikes inside the burst was 6, the minimum burst duration was 100ms, and the minimum inter-burst interval was 500ms. In parallel, we detected the bursts based on the overall vector of ISI (from all channels combined), there the minimum detection threshold was 10ms, the minimum of spikes inside the burst was 30, the minimum burst duration was 100ms, the minimum inter-burst interval was 1000ms. We then look for an overlap between the detected bursts. We obtain binary vectors for each channel and for the population, with 1 indicating burst and 0 the absence of burst with 1ms precision (the vector length is the duration of the recording in ms). We calculate a weighted sum over these vectors (1 for a single channel, 10 for the population detection) and get the population bursting vector. In this vector we threshold all activity that has an onset value above 8 and peak value above 12. The detection was then validated by visual inspection.

In contrast to the simple thresholding in Chapter 2, the method allows the detection of bursts

that have a complex structure and different intensities of spiking within a single population event.

Detection of bursts in the reduced model

To detect the bursts in the reduced model we used the population ISI-thresholding methods defined in the previous section. The bursts are detected using dynamics of the rate variable x. We take only the positive part of the simulated signal and normalize it between 0 to 10 and round the values to the closest integers. Then create a vector of times of x above zero. After that, we pass the vector of time to the modified version of the max interval algorithm B with the following parameters: threshold – 10ms, the minimum number of spikes in the burst – 5, minimal burst duration – 20ms, and minimal inter-burst interval – 20ms. The values here are arbitrary and were found by evaluating the detection for multiple parameter combinations.

Detection of bursts in Ca-signal

In **Chapter 2**, calcium recordings were detrended by fitting lines into the manually selected windows of upwards and downward drifts. The traces were then divided by original recording with these lines (Sukenik* , Vinogradov* et al., 2021). We calculated the derivates of detrended traces and found the burst onsets with a threshold. The threshold is found by analyzing the distribution of the derivative of the network activity (binned with 1 std of the derivative). This description typically has 2 peaks - one is noise and the other one is burst onsets. We set the threshold to b 70% of the distance between two peaks. Next, we calculate the mean and standard deviation of a baseline for each burst with five data points prior to the burst onset. Then we check whether activity in the burst time window exceeds three times the baseline noise level. Finally, burst detection is followed by manual inspection (done by Nirit Sukenik), to ensure that all bursts are included.

To find the bursts in Ca-traces used in **Chapter 3** we first detrended the traces by fitting a polynomial of 4th degree to the raw data. The baseline drift is then subtracted from the raw data. The burst onset was found based on the derivative of the detrended data. We then check if the amplitude of Ca-activity reaches the 3 std of the burst baseline (300ms before the burst onset).

Appendix C
Network simulations paramters

Parameters	Values	Description
$V_\theta - V_{\text{rest}}$	20.0mV	Difference between the firing threshold and resting potential.
$V_\theta - V_{\text{reset}}$	10.0mV	Difference between the firing threshold and reset potential
C_m	250pF	Membrane capacitance
τ_m	20 ms	Membrane timescale
b	12.5pA	Spike-triggered adaptation increment
τ_w	5000 ms	Adaptation time
τ_{rf}	2 ms	Neuron's refractory period
v_{ext}	850 Hz	Rate of external Poisson input (free parameter)
J_{ext}	1.0mV	Strength of external synapses
N	10000	Number of neurons
ε	0.2	Fraction of inhibitory neurons
J	1.4mV	Excitatory synaptic strength
g	4	Relative inhibitory strength $\left(J^I = 5.6mV\right)$
D	3.5 ms	Synaptic delay
K^E	80	Number of excitatory in-degrees
K^I	20	Number of inhibitory in-degrees

Table S4. Model parameters used for comparison with the mean-field analysis. Parameters used for the Fig 2.12.

Parameters	Values	Physiological Range	Description
$V_\theta - V_{\text{rest}}$	20.0mV	$6.8 - 24.4*$ (Ferguson et al., 2015; Fricker et al., 1999)	Difference between the firing threshold and resting potential
$V_\theta - V_{\text{reset}}$	10.0mV	$8.8 - 12^{**}$ (Ferguson et al., 2015; Fricker et al., 1999)	Difference between the firing threshold and reset potential
C_m	250pF^{***}	$115 - 300\text{pF}$ (Ferguson et al., 2015)	Membrane capacitance
τ_m	20 ms	$13.1 - 38.5$ (Brette & Gerstner, 2005)	Membrane timescale
b	12.5pA	$5 - 80\text{pA}$ (Brette & Gerstner, 2005; Ferguson et al., 2015)	Spike-triggered adaptation increment
τ_w	8000 ms	$833 - 12500$ ms (Ferguson et al., 2015)	Adaptation time
τ_{rf}	2 ms	around $1.7 - 2.5$ ms $* **$	Neuron's refractory period
v_{ext}	Free parameter $(490-$ 750 Hz)	n.a.	Rate of external Poisson input (free parameter)
J_{ext}	1mV	n.a.	Strength of external synapses
ε	$0.05; 0.1; 0.2; 0.3; 0.5$ $0.7; 0.8; 0.92$	$0.2 - 0.3$	Fraction of inhibitory neurons
J	2.0mV	$\sim 2.2\text{mV}$ (Barral & D Reyes, 2016)	Excitatory synaptic strength
g	4	3.28^{****} (G. Liu, 2004)	Relative inhibitory strength
D	3.5 ms	2-10ms (AMPA) (Destexhe et al., 1995)	Synaptic delay
N	1000	~ 950 cells per area of measurements	Number of cells
$K^E(\varepsilon)$	$\lfloor 0.078N^E + 17.6 \rfloor$ $K^E(0.2) = 80$	$60 - 120$ (Soriano et al., 2008)	Number of excitatory degrees (fit from the patch clamp experiments); The number of incoming connections depend on the density
K^I	Free parameter	$18 - 36$ (Soriano et al., 2008)	Number of inhibitory connections per neuron

Table S3 Parameters of the network model used to fit the data and their physiological range
* In our measurements the resting potential is- 55mV
** Difference between the firing threshold and AHP
*** The absolute refractory period of a single AP is roughly equal to the duration of the AP
**** at 18 DIV

Appendix D
Population bursting in a model without adaptation

A classical random network of leaky-intergate-and-fire neurons can produce a wide range of dynamical states (N. Brunel & Hakim, 2008). Recently, it has been shown that it also can produce slow-fast dynamics that can approximate the up and down state in the cortical networks in vivo (Tartaglia & Brunel, 2017). Such dynamics indeed can also replicate some of the features of the network bursting in vitro.

Bursting occurs as one of the spontaneous dynamical states of a sparse random network of excitatory and inhibitory leaky integrate-and-fire neurons with delta synapses. Bursting typically emerges in networks with strong synapses and slow external Poisson input that approximates spontaneous activations of neurons. The network spontaneously fluctuates between an almost quiescent asynchronous state and fast synchronous firing. This transition closely resembles the network bursting in neuronal cultures. The bursting in this case appears in the bistable state: with two stable fixed points and an unstable fixed point in between. Increasing the synaptic coupling extends this region as shown in Fig. D.2.

The distribution of inter-burst intervals in the model without adaptation strongly deviates from the IBI distribution observed *in vitro*. Thus, the LIF network without spike-frequency adaptation can only produce an exponential distribution of the IBIs, thus following a memoryless random process with CV=1 (Fig. D.3). As shown in Chapter 2, the model is sensitive to blocking of inhibitory synapses and quickly transitions to a high firing rate state (Fig. 2.13), which corresponds to decreasing g values in (Fig D.2). Overall, we conclude that the model without adaptation is not sufficient to reproduce *in vitro* bursting.

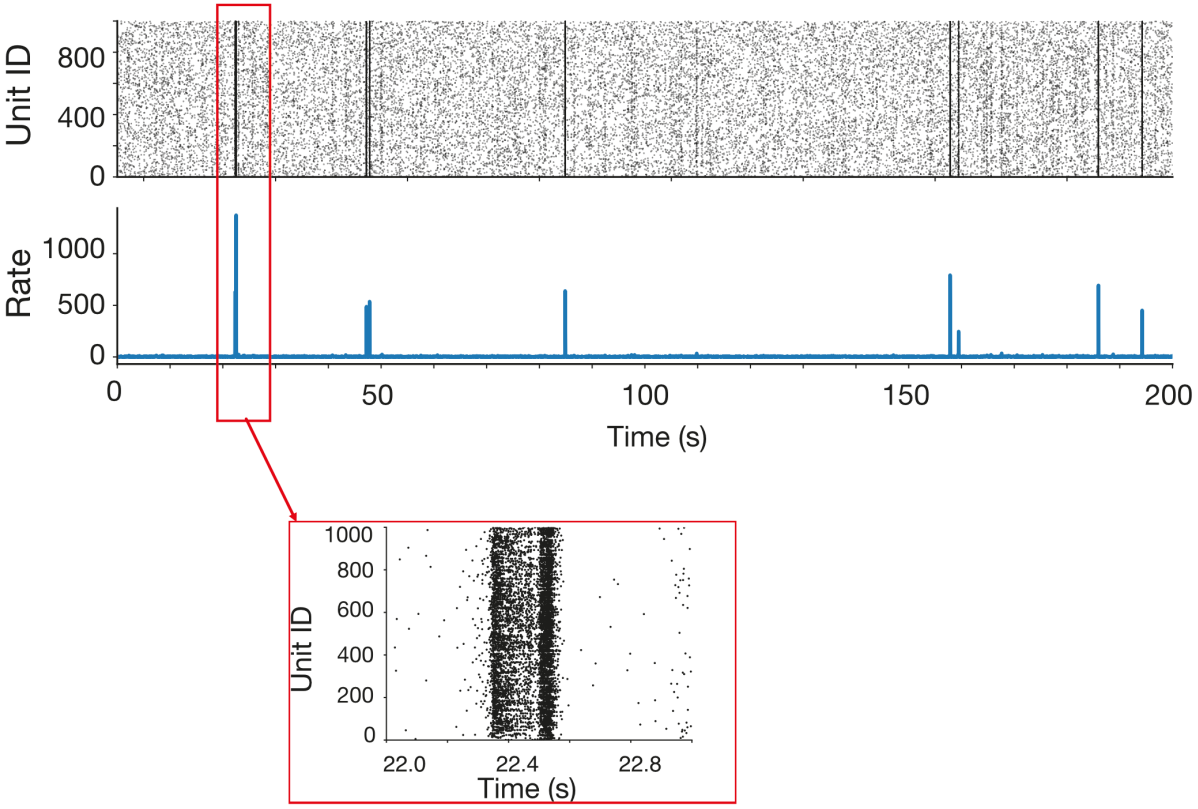


Figure D.1: Population activity in a sparse random network of excitatory and inhibitory neurons without spike-frequency adaptation. The network parameters can be fine-tuned to produce a variety of IBI, while CV of IBI stays close to one.

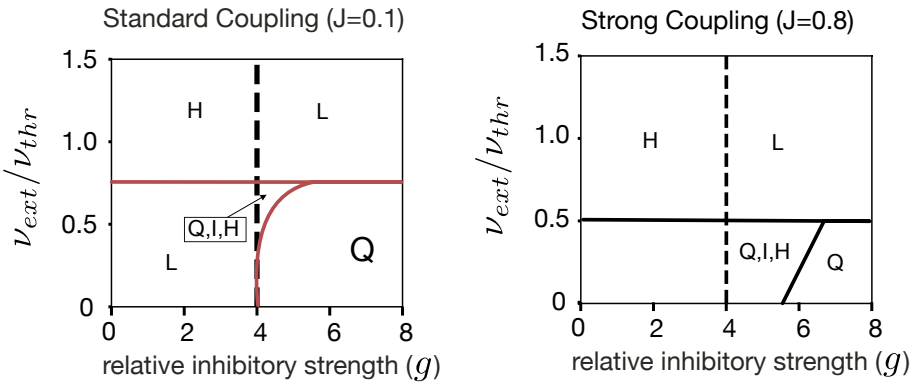


Figure D.2: Saddle node bifurcations of the sparse random network of LIF neurons with different coupling strengths. Increasing the coupling expands the bistable region.

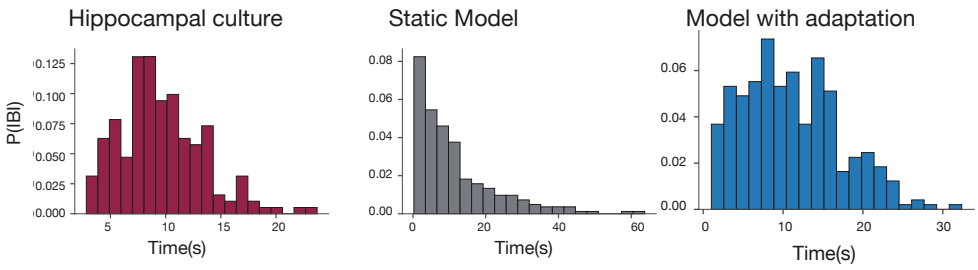


Figure D.3: The distribution of IBIs cultures of dissociated hippocampal neurons compared to the IBI distributions in the sparse random network of LIF neurons with and without adaptation.

Appendix E
Statistical procedures

In this section, I describe the methodological details of the statistical analysis of experimental and modeling data used in Chapter 2.

Burst features

We use a non-parametric one-way ANOVA with permutations (n=10,000) for statistical inference on the burst features and set the alpha level at 0.05. The permutation tests are used because the distributions of the burst features are skewed. For pairwise comparisons, we performed the same permutation test, controlled the false discovery rate (FDR) at 0.05, and reported adjusted p-values. We find a significant main effect of the mean IBIs across networks with different fractions of inhibitory neurons ($p<0.0001$). Further, posthoc comparisons show no significant differences between the mean IBIs in control cultures and in cultures with 10-80% inhibitory neurons ($p>0.05$). Cultures with 0% and 100% inhibitory neurons have significantly longer IBIs compared to the control ($p<0.0001$ and $p=0.0348$). A positive trend of mean IBIs is observed in between 10 and 80% inhibitory cultures (Wilcoxon type test for trend, $p=0.039$, z-score=2.062) when 0% and 100% cultures are excluded. The Coefficient of Variation (CV) of the IBIs grows linearly with an inhibitory percentage from 0.29±0.03 to 0.75±0.04 (Wilcoxon type test for trend, p=5.4x10-13, z-score=7.214).

One-way permutation ANOVA on burst amplitudes shows a significant main effect ($p<0.0001$); Burst amplitudes are significantly higher in 0% networks compared to the control (Post-hoc permutation ANOVA between 0% and control, FDR controlled at 0.05 $p<0.001$). Networks with 100% inhibitory neurons have significantly smaller burst amplitudes compared to control ($p=0.044$). We did not find significant differences between the mean amplitudes of bursts in 10-80% inhibitory cultures compared to control ($p>0.05$). Additionally, we found a significant negative trend of the mean burst amplitudes as the fraction of inhibitory cells increases from 10 to 100% (excluding 0%; Wilcoxon type test for trend, $p=0.039$, z-score = 2.062). The coefficient of variation of the burst amplitudes is significantly smaller in 0%, 50%, 70%, 80%, and 100% cultures compared to the control (Main effect in permutation ANOVA $p<0.0001$, post hoc tests

p<0.0001, p=0.0173, p=0.0098, p<0.0001, p=0.0002). We did not find significant differences between the CV of burst amplitudes of 10%-30% networks and control cultures (p>0.05).

Permutation ANOVA shows the significant main effect of burst durations (p<0.0001). The post-hoc comparisons show that cultures with 0% and 80% inhibitory neurons have significantly longer burst duration compared to the control (p<0.0001, p=0.02). We could not reject the null hypothesis (p>0.05) in all other pairwise comparisons. The CV of durations is lower in the 0% inhibitory cultures compared to the control (p<0.0001) and not significantly different otherwise (p>0.05).

Analysis of the patch clamp data

For the statistical analysis of patch clamp data, we use parametric 2-way ANOVA with percentage (3 levels: 20%, 50%, and 80%) and cell type (excitatory and inhibitory) as two factors. The amount of inhibitory cells included in the analysis is N=12, N=7, and N=7 for 20%, 50%, and 80% inhibitory cultures respectively and the amount of excitatory cells is N=10, N=12, N=22 for 20%, 50% and 80% inhibitory cultures respectively. While we could not reject the null hypothesis that the mean size of PSC in networks with different inhibitory percentages is the same (F(2,64)=0.0156, p=0.984), the size of PSC is significantly different in inhibitory and excitatory neurons (F(1,64)=6.95, p=0.01). The decay is also significantly faster in inhibitory cells (F(1,64)=5.68, p=0.02), and again we did not find significant differences in networks with different inhibitory percentages (F(2,64)=0.24 p = 0.786). The frequencies of PSC decrease significantly with the increasing fraction of inhibitory neurons (F(2,64)=6.348, p=0.003) and we have also found significant differences in frequencies of PSC between inhibitory and excitatory neurons (F(1,64)=60.21, p=8.56x10-11).

Analysis of mEPSC Mean rates, amplitudes, and decay time of mEPSC are measured in 20% and 80% inhibitory cultures. The measurements are done in fluorescent inhibitory and non-fluorescent excitatory cells with the following number of measurements: in 20% inhibitory cultures – inhibitory cells N=10 and excitatory cells N=10. In 80% inhibitory cultures- inhibitory cells N=9 and excitatory cells N=7. The mean rates are significantly different in the groups (ANOVA F=12.9, p<0.0001) and individual comparisons are also significant (p<0.01 in all). Comparisons among mean amplitudes of the different groups, ANOVA, F=8.15, p<0.0001 with significant differences between inhibitory and excitatory cells in both 20% and 80% inhibitory cultures. The decay time of the same groups, ANOVA F=4.36, p<0.01, significant differences only between inhibitory and excitatory cells in 20% inhibitory cultures.

We also compared mEPSC properties between neurons that express a YFP marker (i.e. CRE-OFF, non-GABAergic cells) and neurons that express a tdTomato (i.e. CRE-GAD, GABAergic cells). The size and rate of mEPSC, N=10 neurons in each group, is significantly higher in the GABAergic, tdTomato cells with t=5.76, p<0.001, and t=4.74, p<0.001 respectively. The decay time is not significantly different between the tdTomato and YFP stained groups.

Responses to bicuculline

Analysis of IBIs' response to bicuculline. We compare the mean IBIs of networks with different inhibitory percentages (number of independent samples 0% N=5, 10% N=6, 20% N=3, 30% N=5, 50% N=3, 70% N=6, 80% N=4, 100% N=7, control N=3) repeatedly recorded for at least 20 minutes under concentrations of bicuculline that range over 7 levels from 0.5 to 40μM. The IBIs are normalized to the values at 0μM bicuculline. Because of the complex unbalanced within-between-group experimental design that includes missing observation (in 100% inhibition cultures under high concentration of bicuculline), we use a linear mixed effect model to study the effects of inhibitory percentage and of concentration on the normalized IBIs. We apply a natural logarithm to correct the right-skewness of the normalized IBIs. We fit the linear mixed effect model to the log-transformed normalized IBIs using the bicuculline concentration and inhibitory percentage as fixed effects and individual intercepts for each culture (random effect) assuming Gaussian noise. When all inhibitory percentages are included the residuals of the fitted model show deviation from a normal distribution. We further exclude from the analysis 100% inhibitory cultures, which show the highest variability and include most of the missing observations (only 2 of the 7 100% inhibitory cultures yielded observable IBIs for 10 and 40[μM] bicuculline). This allows for a good model fit with the normal distribution of the residuals (fixed and random effect R2 =0.786). We report the statistics for a model that does not include 100% inhibitory cultures. We find a significant main effect of percentage ($F(7, 87.837) = 3.65$, $p = 0.000167$, degrees of freedom approximated with Kenward-Roger's method), concentration ($F(1,161.01)=343.074$, $p<2.2\text{x}10\text{-}16$), and an interaction between the two ($F(7,161.011)=9.166$, $p=1.623\text{x}10\text{-}9$) on the normalized IBIs. We further compare the interaction of percentage and concentration between every inhibitory percentage and control. We have found significant interaction when comparing 0% and control (beta=-0.63±0.19 s.e.m, $t(27.5) = -3.313$, $p= 0.0181$). We could not reject the null hypothesis in the case of all other pairwise comparisons of the interaction effect. Fitting of the mixed effect model was done in R 3.6 using the lme4 toolbox.

Appendix F
Analysis of the reduced model

F.1 Stability analysis of the reduced model

We study the phase portrait of the model, using a standard stability analysis for the deterministic case (Strogatz, 2018). We start by computing the fixed points and nullclines of the model. Here we ignore the coupling c and set it to 1. The nullclines are given by

$$0 = bx - w \tag{F.1}$$

$$0 = w - \frac{1}{a}\left(a\theta + \ln\left(\frac{w}{J-x}\right)\right) \tag{F.2}$$

The fixed points x^*, w^* are given by the intersection of two lines and can only be found numerically. The solutions are typically 1 or 3 fixed points (Fig.3.1c). The stability of the fixed points can be found by linearizing the system around its fixed points:

$$\text{Jac} = \begin{bmatrix} \frac{\partial \dot{x}}{\partial x} & \frac{\partial \dot{x}}{\partial w} \\ \frac{\partial \dot{w}}{\partial x} & \frac{\partial \dot{w}}{\partial w} \end{bmatrix}_{[x^*,w^*]} = \begin{bmatrix} -\frac{1}{\tau} + \frac{Jae^{-a(\theta-w+x)}}{\left(1+e^{-a(\theta-w+x)}\right)^2\tau} & -\frac{Jae^{-a(\theta-w+x)}}{\left(1+e^{-a(\theta-w+x)}\right)^2\tau} \\ b/\tau_w & -1/\tau_w \end{bmatrix}_{[x^*,w^*]} \tag{F.3}$$

Note that we can rewrite the derivative of the sigmoid as $\phi'(z) = \frac{e^{-z}}{(1+e^{-z})^2}$ and rewrite the Jacobian as

$$\text{Jac} = \begin{bmatrix} -\frac{1}{\tau} + \frac{Ja}{\tau}\phi'(-a(\theta-w+x)) & -\frac{Ja}{\tau}\phi'(-a(\theta-w+x))) \\ b/\tau_w & -1/\tau_w \end{bmatrix}_{[x^*,w^*]} \tag{F.4}$$

The signs of the eigenvalues of the Jacobian will then determine the stability of fixed points.

$$\lambda_\pm = \frac{1}{2}\left(\text{tr(Jac)} + \sqrt{\text{tr(Jac)}^2 - 4\text{det(Jac)}}\right) \tag{F.5}$$

Where tr is a matrix trace and det stands for the matrix determinant. The determinant of the matrix is given by

$$\text{det(Jac)} = \frac{1}{\tau \cdot \tau_w} Ja\left(\phi'(-a(\theta - w + x)) - 1\right) \tag{F.6}$$

And the trace is given by

$$\text{tr (Jac)} = -\frac{1}{\tau} - \frac{1}{\tau_w} Ja\phi'(-a(\theta - w + x)) \tag{F.7}$$

By analyzing the fixed point and their stability we can find that there are three types of dynamical states:

1. Stable single fixed point (either low or high)
2. Two stable, one unstable fixed points (bistability)
3. Oscillations

We can find two bifurcations of the system, where it transitions between the states. The Saddle-node bifurcation-node bifurcation indicates a transition from one stable fixed point to two stable fixed points and one unstable in-between. Its location can be identified by the points where det(Jac) $= 0$. Hopf bifurcation-bifurcation is a transition from 1 stable fixed point to a limit cycle and is indicated by tr (Jac) $= 0$. Note that as long as $\tau \ll \tau_w$ the location of the bifurcations is not strongly affected by the timescales of the system. In practice, we use numerical bifurcation continuation using XPPauto (Ermentrout, 2002) to obtain the bifurcation diagrams. Fig. 3.1d shows an example of the bifurcation diagram for θ and b.

F.2 Analytical description of the invariance in a piece-wise linear equivalent model

Next, we asked if the relationships that allow for the invariance of bursting dynamics can be treated analytically. To simplify the analysis, we will substitute the sigmoid non-linearity with an equivalent piece-wise nonlinear function. The maximum slope of the sigmoid we use is given $0.25aJ$. We scale it by 0.5 and get:

$$f_{\text{pwl}}(x) = \begin{cases} 0 & \text{for } x < \theta - \frac{4}{a} \\ 0.125aJ(x - \theta) + \frac{J}{2} & \text{for } \theta - \frac{4}{a} < x < \theta + \frac{4}{a} \\ J & \text{for } x \geq \theta + \frac{4}{a} \end{cases} \tag{F.8}$$

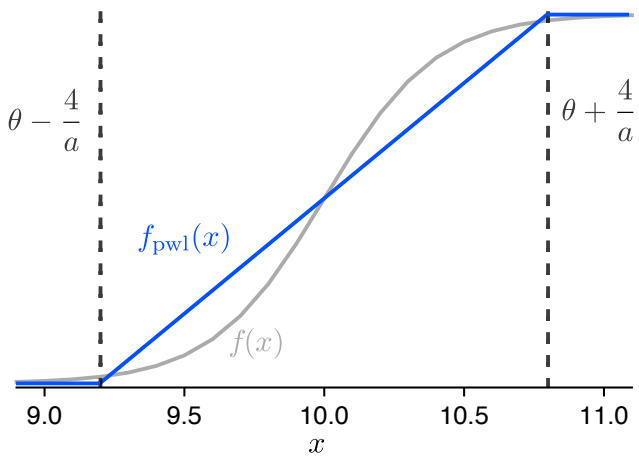


Figure F.1: Piece-wice linear approximation ($f_{\text{pwl}}(x)$) of the original sigmoid ($f(x)$).

Assuming large values of a, that control the steepness of the sigmoid function, this approximation is very accurate. Now we can explicitly find the fixed points The nullclines are now given by

$$0 = \frac{dx}{dt} = x - \theta - \frac{x - J/2}{0.125aJ} \tag{F.9}$$

$$0 = \frac{dw}{dt} = \frac{w}{b} \tag{F.10}$$

We can also explicitly solve for the fixed points:

$$x^* = \frac{J(4 - a\theta)}{(aJ(b-1) + 8)} \tag{F.11}$$

$$w^* = bx^* \tag{F.12}$$

We can also find the maximum and minimum of the $\frac{dx}{dt} = 0$ null-cline. These values give us the boundaries of the bistability in the case of the time scale separation. The boundaries are given by the intersection of $y = x$ line and the $y = f_{\text{pwl}}(x)$, which is gives by x^* and now we need to consider x in the boundaries

$$\theta + 4/a > (J(-4 + a\theta))/(-8 + aJ) \tag{F.13}$$

$$\theta - 4/a < (J(-4 + a\theta))/(-8 + aJ) \tag{F.14}$$

And therefore

$$w^- := (4 - a\theta)/a \tag{F.15}$$

$$w^+ := \frac{-4 + aJ}{a} - \theta \tag{F.16}$$

These two values are the boundaries between which the system transitions between the states.

Invariance in the deterministic case

The most simple case of transition invariance can be found in the slow deterministic oscillations in the model. In this case, the period is always the same, therefore CV(IBI)=0. For this case, we can assume that the system simply jumps between two branches and slowly relaxes to w^- and w^+. For the piece-wise linear approximation, the $x(t)$ then always stays at its fixed point given by 0 and J. Note that it is also a very good approximation of the original model, when a and J are sufficiently large. We can then only analyze the slow variable, assuming that the x is fixed. The general solution of the initial value problem (IVP) for $w(t)$ is

$$w(t) = bx - (w_0 - bx)e^{-t/\tau_w} \tag{F.17}$$

where w_0 is the initial condition. Now we can solve for both b and τ_w as a function of θ, given fixed IBI and burst durations.

The time it takes to relax to w^- after the jump from the upper branch is given by IBI, therefore $w(t = T^{\text{IBI}}) = w^-$. And we can now explicitly solve for τ_w

$$\tau_w = -\frac{T^{\text{IBI}}}{\log(\frac{w^+ - bx}{w^- - bx})} \tag{F.18}$$

Note that here $bx = 0$ and given a fixed IBI the τ_w therefore depend only on w^- and w^+, which are functions of θ as defined above. Correspondingly in the deterministic case, the IBIs only depend on τ_w. Furthermore, we can plug in w^+ and w^- and rewrite the equation above as

$$\tau_w = -\frac{T^{\text{IBI}}}{\log(1 + \frac{aJ-8}{5\theta+4-aJ})} \tag{F.19}$$

Now we can perturb $log(1 + x)$ up to the first order and get the equation that defines the slope of the dependency between θ and τ_w, assuming that the IBIs are given

$$-\frac{5T^{\text{IBI}}\theta}{aJ - 8} \tag{F.20}$$

If θ is sufficiently large, the full expression Eq.F.18 for the τ_w will approach this line. Note that there is also a constant offset that depends on the higher-order terms. To fix the duration of each burst, we need to set the value of adaptation strength b. To this end, we can solve the

equation F.17 for b, assuming that $w(t = T^{\text{dur}}) = w^+$, which is given by

$$b = \frac{w^+ + w^- e^{T^{\text{dur}}/\tau_w}}{x(e^{-T^{\text{dur}}/\tau_w} - 1)} \tag{F.21}$$

We can expand this up to the second order as $e^{\frac{x}{k}} = 1 + \frac{x}{k} + \mathcal{O}(2)$ and insert the w^+ and w^- and get

$$b = \frac{\frac{-4+aJ}{a} - \theta + (1 + \frac{-T^{\text{dur}}}{\tau})(4 - a\theta)/a}{-x\frac{-T^{\text{dur}}}{\tau}} \tag{F.22}$$

from which we get a dependency between the τ_w and θ

$$\theta = x\frac{T^{\text{dur}}}{2\tau} + \frac{J}{2} \tag{F.23}$$

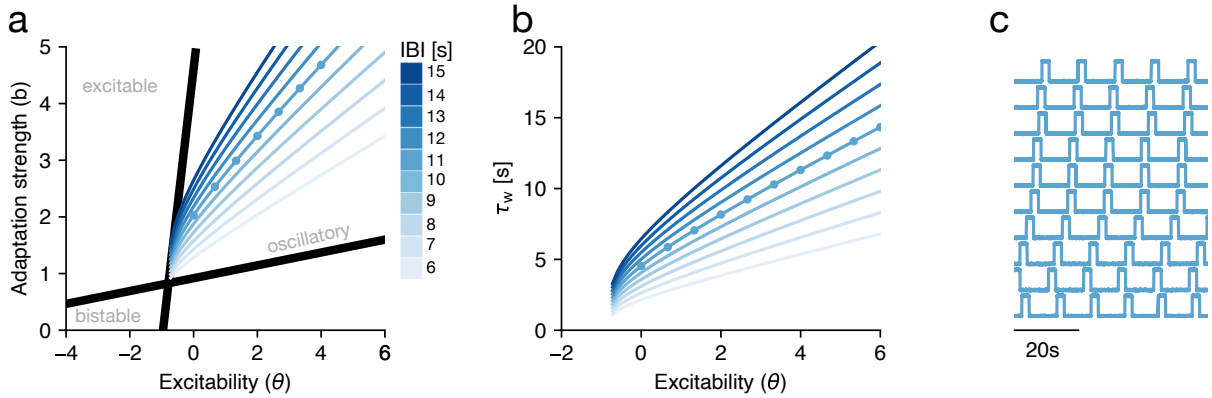


Figure F.2: The invariance of bursting activity in an oscillatory regime without noise. **a** Solution of the adaptation strength of a function of θ for different fixed values of IBI and 2.5 burst durations **b** Solutions for the τ_w as a function of θ. **c** Examples of traces with the IBI=10s and low noise ($\sigma = 0.1$)

F.3 Stochastic methods in the analysis of the IBI and burst duration distributions

In Chapter 2 I introduced the idea of modeling network bursting as a noise-driven transition between two stable fixed points. We assume that the slow variable, e.g. spike firing adaptation, is much slower than the membrane timescale of a single neuron ($\lim_{x\to\infty} \tau_{\text{fast}}/\tau_{\text{slow}} = 0$). This assumption, however, does not always hold and the deviations from this theory can be substantial (Fig. 2.13). We develop a phenomenological model of population activity with fast and slow

variables. Using a standard bifurcation analysis and simulation-based inference of the model, we show that it can exhibit bursting-like activity for a wide range of parameters. For some of the cases in which timescale separation does not fully hold to get a more complete picture, we use the bifurcation analysis and simulation-based inference techniques (see Chapter 3).

This model can also be analyzed using the timescale separation approach and we can then directly obtain a potential of the fast component of the system.

Here we follow an approach developed by Lim and Rinzel, 2010 and use stochastic methods together with survival analysis to obtain the distribution of the IBI and burst durations. Despite the simplicity of the system that we are studying, the application of stochastic methods is quite limited since the full two-dimensional model does not have a potential function. Therefore, one approach is to consider a timescale separation limit, where $\lim_{x\to\infty} \tau_{\text{fast}}/\tau_{\text{slow}} = 0$.

The timescale of adaptation that we inferred with the simulation-based methods (see Chapter 3), is typically a few orders of magnitude larger than the timescale of the rate variable, so this assumption is quite justified.

We can split the analysis into three parts: (1) analysis of the mean escape times with timescale separation, and (2) survival analysis to obtain the distribution of the adaptation values $p(w)$ for which the transitions occur, and (3) calculation of the IBI and duration using the transitions distributions. Each part has its own set of assumptions that will be discussed below. Fig. F.4 shows the full diagram of the analytical approach.

Mean escape time analysis

We assume that w changes very slowly and can be treated as constant. This way we can write an energy function of the model as

$$U(x, w^*) = -\frac{1}{a} J \ln\left(\mathrm{e}^{a(\theta+w-x)} + 1\right) + \frac{x^2}{2} - Jx \tag{F.24}$$

Now we can write the equivalent Fokker-Plank equation:

$$\partial_t p(x,t) = \partial_x[\frac{dU}{dt} p(x,t)] + D\partial_x^2 p(x,t) \tag{F.25}$$

The mean transition time is given by (Gardiner, 2009):

$$M(x_0 \to x_1) = \frac{1}{D} \int_{x_0}^{x_1} e^{U(y)/D} \Big[\int_{-\infty}^{y} e^{-U(z)/D} dz \Big] dy \tag{F.26}$$

The nice part about M is that it gives us the escape rate precisely, a not-so-nice part is that the integrals are impossible to evaluate analytically. To simplify the computations, I assume that the burst emission as a function of w is given by

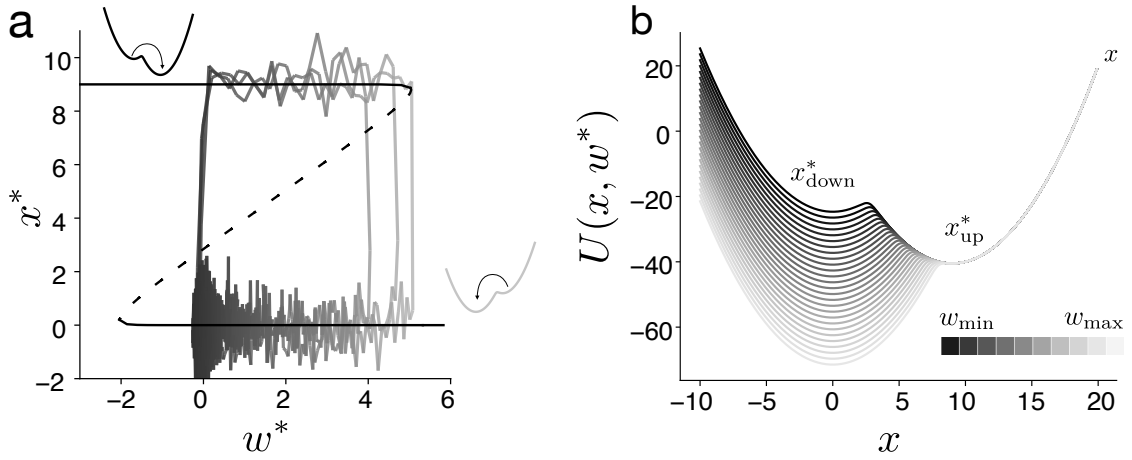


Figure F.3: Network bursts as noise-driven transitions in changing double well-potential under timescale separation ($\tau_m \ll \tau_w$). **a**. Steady-state solutions for the fast variable assuming fixed values of adaptation w^* and trajectories of the transitions in the real simulation of the model (overlaid traces, shade of gray show different values of w). **b**. The double well potential of the system assumes timescale separation (note that the full system does not have a potential). For small values of adaptation, the potential is asymmetric, and the high fixed point x^*_{up} has lower energy. Therefore, the system driven by noise quickly transitions to the high fixed point. When the adaptation is strong (palest grey), the system is more likely to transition to the low fixed point x^*_{down}.

$$\hat{\nu}(w) = c_1 e^{-c_2 w} \tag{F.27}$$

In practice, this approximation works reasonably well and will allow us to proceed further using analytical techniques. The drawback is that we lose an explicit relationship with the model parameters and need to numerically estimate the coefficients for a given set of parameters for the non-linearly (θ, a, J).

Survival analysis for transition probabilities

Now we can treat the value of adaptation as a random variable W at which the transitions between states occur. We can use the survival analysis to compute the probability density at which transitions happen. Survival function (Kleinbaum & Klein, 1996) is defined as $S(w) = \Pr(W > t) = 1 - F(w)$, where $F(w)$ is the cumulative distribution. We then define the hazard function as

$$h(w) = \lim_{h \to 0} \frac{\Pr(w \leq w < w + h \mid w \geq w)}{h} = \frac{f(w)}{S(w-)} \tag{F.28}$$

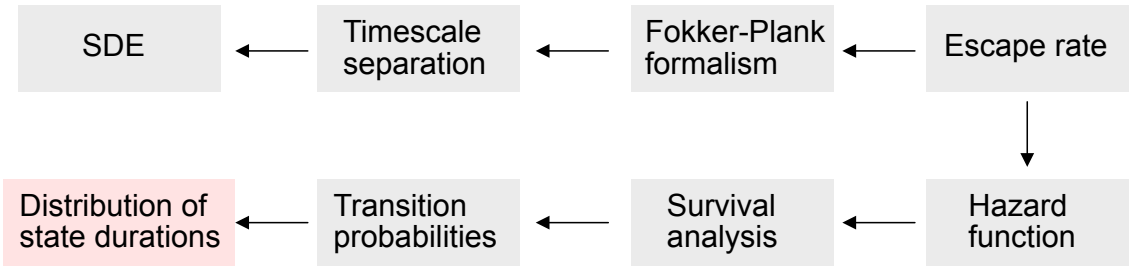


Figure F.4: A diagram of the stochastic methods approach for calculating the distribution of event times.

In our case, the hazard is given by the escape rate, so we set $h(t) = \hat{\nu}(w)$. Next, from the properties of the survival function (Kleinbaum & Klein, 1996), we get

$$\frac{dS}{dw}\frac{dw}{dt} = -S(w)\hat{\nu}(w) \tag{F.29}$$

The survival function is then given by:

$$S(w) = \exp\Big\{ \int \frac{-\hat{\nu}(w)}{dw/dt} dw \Big\} \tag{F.30}$$

where $d = 1/\tau_w$ and k is set to one of the two fixed points of the adaptation variable w^+ or w^-.

Now we can finally evaluate the transition probability:

$$\rho(w) = -\frac{dS}{dw} = \frac{\hat{\nu}(w)}{d(-w' + k)} \exp\Big\{ - \int \frac{\hat{\nu}(w')}{d(-w' + k)} dw' \Big\} \tag{F.31}$$

Since $\hat{\nu}$ is a single exponential, we can expand this equation further and write it in general form as

$$\rho_{a,b}(w) = \frac{c_1}{d(k - w)} \exp\Big\{ - c_2 k + c_2(k - w) + \frac{c_1}{d} e^{-c_2 k} \mathrm{Ei}(c_2(k - w)) \Big\} \tag{F.32}$$

where

$$\mathrm{Ei}(w) = \int_{-\infty}^{w} \frac{e^x}{x} dt \tag{F.33}$$

and a,b are transition up $\rightarrow$ down or down $\rightarrow$ up. Thus, $\rho_{a,b}(w)$ is the probability density for the values of W at which transition happens. Note that this approach requires k to be defined, so we apply it only for the case of bistable dynamics (Fig F.5).

Calculating the probability of IBIs

Let's assume that the w changes exponentially between bursts, and then the time between bursts (IBI) is linked with W as

$$T = -\log(\frac{W^+}{W^-})\tau_w = -\left(\log(W^+) + \log(W^-)\right)\tau_w \tag{F.34}$$

where W and W_0 are random variables.

We can analytically compute the distribution of ratios by applying the change of variables formula. If $W \sim \rho(w)$ then the distribution of $\tau_w log(W)$ is given by

$$\hat{\rho}_{a,b} = d\rho(e^{wd})e^{wd} \tag{F.35}$$

Now the probability density functions of the difference between two random variables are given by the convolution

$$p_{IBI}(t) = \int_{-\infty}^{\infty} \hat{\rho}_{du}(z)\hat{\rho}_{ud}(t+z)dz \tag{F.36}$$

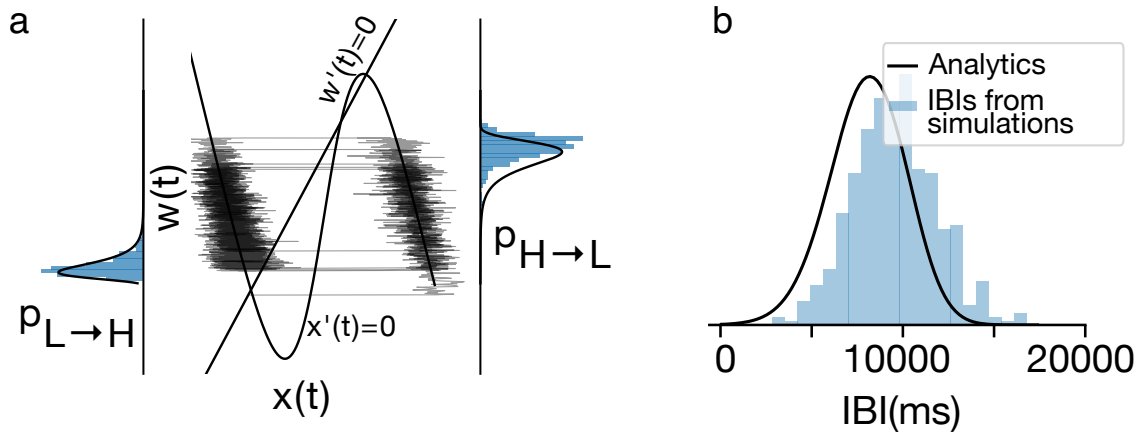


Figure F.5: Estimation of the transition probabilities for the reduced model in a bistable state. **(a)** Distribution of transition values w estimated with survival analysis. **(b)** Estimated distribution of the IBI.

Appendix G
Methods Chapter 4

G.1 Experimental details

All details of the experimental procedures will be available in a consequent publication (Vasiliev et al., in preparation). Here I briefly review the key details. All procedures were carried out with the prior approval of local authorities and in compliance with the European Community Guidelines for the Care and Use of Laboratory Animals.

In all experiments, we used male rats (N=43). The head-fixation chamber and electrodes (in 4 rats) were implanted after 7 days acclimation period. The details of the surgical procedures are described in Vasilev et al., 2022. Here the main difference is that in addition to scalp EEG, Cambridge Neurotech H7 or H9 silicone probe electrodes were implanted into ACC.

Behavioural task and recording setup

Head-fixed rats were trained to discriminate stimuli by running on a treadmill past a distance threshold (Go) or remaining immobile (NoGo). The treadmill was a freely rotating (backward and forward) cylinder on a metal axis. As stimuli, we used black-and-white gratings with a spatial frequency of 0.005 cycles/pixel. The visual Go stimuli were typically rotated by $340\mathrm{deg}$ and NoGo stimuli were $50\mathrm{deg}$. Habituation and training protocols as well as the details of water restriction are described in detail in (Vasilev et al., 2022). Upon a successful running to a Go stimulus past a running threshold, three drops of water were delivered. After running to NoGo stimuli past a threshold no water was delivered and animals received brown noise as negative feedback. The same number of Go and NoGo stimuli were presented in each recording session. In some NoGo stimulus trials, rats began running shortly after NoGo stimulus onset but made a change of mind (CoM) and chose to return to immobility before crossing the threshold. As discussed in the main text, the time of CoM was defined by the peak velocity of the treadmill when the rat stopped running and began slowing down.

Recording and spike sorting

The neural activity in ACC was recorded in 4 rats (between 14-30 recording sessions per animal). The neural activity was recorded by Ryo Iwaio, Dmitriy Vasiliev, and Nelson Totah. Spike sorting was performed using Kilsort2 (https://github.com/jamesjun/Kilosort2), an updated version of Kilosort (Pachitariu et al., 2016). Detected spikes were also manually curated by two human experts. The running velocity was recorded using the analog signal from the rotation of the treadmill axis using a rotary encoder. The signal was then digitalized and processed to obtain the angular velocity downsampled to 100Hz (Vasilev et al., 2022).

G.2 Data analysis

For each trial type (HT, FA, CR) we calculated the peak velocity (MATLAB function 'peakdetect'). Then we sorted the max velocity for each trial type into 3 tertiles: low, medium, and high.

ANOVA

To test the changes in firing rates, we aligned the spikes according to the peak velocity in CR and HT trials. The spike counts were then calculated with a 100ms time bin for each trial type and velocity tertile. We tested the changes in the firing rate using one-way ANOVA with 3-speed tertiles bin-wise ('anova1'). The neuron firing rate changes were considered when there were significant changes in 3 consequent bins (with a significance threshold of 0.05). The analysis was run separately for CR and HT trials. For this analysis, we considered all recorded neurons (603 units in total).

G.3 Linear encoding model

Model setup and regressor types

We designed a multiple regression model to relate the neuronal responses to the level of conflict and simultaneously to the behavioral and task-related variables. We closely follow the methodology from (Engelhard et al., 2019). For this analysis, we consider the firing rate in all recording sessions. We cut the spiking activity for every trial from the stimulus onset to 1.8s afterward. This window includes both behavioral responses and feedback presentation in all trials.

The spike times for every neuron were binned with 10ms and convolved with a 50ms Gaussian kernel to approximate the instantaneous firing rate[1].

Next, we design the GLM with the instantaneous firing rate for each neuron as a dependent variable and behavioral and task variables as independent variables. The regressors can be split into 3 categories: 1. trial-type variables, 2. continuous variables, and 3. event variables (Fig. 5.4). We also separate the regressors into groups based on the types of predictors. In the final model, we had 6 groups: trial type, stimulus, running velocity, feedback, and conflict. Some of the groups (i.e. velocity group) contained both event-linked and continuous regressors.

For the trial-type group was comprised of 3 regressors that were 0 or 1 depending on the trial type (for HT, FA, and CR trial types).

Event regressors were generated by convolving the event delta function with a set of splines. Events were stimulus onset, feedback onset, and peak velocity onset. In general, we generated B-splines with 10 degrees of freedom using the Matlab version of the fda package https://www.psych.mcgill.ca/misc/fda/software.html (Silverman & Ramsay, 2002). There were small differences depending on which event we considered.

- For stimulus onset, we used a set of 10 splines, that span the whole trial and had break points at $t_1 = 0, t_2 = 0.9, t_3 = 1.8$s from the stimulus onset. Thus, we have 10 progressors in the stimulus group.

- The feedback regressor group was identical to the stimulus group but convolved with the feedback onset separately for the positive and negative feedback. Therefore, the whole regressor group included 20 regressors.

- Velocity spines were analogous to the stimulus spines, but scaled with the peak velocity of each trial.

- The conflict group covered the time around the peak velocity in CR trials. It was comprised of 7 spines that go forward from the peak onset (with breakpoints $t_1 = 0, t_2 = 0.9, t_3 = 1.5$ from the peak), and 3 splines that go backward from the peak velocity (10 splines in total). These splines were also scaled with the peak velocity for each CR trial.

[1]We validated the results with shorted and longer kernels (up to 200ms) and did not see a qualitative change in the inferred clusters and regressor contributions, the variance explained in the full model is, however, affected by the smoothing and ranges from 10 to 25%. Similarly, we obtain qualitatively consistent results when the neural activity is binned with 50ms bins without smoothing.

A continuous variable that we considered was the running velocity in each trial. We down-sampled the velocity to 100Hz to match the firing rate discretization. The velocity group, thus, had only one continuous velocity regressor and 10 spline-based velocity regressors (see above).

Additionally, we include a session-wide spine basis, 5 B-splines with 10 degrees of freedom that span the whole session and account for the baseline firing drift. The model also included a constant offset for every neuron.

The final encoding model can be written as:

$$Y(t)_n = \beta_n + \sum_{k=1}^{K_{\text{event}}} \sum_{j=1}^{N_{\text{spline}}^{\text{event type}}} \beta_{n,k,j}^{\text{event}} \mathbf{e}_{k,s}(t) + \sum_{k=1}^{K_{\text{trial type}}} \beta_{n,k}^{\text{trial type}} \mathbf{c}_{k,s} + \beta_n^{\text{vel}} \mathbf{v}_s + \sum_{j=1}^{N_{\text{baseline}}} \beta_{n,j}^{\text{baseline}} \mathbf{d}_{j,s} + \varepsilon \tag{G.1}$$

$Y(t)_n$ is the instantaneous firing rate of a neuron n. The first sum is over the discrete (convolved with splines) events where, k is the index that goes through different event types (stimulus onset, feedback onset, peak velocity onset), and j indexes the spline number of each spline type. The vector of spline basis convolved with the delta for every event $\mathbf{e}_{k,s}(t)$ was unique for each session s. The second term is for trial-wide binary regressors ($\mathbf{c}_{k,s}$) that take 0 or 1 depending on the trial type (HT, FA, CR, $K_{\text{trial type}} = 3$). Next, there is a single velocity vector for each session $\mathbf{v}_s$ regressor. Finally, the last sum is over the spline basis for the baseline drift ($\mathbf{d}_{j,s}$) and ε is Gaussian noise. We fit the model in Matlab using the least squares solution with an extra L_2 penalty with $\lambda = 0.0001$.

We also tested feedback expectation spines and rule splines that were not included in the final model. The rule spines were the interaction of the stimulus spline and the stimulus type (Go or NoGo). The feedback expectation splines were identical to feedback splines but went backward from the feedback onset.

Model testing

To test the model we excluded separate regressor groups one by one and computed the 5-fold cross-validation over trials for every unit. We tested 6 regressor groups (m) instead of testing 57 single regressors β in the final model. The decrease of variance explained was then checked with paired t-test for every regressor group with Bonferroni-Holm adjusted p-values for multiple comparisons.

Regressor contributions

The contribution g of every regressor group was quantified $g_{\text{m}} = 1 - \frac{r^2_{\text{m}}}{r^2_{\text{full}}}$, where r^2_{m} variance explained in the model without a regressor group (m) and r^2_{full} is the variance explained of the full model. For all regressor groups apart from the baseline, we normalize the contributions by their sum $\frac{g_{\text{m}}}{\sum_m g_m}$. In some cases, when the model performance is worse than a prediction by a constant offset (which might be the case for neurons with very few spikes given that we compute prediction on the test trials). In such cases, we set negative contributions to zero.

Clustering

We cluster the neurons based on the regressor group contributions (g) using a Gaussian Mixture Model (GMM). We learn the means and covariance of the GMM using Matlab 'fitgmdist' function with full covariance matrix and L_2 regularization=0.01. To decide on the number of clusters for GMM, we identify the minimum of BIC between 1 and 20 clusters. Then we test the model by randomly shuffling the contribution per neuron 10000 times and comparing the negative log-likelihood of the identified model.

G.4 Population analysis

A detailed description of the demixed PCA algorithm ad its relationship to PCA can be found in H. Here I describe the major steps in the application of the method to spiking activity in ACC.

Preprocessing. We focused on the activity only from CR and HT trials to contrast the conflict. We expect the clear conflict signal to appear in correct rejection trials where animals self-corrected their behavior. We do not expect to see conflict in hit trials where animals correctly decided to run when they saw a go stimulus. To quantify the speed magnitude, we find the point of maximum running velocity in each trial and compute trial-specific tertiles. Thus, we have 2X3 ANOVA-like conditions: 3 levels of speed with 2 different trial types (HT and CR).

In every trial, we aligned the spike times with respect to the peak velocity. Next, we compute the spike counts in the time window 500ms before the event onset and 1500ms after the event onset using a 10ms time bin. Finally, for every neuron, we convolve the spiking activity a Gaussian kernel $\sigma = 50$ms. We excluded neurons whose average firing rate was less than 1Hz and also removed the neurons from the recording session with less than 2 trials in each condition. Thus, we used 426 single units pooled across 4 rats. The data is then organized in a tensor with dimensions $N \times S \times Q \times T \times K$, where N is the number of neurons, S is the speed tertile, Q is

the trial type, T is the trial time (-0.5 to 1.5s around peak velocity), K is the number of trials. Note that since we pool together the data from different recording session sessions, the tensor contains neurons that were not recorded simultaneously in the same trials, and therefore some trials would be empty (filled with NaNs).

To compute the demixed components we dPCA toolbox https://github.com/machenslab/dPCA that implements the algorithm from (Kobak et al., 2016). Briefly, we compute the average responses for every neuron in 2 trial types and 3 velocity tertiles.

From the tensor of average responses, we compute the marginalization for every condition: trial types, velocity tertile as well as for the interaction between conditions and condition-independent marginalization for all conditions (Fig. 5.10). Afterward, linear regression weights are trained to predict the marginalized responses. Then we compute principal components of the covariance of these responses are calculated and factorized into a decoder and encoder matrix (Kobak et al., 2016 and see also H). We fit dPCA with extra covariance terms estimated from individual trials as well as small regularization $\lambda = 10^{-4}$.

To compute the classification accuracy, we used the decoding axes for different condition terms to predict the type of trial at every time point. We separately calculate the classification accuracy for every condition. For instance, for the trial-type condition, we have 2 possible classes (HT or CR) and we would like to test the first dPC for the trial type. We first project the average response for each condition to the decoder axis of the 1st dPC. Then we can project individual trials onto the same decoder axis and calculate the closest average response to it. This minimum can be used as a trial-type prediction.

Following Kobak et al., 2016, we run this procedure with Monte Carlo leave-group-out cross-validation. For every condition, we randomly exclude one trial for each neuron (test trial). Then we compute the dPCA on the remaining trials (train trials) and calculate the prediction for this trial. We repeat the procedure 100 times and calculate the average accuracy (Fig. 5.11c). Next, to check if the average accuracy is above the chance level, we shuffle the trails between conditions for every neuron and compute the predictions for averages based on the shuffled dat

Appendix H
Demixed PCA in details

Demixed PCA (Kobak et al., 2016) is a generalization of reduced rank regression and principal component analysis that allows to extract latent population-wide components. Similarly to the classical definition of PCA, the technique searches for a number of latent components that explain the maximum variance in the recorded data. However, unlike classical PCA the components are optimized such that they capture the largest amount of variance in specific and pre-specified conditions. Loosely speaking, the technique combines dimensionality reduction with regression analysis. In this appendix, we review the basic definition of PCA and reduced rank regression and then define the basic methodology of demixed PCA following Kobak et al., 2016

Principal Component Analysis

Let us first revisit the basic definitions behind the Principal Component Analysis with specific application to N-dimensional recordings of neurons in mind. Let $\mathbf{X}$ be a matrix of the size $N \times T$, where raws are recorded neurons and values in columns are individual samples, e.g. activity recorded at each time t. The main objective of PCA is to find a linear transformation $\mathbf{P}$ for $\mathbf{X}$

$$\mathbf{PX} = \mathbf{Y} \tag{H.1}$$

where $\mathbf{Y}$ is the new representation of the data such that its covariance $\mathbf{C_Y} = \frac{1}{n}\mathbf{YY}^T$ is a diagonal matrix. The rows for $\mathbf{Y}$ then correspond to the principal components of the data.

The most practical way to obtain $\mathbf{P}$ includes the decomposition of the covariance of the original data $\mathbf{C_X}$ either with eigenvectors or by applying singular value decomposition (see Shlens, 2014 for more details). However, it is also possible to write the PCA as a constrained optimization problem. In this case, one can write down a PCA loss as

$$L_{\text{PCA}} = \left\|\mathbf{X} - \mathbf{P}^\top\mathbf{PX}\right\|^2 = \left\|\mathbf{X} - \mathbf{YP}^\top\right\|^2 = \sum_{i=1}^{N}\sum_{j=1}^{T}(x_{ij} - \sum_{k=1}^{K} Y_{ik}P_{ik}) \tag{H.2}$$

where k is the component number. We then need to minimize this loss subject to $\mathbf{P}^\top\mathbf{P} = \mathbf{I}$, where $\mathbf{I}$ is an identity matrix (i.e. assuring that the components are orthogonal to each other). In

other words, we would minimize the residuals between the projected and original data. Note that, this view is equivalent to several other definitions of PCA as an optimization problem, for instance maximizing the variance of the projection of $\mathbf{X}$ onto the principal components (Udell et al., 2016; Vidal et al., 2016 and also an excellent blog post on PCA as optimization from Alex Williams).

Intuitively, PCA linearly maps the neuronal activity onto a set of orthogonal components. The matrix $\mathbf{P}$ can be then viewed as a decoder, whereas $\mathbf{P}^\top$ is an encoder. Even though the top principal components will capture most of the variance in the data, it usually does not separate task-related variance into separate components (Machens, 2010).

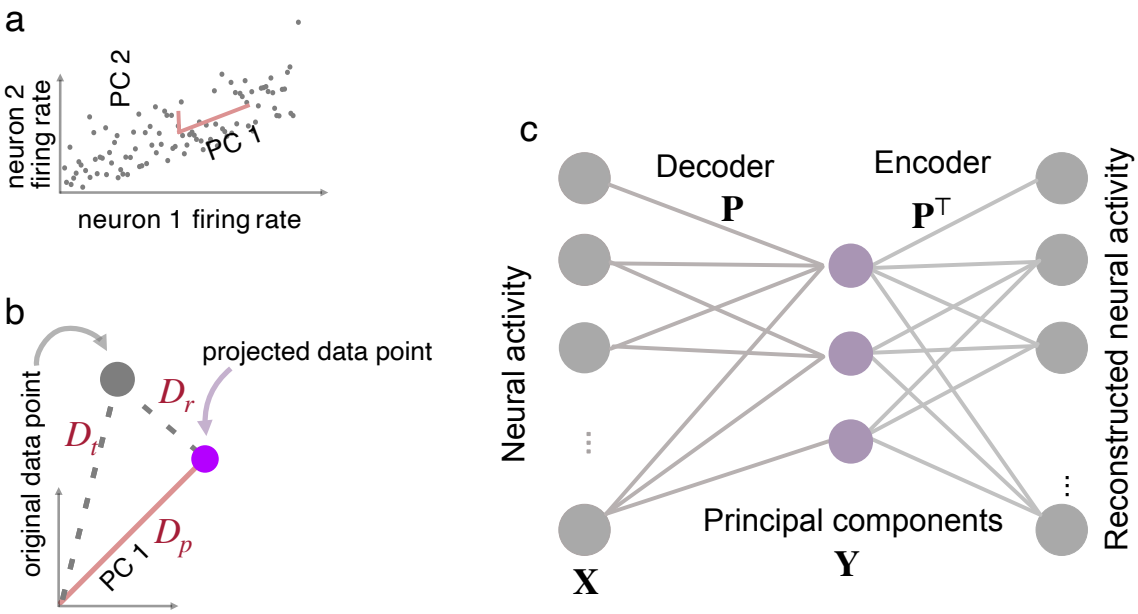


Figure H.1: **(a)**. Toy data of two neurons, which drift in time and receive Gaussian white noise as inputs. The first axis of the principal components captures the drift. Note that the PC axis here is rows of eigenvectors of the covariance matrix scaled by the square root of its eigenvalue. The second PC captures shared Gaussian white noise. **(b)** PCA objective: either maximizing the variance of the PC (D_p) or minimizing the residuals between the original and projected data (D_r). $D_t^2 = D_r^2 + D_p^2$. **(c)** PCA as decoding and encoding of neuronal data.

Demixed PCA

The idea behind demixed PCA is to allow the learning of principal components that depend on experimental variables. The main steps of dPCA could be described as follows:

1. Compute marginalisations $\mathbf{X}_\phi$ of the $\mathbf{X}$ that correspond to the experimental variables of interest

2. Learning linear regression weights $\mathbf{C}$ to predict marginalisation $\mathbf{X}_\phi$ using the whole data as predictors $\mathbf{X}$ and get the predictions $\mathbf{M} = \mathbf{CX}$.

3. Compute eigenvectors of the covariance of this predictions

4. Factorise the top principal components into predictors and encoder.

The first step of the algorithm is computing the marginalization of average neuronal responses in each experimental condition (either task or behavior related in our case). Intuitively, one can see that an average activity of a single neuron can be factorized into

$$\mathbf{X} = \sum_\phi \mathbf{X} + \mathbf{X}_{\mathbf{noise}} \tag{H.3}$$

where, following the notation in Kobak et al., 2016, ϕ is the index of task-relevant parameter. For instance, if we recorded neuronal activity during a presentation of a particular stimulus, then the whole recorded activity can be decomposed into the average activity during the stimulation and noise.

More precisely, let $\bar{x}_\phi$ be a marginalisation of the average single neuron activity for $\phi \in \Phi$ is a set of task-relevant parameters, then

$$\hat{x}_\phi = \langle x - \sum_{\tau \in \phi} \bar{x}_\tau \rangle_{\Phi \setminus \phi} \tag{H.4}$$

$\dot{\langle\rangle}_{\Phi\setminus\phi}$ denotes averaging over all the possible task-factors that are not ϕ. Importantly, both covariance and correlations between the marginalizations are zero (for derivations see Kobak et al., 2016). Therefore,

$$\mathbf{var}[x] = \sum_\phi \mathbf{var}[\hat{x}_\phi] + \mathbf{var}[x_{\text{noise}}] \tag{H.5}$$

And in the multivariate case

$$\mathbf{C} = \sum_{\phi} (\mathbf{C}_{\phi} + BC_{\text{noise}}) \tag{H.6}$$

The marginalizations are useful to find task-specific principal components in the data. Thus, now we can write the dPCA loss as

$$\mathbf{L} = \sum_{\phi} \|\mathbf{X}_{\phi} - \mathbf{F}_{\phi}\mathbf{D}_{\phi}\mathbf{X}\|^2 \tag{H.7}$$

Note that now instead of reconstructing the whole data, the objective includes reconstructing each marginalization. Furthermore, here we explicitly write $\mathbf{F}_{\phi}$ and $\mathbf{D}_{\phi}$ for the encoder/decoder pair that corresponds to each marginalization.

Each loss term thus increases the variance from a specific marginalization and decreases the variance from all the other marginalizations. To see this one can rewrite the loss as follows

$$\mathbf{L}_{\phi} = \|\mathbf{X}_{\phi} - \mathbf{F}_{\phi}\mathbf{D}_{\phi}\mathbf{X}_{\phi}\|^2 + \|\mathbf{F}_{\phi}\mathbf{D}_{\phi}(\mathbf{X} - \mathbf{X}_{\phi})\|^2 \tag{H.8}$$

The second term here corresponds to the variance from all the other marginalizations, while the first one is similar to the standard PCA loss and decreases the variance from the residuals in the selected marginalization. Now, if $\mathbf{F}_{\phi}\mathbf{D}_{\phi} = \mathbf{A}$ and we do not have any constraints on $\mathbf{A}$ then the solution to H.7 can be found analytically with ordinary least squares equation (OLS), we call this solution $\mathbf{A}_{\text{OLS}}$. However, given that we expect $\mathbf{A}$ to have a pre-specified rank, the solution renders into the solution of the reduced rank regression, which is equivalent to performing the PCA on the $\mathbf{A}_{\text{OLS}}\mathbf{X}$. To see that, let us decompose the loss into two terms

$$\|\mathbf{X}_{\phi} - \mathbf{A}_{\mathbf{q}}\mathbf{X}\|^2 = \|\mathbf{X}_{\phi} - \mathbf{A}_{\text{OLS}}\mathbf{X}\|^2 + \|\mathbf{A}_{\text{OLS}}\mathbf{X} - \mathbf{A}_{\text{q}}\mathbf{X}\|^2 \tag{H.9}$$

where $\mathbf{A}_q$ is the rank q approximation of $\mathbf{A}$. Now we can see that the loss only depends on the second term and the best rank q approximation of $\mathbf{A}_{\text{OLS}}\mathbf{X}$ is given by q principal components. So finally, we arrive at $\mathbf{U_q}\mathbf{U_q}^{\top}\mathbf{A}_{\text{OLS}}$, where $\mathbf{U_q}$ is the matrix of top q components. Now we can set decoder to be $\mathbf{D} = \mathbf{U}_q^{\top}\mathbf{A}_{\text{OLS}}$ and encoder to be given by the q first principal components of $\mathbf{A}_{\text{OLS}}\mathbf{X}$, which is $\mathbf{U_q}$. One can run the procedure iteratively for all marginalizations of interest and eventually reach a good reconstruction of the data (see Fig. 5.11). The main benefit of dPCA is that now all the top q demixed components would correspond to the maximum variance only in a task-specific response and therefore might give a meaningful interpretation of population dynamics.

Note that here, the constraint on the orthogonality of the extracted components for different marginalizations is lost. Therefore, top demixed components may be correlated or non-orthogonal. Furthermore, the implementation of dPCA based on reduced rank regression also does not impose the orthogonality of encoder/decoder axes. According to Kobak et al., 2016, this allows for better isolation of task-related components.

Regularisation and unbalanced data

So far we considered only the average activity in each marginalization and assumed that the conditions have an equal number of observations. However, it is not the case in the dataset for conflict and error encoding in ACC. Therefore, we need to add an extra correction term that would account for the differences in the residual noise coming from different marginalizations are rewrite the loss as

$$\mathbf{L}_\phi = \|\mathbf{X}_\phi - \mathbf{F}\mathbf{D}\mathbf{X}_{\text{PSTH}}\|^2 + KSQT \left\|\mathbf{F}\mathbf{D}\mathbf{C}_{\text{noise}}^{1/2}\right\|^2 + \mu \|\mathbf{F}\mathbf{D}\|^2 \tag{H.10}$$

the first term includes only the PSTH without the noise $\mathbf{X}_{\text{PSTH}} = \mathbf{X} - \mathbf{X}_{\text{noise}}$, the second term introduces the covariance matrix of the residual noise $\mathbf{C}_{\text{noise}} = \frac{1}{n}\mathbf{X}_{\text{noise}}\mathbf{X}_{\text{noise}}^\top$ for originally multiplied by the number of trials in each condition.

Bibliography

Abbott, L. F., & Nelson, S. B. (2000). Synaptic plasticity: Taming the beast. *Nature Neuroscience*, *3*(11), 1178–1183. https://doi.org/10.1038/81453

Ahtiainen, A., Genocchi, B., Tanskanen, J. M. A., Barros, M. T., Hyttinen, J. A. K., & Lenk, K. (2021). Astrocytes Exhibit a Protective Role in Neuronal Firing Patterns under Chemically Induced Seizures in Neuron–Astrocyte Co-Cultures. *International Journal of Molecular Sciences*, *22*(23).

Alreja, A., Nemenman, I., & Rozell, C. J. (2022). Constrained brain volume in an efficient coding model explains the fraction of excitatory and inhibitory neurons in sensory cortices. *PLOS Computational Biology*, *18*(1), e1009642. https://doi.org/10.1371/journal.pcbi.1009642

Antolík, J., Hofer, S. B., Bednar, J. A., & Mrsic-Flogel, T. D. (2016). Model Constrained by Visual Hierarchy Improves Prediction of Neural Responses to Natural Scenes. *PLOS Computational Biology*, *12*(6). https://doi.org/10.1371/journal.pcbi.1004927

Aoi, M., & Pillow, J. W. (2018). Model-based targeted dimensionality reduction for neuronal population data. *Advances in neural information processing systems*, *31*.

Aponte, D. A., Handy, G., Kline, A. M., Tsukano, H., Doiron, B., & Kato, H. K. (2021). Recurrent network dynamics shape direction selectivity in primary auditory cortex. *Nature Communications*, *12*(1), 314. https://doi.org/10.1038/s41467-020-20590-6

Aristotle. (n.d.). *Aristotle's physics: Books i and ii*. https://classics.mit.edu/Aristotle/physics.html

Badin, A.-S., Fermani, F., & Greenfield, S. A. (2017). The features and functions of neuronal assemblies: Possible dependency on mechanisms beyond synaptic transmission. *Frontiers in Neural Circuits*, *10*. https://doi.org/10.3389/fncir.2016.00114

Bal, T., Nagy, F., & Moulins, M. (1988). The pyloric central pattern generator in crustacea: A set of conditional neuronal oscillators. *Journal of Comparative Physiology A*, *163*, 715–727.

Barack, D. L., & Krakauer, J. W. (2021). Two views on the cognitive brain. *Nature Reviews Neuroscience*, *22*(6).

Barak, O. (2017). Recurrent neural networks as versatile tools of neuroscience research. *Current opinion in neurobiology*, *46*, 1–6.

Barral, J., & D Reyes, A. (2016). Synaptic scaling rule preserves excitatory–inhibitory balance and salient neuronal network dynamics. *Nature Neuroscience*, *19*(12), 1690–1696. https://doi.org/10.1038/nn.4415

Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., & Robert, C. P. (2009). Adaptive approximate Bayesian computation. *Biometrika*, *96*(4), 983–990. https://doi.org/10.1093/biomet/asp052

Beck, J., Deistler, M., Bernaerts, Y., Macke, J. H., & Berens, P. (2022). Efficient identification of informative features in simulation-based inference. *Advances in Neural Information Processing Systems*, *35*, 19260–19273.

Beggs, J. M., & Plenz, D. (2003). Neuronal Avalanches in Neocortical Circuits. *Journal of Neuroscience*, *23*(35). https://doi.org/10.1523/JNEUROSCI.23-35-11167.2003

Bellec, G., Scherr, F., Subramoney, A., Hajek, E., Salaj, D., Legenstein, R., & Maass, W. (2020). A solution to the learning dilemma for recurrent networks of spiking neurons. *Nature Communications*, *11*. https://doi.org/10.1038/s41467-020-17236-y

Benda, J., & Herz, A. V. M. (2003). A Universal Model for Spike-Frequency Adaptation. *Neural Computation*, *15*(11), 2523–2564. https://doi.org/10.1162/089976603322385063

Benoit, C.-E., Solopchuk, O., Borragán, G., Carbonnelle, A., Van Durme, S., & Zénon, A. (2019). Cognitive task avoidance correlates with fatigue-induced performance decrement but not with subjective fatigue. *Neuropsychologia*, *123*, 30–40. https://doi.org/10.1016/j.neuropsychologia.2018.06.017

Berens, P., Keliris, G., Ecker, A., Logothetis, N., & Tolias, A. (2008). Comparing the feature selectivity of the gamma-band of the local field potential and the underlying spiking activity in primate visual cortex. *Frontiers in Systems Neuroscience*, *2*. Retrieved May 4, 2023, from https://www.frontiersin.org/articles/10.3389/neuro.06.002.2008

Berg, R. W., Alaburda, A., & Hounsgaard, J. (2007). Balanced Inhibition and Excitation Drive Spike Activity in Spinal Half-Centers. *Science*, *315*(5810), 390–393. https://doi.org/10.1126/science.1134960

Bi, G.-q., & Poo, M.-m. (1999). Distributed synaptic modification in neural networks induced by patterned stimulation. *Nature*, *401*(6755). https://doi.org/10.1038/44573

Bialek, W. (2018). Perspectives on theory at the interface of physics and biology. *Reports on Progress in Physics*, *81*(1), 012601. https://doi.org/10.1088/1361-6633/aa995b

Bialek, W., Van Steveninck, R. R. D. R., & Tishby, N. (2006). Efficient representation as a design principle for neural coding and computation. *2006 IEEE international symposium on information theory*, 659–663.

Bienenstock, E., Cooper, L., & Munro, P. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *Journal of Neuroscience*, *2*(1), 32–48. https://doi.org/10.1523/JNEUROSCI.02-01-00032.1982

Bishop, C. M. (1994). Mixture density networks.

Blais, B. S., Intrator, N., Shouval, H., & Cooper, L. N. (1998). Receptive Field Formation in Natural Scene Environments: Comparison of Single-Cell Learning Rules. *Neural Computation*, *10*(7), 1797–1813. https://doi.org/10.1162/089976698300017142

Blakemore, C., Carpenter, R. H. S., & Georgeson, M. A. (1970). Lateral Inhibition between Orientation Detectors in the Human Visual System. *Nature*, *228*(5266), 37–39. https://doi.org/10.1038/228037a0

Boksem, M. A. S., Meijman, T. F., & Lorist, M. M. (2006). Mental fatigue, motivation and action monitoring. *Biological Psychology*, *72*(2), 123–132. https://doi.org/10.1016/j.biopsycho.2005.08.007

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological review*, *108*(3), 624.

Botvinick, M. M., & Cohen, J. D. (2014). The Computational and Neural Basis of Cognitive Control: Charted Territory and New Frontiers. *Cognitive Science*, *38*(6). https://doi.org/10.1111/cogs.12126

Breskin, I., Soriano, J., Moses, E., & Tlusty, T. (2006). Percolation in Living Neural Networks. *Physical Review Letters*, *97*(18). https://doi.org/10.1103/PhysRevLett.97.188102

Brette, R., & Gerstner, W. (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *Journal of Neurophysiology*, *94*(5), 3637–3642. https://doi.org/10.1152/jn.00686.2005

Brito, C. S. N., & Gerstner, W. (2016). Nonlinear hebbian learning as a unifying principle in receptive field formation. *PLOS Computational Biology*, *12*(9), 1–24. https://doi.org/10.1371/journal.pcbi.1005070

Brunel. (2000). Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *Journal of Computational Neuroscience*, *8*. https://doi.org/10.1023/A:1008925309027

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

Brunel, N., & Hakim, V. (2008). Sparsely synchronized neuronal oscillations. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *18*(1), 015113. https://doi.org/10.1063/1.2779858

Buzsáki, G., & Wang, X.-J. (2012). Mechanisms of gamma oscillations. *Annual review of neuroscience*, *35*, 203–225.

Cabrera-Garcia, D., Warm, D., de la Fuente, P., Fernández-Sánchez, M. T., Novelli, A., & Villanueva-Balsera, J. M. (2021). Early prediction of developing spontaneous activity in cultured neuronal networks. *Scientific Reports*, *11*(1).

Cadena, S. A., Denfield, G. H., Walker, E. Y., Gatys, L. A., Tolias, A. S., Bethge, M., & Ecker, A. S. (2019). Deep convolutional models improve predictions of macaque v1 responses to natural images. *PLoS computational biology*, *15*(4), e1006897.

Callegari, F., Brofiga, M., Poggio, F., & Massobrio, P. (2022). Stimulus-Evoked Activity Modulation of In Vitro Engineered Cortical and Hippocampal Networks. *Micromachines*, *13*(8). https://doi.org/10.3390/mi13081212

Capano, V., Herrmann, H. J., & de Arcangelis, L. (2015). Optimal percentage of inhibitory synapses in multi-task learning. *Scientific Reports*, *5*, 9895. https://doi.org/10.1038/srep09895

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior Cingulate Cortex, Error Detection, and the Online Monitoring of Performance. *Science*, *280*(5364), 747–749. https://doi.org/10.1126/science.280.5364.747

Charlesworth, P., Cotterill, E., Morton, A., Grant, S. G., & Eglen, S. J. (2015). Quantitative differences in developmental profiles of spontaneous activity in cortical and hippocampal cultures. *Neural Development*, *10*(1), 1. https://doi.org/10.1186/s13064-014-0028-0

Chen, R. T., Rubanova, Y., Bettencourt, J., & Duvenaud, D. K. (2018). Neural ordinary differential equations. *Advances in neural information processing systems*, *31*.

Choudhary, A., Peles, D., Nayak, R., Mizrahi, L., & Stern, S. (2022). Current progress in understanding schizophrenia using genomics and pluripotent stem cells: A meta-analytical overview. *Schizophrenia Research*. https://doi.org/10.1016/j.schres.2022.11.001

Clopath, C., Vogels, T. P., Froemke, R. C., & Sprekeler, H. (2016). Receptive field formation by interacting excitatory and inhibitory synaptic plasticity. *bioRxiv*. https://doi.org/10.1101/066589

Cohen, D., & Segal, M. (2011). Network bursts in hippocampal microcultures are terminated by exhaustion of vesicle pools. *Journal of Neurophysiology*, *106*(5), 2314–2321. https://doi.org/10.1152/jn.00969.2010

Cohen, M. X., & van Gaal, S. (2014). Subthreshold muscle twitches dissociate oscillatory neural signatures of conflicts from errors. *NeuroImage*, *86*, 503–513. https://doi.org/10.1016/j.neuroimage.2013.10.033

Cotterill, E., Charlesworth, P., Thomas, C. W., Paulsen, O., & Eglen, S. J. (2016). A comparison of computational methods for detecting bursts in neuronal spike trains and their application

to human stem cell-derived neuronal networks. *Journal of neurophysiology*, *116*(2), 306–321.

Cranmer, K., Brehmer, J., & Louppe, G. (2020). The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 201912789. https://doi.org/10.1073/pnas.1912789117

Deistler, M., Macke, J. H., & Gonçalves, P. J. (2022). Energy-efficient network activity from disparate circuit parameters. *Proceedings of the National Academy of Sciences*, *119*(44), e2207632119.

Deneve, S., & Machens, C. (2016). Efficient codes and balanced networks. *Nature Neuroscience*, *19*, 375–382. https://doi.org/10.1038/nn.4243

Destexhe, A., Mainen, Z. F., & Sejnowski, T. J. (1995). Fast kinetic models for simulating ampa, nmda, gaba a and gaba b receptors. *The Neurobiology of Computation: Proceedings of the Third Annual Computation and Neural Systems Conference*, 9–14.

Dichter, M., & Ayala, G. (1987). Cellular mechanisms of epilepsy: A status report. *Science*, *237*(4811), 157–164. https://doi.org/10.1126/science.3037700

Ding, Z., Fahey, P. G., Papadopoulos, S., Wang, E. Y., Celii, B., Papadopoulos, C., Kunin, A., Chang, A., Fu, J., Ding, Z., Patel, S., Ponder, K., Muhammad, T., Bae, J. A., Bodor, A. L., Brittain, D., Buchanan, J., Bumbarger, D. J., Castro, M. A., … Tolias, A. S. (2023). Functional connectomics reveals general wiring rule in mouse visual cortex. https://doi.org/10.1101/2023.03.13.531369

D'Souza, R. D., Bista, P., Meier, A. M., Ji, W., & Burkhalter, A. (2019). Spatial clustering of inhibition in mouse primary visual cortex. *Neuron*, *104*(3), 588–600.e5. https://doi.org/10.1016/j.neuron.2019.09.020

Dubreuil, A., Valente, A., Beiran, M., Mastrogiuseppe, F., & Ostojic, S. (2022). The role of population structure in computations through neural dynamics. *Nature Neuroscience*, *25*(6). https://doi.org/10.1038/s41593-022-01088-4

Durkan, C., Bekasov, A., Murray, I., & Papamakarios, G. (2019). Neural spline flows. *Advances in neural information processing systems*, *32*.

Ebitz, R. B., & Platt, M. L. (2015). Neuronal activity in primate dorsal anterior cingulate cortex signals task conflict and predicts adjustments in pupil-linked arousal. *Neuron*, *85*(3), 628–640.

Ebitz, R. B., Smith, E. H., Horga, G., Schevon, C. A., Yates, M. J., McKhann, G. M., Botvinick, M. M., Sheth, S. A., & Hayden, B. Y. (2020). Human dorsal anterior cingulate neurons signal conflict by amplifying task-relevant information. *BioRxiv*, 2020–03.

Ecker, A. S., & Tolias, A. S. (2014). Is there signal in the noise? *Nature neuroscience*, *17*(6), 750–751. Retrieved July 3, 2017, from https://www.nature.com/articles/nn.3722

Eckmann, J.-P., Jacobi, S., Marom, S., Moses, E., & Zbinden, C. (2008). Leader neurons in population bursts of 2D living neural networks. *New Journal of Physics*, *10*(1), 015011.

Eckmann, S., & Gjorgjieva, J. (2022). Synapse-type-specific competitive hebbian learning forms functional recurrent networks. *bioRxiv*. https://doi.org/10.1101/2022.03.11.483899

Effenberger, F., Jost, J., & Levina, A. (2015). Self-organization in Balanced State Networks by STDP and Homeostatic Plasticity. *PLOS Computational Biology*, *11*(9), e1004420. https://doi.org/10.1371/journal.pcbi.1004420

Engelhard, B., Finkelstein, J., Cox, J., Fleming, W., Jang, H. J., Ornelas, S., Koay, S. A., Thiberge, S. Y., Daw, N. D., Tank, D. W., & Witten, I. B. (2019). Specialized coding of sensory, motor and cognitive variables in VTA dopamine neurons. *Nature*, *570*(7762). https://doi.org/10.1038/s41586-019-1261-9

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Ermentrout, B. (2002). *Simulating, analyzing, and animating dynamical systems*. Society for Industrial; Applied Mathematics. https://doi.org/10.1137/1.9780898718195

Estévez-Priego, E., Moreno-Fina, M., Monni, E., Kokaia, Z., Soriano, J., & Tornero, D. (2023). Long-term calcium imaging reveals functional development in hiPSC-derived cultures comparable to human but not rat primary cultures. *Stem Cell Reports*, *18*(1), 205–219. https://doi.org/10.1016/j.stemcr.2022.11.014

Eytan, D., & Marom, S. (2006). Dynamics and Effective Topology Underlying Synchronization in Networks of Cortical Neurons. *Journal of Neuroscience*, *26*(33), 8465–8476. https://doi.org/10.1523/JNEUROSCI.1627-06.2006

Fardet, T. (2018). *Growth and activity of neuronal cultures : Emergence of organized behaviors* (These de doctorat). Sorbonne Paris Cité. Retrieved April 13, 2023, from https://www.theses.fr/2018USPCC002

Fardet, T., Ballandras, M., Bottani, S., Métens, S., & Monceau, P. (2018). Understanding the Generation of Network Bursts by Adaptive Oscillatory Neurons. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00041

Fechner, G. T. (1860). *Elemente der psychophysik* (Vol. 2). Breitkopf u. Härtel.

Feinerman, O., Segal, M., & Moses, E. (2005). Signal Propagation Along Unidimensional Neuronal Networks. *Journal of Neurophysiology*, *94*(5), 3406–3416. https://doi.org/10.1152/jn.00264.2005

Ferguson, K. A., Njap, F., Nicola, W., Skinner, F. K., & Campbell, S. A. (2015). Examining the limits of cellular adaptation bursting mechanisms in biologically-based excitatory networks of the hippocampus. *Journal of Computational Neuroscience*, *39*(3), 289–309.

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: From behavior to neural representations. *Trends in cognitive sciences*, *14*(3), 119–130.

Fisher, D., Olasagasti, I., Tank, D. W., Aksay, E. R., & Goldman, M. S. (2013). A modeling framework for deriving the structural and functional architecture of a short-term memory microcircuit. *Neuron*, *79*(5), 987–1000.

Frega, M., Tedesco, M., Massobrio, P., Pesce, M., & Martinoia, S. (2014). Network dynamics of 3d engineered neuronal cultures: A new experimental model for in-vitro electrophysiology. *Scientific reports*, *4*(1), 5489.

Fricker, D., Verheugen, J. A., & Miles, R. (1999). Cell-attached measurements of the firing threshold of rat hippocampal neurones. *The Journal of physiology*, *517*(3), 791–804.

Froemke, R. C. (2015). Plasticity of cortical excitatory-inhibitory balance. *Annual review of neuroscience*, *38*, 195–219.

Fu, Z., Beam, D., Chung, J. M., Reed, C. M., Mamelak, A. N., Adolphs, R., & Rutishauser, U. (2022). The geometry of domain-general performance monitoring in the human medial frontal cortex. *Science*, *376*(6593), eabm9922.

Fu, Z., Sajad, A., Errington, S. P., Schall, J. D., & Rutishauser, U. (2023). Neurophysiological mechanisms of error monitoring in human and non-human primates. *Nature Reviews Neuroscience*, *24*(3), 153–172. https://doi.org/10.1038/s41583-022-00670-w

Gal, A., Eytan, D., Wallach, A., Sandler, M., Schiller, J., & Marom, S. (2010). Dynamics of Excitability over Extended Timescales in Cultured Cortical Neurons. *Journal of Neuroscience*, *30*(48), 16332–16342. https://doi.org/10.1523/JNEUROSCI.4859-10.2010

Gardiner, C. (2009). *Stochastic methods: A handbook for the natural and social sciences*. Springer.

Georges, A., Kotliar, G., Krauth, W., & Rozenberg, M. J. (1996). Dynamical mean-field theory of strongly correlated fermion systems and the limit of infinite dimensions. *Reviews of Modern Physics*, *68*(1), 13.

Germain, M., Gregor, K., Murray, I., & Larochelle, H. (2015). Made: Masked autoencoder for distribution estimation. *International conference on machine learning*, 881–889.

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

Giannakakis, E., Hutchings, F., Papasavvas, C. A., Han, C. E., Weber, B., Zhang, C., & Kaiser, M. (2020). Computational modelling of the long-term effects of brain stimulation on the local and global structural connectivity of epileptic patients. *PLOS ONE*, *15*(2), 1–21. https://doi.org/10.1371/journal.pone.0221380

Giannakakis, E., Vinogradov, O., Buendia, V., & Levina, A. (2023). Recurrent connectivity structure controls the emergence of co-tuned excitation and inhibition. *bioRxiv*, 2023–02.

Gigante, G., Deco, G., Marom, S., & Del Giudice, P. (2015). Network Events on Multiple Space and Time Scales in Cultured Neural Networks and in a Stochastic Rate Model. *PLOS Computational Biology*, *11*(11), e1004547. https://doi.org/10.1371/journal.pcbi.1004547

Giugliano, M., Darbon, P., Arsiero, M., Lüscher, H.-R., & Streit, J. (2004). Single-Neuron Discharge Properties and Network Activity in Dissociated Cultures of Neocortex. *Journal of Neurophysiology*, *92*(2), 977–996. https://doi.org/10.1152/jn.00067.2004

Gjorgjieva, J., Clopath, C., Audet, J., & Pfister, J.-P. (2011). A triplet spike-timing–dependent plasticity model generalizes the bienenstock–cooper–munro rule to higher-order spatiotemporal correlations. *Proceedings of the National Academy of Sciences*, *108*(48), 19383–19388. https://doi.org/10.1073/pnas.1105933108

Gonçalves, P. J., Lueckmann, J.-M., Deistler, M., Nonnenmacher, M., Öcal, K., Bassetto, G., Chintaluri, C., Podlaski, W. F., Haddad, S. A., Vogels, T. P., Greenberg, D. S., & Macke, J. H. (2020). Training deep neural density estimators to identify mechanistic models of neural dynamics. *eLife*, *9*, e56261. https://doi.org/10.7554/eLife.56261

Greenberg, D., Nonnenmacher, M., & Macke, J. (2019). Automatic posterior transformation for likelihood-free inference. *International Conference on Machine Learning*, 2404–2414.

Gu, X., & Spitzer, N. C. (1995). Distinct aspects of neuronal differentiation encoded by frequency of spontaneous ca2+ transients. *Nature*, *375*(6534), 784–787.

Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., & Sethna, J. P. (2007). Universally Sloppy Parameter Sensitivities in Systems Biology. *PLoS Computational Biology*, *3*(10). https://doi.org/10.1371/journal.pcbi.0030189

Haefner, R. M., Berkes, P., & Fiser, J. (2016). Perceptual Decision-Making as Probabilistic Inference by Neural Sampling. *Neuron*, *90*(3), 649–660. https://doi.org/10.1016/j.neuron.2016.03.020

Harris-Warrick, R. M., Marder, E., Selverston, A. I., Moulins, M., Sejnowski, T. J., & Poggio, T. A. (1992). *Dynamic biological networks: The stomatogastric nervous system*. MIT press.

He, L. S., Rue, M. C., Morozova, E. O., Powell, D. J., James, E. J., Kar, M., & Marder, E. (2020). Rapid adaptation to elevated extracellular potassium in the pyloric circuit of the crab, *Cancer borealis*. *Journal of Neurophysiology*, *123*(5), 2075–2089. https://doi.org/10.1152/jn.00135.2020

Hellyer, P. J., Jachs, B., Clopath, C., & Leech, R. (2016). Local inhibitory plasticity tunes macroscopic brain dynamics and allows the emergence of functional brain networks. *NeuroImage*, *124*, 85–95. https://doi.org/10.1016/j.neuroimage.2015.08.069

Hennequin, G., Agnes, E. J., & Vogels, T. P. (2017). Inhibitory plasticity: Balance, control, and codependence. *Annual review of neuroscience*, *40*, 557–579.

Hester, R., Fassbender, C., & Garavan, H. (2004). Individual Differences in Error Processing: A Review and Reanalysis of Three Event-related fMRI Studies Using the GO/NOGO Task. *Cerebral Cortex*, *14*(9), 986–994. https://doi.org/10.1093/cercor/bhh059

Hilgetag, C., & Kaiser, M. (2004). Clustered organization of cortical connectivity. *Neuroinformatics*, *2*, 353–60. https://doi.org/10.1385/NI:2:3:353

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

Hoffmann, A., Ziller, M., & Spengler, D. (2019a). Progress in iPSC-Based Modeling of Psychiatric Disorders. *International Journal of Molecular Sciences*, *20*(19). https://doi.org/10.3390/ijms20194896

Hoffmann, A., Ziller, M., & Spengler, D. (2019b). Progress in iPSC-Based Modeling of Psychiatric Disorders. *International Journal of Molecular Sciences*, *20*(19), 4896. https://doi.org/10.3390/ijms20194896

Huh, D., & Sejnowski, T. J. (2017). Gradient Descent for Spiking Neural Networks. *arXiv:1706.04698*. Retrieved February 8, 2018, from http://arxiv.org/abs/1706.04698

Humeau, Y., & Choquet, D. (2019). The next generation of approaches to investigate the link between synaptic plasticity and learning. *Nature neuroscience*, *22*(10), 1536–1543.

Hurwitz, C., Kudryashova, N., Onken, A., & Hennig, M. H. (2021). Building population models for large-scale neural recordings: Opportunities and pitfalls. *Current opinion in neurobiology*, *70*, 64–73.

Hyvärinen, T., Hyysalo, A., Kapucu, F. E., Aarnos, L., Vinogradov, A., Eglen, S. J., Ylä-Outinen, L., & Narkilahti, S. (2019). Functional characterization of human pluripotent stem cell-derived cortical networks differentiated on laminin-521 substrate: Comparison to rat cortical cultures. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-53647-8

Iannaccone, R., Hauser, T. U., Staempfli, P., Walitza, S., Brandeis, D., & Brem, S. (2015). Conflict monitoring and error processing: New insights from simultaneous EEG–fMRI. *NeuroImage*, *105*, 395–407. https://doi.org/10.1016/j.neuroimage.2014.10.028

Isaacson, J. S., & Scanziani, M. (2011). How Inhibition Shapes Cortical Activity. *Neuron*, *72*(2), 231–243. https://doi.org/10.1016/j.neuron.2011.09.027

Ito, S., Stuphorn, V., Brown, J. W., & Schall, J. D. (2003). Performance Monitoring by the Anterior Cingulate Cortex During Saccade Countermanding. *Science*, *302*(5642), 120–122. https://doi.org/10.1126/science.1087847

Ivenshitz, M., & Segal, M. (2010). Neuronal Density Determines Network Connectivity and Spontaneous Activity in Cultured Hippocampus. *Journal of Neurophysiology*, *104*(2), 1052–1060. https://doi.org/10.1152/jn.00914.2009

Izhikevich, E. M. (2007). *Dynamical systems in neuroscience*. MIT press.

Jercog, D., Roxin, A., Barthó, P., Luczak, A., Compte, A., & de la Rocha, J. (2017). UP-DOWN cortical dynamics reflect state transitions in a bistable network. *eLife*, *6*. https://doi.org/10.7554/eLife.22425

Jutras, M. J., & Buffalo, E. A. (2010). Synchronous neural activity and memory formation. *Current opinion in neurobiology*, *20*(2), 150–155.

Kagan, B. J., Kitchen, A. C., Tran, N. T., Habibollahi, F., Khajehnejad, M., Parker, B. J., Bhat, A., Rollo, B., Razi, A., & Friston, K. J. (2022). In vitro neurons learn and exhibit sentience when embodied in a simulated game-world. *Neuron*, *110*(23), 3952–3969.e8. https://doi.org/10.1016/j.neuron.2022.09.001

Kapucu, F. E., Vinogradov, A., Hyvärinen, T., Ylä-Outinen, L., & Narkilahti, S. (2022). Comparative microelectrode array data of the functional development of hPSC-derived and rat neuronal networks. *Scientific Data*, *9*(1). https://doi.org/10.1038/s41597-022-01242-4

Karagiannis, P., Takahashi, K., Saito, M., Yoshida, Y., Okita, K., Watanabe, A., Inoue, H., Yamashita, J. K., Todani, M., Nakagawa, M., et al. (2019). Induced pluripotent stem cells and their use in human models of disease and development. *Physiological reviews*, *99*(1), 79–114.

Kaufman, M., Reinartz, S., & Ziv, N. E. (2014). Adaptation to prolonged neuromodulation in cortical cultures: An invariable return to network synchrony, 22.

Khajehabdollahi, S., Giannakakis, E., Prosi, J., & Levina, A. (2021). *Reservoir computing with self-organizing neural oscillators* (Vol. ALIFE 2021: The 2021 Conference on Artificial Life). https://doi.org/10.1162/isal_a_00409

Kidger, P., Foster, J., Li, X., & Lyons, T. (2021). Efficient and Accurate Gradients for Neural SDEs. *Advances in Neural Information Processing Systems*.

Kim, D. H., Kim, J., Marques, J. C., Grama, A., Hildebrand, D. G. C., Gu, W., Li, J. M., & Robson, D. N. (2017). Pan-neuronal calcium imaging with cellular resolution in freely swimming zebrafish. *Nature Methods*, *14*(11). https://doi.org/10.1038/nmeth.4429

Kirillov, A. (n.d.). NeuroExplorer Manual, 373.

Kleinbaum, D. G., & Klein, M. (1996). *Survival analysis a self-learning text*. Springer.

Klindt, D., Ecker, A. S., Euler, T., & Bethge, M. (2017). Neural system identification for large populations separating " what" and " where". *Advances in Neural Information Processing Systems*, *30*.

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C. E., Kepecs, A., Mainen, Z. F., Qi, X.-L., Romo, R., Uchida, N., & Machens, C. K. (2016). Demixed principal component analysis of neural population data. *eLife*, *5*, e10989. https://doi.org/10.7554/eLife.10989

Kriegeskorte, N., & Wei, X.-X. (2021). Neural tuning and representational geometry. *Nature Reviews Neuroscience*, *22*(11), 703–718. https://doi.org/10.1038/s41583-021-00502-3

Kullmann, D. M. (2011). Interneuron networks in the hippocampus. *Current Opinion in Neurobiology*, *21*(5), 709–716. https://doi.org/10.1016/j.conb.2011.05.006

Laboratory, I. B., Banga, K., Benson, J., Bonacchi, N., Bruijns, S. A., Campbell, R., Chapuis, G. A., Churchland, A. K., Davatolhagh, M. F., Lee, H. D., et al. (2022). Reproducibility of in-vivo electrophysiological measurements in mice. *bioRxiv*, 2022–05.

Larisch, R., Gönner, L., Teichmann, M., & Hamker, F. H. (2021). Sensory coding and contrast invariance emerge from the control of plastic inhibition over emergent selectivity. *bioRxiv*. https://doi.org/10.1101/2020.04.07.029157

LeMasson, G., Marder, E., & Abbott, L. (1993). Activity-dependent regulation of conductances in model neurons. *Science*, *259*(5103), 1915–1917.

Levenstein, D., Alvarez, V. A., Amarasingham, A., Azab, H., Chen, Z. S., Gerkin, R. C., Hasenstaub, A., Iyer, R., Jolivet, R. B., Marzen, S., Monaco, J. D., Prinz, A. A., Quraishi, S., Santamaria, F., Shivkumar, S., Singh, M. F., Traub, R., Nadim, F., Rotstein, H. G., & Redish, A. D. (2023). On the Role of Theory and Modeling in Neuroscience. *The Journal of Neuroscience*, *43*(7), 1074–1088. https://doi.org/10.1523/JNEUROSCI.1179-22.2022

Levenstein, D., Buzsáki, G., & Rinzel, J. (2019). NREM sleep in the rodent neocortex and hippocampus reflects excitable dynamics. *Nature Communications*, *10*(1), 2478. https://doi.org/10.1038/s41467-019-10327-5

Li, X., Wong, T.-K. L., Chen, R. T., & Duvenaud, D. K. (2020). Scalable gradients and variational inference for stochastic differential equations. *Symposium on Advances in Approximate Bayesian Inference*, 1–28.

Lim, S., & Rinzel, J. (2010). Noise-induced transitions in slow wave neuronal dynamics. *Journal of Computational Neuroscience*, *28*(1), 1–17. https://doi.org/10.1007/s10827-009-0178-y

Lindner, B., & Schimansky-Geier, L. (1999). Analytical approach to the stochastic FitzHugh-Nagumo system and coherence resonance. *Physical Review E*, *60*(6), 7270–7276. https://doi.org/10.1103/PhysRevE.60.7270

Litwin-Kumar, A., & Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature Neuroscience*, *15*(11), 1498–1505. https://doi.org/10.1038/nn.3220

Litwin-Kumar, A., & Doiron, B. (2014). Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nature communications*, *5*, 5319. https://doi.org/10.1038/ncomms6319

Liu, G. (2004). Local structural balance and functional interaction of excitatory and inhibitory synapses in hippocampal dendrites. *Nature Neuroscience*, *7*(4), 373–379. https://doi.org/10.1038/nn1206

Liu, Y., Schubert, J., Sonnenberg, L., Helbig, K. L., Hoei-Hansen, C. E., Koko, M., Rannap, M., Lauxmann, S., Huq, M., Schneider, M. C., et al. (2019). Neuronal mechanisms of mutations in scn8a causing epilepsy or intellectual disability. *Brain*, *142*(2), 376–390.

Lonardoni, D., Amin, H., Marco, S. D., Maccione, A., Berdondini, L., & Nieus, T. (2017). Recurrently connected and localized neuronal communities initiate coordinated spontaneous activity in neuronal networks. *PLOS Computational Biology*, *13*(7), e1005672. https://doi.org/10.1371/journal.pcbi.1005672

Lorist, M. M., Boksem, M. A., & Ridderinkhof, K. R. (2005). Impaired cognitive control and reduced cingulate activity during mental fatigue. *Cognitive Brain Research*, *24*(2), 199–205.

Lotila, J., Hyvärinen, T., Skottman, H., Airas, L., Narkilahti, S., & Hagman, S. (2022). Establishment of a human induced pluripotent stem cell line (TAUi008-A) derived from a multiple sclerosis patient. *Stem Cell Research*, *63*, 102865. https://doi.org/10.1016/j.scr.2022.102865

Lueckmann, J.-M., Boelts, J., Greenberg, D. S., Gonçalves, P. J., & Macke, J. H. (2021). Benchmarking Simulation-Based Inference. *arXiv:2101.04653*.

Lueckmann, J.-M., Goncalves, P. J., Bassetto, G., Öcal, K., Nonnenmacher, M., & Macke, J. H. (2017). Flexible statistical inference for mechanistic models of neural dynamics. *Advances in neural information processing systems*, *30*.

Luz, Y., & Shamir, M. (2012). Balancing feed-forward excitation and inhibition via hebbian inhibitory synaptic plasticity. *PLOS Computational Biology*, *8*(1), 1–12. https://doi.org/10.1371/journal.pcbi.1002334

Machens, C. K. (2010). Demixing population activity in higher cortical areas. *Frontiers in computational neuroscience*, *4*, 126.

Mackwood, O., Naumann, L. B., & Sprekeler, H. (2021). Learning excitatory-inhibitory neuronal assemblies in recurrent networks. *eLife*, *10*, e59715. https://doi.org/10.7554/eLife.59715

Maeda, E., Robinson, H., & Kawana, A. (1995). The mechanisms of generation and propagation of synchronized bursting in developing networks of cortical neurons. *Journal of Neuroscience*, *15*(10), 6834–6845.

Mainen, Z. F., & Sejnowski, T. J. (1996). Influence of dendritic structure on firing pattern in model neocortical neurons. *Nature*, *382*(6589), 363–366.

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, *503*(7474), 78–84. https://doi.org/10.1038/nature12742

Marder, E., & Goaillard, J.-M. (2006). Variability, compensation and homeostasis in neuron and network function. *Nature Reviews Neuroscience*, *7*(7), 563–574.

Masquelier, T., & Deco, G. (2013). Network Bursting Dynamics in Excitatory Cortical Neuron Cultures Results from the Combination of Different Adaptive Mechanism. *PLoS ONE*, *8*(10), e75824. https://doi.org/10.1371/journal.pone.0075824

Mattia, M., & Sanchez-Vives, M. V. (2012). Exploring the spectrum of dynamical regimes and timescales in spontaneous cortical activity. *Cognitive Neurodynamics*, *6*(3), 239–250. https://doi.org/10.1007/s11571-011-9179-4

McCormick, D. A., Shu, Y., & Yu, Y. (2007). Hodgkin and huxley model—still standing? *Nature*, *445*(7123), E1–E2.

Miehl, C., & Gjorgjieva, J. (2022). Stability and learning in excitatory synapses by nonlinear inhibitory plasticity. *bioRxiv*. https://doi.org/10.1101/2022.03.28.486052

Mongillo, G., Rumpel, S., & Loewenstein, Y. (2018). Inhibitory connectivity defines the realm of excitatory plasticity. *Nature neuroscience*, *21*(10), 1463–1470.

Monosov, I. E., Haber, S. N., Leuthardt, E. C., & Jezzini, A. (2020). Anterior Cingulate Cortex and the Control of Dynamic Behavior in Primates. *Current Biology*, *30*(23), R1442–R1454. https://doi.org/10.1016/j.cub.2020.10.009

Montalà-Flaquer, M., López-León, C. F., Tornero, D., Houben, A. M., Fardet, T., Monceau, P., Bottani, S., & Soriano, J. (2022). Rich dynamics and functional organization on topographically designed neuronal networks in vitro. *Iscience*, *25*(12), 105680.

Nakamura, K., Roesch, M. R., & Olson, C. R. (2005). Neuronal Activity in Macaque SEF and ACC During Performance of Tasks Involving Conflict. *Journal of Neurophysiology*, *93*(2), 884–908. https://doi.org/10.1152/jn.00305.2004

Oesterle, J., Krämer, N., Hennig, P., & Berens, P. (2022). Probabilistic solvers enable a straightforward exploration of numerical uncertainty in neuroscience models. *Journal of Computational Neuroscience*, *50*(4), 485–503.

Okujeni, S., Kandler, S., & Egert, U. (2017). Mesoscale Architecture Shapes Initiation and Richness of Spontaneous Network Activity. *The Journal of Neuroscience*, *37*(14), 3972–3987. https://doi.org/10.1523/JNEUROSCI.2552-16.2017

Okun, M., & Lampl, I. (2008). Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature Neuroscience*, *11*(5), 535–537. https://doi.org/10.1038/nn.2105

O'Leary, T., Sutton, A. C., & Marder, E. (2015). Computational models in the age of large datasets. *Current Opinion in Neurobiology*, *32*, 87–94. https://doi.org/10.1016/j.conb.2015.01.006

O'Leary, T., van Rossum, M. C. W., & Wyllie, D. J. A. (2010). Homeostasis of intrinsic excitability in hippocampal neurones: Dynamics and mechanism of the response to chronic depolarization: Homeostatic regulation of intrinsic excitability. *The Journal of Physiology*, *588*(1), 157–170. https://doi.org/10.1113/jphysiol.2009.181024

Ori, H., Hazan, H., Marder, E., & Marom, S. (2020). Dynamic clamp constructed phase diagram for the hodgkin and huxley model of excitability. *Proceedings of the National Academy of Sciences*, *117*(7), 3575–3582.

Orlandi, J. G., Soriano, J., Alvarez-Lacalle, E., Teller, S., & Casademunt, J. (2013). Noise focusing and the emergence of coherent activity in neuronal cultures. *Nature Physics*, *9*(9), 582–590. https://doi.org/10.1038/nphys2686

Ostojic, S. (2014). Two types of asynchronous activity in networks of excitatory and inhibitory spiking neurons. *Nature Neuroscience*, *17*(4). https://doi.org/10.1038/nn.3658

Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., & D, H. K. (2016). Kilosort: Realtime spike-sorting for extracellular electrophysiology with hundreds of channels. https://doi.org/10.1101/061481

Papamakarios, G. (2019). Neural density estimation and likelihood-free inference. *arXiv preprint arXiv:1910.13233*.

Papamakarios, G., Pavlakou, T., & Murray, I. (2017). Masked autoregressive flow for density estimation. *Advances in neural information processing systems*, *30*.

Parra, P., Gulyás, A. I., & Miles, R. (1998). How many subtypes of inhibitory cells in the hippocampus? *Neuron*, *20*(5), 983–993.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., & Lerer, A. (2017). Automatic differentiation in pytorch.

Payeur, A., Guerguiev, J., Zenke, F., Richards, B. A., & Naud, R. (2021). Burst-dependent synaptic plasticity can coordinate learning in hierarchical circuits. *Nature neuroscience*, *24*(7), 1010–1019.

Pelkey, K. A., Chittajallu, R., Craig, M. T., Tricoire, L., Wester, J. C., & McBain, C. J. (2017). Hippocampal gabaergic inhibitory interneurons. *Physiological reviews*, *97*(4), 1619–1747.

Penn, Y., Segal, M., & Moses, E. (2016). Network synchronization in hippocampal neurons. *Proceedings of the National Academy of Sciences*, *113*(12), 3341–3346. https://doi.org/10.1073/pnas.1515105113

Pfister, J.-P., & Gerstner, W. (2006). Triplets of spikes in a model of spike timing-dependent plasticity. *Journal of Neuroscience*, *26*(38), 9673–9682. https://doi.org/10.1523/JNEUROSCI.1425-06.2006

Pfister, R., Schwarz, K., Janczyk, M., Dale, R., & Freeman, J. (2013). Good things peak in pairs: A note on the bimodality coefficient. *Frontiers in Psychology*, *4*. Retrieved May 10, 2023, from https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00700

Pozzorini, C., Naud, R., Mensi, S., & Gerstner, W. (2013). Temporal whitening by power-law adaptation in neocortical neurons. *Nature Neuroscience*, *16*(7), 942–948. https://doi.org/10.1038/nn.3431

Prinz, A. A., Bucher, D., & Marder, E. (2004). Similar network activity from disparate circuit parameters. *Nature Neuroscience*, *7*(12), 1345–1352. https://doi.org/10.1038/nn1352

Quadrato, G., & Arlotta, P. (2017). Present and future of modeling human brain development in 3d organoids. *Current opinion in cell biology*, *49*, 47–52.

Rabbitt, P. A. (1966). Errors and error correction in choice-response tasks. *Journal of experimental psychology*, *71*(2), 264.

Renart, A., Rocha, J. d. l., Bartho, P., Hollender, L., Parga, N., Reyes, A., & Harris, K. D. (2010). The Asynchronous State in Cortical Circuits. *Science*, *327*(5965), 587–590. https://doi.org/10.1126/science.1179850

Renault, R., Sukenik, N., Descroix, S., Malaquin, L., Viovy, J.-L., Peyrin, J.-M., Bottani, S., Monceau, P., Moses, E., & Vignes, M. (2015). Combining Microfluidics, Optogenetics and Calcium Imaging to Study Neuronal Communication In Vitro. *PLOS ONE*, *10*(4), e0120680. https://doi.org/10.1371/journal.pone.0120680

Rhoades, B. K., & Gross, G. W. (1994). Potassium and calcium channel dependence of bursting in cultured neuronal networks. *Brain Research*, *643*(1-2), 310–318. https://doi.org/10.1016/0006-8993(94)90039-6

Rienecker, K. D. A., Poston, R. G., & Saha, R. N. (2020). Merits and Limitations of Studying Neuronal Depolarization-Dependent Processes Using Elevated External Potassium. *ASN Neuro*, *12*, 175909142097480. https://doi.org/10.1177/1759091420974807

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature neuroscience*, *2*, 1019–25. https://doi.org/10.1038/14819

Robinson, H. P., Kawahara, M., Jimbo, Y., Torimitsu, K., Kuroda, Y., & Kawana, A. (1993). Periodic synchronized bursting and intracellular calcium transients elicited by low magnesium in cultured cortical neurons. *Journal of Neurophysiology*, *70*(4), 1606–1616. https://doi.org/10.1152/jn.1993.70.4.1606

Rosh, I., Tripathi, U., Hussein, Y., Rike, W. A., Manole, A., Cordeiro, D., Houlden, H., Winkler, J., Gage, F., & Stern, S. (2023). Synaptic dysfunction and dysregulation of extracellular matrix-related genes in dopaminergic neurons derived from Parkinson's disease sporadic patients and with GBA1 mutations. *bioRxiv*. https://doi.org/10.1101/2023.04.10.536143

Sahara, S., Yanagawa, Y., O'Leary, D. D. M., & Stevens, C. F. (2012). The Fraction of Cortical GABAergic Neurons Is Constant from Near the Start of Cortical Neurogenesis to Adulthood. *Journal of Neuroscience*, *32*(14), 4755–4761. https://doi.org/10.1523/JNEUROSCI.6412-11.2012

Sajad, A., Errington, S. P., & Schall, J. D. (2022). Functional architecture of executive control and associated event-related potentials in macaques. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-33942-1

Sanchez-Vives, M. V., Mattia, M., Compte, A., Perez-Zabalza, M., Winograd, M., Descalzo, V. F., & Reig, R. (2010). Inhibitory Modulation of Cortical Up States. *Journal of Neurophysiology*, *104*(3), 1314–1324. https://doi.org/10.1152/jn.00178.2010

Sanchez-Vives, M. V., Nowak, L. G., & McCormick, D. A. (2000). Cellular Mechanisms of Long-Lasting Adaptation in Visual Cortical Neurons *In Vitro*. *The Journal of Neuroscience*, *20*(11), 4286–4299. https://doi.org/10.1523/JNEUROSCI.20-11-04286.2000

Sancristóbal, B., Rebollo, B., Boada, P., Sanchez-Vives, M. V., & Garcia-Ojalvo, J. (2016). Collective stochastic coherence in recurrent neuronal networks. *Nature Physics*, *12*(9), 881–887.

Schmiester, L., Weindl, D., & Hasenauer, J. (2021). Efficient gradient-based parameter estimation for dynamic models using qualitative data. *Bioinformatics*, *37*(23), 4493–4500.

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). Nih image to imagej: 25 years of image analysis. *Nature methods*, *9*(7), 671–675.

Schneider, S., Lee, J. H., & Mathis, M. W. (2023). Learnable latent embeddings for joint behavioural and neural analysis. *Nature*, 1–9.

Schottdorf, M. (2017). *The reconstitution of visual cortical feature selectivity in vitro* (Doctoral dissertation). Georg-August-Universität Göttingen.

Seeman, S. C., Campagnola, L., Davoudian, P. A., Hoggarth, A., Hage, T. A., Bosma-Moody, A., Baker, C. A., Lee, J. H., Mihalas, S., Teeter, C., Ko, A. L., Ojemann, J. G., Gwinn, R. P., Silbergeld, D. L., Cobbs, C., Phillips, J., Lein, E., Murphy, G., Koch, C., … Jarsky, T. (2018). Sparse recurrent excitatory connectivity in the microcircuit of the adult mouse and human cortex. *eLife*, *7*, e37349. https://doi.org/10.7554/eLife.37349

Segal, M. (2005). Dendritic spines and long-term plasticity. *Nature Reviews Neuroscience*, *6*(4), 277–284.

Segal, M., Greenberger, V., & Korkotian, E. (2003). Formation of dendritic spines in cultured striatal neurons depends on excitatory afferent activity: Afferents determine dendritic spine formation. *European Journal of Neuroscience*, *17*(12), 2573–2585. https://doi.org/10.1046/j.1460-9568.2003.02696.x

Segev, R., Shapira, Y., Benveniste, M., & Ben-Jacob, E. (2001). Observations and modeling of synchronized bursting in two-dimensional neural networks. *Physical Review E*, *64*(1). https://doi.org/10.1103/PhysRevE.64.011920

Shahaf, G., & Marom, S. (2001). Learning in Networks of Cortical Neurons. *The Journal of Neuroscience*, *21*(22), 8782–8788. https://doi.org/10.1523/JNEUROSCI.21-22-08782.2001

Sheth, S. A., Mian, M. K., Patel, S. R., Asaad, W. F., Williams, Z. M., Dougherty, D. D., Bush, G., & Eskandar, E. N. (2012). Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. *Nature*, *488*(7410).

Shlens, J. (2014). A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*.

Silverman, B., & Ramsay, J. (2002). Applied functional data analysis: Methods and case studies.

Smith, E. H., Horga, G., Yates, M. J., Mikell, C. B., Banks, G. P., Pathak, Y. J., Schevon, C. A., McKhann, G. M., Hayden, B. Y., Botvinick, M. M., & Sheth, S. A. (2019). Widespread temporal coding of cognitive control in the human prefrontal cortex. *Nature Neuroscience*, *22*(11). https://doi.org/10.1038/s41593-019-0494-0

Soriano, J. (2023). Neuronal cultures: Exploring biophysics, complex systems, and medicine in a dish. *Biophysica*, *3*(1), 181–202.

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

Sprekeler, H. (2017). Functional consequences of inhibitory plasticity: Homeostasis, the excitation-inhibition balance and beyond. *Current opinion in neurobiology*, *43*, 198–203.

Stern, S., Santos, R., Marchetto, M. C., Mendes, A. P. D., Rouleau, G. A., Biesmans, S., Wang, Q.-W., Yao, J., Charnay, P., Bang, A. G., Alda, M., & Gage, F. H. (2018). Neurons derived from patients with bipolar disorder divide into intrinsically different sub-populations of neurons, predicting the patients' responsiveness to lithium. *Molecular Psychiatry*, *23*(6), 1453–1465. https://doi.org/10.1038/mp.2016.260

Sterratt, D., Graham, B., Gillies, A., & Willshaw, D. (2011). *Principles of computational modelling in neuroscience*. Cambridge University Press.

Strogatz, S. H. (2018). *Nonlinear dynamics and chaos with student solutions manual: With applications to physics, biology, chemistry, and engineering*. CRC press.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: General*, *18*, 643–662. https://doi.org/10.1037/h0054651

Sukenik* , Vinogradov*, N. O., Weinreb, E., Segal, M., Levina, A., & Moses, E. (2021). Neuronal circuits overcome imbalance in excitation and inhibition by adjusting connection numbers. *Proceedings of the National Academy of Sciences*, *118*(12). https://doi.org/10.1073/pnas.2018459118

Sunnåker, M., Busetto, A. G., Numminen, E., Corander, J., Foll, M., & Dessimoz, C. (2013). Approximate Bayesian Computation. *PLoS Computational Biology*, *9*(1), e1002803. https://doi.org/10.1371/journal.pcbi.1002803

Suresh, J., Radojicic, M., Pesce, L. L., Bhansali, A., Wang, J., Tryba, A. K., Marks, J. D., & van Drongelen, W. (2016). Network burst activity in hippocampal neuronal cultures: The role of synaptic and intrinsic currents. *Journal of Neurophysiology*, *115*(6). https://doi.org/10.1152/jn.00995.2015

Sussillo, D., Jozefowicz, R., Abbott, L., & Pandarinath, C. (2016). Lfads-latent factor analysis via dynamical systems. *arXiv preprint arXiv:1608.06315*.

Taniguchi, H., He, M., Wu, P., Kim, S., Paik, R., Sugino, K., Kvitsani, D., Fu, Y., Lu, J., Lin, Y., Miyoshi, G., Shima, Y., Fishell, G., Nelson, S. B., & Huang, Z. J. (2011). A Resource of Cre Driver Lines for Genetic Targeting of GABAergic Neurons in Cerebral Cortex. *Neuron*, *71*(6), 995–1013. https://doi.org/10.1016/j.neuron.2011.07.026

Tartaglia, E. M., & Brunel, N. (2017). Bistability and up/down state alternations in inhibition-dominated randomly connected networks of LIF neurons. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-12033-y

Taub, A. H., Katz, Y., & Lampl, I. (2013). Cortical Balance of Excitation and Inhibition Is Regulated by the Rate of Synaptic Activity. *Journal of Neuroscience*, *33*(36), 14359–14368. https://doi.org/10.1523/JNEUROSCI.1748-13.2013

Taylor, S. F., Kornblum, S., Lauber, E. J., Minoshima, S., & Koeppe, R. A. (1997). Isolation of Specific Interference Processing in the Stroop Task: PET Activation Studies. *NeuroImage*, *6*(2), 81–92. https://doi.org/10.1006/nimg.1997.0285

Tejero-Cantero, A., Boelts, J., Deistler, M., Lueckmann, J.-M., Durkan, C., Gonçalves, P. J., Greenberg, D. S., & Macke, J. H. (2020). SBI – A toolkit for simulation-based inference. https://doi.org/10.48550/arXiv.2007.09114

Tolley, N., Rodrigues, P. L., Gramfort, A., & Jones, S. R. (2023). Methods and considerations for estimating parameters in biophysically detailed neural models with simulation based inference. *bioRxiv*, 2023–04.

Transtrum, M. K., Machta, B. B., Brown, K. S., Daniels, B. C., Myers, C. R., & Sethna, J. P. (2015). Perspective: Sloppiness and emergent theories in physics, biology, and beyond. *The Journal of Chemical Physics*, *143*(1). https://doi.org/10.1063/1.4923066

Trujillo, C. A., Gao, R., Negraes, P. D., Gu, J., Buchanan, J., Preissl, S., Wang, A., Wu, W., Haddad, G. G., Chaim, I. A., Domissy, A., Vandenberghe, M., Devor, A., Yeo, G. W., Voytek, B., & Muotri, A. R. (2019). Complex Oscillatory Waves Emerging from Cortical Organoids Model Early Human Brain Network Development. *Cell Stem Cell*, *25*(4). https://doi.org/10.1016/j.stem.2019.08.002

Turrigiano, G. (2011). Too Many Cooks? Intrinsic and Synaptic Homeostatic Mechanisms in Cortical Circuit Refinement. *Annual Review of Neuroscience*, *34*(1), 89–103. https://doi.org/10.1146/annurev-neuro-060909-153238

Turrigiano, G. (2012). Homeostatic Synaptic Plasticity: Local and Global Mechanisms for Stabilizing Neuronal Function. *Cold Spring Harbor Perspectives in Biology*, *4*(1). https://doi.org/10.1101/cshperspect.a005736

Udell, M., Horn, C., Zadeh, R., Boyd, S., et al. (2016). Generalized low rank models. *Foundations and Trends® in Machine Learning*, *9*(1), 1–118.

Umbach, G. S., Tan, R. J., Jacobs, J., Pfeiffer, B. E., & Lega, B. C. (2021). Flexibility of functional neuronal assemblies supports human memory. *Nature Communications*, *13*.

Uysal, B., Löffler, H., Rosa, F., Lerche, H., & Schwarz, N. (2019). Generation of an induced pluripotent stem cell (iPSC) line (HIHDNEi003-A) from a patient with developmental and epileptic encephalopathy carrying a KCNA2 (p.Thr374Ala) mutation. *Stem Cell Research*, *40*, 101543. https://doi.org/10.1016/j.scr.2019.101543

vanPelt, J., Wolters, P., Corner, M., Rutten, W., & Ramakers, G. (2004). Long-Term Characterization of Firing Dynamics of Spontaneous Bursts in Cultured Neural Networks. *IEEE Transactions on Biomedical Engineering*, *51*(11), 2051–2062. https://doi.org/10.1109/TBME.2004.827936

Vasilev, D., Watanabe, M., Logothetis, N. K., & Totah, N. K. (2022). Focusing perceptual attention in one modality constrains subsequent learning in another modality. *Biorxiv*, 2022–01. https://doi.org/10.1101/2022.01.22.477334

Vidal, R., Ma, Y., & Sastry, S. (2016). *Generalized Principal Component Analysis* (Vol. 40). Springer. https://doi.org/10.1007/978-0-387-87811-9

Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C., & Gerstner, W. (2011). Inhibitory Plasticity Balances Excitation and Inhibition in Sensory Pathways and Memory Networks. *Science*, *334*(6062), 1569–1573. https://doi.org/10.1126/science.1211095

Vogels, T., & Abbott, L. (2009). Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nature neuroscience*, *12*, 483–91. https://doi.org/10.1038/nn.2276

Vreeswijk, C. v., & Sompolinsky, H. (1998). Chaotic Balanced State in a Model of Cortical Circuits. *Neural Computation*, *10*(6), 1321–1371. https://doi.org/10.1162/089976698300017214

Wagenaar, D. A., Madhavan, R., Pine, J., & Potter, S. M. (2005). Controlling Bursting in Cortical Cultures with Closed-Loop Multi-Electrode Stimulation. *The Journal of Neuroscience*, *25*(3), 680–688. https://doi.org/10.1523/JNEUROSCI.4209-04.2005

Wagenaar, D. A., Pine, J., & Potter, S. M. (2006). An extremely rich repertoire of bursting patterns during the development of cortical cultures. *BMC neuroscience*, *7*(1), 11. https://doi.org/10.1186/1471-2202-7-11

Wang, H.-X., Gerkin, R. C., Nauen, D. W., & Bi, G.-Q. (2005). Coactivation and timing-dependent integration of synaptic potentiation and depression. *Nature Neuroscience*, *8*(2). https://doi.org/10.1038/nn1387

Wehr, M., & Zador, A. M. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature*, *426*(6965), 442–446. https://doi.org/10.1038/nature02116

Weinreb, E., & Moses, E. (2022). Mechanistic insights into ultrasonic neurostimulation of disconnected neurons using single short pulses. *Brain stimulation*, *15*(3), 769–779.

Wildenberg, G. A., Rosen, M. R., Lundell, J., Paukner, D., Freedman, D. J., & Kasthuri, N. (2021). Primate neuronal connections are sparse in cortex as compared to mouse. *Cell Reports*, *36*(11), 109709. https://doi.org/10.1016/j.celrep.2021.109709

Williamson, R. C., Doiron, B., Smith, M. A., & Yu, B. M. (2019). Bridging large-scale neuronal recordings and large-scale network models using dimensionality reduction. *Current Opinion in Neurobiology*, *55*, 40–47. https://doi.org/10.1016/j.conb.2018.12.009

Wilson, N. R., Ty, M. T., Ingber, D. E., Sur, M., & Liu, G. (2007). Synaptic reorganization in scaled networks of controlled size. *Journal of Neuroscience*, *27*(50), 13581–13589.

Wu, G. K., Arbuckle, R., Liu, B.-h., Tao, H. W., & Zhang, L. I. (2008). Lateral Sharpening of Cortical Frequency Tuning by Approximately Balanced Inhibition. *Neuron*, *58*(1). https://doi.org/10.1016/j.neuron.2008.01.035

Wu, Y. K., Miehl, C., & Gjorgjieva, J. (2022). Regulation of circuit organization and function through inhibitory synaptic plasticity. *Trends in Neurosciences*, *45*(12), 884–898. https://doi.org/10.1016/j.tins.2022.10.006

Yamamoto, H., Spitzner, F. P., Takemuro, T., Buendía, V., Morante, C., Konno, T., Sato, S., Hirano-Iwata, A., Priesemann, V., & Muñoz, M. A. (2022). Modular architecture facilitates noise-driven control of synchrony in neuronal networks. *arXiv preprint arXiv:2205.10563*.

Yeung, N. (2013). *Conflict Monitoring and Cognitive Control*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199988709.013.0018

Yizhar, O., Fenno, L. E., Prigge, M., Schneider, F., Davidson, T. J., O'Shea, D. J., Sohal, V. S., Goshen, I., Finkelstein, J., Paz, J. T., Stehfest, K., Fudim, R., Ramakrishnan, C., Huguenard, J. R., Hegemann, P., & Deisseroth, K. (2011). Neocortical excitation/inhibition balance in information processing and social dysfunction. *Nature*, *477*(7363), 171–178. https://doi.org/10.1038/nature10360

Yu, B. M., Cunningham, J. P., Santhanam, G., Ryu, S. I., Shenoy, K. V., & Sahani, M. (2009). Gaussian-Process Factor Analysis for Low-Dimensional Single-Trial Analysis of Neural Population Activity. *Journal of Neurophysiology*, *102*(1), 614–635. https://doi.org/10.1152/jn.90941.2008

Zenke, F., Agnes, E., & Gerstner, W. (2015). Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nature Communications*, *6*, 6922. https://doi.org/10.1038/ncomms7922

Zenke, F., & Vogels, T. P. (2021). The Remarkable Robustness of Surrogate Gradient Learning for Instilling Complex Function in Spiking Neural Networks. *Neural Computation*, *33*(4), 899–925. https://doi.org/10.1162/neco_a_01367

Zhou, D., & Wei, X.-X. (2020). Learning identifiable and interpretable latent models of high-dimensional neural activity using pi-vae. *Advances in Neural Information Processing Systems*, *33*, 7234–7247.

Zierenberg, J., Wilting, J., & Priesemann, V. (2018). Homeostatic plasticity and external input shape neural network dynamics. *Physical Review X*, *8*(3). https://doi.org/10.1103/PhysRevX.8.031018

Zimnik, A. J., & Churchland, M. M. (2021). Independent generation of sequence elements by motor cortex. *Nature Neuroscience*, *24*(3), 412–424. https://doi.org/10.1038/s41593-021-00798-5