Discovering structure in multi-neuron populations through network modeling

Carsen Stringer

Dissertation submitted for the degree of

Doctor of Philosophy

in

Computational Neuroscience

University College London

2017

Declaration

I, Carsen Stringer, declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Carsen Stringer

October 23, 2017

Abstract

Our brains contain billions of neurons, which are continually producing electrical signals to relay information around the brain. Yet most of our knowledge of how the brain works comes from studying the activity of one neuron at a time. Recently, studies of multiple neurons have shown that they tend to be active together. These coordinated dynamics vary across brain states and impact the way that external sensory information is processed.

To investigate the network-level mechanisms that underlie these dynamics, we developed novel computational techniques to fit a deterministic spiking network model directly to multi-neuron recordings from different rodent species, sensory modalities, and behavioral states. We found that inhibition modulates the interactions between intrinsic dynamics and sensory inputs to control the reliability of sensory representations.

We next recorded from awake mice using calcium imaging techniques, and acquired activity from 10,000 neurons simultaneously in visual cortex while presenting 2,800 different natural images. In awake mice, these intrinsic population-wide fluctuations were suppressed and responses to visual stimuli were reliable. The stimulus-related information was stored in a high-dimensional neural space: 1,000 dimensions of neural activity accounted for 90% of the variance.

Although awake mice lacked large population-wide fluctuations in activity, we observed several dozen dimensions of spontaneous activity. These dimensions of spontaneous activity were not spatially organized in cortex. Instead they were related to the orofacial behaviors of the mouse: over 50% of the shared variability of the network could be predicted from the facial movements of the mouse.

In simulations of high-dimensional network activity, flexible patterns of activity were reproduced only if the network contained multiple dimensions of inhibitory activity. We tested this hypothesis in our recordings and found that inhibitory neuron activity did track excitatory neuron activity across multiple dimensions.

Acknowledgments

I have learned an incredible amount throughout my PhD and I'd like to thank all the people who supported me both intellectually and personally.

First, I'd like to thank my family, in particular my sister, my mother, and my aunt Tracey and uncle Benny. They could have pressured me to stay in the US for my graduate degree, and they were very supportive of my decision to go abroad. And they have listened to all my troubles halfway around the world and I'm so grateful to have their continual support.

Second, I'd like to thank Marius Pachitariu for his intellectual and personal support during my PhD. It has been an absolute blast working together these past four years. I came to London with only a vague scientific goal in mind, and by combining his machine learning skills with my applied mathematics experience, we have been able to develop interesting neural hypotheses.

I'd like to thank all of my colleagues at Gatsby for their intellectual support as we suffered through the many obstacles necessary to get where we are today. I am greatly indebted to Peter Dayan for the scientific and moral support he provided throughout my PhD. I'd also like to thank all of the members of the Carandini-Harris lab who have warmly adopted me into the lab. Everyone is incredibly nice and very knowledgeable.

Last but not least, I'd like to thank my advisors, Kenneth Harris and Matteo Carandini, for their awesome scientific guidance. They are excellent scientists and their passion for science is really inspiring as a young scientist. They've helped to make me a more critical thinker and taught me how to focus my research goals. I'd like to thank Kenneth Harris for helping me to combine theory and experiment – what I've learned will help me throughout my career and direct my future research.

Contents

List of Figures

1

Introduction

Our brains contain billions of neurons, which are continually producing electrical signals to relay information. Yet much of our knowledge of how the brain works comes from studying the activity of one neuron at a time. Single neuron patterns of activity have been shown to be diverse across brain areas and brain states [Sanchez-Vives and McCormick, 2000, Nowak et al., 2003, Hattox and Nelson, 2007]. However, inferring the workings of the brain one neuron at a time is an arduous task. A single neuron's activity is but a shadow of the highly-complex dynamics of the population activity. We instead turned to multi-neuron recordings to more directly investigate the types of neural responses produced by local networks in the brain. In this thesis, I will investigate how these population responses vary across brain states and across external stimuli, and try to understand the neural data using a variety of computational techniques, ranging from neural network simulations to dimensionality reduction approaches and behavioral quantification.

Internal brain dynamics

To understand how the brain generates behavior, perhaps we can start by studying it during sleep, when there is no behavior to produce. However, even during sleep, brains are highly active, and produce large signals that can be easily read at the surface of the skull. These signals are dominated by low frequency fluctuations, which were first observed in 1938 using electroencephalography (EEG) [Collura, 1993]. In deep sleep, brain activity synchronizes in slow waves of 0.5 - 2 Hz frequency. However, during wakefulness, the low-frequency oscillations have much

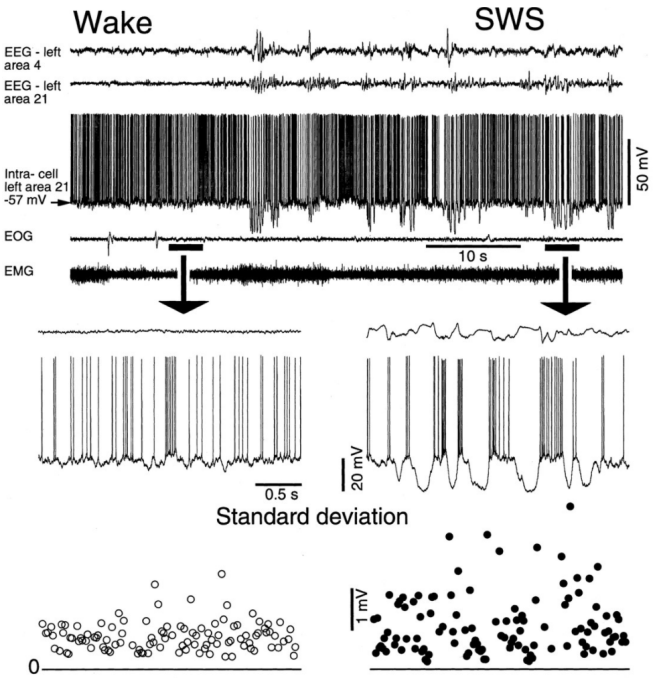

Figure 1.1: Intracellular recording during wakefulness and sleep. Figure 7 from [Steriade et al., 2001]. It depicts an in vivo intracellular recording from higher-order cortical area 21 in cat. EEG signals were acquired in area 4 (motor cortex) and area 21 simultaneously (top traces). The cell's membrane potential is more constant in the awake state than the sleep state. During sleep the membrane potential oscillates between periods of hyperpolarization and depolarization.

lower amplitudes, and instead the signal is dominated by higher frequencies. Slow oscillations can also be detected at the level of single neurons under anesthesia or during sleep, and, similar to the EEG, they are less pronounced during wakefulness [Steriade et al., 2001, Constantinople and Bruno, 2011] (Figure 1.1).

More recently, multi-neuron recordings in cortex have shown that most neurons in a population synchronize their activity at slow frequencies when an animal is under anesthesia or in deep sleep [Haider et al., 2006, Berkes et al., 2011, Luczak et al., 2009, Sakata and Harris, 2009, Pachitariu et al., 2015, Schölvinck et al., 2015, Constantinople and Bruno, 2011]. In anesthetized and sleeping states, neurons tend to be active together in short bursts called "up" states, which are followed by periods in which they are less active called "down" states. During "down" states most neurons do not fire. These "up" and "down" states are low-frequency oscillations in the range of 0.5 - 4 Hz [Pachitariu et al., 2015] (Figure 1.2).

When animals are awake, low-frequency spontaneous fluctuations are less pronounced, but are still present [Steriade et al., 2001, Luczak et al., 2009, Constantinople and Bruno, 2011, Okun et al., 2015, Schölvinck et al., 2015].

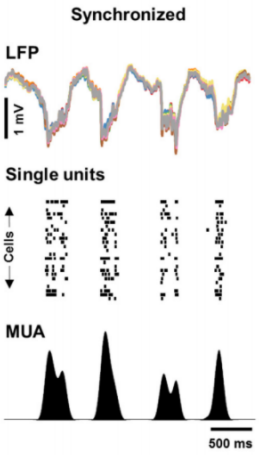

Figure 1.2: Up and down states in a gerbil under ketamine/xylaxine anesthesia. The LFP shows oscillations at 1-2 Hz in the neural activity. The neural activity raster shows highly synchronized activity on "up" states and zero activity during "down" states. Figure from [Pachitariu et al., 2015].

Also, periods in which no neurons fire ("down" states) rarely occur. Nonetheless, fluctuations in population-wide activity do persist in awake states, which has been recently emphasized by rodent and primate studies [Tan et al., 2014, Vinck et al., 2015, McGinley et al., 2015a, Engel et al., 2016].

These fluctuations in neural activity are generated by local networks of neurons, in both awake and anesthetized animals. To show this, [Shapcott et al., 2016, Schmid et al., 2013] lesioned primary visual cortex (V1) in monkeys, and looked at the activity of a higher order visual area (V4) which receives its primary input from V1. After the removal of the feedforward input (V1), V4 neurons were slightly more correlated to each other, despite having lower stimulus responses on average. This shows that the removal of feedforward inputs does not remove the correlated fluctuations among V4 neurons. In a different study, [Malina et al., 2016] optogenetically silenced cortical firing in anesthetized mice by activating inhibitory neurons in barrel cortex. This had no effect on stimulus-evoked firing rates in thalamus. However, membrane potential correlations among neurons in layers 4/5 in cortex were significantly reduced compared to normal activity. These correlations are therefore not the consequence of shared feedforward input from thalamus, but appear to be generated by the local circuits.

Behavioral states modulate neural activity

Internally-generated activity can be modulated by the behavioral state of the animal. The behavioral state of the animal may be a combination of the animal's actions and its environmental context. For example, locomotion naturally induces an aroused state in which pupil area increases [Vinck et al., 2015]. An air puff delivered by the experimenter can also induce arousal, as measured by the pupil area of the animal, but this may be an arousal state associated with fear [Vinck et al., 2015]. Task engagement or attention may induce a physiologically different state from locomotive or fear-induced arousal states. Whisking may induce yet another arousal state which is associated with exploration [Kurnikova et al., 2017]. Are there even more brain states that influence neural activity? We investigate this question in Chapter 5. First, we focus on the one-dimensional modulation of neural activity by arousal.

The arousal state induced by locomotion has been well studied recently. Locomotion increases firing rates in visual cortex, depolarizes the membrane potential of neurons, and reduces low-frequency fluctuations [Niell and Stryker, 2010, Bennett et al., 2013, Polack et al., 2013, Vinck et al., 2015]. In addition, during locomotion, responses to visual stimuli become more reliable and more easily discriminated [Erisken et al., 2014, Vinck et al., 2015, Dadarlat and Stryker, 2017]. In the upper layers of auditory cortex, low-frequency fluctuations were also reduced during running [Schneider et al., 2014]. However, unlike in visual cortex, spontaneous excitatory neuron firing rates decreased and the membrane potentials were more hyperpolarized. This decrease in excitatory firing rates may be restricted to layers 2/3 of auditory cortex, because [McGinley et al., 2015a] saw an increase in spontaneous firing rates in auditory cortex in layer 5. It would appear that in different systems firing rates go down or up during locomotion, but this is still controversial. However, it is not controversial that in all reports so far, locomotion profoundly diminished low-frequency fluctuations, reducing the prominence of "up" and "down" states.

Alertness evoked by a light air puff has also been shown to reduce low-frequency fluctuations, but spontaneous firing rates decrease instead of increase such as during running [Vinck et al., 2015]. Task-engagement alters neural activity,

reducing spontaneous firing rates in stationary animals [McGinley et al., 2015a, Otazu et al., 2009, Buran et al., 2014] (but not all studies find a decrease in membrane potential during task engagement [Sachidhanandam et al., 2013]). Attention also decreases fluctuations in population-wide activity [Mitchell et al., 2009, Cohen and Maunsell, 2009, McGinley et al., 2015a, Engel et al., 2016].

The effect of arousal may be different in active senses, such as touch. Whisker movement enhances the firing rates of some neurons in barrel cortex [de Kock and Sakmann, 2009, Gentet et al., 2010, Peron et al., 2015]. [de Kock and Sakmann, 2009] found neurons strongly modulated by whisking in layer 5A. [Gentet et al., 2012] found that whisking suppressed the firing rates of fast-spiking inhibitory neurons and somatostatin-positive (SOM+) inhibitory neurons, enhanced non-fast-spiking and non-SOM+ inhibitory neuron firing rates, and did not affect excitatory neuron firing rates. These effects are not entirely consistent with the effect of arousal in other sensory modalities. Nonetheless, consistently in all neuron classes, subthreshold membrane fluctuations decreased during whisking [Gentet et al., 2010], similar to the effect of other arousal stimuli in visual cortex and auditory cortex.

What are the biophysical mechanisms by which different behavioral states modulate cortical activity?

Biophysical mechanisms underlying different brain states

Behavioral states produce distinct changes in neural activity. What are the biophysical mechanisms underlying cortical activity in different behavioral states? One candidate mechanism for altering neural activity across states is inhibitory neuron modulation.

First we discuss how inhibitory activity naturally changes across behavioral states. Next we explore the influence of perturbing inhibition on local circuit activity. Then we discuss mechanisms by which the brain may modulate inhibition across behavioral states.

State-dependent modulation of inhibition in cortex

The inhibitory conductances observed in excitatory neurons changes depending on the animal's state. Relative to anesthetized states, wakefulness strongly increases the stimulus-evoked inhibitory conductances [Haider et al., 2013]. This strong inhibition

in awake animals quickly shunts the excitatory drive and results in sharper tuning and sparser firing than the balanced excitatory and inhibitory conductances observed under anesthesia [Wehr and Zador, 2003, Haider et al., 2013].

This may be mediated by the increased firing of local inhibitory neurons, which significantly increase their firing rates during active wakefulness, such as during locomotion or in a task [Schneider et al., 2014, Kato et al., 2013, Kuchibhotla et al., 2017]. Local putative inhibitory neurons in extrastriate (V4) cortex have been shown to increase their firing rates relative to excitatory firing rates when an animal is attending versus not [Snyder et al., 2016]. Locomotion may also modulate inhibitory firing rates. There is controversy in the literature as to whether SOM+ inhibitory neurons increase their activity during running, but several studies have found an increase in putative PV+ inhibitory neuron firing during running [Niell and Stryker, 2010, Polack et al., 2013, Schneider et al., 2014, Vinck et al., 2015, Pakan et al., 2016], consistent with our results (see Chapter 2, Figure 2.15).

Can perturbations in inhibitory neuron activity alone produce the changes in cortical activity observed during changes in behavioral state?

Changes in cortical activity from perturbations of inhibition

The effects of inhibition on neural activity have been tested directly using pharmacological and optogenetic manipulations. Pharmacologically reducing inhibition in vitro increases the strength of the correlations between excitatory neurons in a graded manner [Sippy and Yuste, 2013]. The resulting neural state without inhibition resembles "up"/"down" fluctuations. In vivo optogenetic suppression of inhibitory neurons increases low-frequency fluctuations [Zhu et al., 2015, Chen et al., 2015]. This suggests that less inhibitory activity may be present in states with more low-frequency fluctuations, which occur in less aroused or anesthetized behavioral states.

Optogenetic activation of inhibitory interneurons generally results in sharper tuning, weaker correlations, and enhanced behavioral performance [Wilson et al., 2012, Lee et al., 2012, Chen et al., 2015], suggesting that more inhibitory activity may be present in more aroused behavioral states. In summary, more inhibition in the cortical circuit suppresses spontaneous fluctuations and enhances sensory coding. Inhibition as a switch among brain states may therefore be sufficient to explain the

diversity of neural activity observed.

Mechanisms to modulate inhibitory activity

Increased inhibition of cortical areas during wakefulness may be mediated by direct GABAergic projections from histaminergic neurons in the hypothalamus [Yu et al., 2015]. These neurons project directly to sensory cortical areas and their activity is increased during periods of wakefulness in mice.

Local cortical inhibitory neurons may be influenced by other cortical areas during different behaviors. For instance, axons from motor cortex preferentially target inhibitory neurons in mouse auditory cortex [Schneider et al., 2014]. When the mouse runs, these axons become active and increase the activity of inhibitory neurons in auditory cortex.

Inhibitory neurons in sensory cortex also receive direct input from other brain structures such as thalamus. Thalamus provides strong feedforward input to inhibitory neurons [Yu et al., 2016]. This thalamic input may be modulated depending on the brain state. Thalamic nuclei which innervate barrel cortex change their firing rates depending on the behavioral context: thalamic neurons increase their firing rates during whisking compared to periods of non-whisking [Urbain et al., 2015]. Thalamic inputs to visual cortex are also modulated by behavior, specifically inputs from the lateral genicuate nucleus increase their firing during locomotion [Roth et al., 2016].

Neuromodulators also change in concentration depending on the behavioral state. Increases in acetylcholine (ACh) and norepinephrine (NE) have been observed during wakefulness and arousal [Berridge and Waterhouse, 2003, Jones, 2008], and during periods of cortical desynchronization in which low-frequency oscillations in the LFP are suppressed [Goard and Dan, 2009, Chen et al., 2015, Castro-Alamancos and Gulati, 2014].

These neuromodulators may enhance inhibitory neuron activity. Direct stimulation of the basal forebrain has been shown to produce ACh-mediated increases in the activity of fast-spiking inhibitory neurons and decreases in the variability of evoked responses in cortex [Sakata, 2016, Castro-Alamancos and Gulati, 2014, Goard and Dan, 2009]. In addition, optogenetic activation of cholinergic projections to cortex resulted in increased firing of SOM+ inhibitory neurons and

reduced slow fluctuations [Chen et al., 2015]. The release of NE in cortex through microdialysis had similar effects, increasing fast-spiking inhibitory activity and reducing spontaneous spike rates [Castro-Alamancos and Gulati, 2014]. Indeed, blocking NE receptors strengthened slow fluctuations in the membrane potential of cortical neurons [Constantinople and Bruno, 2011]. More studies are needed to tease apart the effects of different neurotransmitters on pyramidal neurons and interneurons, but much of the existing evidence suggests that they influence internal fluctuations of the network by altering the amount of inhibition in the circuit [Castro-Alamancos and Gulati, 2014, Chen et al., 2015, Sakata, 2016].

Internal dynamics interact with external stimuli

Correlations and low-frequency fluctuations are less pronounced in aroused brain states than passive or anesthetized brain states. Do these correlated fluctuations impair sensory encoding and thus their reduction during awake states is advantageous?

Correlated fluctuations in neural activity can interact with external stimuli. For example, external stimuli can trigger an early ”up” states in which many neurons respond indiscriminately to a stimulus [Luczak et al., 2009, Berkes et al., 2011, Pachitariu et al., 2015]. Thus, the tuning of neurons to stimuli is not sharp, because most neurons seem to respond to any stimulus. On the other hand, if an ”up” state happened recently and the neurons are on a ”down” state, the stimulus may fail to elicit any response at all [Pachitariu et al., 2015, Curto et al., 2009]. Response reliability is therefore also low in the presence of large internal fluctuations. Thus the reduction of fluctuations can enhance sensory reliability [Goard and Dan, 2009, Sakata, 2016], such as the reduction observed during aroused states [Vinck et al., 2015, McGinley et al., 2015a].

This suggests that enhancing inhibitory activity may improve sensory coding by abolishing low-frequency fluctuations in neural activity. [Wilson et al., 2012, Lee et al., 2012] observed enhanced tuning to sensory stimuli during optogenetic activation of inhibitory neurons. [Zhu et al., 2015] observed decreased reliability in stimulus responses during optogenetic suppression of inhibitory neurons, consistent with the hypothesis that less inhibition increases fluctuations and decreases

reliability. We investigate this hypothesis in detail in Chapter 2.

High-dimensional stimulus responses

In the absence of low-frequency fluctuations, high-dimensional responses to stimuli are observed. For instance, stimulus responses in visual cortex can have complex, non-classical receptive fields [Hubel and Wiesel, 1962, Martinez and Alonso, 2001, Vinje and Gallant, 2002]. Complex cells in visual cortex respond to an oriented bar regardless of the position of the bar in their receptive fields [Hubel and Wiesel, 1962]. It is thought that their activity is the summation of multiple simple cells with ON-OFF fields in the receptive field of the complex cell [Martinez and Alonso, 2001]. These responses are multi-dimensional and suggest that higher-order visual features are being computed in visual cortex. We examine visual responses in awake mice and quantify their structure in Chapter 3.

Thesis outline

Mechanisms underlying spontaneous fluctuations

In Chapter 2, we investigate the biophysical mechanisms underlying spontaneous fluctuations in cortical activity. As outlined above, no external inputs are necessary to generate correlated, whole-population activity [Cohen-Kashi Malina et al., 2016, Shapcott et al., 2016]. This intrinsic variability can be reproduced in biophysical spiking network models, even in the absence of external noise [van Vreeswijk and Sompolinsky, 1996, Amit and Brunel, 1997, Renart et al., 2010, Litwin-Kumar and Doiron, 2012, Wolf et al., 2014]. In addition, these networks can be designed to have interpretable parameters, but they have not yet been fit directly to multi-neuron recordings. The inability to fit the networks directly to recordings has made it difficult to identify which of the network features, if any, play an important role in vivo. To overcome this limitation, we used a novel computational approach that allowed us to fit spiking networks directly to individual multi-neuron recordings. We then explored the parameters of the networks fit to the recordings. In particular, we investigated the role of inhibitory activity in quenching low-frequency fluctuations in

network simulations. We then examined inhibitory neuron activity in vivo across different brain states with different levels of internal fluctuations.

High-dimensional cortical activity in awake mice

The remainder of the thesis beyond chapter 2 analyzes neural activity in ∼10,000 simultaneously-recorded neurons. Using calcium imaging, we developed a technique to record ∼10,000 neurons simultaneously [Pachitariu et al., 2016b]. The motion correction techniques used in the recordings are described in Appendices A and B.

During the recording, we presented 2,800 different natural scenes to awake mice (Chapter 3). We investigated the dimensionality of these visual responses, and whether the dimensionality was inherited from the images or a consequence of the neural architecture.

Spontaneous activity and multi-dimensional behaviors

Although awake mice lack large population-wide fluctuations in activity, we still hypothesized some structure in the neural activity may be present spontaneously. We developed a method for estimating the number of significant dimensions of spontaneous activity (Chapter 4).

If locomotion and pupil area are correlated with neural activity changes, then are other behaviors of the mouse also associated with neural activity? To answer this question, we developed a processing pipeline for classifying the orofacial behaviors of head-fixed mice (Chapter 5). We then predicted neural activity from these multi-dimensional orofacial behaviors and investigated the dimensionality of this relationship between activity and behavior.

Network architectures that support high-dimensional dynamics

Finally, we returned to neural network simulations to try to understand the implications of high-dimensional network activity. Simulated networks of neurons can exhibit multi-dimensional excitatory activity, either through recurrent dynamics [Litwin-Kumar and Doiron, 2012] or through feedforward activation [Mochol et al., 2015]. In the latter case, there are two possible options for inhibitory structure: (1) global, one-dimensional inhibition, and (2) structured, high-dimensional inhibition,

for example paralleling the structure of the excitatory activity patterns. We investigated these network structures and their implications for neural activity (Chapter 6). Next, we tested these hypotheses in neural recordings in which all interneurons were labelled with tdTomato. We quantified the dimensionality of the inhibitory activity and its relation to the excitatory neuron activity.

2

Inhibitory control of correlated intrinsic variability

Cortical networks exhibit intrinsic dynamics that drive coordinated, large-scale fluctuations across neuronal populations and create noise correlations that impact sensory coding. To investigate the network-level mechanisms that underlie these dynamics, we developed novel computational techniques to fit a deterministic spiking network model directly to multi-neuron recordings from different species, sensory modalities, and behavioral states. The model generated correlated variability without external noise and accurately reproduced the wide variety of activity patterns in our recordings. Analysis of the model parameters suggested that differences in noise correlations across recordings were due primarily to differences in the strength of feedback inhibition. Further analysis of our recordings confirmed that putative inhibitory neurons were indeed more active during desynchronized cortical states with weak noise correlations, such as during running. Our results demonstrate that network models with intrinsically-generated variability can accurately reproduce the activity patterns observed in multi-neuron recordings and suggest that inhibition modulates the interactions between intrinsic dynamics and sensory inputs to control the strength of noise correlations.

Introduction

[1] The patterns of cortical activity evoked by sensory stimuli provide the internal representation of the outside world that underlies perception. However, these patterns are driven not only by sensory inputs, but also by the intrinsic dynamics of the underlying cortical network. These dynamics can create correlations in the activity of neuronal populations with important consequences for coding and computation [Shadlen et al., 1996, Abbott and Dayan, 1999, Averbeck et al., 2006]. The correlations between pairs of neurons have been studied extensively [Cohen and Kohn, 2011, Ecker et al., 2010, Averbeck et al., 2006], and recent studies have demonstrated that they are driven by dynamics involving coordinated, large-scale fluctuations in the activity of many cortical neurons [Sakata and Harris, 2009, Pachitariu et al., 2015, Okun et al., 2015]. Inactivation of the cortical circuit suppresses these synchronized fluctuations at the level of the membrane potential, in both awake and anesthetized animals, suggesting that this synchronization is cortical in origin [Malina et al., 2016]. Importantly, the nature of these dynamics and the correlations that they create are dependent on the state of the underlying network; it has been shown that various factors modulate the strength of correlations, such as anaesthesia [Harris and Thiele, 2011, Schölvinck et al., 2015, Constantinople and Bruno, 2011], attention [Cohen and Maunsell, 2009, Mitchell et al., 2009, Buran et al., 2014], locomotion [Schneider et al., 2014, Erisken et al., 2014], and alertness [Vinck et al., 2015, McGinley et al., 2015a]. In light of these findings, it is critical that we develop a deeper understanding of the origin and coding consequences of correlations at the biophysical network level.

While a number of modeling studies have explored the impact of correlations on sensory coding [Shadlen et al., 1996, de la Rocha et al., 2007, Averbeck et al., 2006, Pillow et al., 2008, Ecker et al., 2011, Moreno-Bote et al., 2014], there have been few efforts to identify their biophysical origin; the standard assumption that correlations arise from common input noise [de la Rocha et al., 2007, Doiron et al., 2016, Lyamzin et al., 2015] simply pushes the correlations from spiking to the

[1] The work described in this chapter is published in [Stringer et al., 2016]. This work was done in collaboration with Marius Pachitariu and Nicholas Lesica. Marius Pachitariu, Nicholas Lesica and I designed the experiments, performed the analyses, and wrote the manuscript. The neural recordings were performed by Nicholas Lesica, Nicholas Steinmetz, Michael Okun, and Peter Bartho. Kenneth Harris, Maneesh Sahani, Nicholas Steinmetz, Michael Okun, and Peter Bartho provided comments on the manuscript.

membrane voltage without providing insight into their genesis. Models that use external noise to create correlations have been used in theoretical investigations of how network dynamics can transform correlations [Doiron et al., 2016], but no physiological source for the external noise used in these models has yet been identified. However, no external noise is needed to generate the correlated activity that is observed in vivo; in vitro experimental studies have shown that cortical networks are capable of generating large-scale fluctuations intrinsically [Sanchez-Vives et al., 2010, Sanchez-Vives and McCormick, 2000], and in vivo results suggest that the majority of cortical fluctuations arise locally [Malina et al., 2016]. If the major source of the correlations in cortical networks is, in fact, internal, then the network features that control these correlations may be different from those that control correlations in model networks with external noise.

We demonstrate here that network models with intrinsic variability are indeed capable of reproducing the activity patterns that are observed in vivo, and then proceed to use a large number of multi-neuron recordings and a model-based analysis to investigate the mechanisms that control intrinsically generated-noise correlations. For our results to provide direct insights into physiological mechanisms, we required a model with several properties: (1) the model must be able to internally generate the complex intrinsic dynamics of cortical networks, (2) it must be possible to fit the model parameters directly to spiking activity from individual multi-neuron recordings, and (3) the model must be biophysically interpretable and enable predictions that can be tested experimentally. No existing model satisfies all of these criteria; the only network models that have been fit directly to multi-neuron recordings have relied on either abstract dynamical systems [Curto et al., 2009] or probabilistic frameworks in which variability is modelled as stochastic and correlated variability arises through abstract latent variables whose origin is assumed to lie either in unspecified circuit processes [Ecker et al., 2014, Macke et al., 2011, Pachitariu et al., 2013, Pillow et al., 2008] or elsewhere in the brain [Goris et al., 2014, de la Rocha et al., 2007]. While these models are able to accurately reproduce many features of cortical activity and provide valuable summaries of the phenomenological and computational properties of cortical networks, their parameters are difficult to interpret at a biophysical level.

One alternative to these abstract stochastic models is a biophysical spiking

network, [van Vreeswijk and Sompolinsky, 1996, Amit and Brunel, 1997, Renart et al., 2010, Litwin-Kumar and Doiron, 2012, Wolf et al., 2014]. These networks can be designed to have interpretable parameters, but have not been shown to internally generate large-scale fluctuations and noise correlations of the kind routinely seen in multi-neuron recordings. Networks with structured connectivity have been shown to generate correlated activity in small groups containing less than 5% of all neurons[Litwin-Kumar and Doiron, 2012], but not in the entire network. Furthermore, large-scale neural network models have not yet been fit directly to multi-neuron recordings and, thus, their use has been limited to attempts to explain qualitative features of cortical dynamics through manual tuning of network parameters. This inability to fit the networks directly to recordings has made it difficult to identify which of these network features, if any, play an important role in vivo. To overcome this limitation, we used a novel computational approach that allowed us to fit spiking networks directly to individual multi-neuron recordings. By taking advantage of the computational power of graphics processing units (GPUs), we were able to simulate the network with millions of different parameter values for 800 seconds each to find those that best reproduced the structure of the activity in a given recording.

We developed a novel biophysical spiking network with intrinsic variability and a small number of parameters that was able to capture the apparently "doubly chaotic" structure of cortical activity [Churchland and Abbott, 2012]. Previous models with intrinsic variability have been successful in capturing both the microscopic trial-to-trial variability in spike timing and macroscopic long-timescale fluctuations in spike rate in individual neurons [van Vreeswijk and Sompolinsky, 1996, Amit and Brunel, 1997, Vogels and Abbott, 2005], but none of these models have been able to capture the coordinated, large-scale fluctuations that are shared across neurons. By combining spike-frequency adaptation with high excitatory connectivity, our network is able to generate intrinsic global fluctuations that are of variable duration, arise at random times, and do not necessarily phase-lock to external input, thus creating noise correlations in evoked responses. This correlated intrinsic variability distinguishes our model from previous rate or spiking network models [Parga and Abbott, 2007, Renart et al., 2010, Wolf et al., 2014, Doiron et al., 2016], as well as from phenomenological dynamical systems [Macke et al.,

2011, Pachitariu et al., 2013], all of which create noise correlations by injecting common noise into all neurons, an approach which, by construction, provides little insight into the biophysical mechanisms that generate the noise [Doiron et al., 2016].

To gain insight into the mechanisms that control noise correlations in vivo, we took the following approach: (1) we assembled multi-neuron recordings from different species, sensory modalities, and behavioral states to obtain a representative sample of cortical dynamics; (2) we generated activity from the network model to understand how each of its parameters controls its dynamics, and we verified that it was able to produce a variety of spike patterns that were qualitatively similar to those observed in vivo; (3) we fit the model network directly to the spontaneous activity in each of our recordings, and we verified that the spike patterns generated by the network quantitatively matched those in each recording; (4) we examined responses to sensory stimuli to determine which of the model parameters could account for the differences in noise correlations across recordings – the results of this analysis identified the strength of feedback inhibition as a key parameter and predicted that the activity of inhibitory interneurons should vary inversely with the strength of noise correlations; (5) we confirmed this prediction through additional analysis of our recordings showing that the activity of putative inhibitory neurons is increased during periods of cortical desynchronization with weak noise correlations in both awake and anesthetized animals; (6) we repeated all of the above analyses in recordings from mice during periods of locomotion to show that our results also apply to the cortical state transitions that are induced by natural behavior. Our results suggest that weak inhibition allows activity to be dominated by coordinated, large-scale fluctuations that cause the state of the network to vary over time and, thus, create variability in the responses to successive stimuli that is correlated across neurons. In contrast, when inhibition is strong, these fluctuations are suppressed and the network state remains constant over time, allowing the network to respond reliably to successive stimuli and eliminating noise correlations.

Results

Cortical networks exhibit a wide variety of intrinsic dynamics

To obtain a representative sample of cortical activity patterns, we collected multi-neuron recordings from different species (mouse, gerbil, or rat), sensory modalities (A1 or V1), and behavioral states (awake or under one of several anesthetic agents). We compiled recordings from a total of 59 multi-neuron populations across 6 unique recording types (i.e. species/modality/state combinations; see Supplementary File 1). The spontaneous activity in different recordings exhibited striking differences not only in overall activity level, but also in the spatial and temporal structure of activity patterns; while concerted, large-scale fluctuations were prominent in some recordings, they were nearly absent in others (Figure 2.1a). In general, large-scale fluctuations were weak in awake animals and strong under anesthesia, but this was not always the case (see further examples in Figure 2.4 and summary statistics for each recording in Figure 2.6).

The magnitude and frequency of the large-scale fluctuations in each recording were reflected in the autocorrelation function of the multi-unit activity (MUA, the summed spiking of all neurons in the population in 15 ms time bins). The autocorrelation function of the MUA decayed quickly to zero for recordings with weak large-scale fluctuations, but had oscillations that decayed slowly for recordings with stronger fluctuations (Figure 2.1b). The activity patterns in recordings with strong large-scale fluctuations were characterized by clear transitions between up states, where most of the population was active, and down states, where the entire population was silent. These up and down state dynamics were reflected in the distribution of the MUA across time bins; recordings with strong large-scale fluctuations had a large percentage of time bins with zero spikes (Figure 2.1c).

To summarize the statistical structure of the activity patterns in each recording, we measured four quantities. We used mean spike rate to describe the overall level of activity, mean pairwise correlations to describe the spatial structure of the activity patterns, and two different measures to describe the temporal structure of the activity patterns – the decay time of the autocorrelation function of the MUA, and the percentage of MUA time bins with zero spikes. While there were some dependencies in the values of these quantities across different recordings (Figure 2.1d), there was

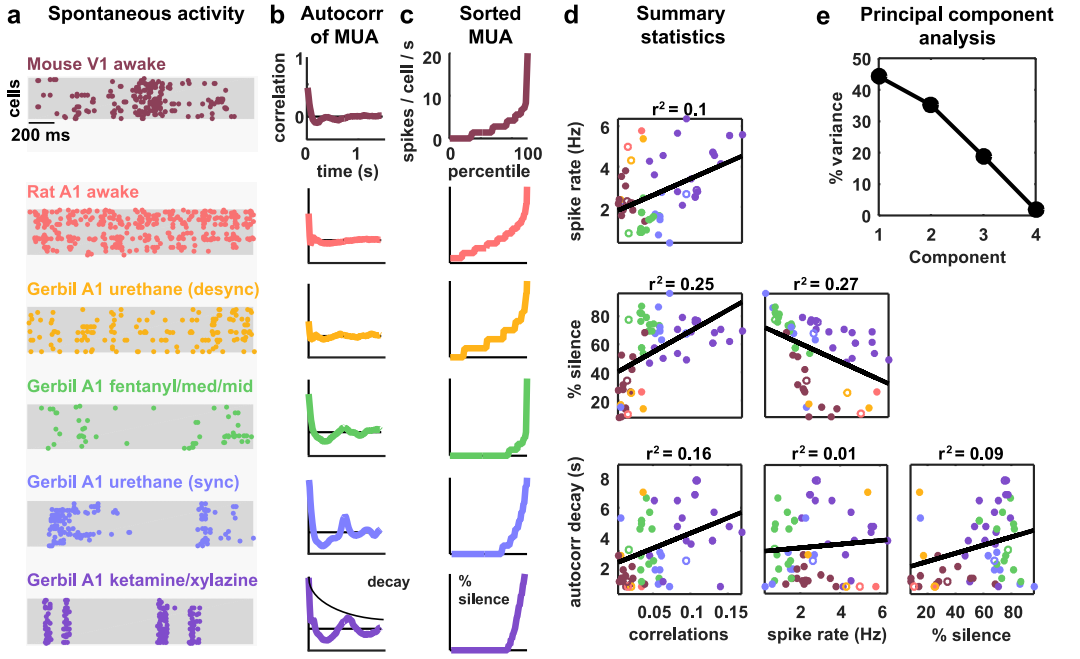

Figure 2.1: Cortical networks exhibit a wide variety of intrinsic dynamics. **(a)** Multi-neuron raster plots showing examples of a short segment of spontaneous activity from each of our recording types. Each row in each plot represents the spiking of one single unit. Note that recordings made under urethane were separated into two different recording types, synchronized ('sync') and desynchronized ('desync'), as described in the Methods. **(b)** The autocorrelation function of the multi-unit activity (MUA, the summed spiking of all neurons in the population in 15 ms time bins) for each example recording. The timescale of the autocorrelation function (the 'autocorr decay') was measured by fitting an exponential function to its envelope as indicated. **(c)** The values of the MUA across time bins sorted in ascending order. The percentage of time bins with zero spikes (the '% silence') is indicated. **(d)** Scatter plots showing all possible pairwise combinations of the summary statistics for each recording. Each point represents the values for one recording. Colors correspond to recording types as in A. The recordings shown in A are denoted by open circles. The best fit line and the fraction of the variance that it explained are indicated on each plot. **(e)** The percent of the variance in the summary statistics across recordings that is explained by each principal component of the values.

also considerable scatter both within and across recording types. This scatter suggests that there is no single dimension in the space of cortical dynamics along which the overall level of activity and the spatial and temporal structure of the activity patterns all covary, but rather that cortical dynamics span a multi-dimensional continuum [Harris and Thiele, 2011]. This was confirmed by principal component analysis; even in the already reduced space described by our summary statistics, three principal components were required to account for the differences in spike patterns across recordings (Figure 2.1e).

A deterministic spiking network model of cortical activity

To investigate the network-level mechanisms that control cortical dynamics, we developed a biophysically-interpretable model that was capable of reproducing the wide range of activity patterns observed in vivo. We constructed a minimal deterministic network of excitatory spiking integrate-and-fire neurons with non-selective feedback inhibition and single-neuron adaptation currents (Figure 2.2a). Each neuron receives constant tonic input, and the neurons are connected randomly and sparsely with 5% probability. The neurons are also coupled indirectly through global, supralinear inhibitory feedback driven by the spiking of the entire network [Rubin et al., 2015], reflecting the near-complete interconnectivity between pyramidal neurons and interneurons in local populations [Hofer et al., 2011, Fino and Yuste, 2011, Packer and Yuste, 2011]. The supralinearity of the inhibitory feedback is a critical feature of the network, as it shifts the balance of excitation and inhibition in favor of inhibition when the network is strongly driven, as has been observed in awake animals [Haider et al., 2013].

The model has five free parameters: three controlling the average strength of excitatory connectivity, the strength of inhibitory feedback, and the strength of adaptation, respectively, and two controlling the strength of the tonic input to each neuron, which is chosen from an exponential distribution. The timescales that control the decay of the excitatory, inhibitory and adaptation currents are fixed at 5 ms, 3.75 ms and 375 ms, respectively. (These timescales have been chosen based on the physiologically known timescales of AMPA, $GABA_A$, and the calcium-dependent afterhyperpolarizing current. We also verified that the qualitative nature of our results did not change when we included slow conductances or clustered connectivity; see Figure 2.3).

Note that no external noise input is required to generate variable activity; population-wide fluctuations over hundreds of milliseconds are generated when the slow adaptation currents synchronize across neurons to maintain a similar state of adaptation throughout the entire network, which, in turn, results in coordinated spiking [Latham et al., 2000, Destexhe, 2009]. The variability in the model arises through chaotic amplification of small changes in initial conditions or small perturbations to the network that cause independent simulations to diverge. In some parameter regimes, the instability of the network is such that the structure of the

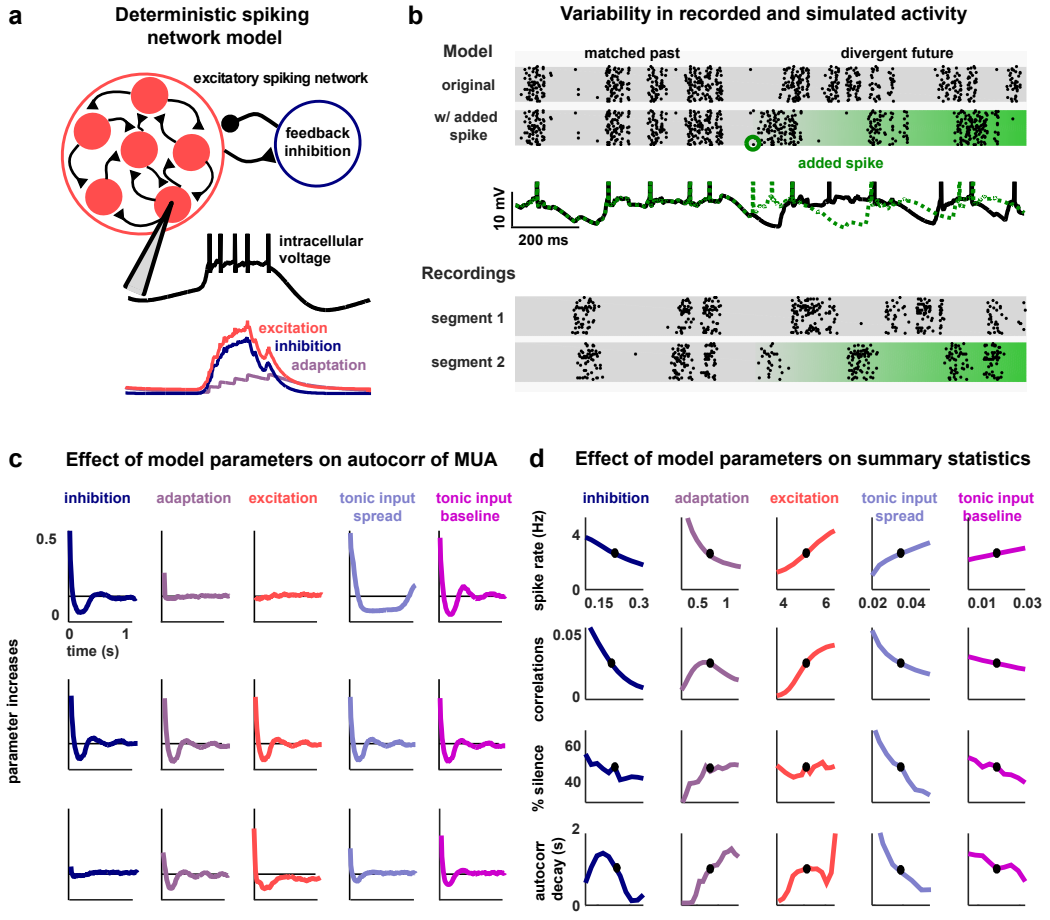

Figure 2.2: A deterministic spiking network model of cortical activity. **(a)** A schematic diagram of our deterministic spiking network model. An example of a short segment of the intracellular voltage of a model neuron is also shown, along with the corresponding excitatory, inhibitory and adaptation currents. **(b)** An example of macroscopic variability in cortical recordings and network simulations. The top two multi-neuron raster plots show spontaneous activity generated by the model. By adding a very small perturbation, in this case one spike added to a single neuron, the subsequent activity patterns of the network can change dramatically. The middle traces show the intracellular voltage of the model neuron to which the spike was added. The bottom two raster plots show a similar phenomenon observed in vivo. Two segments of activity extracted from different periods during the same recording were similar for three seconds, but then immediately diverged. **(c)** The autocorrelation function of the MUA measured from network simulations with different model parameter values. Each column shows the changes in the autocorrelation function as the value of one model parameter is changed while all others are held fixed. The fixed values used were $w_I = 0.22$, $w_A = 0.80$, $w_E = 4.50$, $b_1 = 0.03$, $b_0 = 0.013$. **(d)** The summary statistics measured from network simulations with different model parameter values. Each line shows the changes in the indicated summary statistic as one model parameter is changed while all others are held fixed. Fixed values were as in **c**.

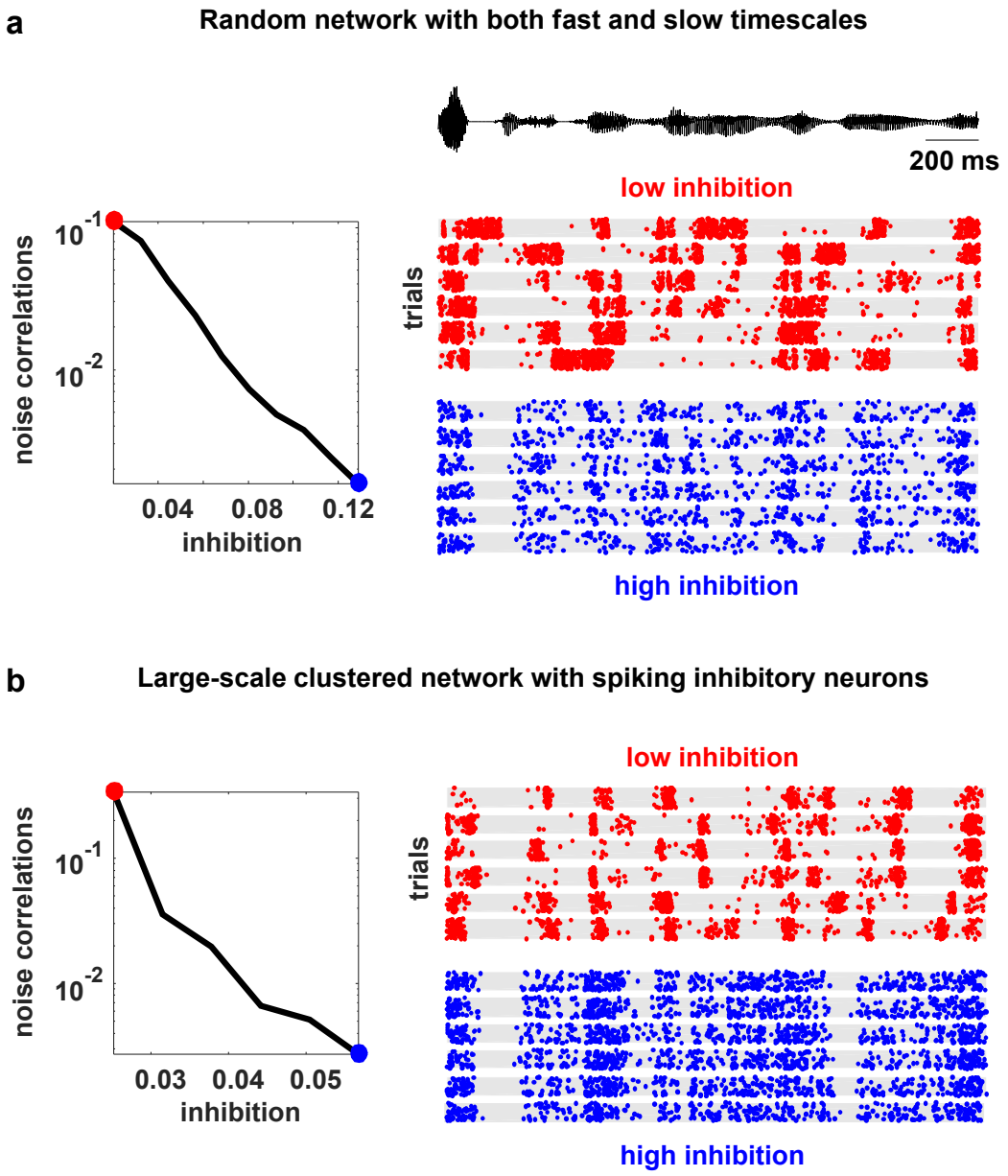

Figure 2.3: Model networks with long timescales and structured architecture. The noise correlations are plotted as a function of the strength of feedback inhibition. Both the multi-timescale network and the clustered network produced deterministic intrinsic correlated variability in response to the IC-derived external input (as in Figure 2.10). This variability was quenched as inhibitory feedback was increased. The details of the networks are given in the Materials and methods.

spike patterns generated by the model is sensitive to changes in the spike times of individual neurons. In fact, a single spike added randomly to a single neuron during simulated activity is capable of changing the time course of large-scale fluctuations, in some cases triggering immediate population-wide spiking (Figure 2.2b, top rows). Similar phenomena have been observed in vivo previously [London et al., 2010] and were also evident in our recordings when comparing different extracts of cortical activity; spike patterns that were similar for several seconds often then began to diverge almost immediately (Figure 2.2b, bottom rows).

Multiple features of the network model can control its dynamics

The dynamical regime of the network model is determined by the interactions between its different features. To determine the degree to which each feature of the network was capable of influencing the structure of its activity patterns, we analyzed the effects of varying the value of each model parameter. We started from a fixed set of parameter values and simulated activity while independently sweeping each parameter across a wide range of values. The results of these parameter sweeps clearly demonstrate that each of the five parameters can exert strong control over the dynamics of the network, as both the overall level of activity and the spatial and temporal structure of the patterns in simulated activity varied widely with changes in each parameter (Figure 2.2c-d).

With the set of fixed parameter values used for the parameter sweeps, the network is in a regime with slow, ongoing fluctuations between up and down states. In this regime, the amplification of a small perturbation results in a sustained, prolonged burst of activity (up state), which, in turn, drives a build-up of adaptation currents that ultimately silences the network for hundreds of milliseconds (down state) until the cycle repeats. These fluctuations can be suppressed by an increase in the strength of feedback inhibition, which eliminates slow fluctuations and shifts the network into a regime with weak, tonic spiking and weak correlations (Figure 2.2c-d, first column); in this regime, small perturbations are immediately offset by the strong inhibition and activity is returned to baseline. Strong inhibition also offsets externally-induced perturbations in balanced networks [Renart et al., 2010], but in our model such perturbations are internally-generated and would result in runaway excitation in the absence of inhibitory stabilization. The fluctuations between up and

down states can also be suppressed by decreasing adaptation (Figure 2.2c-d, second column); without adaptation currents to create slow, synchronous fluctuations across the network, neurons exhibit strong, tonic spiking.

The dynamics of the network can also be influenced by changes in the strength of the recurrent excitation or tonic input. Increasing the strength of excitation results in increased activity and stronger fluctuations, as inhibition is unable to compensate for the increased amplification of small perturbations (Figure 2.2c-d, third column). Increasing the spread or baseline level of tonic input also results in increased activity, but with suppression, rather than enhancement, of slow fluctuations (Figure 2.2c-d, fourth and fifth column). As either the spread or baseline level of tonic input is increased, more neurons begin to receive tonic input that is sufficient to overcome their adaptation current and, thus, begin to quickly reinitiate up states after only brief down states and, eventually, transition to tonic spiking.

The network model reproduces the dynamics observed in vivo

The network simulations demonstrate that each of its features is capable of controlling its dynamics and shaping the structure of its activity patterns. To gain insight into the mechanisms that may be responsible for creating the differences in dynamics observed in vivo, we fit the model to each of our recordings. We optimized the model parameters so that the patterns of activity generated by the network matched those observed in spontaneous activity (Figure 2.4a). We measured the agreement between the simulated and recorded activity by a cost function which was the sum of discrepancies in the autocorrelation function of the MUA, the distribution of MUA values across time bins, and the mean pairwise correlations. Together, these statistics describe the overall level of activity in each recording, as well as the spatial and temporal structure of its activity patterns.

Fitting the model to the recordings required us to develop new computational techniques. The network parametrization is fundamentally nonlinear, and the statistics used in the cost function are themselves nonlinear functions of a dynamical system with discontinuous integrate-and-fire mechanisms. Thus, as no gradient information was available to guide the optimization, we used Monte Carlo simulations to generate activity and measure the relevant statistics with different parameter values. By using GPU computing resources, we were able to design and

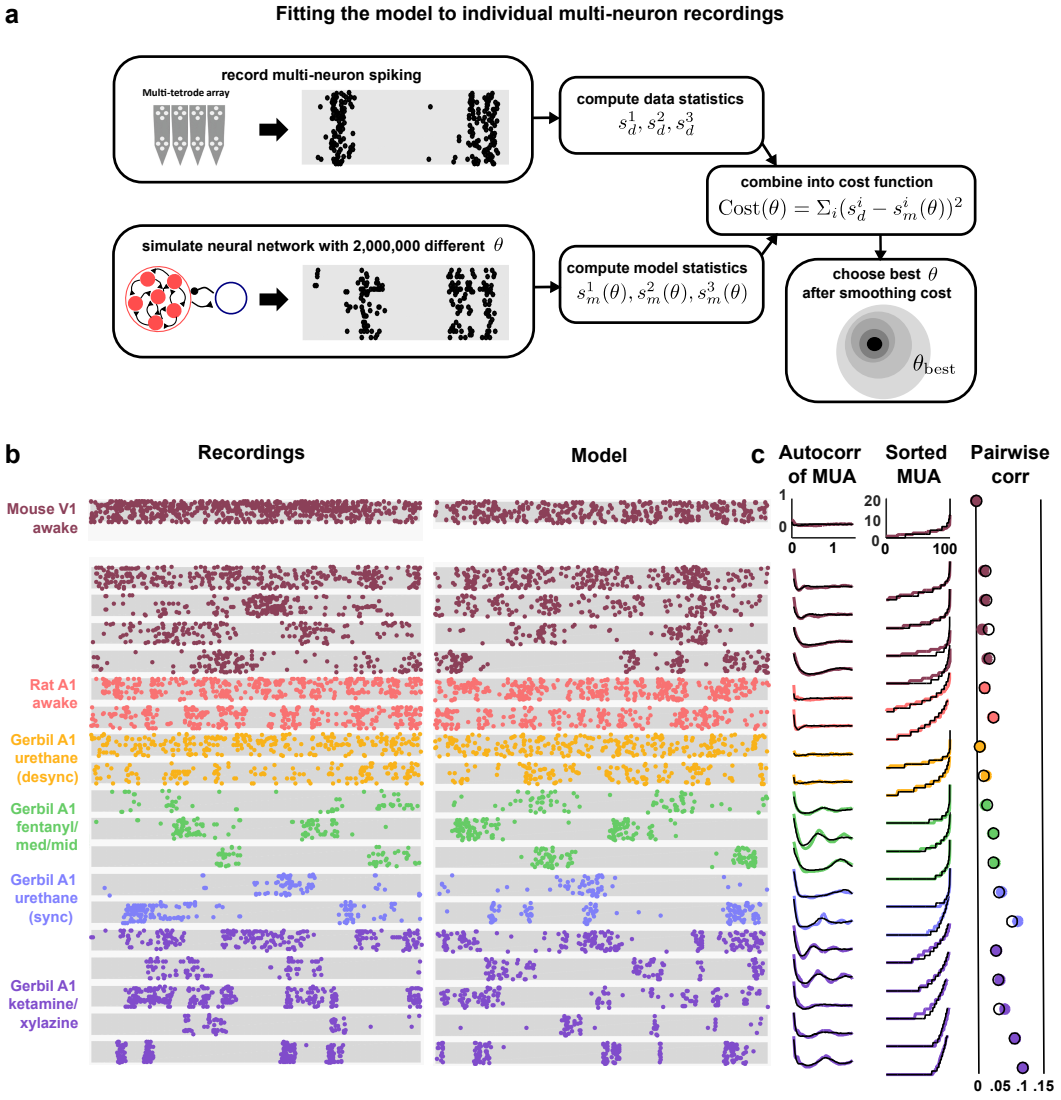

Figure 2.4: Deterministic spiking networks reproduce the dynamics observed in vivo. (a) A schematic diagram illustrating how the parameters of the network model were fit to individual multi-neuron recordings. **(b)** Examples of spontaneous activity from different recordings, along with spontaneous activity generated by the model fit to each recording. **(c)** The left column shows the autocorrelation function of the MUA for each recording, plotted as in Figure 2.1. The black lines show the autocorrelation function measured from spontaneous activity generated by the model fit to each recording. The middle column shows the sorted MUA for each recording along with the corresponding model fit. The right column shows the mean pairwise correlations between the spiking activity of all pairs of neurons in each recording (after binning activity in 15 ms bins). The colored circles show the correlations measured from the recordings and the black open circles show the correlations measured from from spontaneous activity generated by the model fit to each recording.

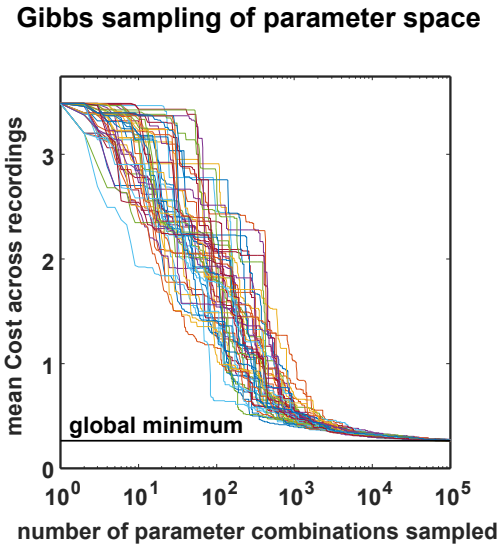

Figure 2.5: Optimization performance of the MCMC procedure. The total cost over all recordings is plotted as a function of sample number for 50 different optimizations, started with different random seeds. All the optimizations found near-global minima for all datasets with less than 100,000 samples. To efficiently conduct this analysis, we restricted the optimizations to the already fully-sampled grid, but the optimization procedure would in general allow sampling any parameter value by using a continuous instead of a discrete proposal distribution.

implement network simulations that ran 10000x faster than real time, making it feasible to sample the cost function with high resolution and locate its global minimum to identify the parameter configuration that resulted in activity patterns that best matched those of each recording. We also verified that the global minimum of the cost function could be identified with 10x fewer samples of simulated activity using a Gibbs sampling optimizer with simulated annealing (Figure 2.5), but the results presented below are based on the global minima identified by the complete sampling of parameter space.

The model was flexible enough to capture the wide variety of activity patterns observed across our recordings, producing both decorrelated, tonic spiking and coordinated, large-scale fluctuations between up and down states as needed (see examples in Figure 2.4b, statistics for all recordings and models in 2.6, and parameter values and goodness-of-fit measures for all recordings in Figure 2.7). The fits were also quantitatively accurate. We found that the median variance explained by the model of the autocorrelation function of the MUA, the distribution of MUA values

across time bins, and the mean pairwise correlations were 82%, 90%, and 97% respectively (Figure 2.8a). In fact, these fits were about as good as possible given the length of our recordings: the fraction of the variance in the statistics of one half of each recording that was explained by the statistics of the other half of the recording were 84%, 98%, and 100% respectively (Figure 2.8b). Because we used a cost function that captured many different properties of the recorded activity while fitting only a very small number of model parameters, the risk of network degeneracies was relatively low [Gutierrez et al., 2013, Marder et al., 2015]. Nonetheless, we also confirmed that analysis of model parameters corresponding to local minima of the cost function did not lead to a different interpretation of our results (see Figure 2.9).

Strong inhibition suppresses noise correlations

Our main interest was in understanding how the different network-level mechanisms that are capable of controlling intrinsic dynamics contribute to the correlated variability in responses evoked by sensory stimuli. The wide variety of intrinsic dynamics in our recordings was reflected in the differences in evoked responses across recording types; while some recordings contained strong, reliable responses to the onset of a stimulus, other recordings contained responses that were highly variable across trials (Figure 2.10a). There were also large differences in the extent to which the variability in evoked responses was correlated across the neurons in each recording; pairwise noise correlations were large in some recordings and extremely weak in others, even when firing rates were similar (Figure 2.10b).

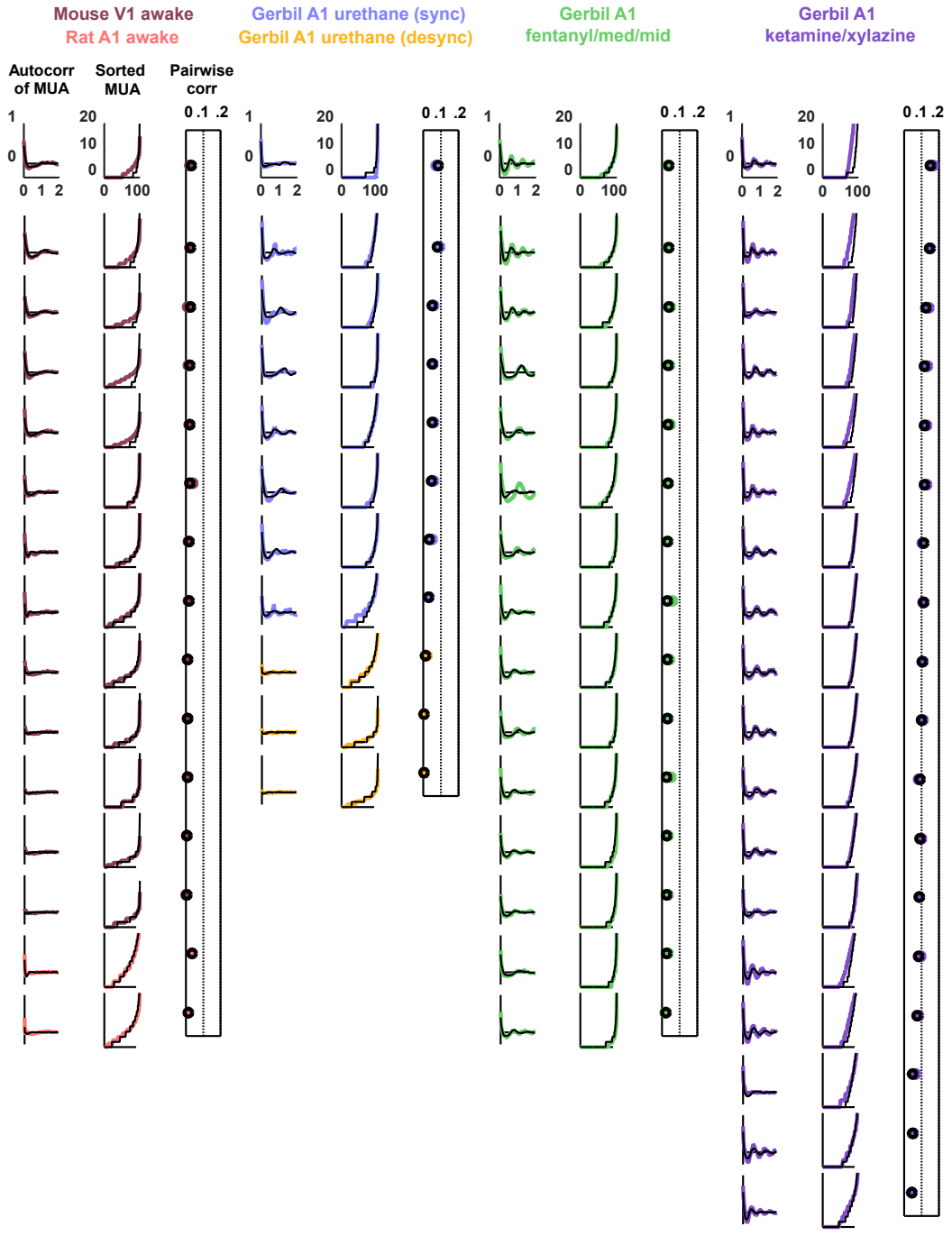

Figure 2.6: Statistics across all recordings. We fit the model to three statistics: (1) the autocorrelation function of the multi-unit activity (MUA, the summed spiking of all neurons in the population in 15 ms time bins), (2) the values of the MUA across time bins sorted in ascending order, and (3) the mean pairwise correlations across all pairs of neurons (in 15 ms time bins). The statistics for all 59 recordings are shown here in color. The model was fit to each of these recordings and the statistics of the activity generated by the model are shown in black.

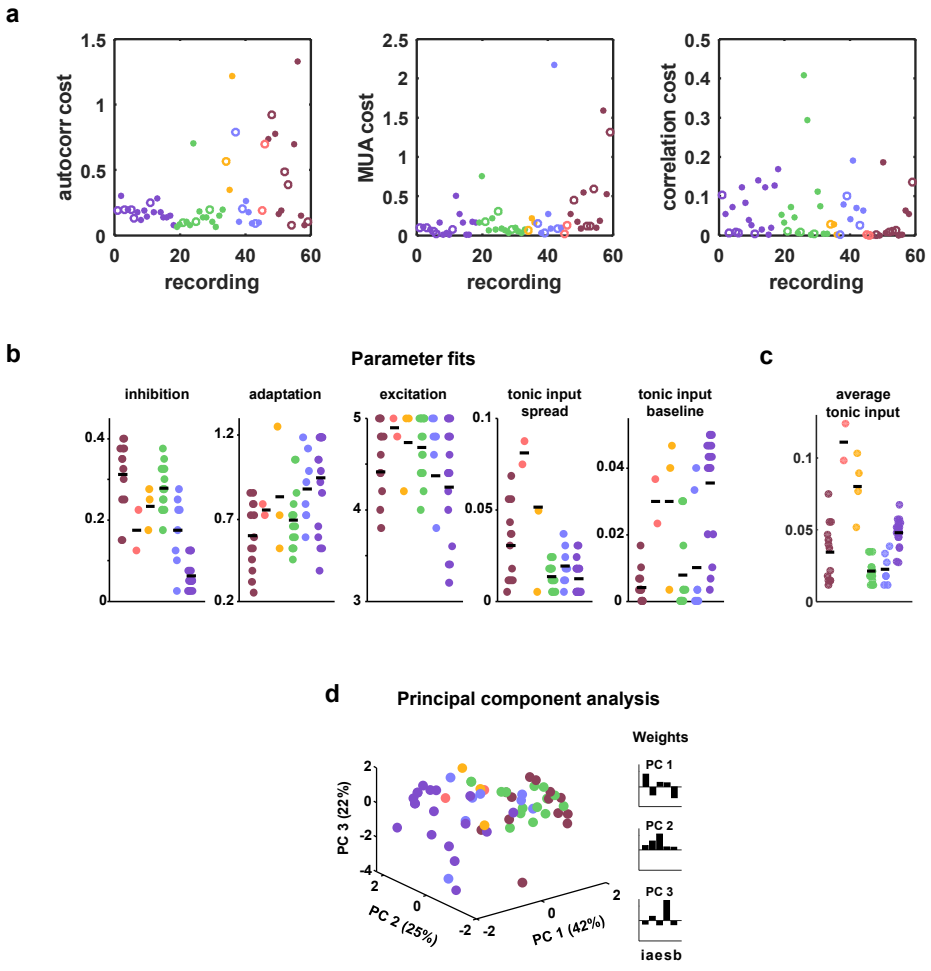

Figure 2.7: Costs and parameter values (a) The three separate terms that combine into our cost function are shown for each recording. Open circles indicate datasets shown in Figure 2.4. **(b)** All parameter fits for each recording. Colors are used to indicate recordings of the same type. A small jitter was added to the horizontal location of each point. Black lines indicate median values for each recording type. **(c)** The average tonic input in the network fit to each recording, computed as the sum of the baseline tonic input and the mean of the exponential distribution from which the random component of the tonic input for each neuron was drawn. To compare urethane desync and urethane sync parameter values, we used a Wilcoxon rank-sum test, and found the significance of the average tonic input in desync versus sync was $p < 10^{-2}$. The significance for adaptation, inhibition, excitation, and tonic input spread were $p = 0.491$; $p = 0.236$; $p = 0.349$, and $p = 0.27$, respectively. **(d)** A 3-dimensional principal component analysis shows that recordings generally cluster by type, but there is considerable variability both across and within recording types. The inset shows the 5-dimensional PCs.

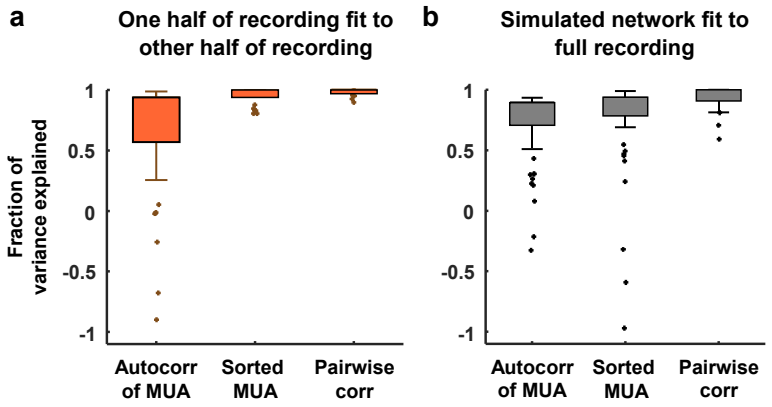

Figure 2.8: Variance explained by model fits. **(a)** Each neural recording was split into two halves (interleaved segments of 4 s each) and the autocorrelation function of the MUA, the distribution of MUA values across time bins, and the mean pairwise correlations were computed for each half. The fraction of the variance in the statistics of one half of each recording that was explained by the statistics of the other half of the recording is shown. The median variance explained for the autocorrelation function of the MUA, the distribution of MUA values across time bins, and the mean pairwise correlations were 84%, 98%, and 100% respectively. **(b)** The amount of variance in the statistics of each full recording that was explained by the model fit is shown. The median variance explained for the autocorrelation function of the MUA, the distribution of MUA values across time bins, and the mean pairwise correlations were 82%, 90%, and 97% respectively.

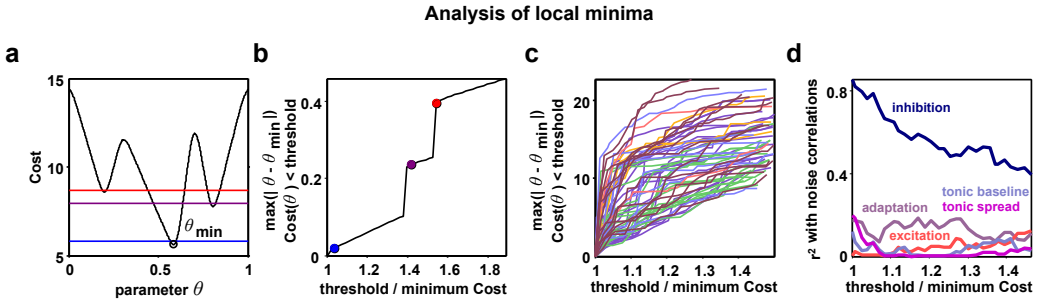

Figure 2.9: Analysis of local minima. Parameter identifiability has recently been raised as a potential problem in interpreting the results of network simulations. To mitigate this problem, we designed our model to have a very small number of parameters and we fit three different functions of the recordings, two of which varied as a function of time or rank. To confirm that the analysis of parameter combinations other than those corresponding to the global minimum of the cost function for each recording would not lead to a different interpretation of our results, we also considered local minima in regions of parameter space that were distant from the global minimum. It is possible that such local minima correspond to parameter regimes that are qualitatively different from the global minimum, yet still capture the statistics of the recordings relatively well. We found the parameters corresponding to local minima did not consistently emphasize the role of any parameter other than inhibition; the strength of inhibitory feedback remained the dominant influence on noise correlations, even for local minima far removed from the global minimum. **(a)** A schematic diagram showing an example nonlinear cost function. Several different threshold values are indicated by the colored lines. All costs below threshold are considered and the parameter q furthest away from the global minimum is chosen to plot in panel b. **(b)** As the threshold value is increased from the global minimum, the distance of the q with Cost (q) $<$ threshold that is furthest from the global minimum is plotted. Discontinuities are visible when the threshold surpasses values at local minima. **(c)** Same as **b**, but for the actual model fits to each recording. The values on the vertical axis are specified in terms of the grid spacing used for the Monte Carlo simulations. While some discontinuities are visible, the functions tend to increase gradually. **(d)** For each threshold value, we computed the r^2 between the value of each of the five model parameters and noise correlations, as in Figure 2.12a. This analysis shows that considering local minima situated far from the global minimum serves only to diminish the relationship between inhibition and noise correlations, without revealing any strong relationships between noise correlations and any other parameter.

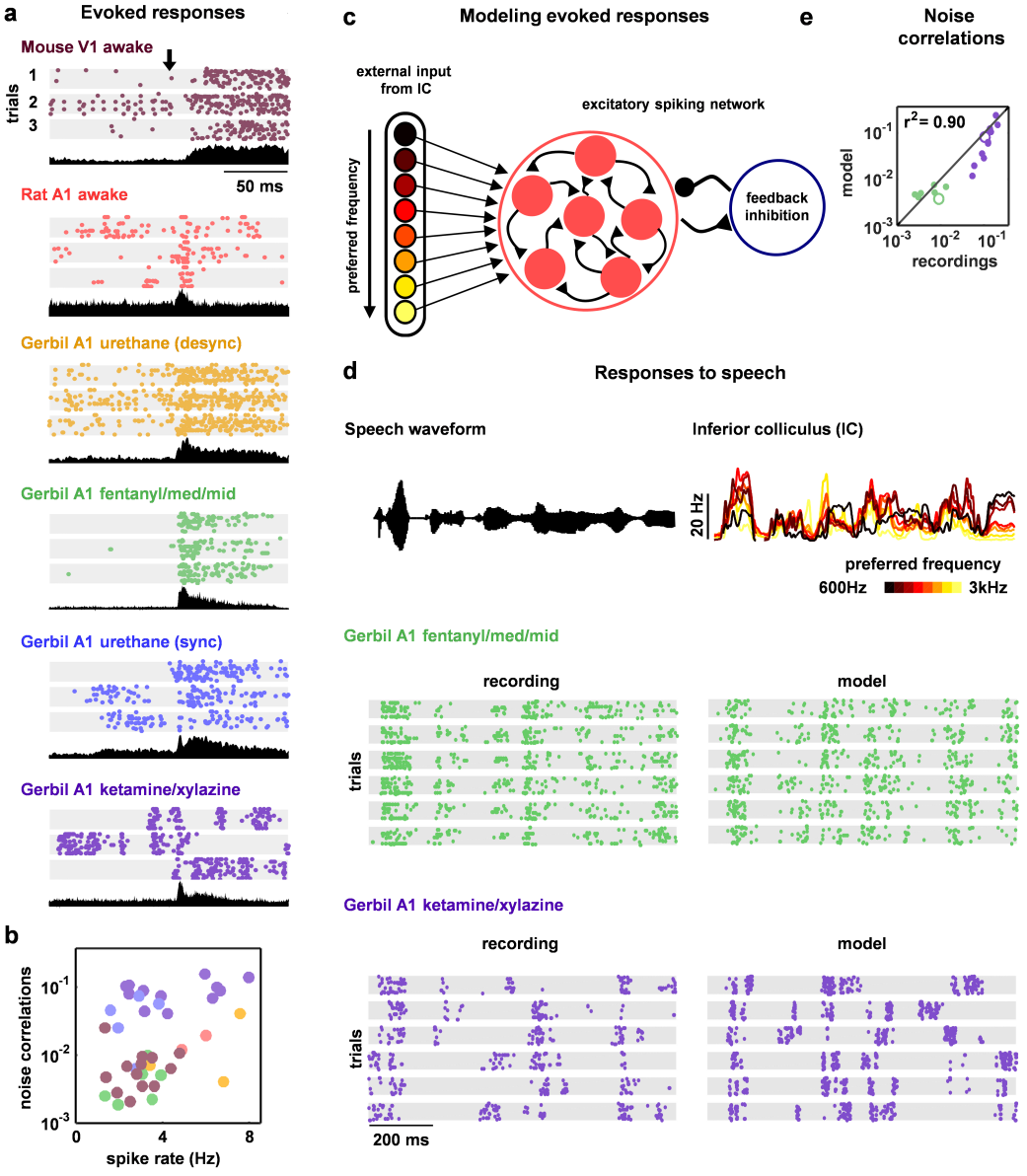

Figure 2.10: Deterministic spiking networks reproduce the noise correlations observed in vivo.

Figure 2.10: Deterministic spiking networks reproduce the noise correlations observed in vivo. (a) Multi-neuron raster plots and PSTHs showing examples of evoked responses from each of our recording types. Each row in each raster plot represents the spiking of one single unit. Each raster plot for each recording type shows the response on a single trial. The PSTH shows the MUA averaged across all presentations of the stimulus. Different stimuli were used for different recording types (see Methods). **(b)** A scatter plot showing the mean spike rates and mean pairwise noise correlations (after binning the evoked responses in 15 ms bins) for each recording. Each point represents the values for one recording. Colors correspond to recording types as in **a**. The recordings shown in **a** are denoted by open circles. Values are only shown for the 38 of 59 recordings that contained both spontaneous activity and evoked responses. **(c)** A schematic diagram illustrating the modelling of evoked responses. We constructed the external input using recordings of responses from more than 500 neurons in the inferior colliculus (IC), the primary relay nucleus of the auditory midbrain that provides the main input to the thalamocortical circuit. We have shown previously that the Fano factors of the responses of IC neurons are close to 1 and the noise correlations between neurons are extremely weak [Garcia-Lazaro et al., 2013], suggesting that the spiking activity of a population of IC neurons can be well described by series of independent, inhomogeneous Poisson processes. To generate the responses of each model network to the external input, we averaged the activity of each IC neuron across trials, grouped the IC neurons by their preferred frequency, and selected a randomly chosen subset of 10 neurons from the same frequency group to drive each cortical neuron. **(d)** The top left plot shows the sound waveform presented in the IC recordings used as input to the model cortical network. The top right plot shows PSTHs formed by averaging IC responses across trials and across all IC neurons in each preferred frequency group. The raster plots show the recorded responses of two cortical populations on successive trials, along with the activity generated by the network model fit to each recording when driven by IC responses to the same sounds. **(e)** A scatter plot showing the noise correlations of responses measured from the actual recordings and from simulations of the network model fit to each recording when driven by IC responses to the same sounds.

Because evoked spike patterns can depend strongly on the specifics of the sensory stimulus, we could not make direct comparisons between experimental responses across different species and modalities; our goal was to identify the internal mechanisms that are responsible for the differences in noise correlations across recordings and, thus, any differences in spike patterns due to differences in external input would confound our analysis. To overcome this confound and enable the comparison of noise correlations across recording types, we simulated the response of the network to the same external input for all recordings. We constructed the external input using recordings of spiking activity from the inferior colliculus (IC), a primary relay nucleus in the subcortical auditory pathway (Figure 2.10c-d). Using the subset of our cortical recordings in which we presented the same sounds that were also presented during the IC recordings, we verified that the noise correlations in the simulated cortical responses were similar to those in the recordings (Figure 2.10e).

The parameter sweeps described in Figure 2.2 demonstrated that there are

multiple features of the model network that can control its intrinsic dynamics, and a similar analysis of the noise correlations in simulated responses to external input produced similar results (Figure 2.11). To gain insight into which of these features could account for the differences in noise correlations across our recordings, we examined the dependence of the strength of the noise correlations in each recording on each of the model parameters. While several parameters were able to explain a significant amount of the variance in noise correlations across recordings, the amount of variance explained by the strength of inhibitory feedback was by far the largest (Figure 2.12a). The predominance of inhibition in the control of noise correlations was confirmed by the measurement of partial correlations (the correlation between the noise correlations and each parameter that remains after factoring out the influence of the other parameters; partial r^2 for inhibition: 0.67, excitation: 0.02, adaptation: 0.08, tonic input spread: 0.17, and tonic input baseline: 0.04). We also performed parameter sweeps to confirm that varying only the strength of inhibition was sufficient to result in large changes in noise correlations in the parameter regime of each recording (Figure 2.12b).

Strong inhibition sharpens tuning and enables accurate decoding

We also examined how different features of the network controlled other aspects of evoked responses. We began by examining the extent to which differences in the value of each model parameter could explain differences in stimulus selectivity across recordings. To estimate selectivity, we drove the model network that was fit to each cortical recording with external inputs constructed from IC responses to tones, and used the simulated responses to measure the width of the frequency tuning curves of each model neuron. Although each model network received the same external inputs, the selectivity of the neurons in the different networks varied widely. The average tuning width of the neurons in each network varied most strongly with the strength of the inhibitory feedback in the network (Figure 2.12c; partial r^2 for inhibition: 0.74, excitation: 0.06, adaptation: 0.48, tonic input spread: 0.01, and tonic input baseline: 0.37), and varying the strength of inhibition alone was sufficient to drive large changes in tuning width (Figure 2.12d). These results are consistent with experiments demonstrating that inhibition can control the selectivity of cortical neurons [Lee et al., 2012], but suggest that this control does not require structured

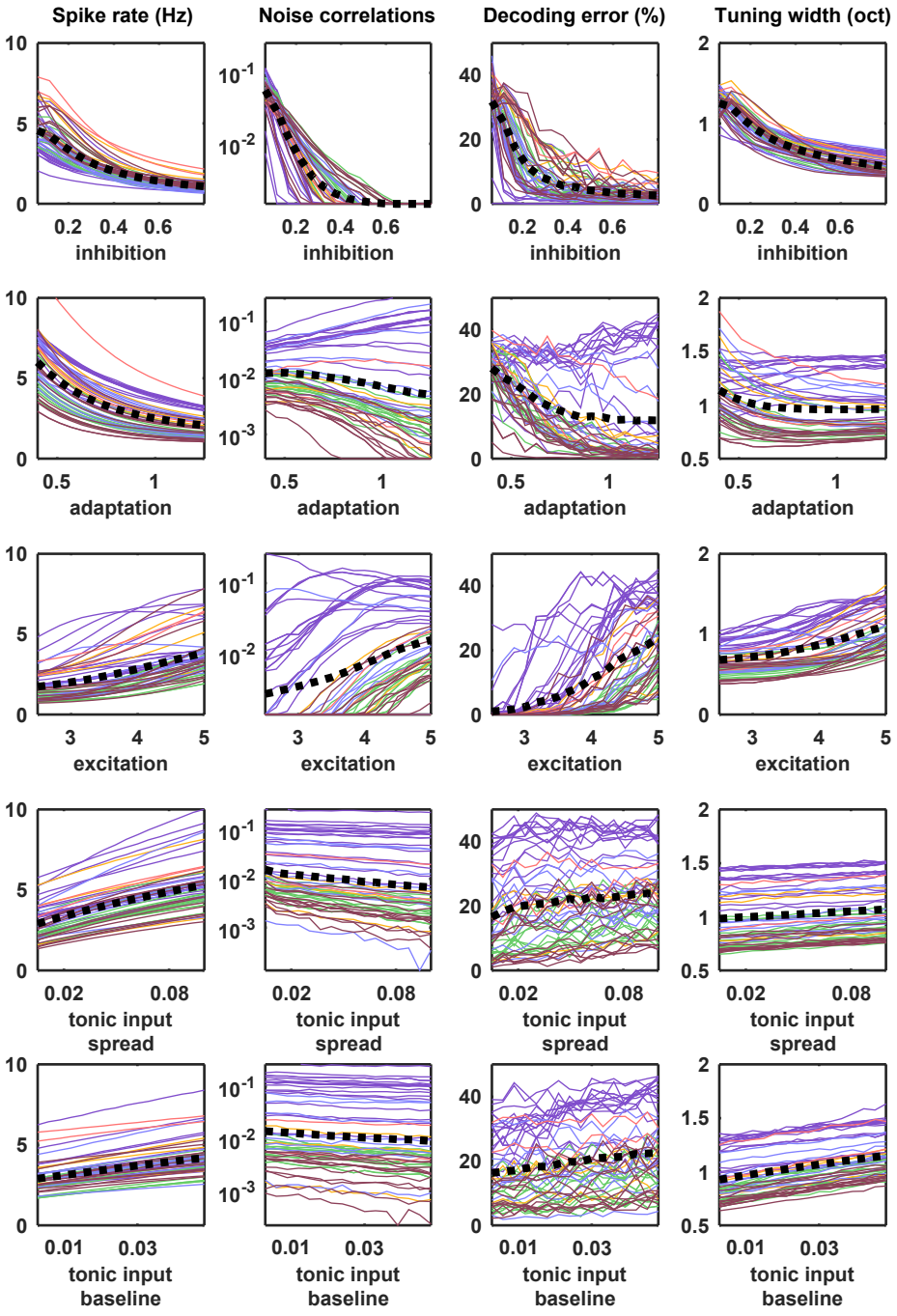

Figure 2.11: Parameter sweeps for responses to external input. Parameter sweeps were performed for each recording and each parameter. Each line corresponds to the model fit to one of the 59 recordings. All parameters were fixed at the values fit to each recording, except for the parameter indicated on the horizontal axis, which was swept across a wide range. Activity was generated from the model with these parameters and driven by the IC-derived external input. The spike rate, noise correlations, tuning width and percent decoding error were computed as described for Figure 2.12.

lateral inhibition.

We also investigated the degree to which the activity patterns generated by the model fit to each cortical recording could be used to discriminate different external inputs. We trained a decoder to infer which of seven possible stimuli evoked a given single-trial activity pattern and examined the extent to which differences in the value of each model parameter could account for the differences in decoder performance across recordings. Again, the amount of variance explained by the strength of inhibitory feedback was by far the largest (Figure 2.12e; partial r^2 for inhibition: 0.5, excitation: 0.16, adaptation 0.27, tonic input spread 0.02, and tonic input baseline 0.03); decoding was most accurate for activity patterns generated by networks with strong inhibition, consistent with the weak noise correlations and high selectivity of these networks. Parameter sweeps confirmed that varying only the strength of inhibition was sufficient to result in large changes in decoder performance (Figure 2.12f).

Activity of fast-spiking (FS) neurons is increased during periods of cortical desynchronization with weak noise correlations

Our model-based analyses suggest an important role for feedback inhibition in controlling the way in which responses to sensory inputs are shaped by intrinsic dynamics. In particular, our results predict that inhibition should be strong in dynamical regimes with weak noise correlations. To test this prediction, we performed further analysis of our recordings to estimate the strength of inhibition in each recorded population. We classified the neurons in each recording based on the width of their spike waveforms (Figure 2.13).

The waveforms for all recording types fell into two distinct clusters, allowing us to separate fast-spiking (FS) neurons from regular-spiking (RS) neurons. In general, more than 90% of FS cortical neurons have been reported to be parvalbumin-positive (PV+) inhibitory neurons [Nowak et al., 2003, Kawaguchi and Kubota, 1997, Barthó et al., 2004, Cho et al., 2010, Madisen et al., 2012, Stark et al., 2013, Cohen and Mizrahi, 2015], and this value approaches 100% in the deep cortical layers where we recorded [Cardin et al., 2009]. While the separation of putative inhibitory and excitatory neurons based on spike waveforms is imperfect (nearly all FS neurons are inhibitory, but a small fraction (less than 20%) of RS neurons are also inhibitory

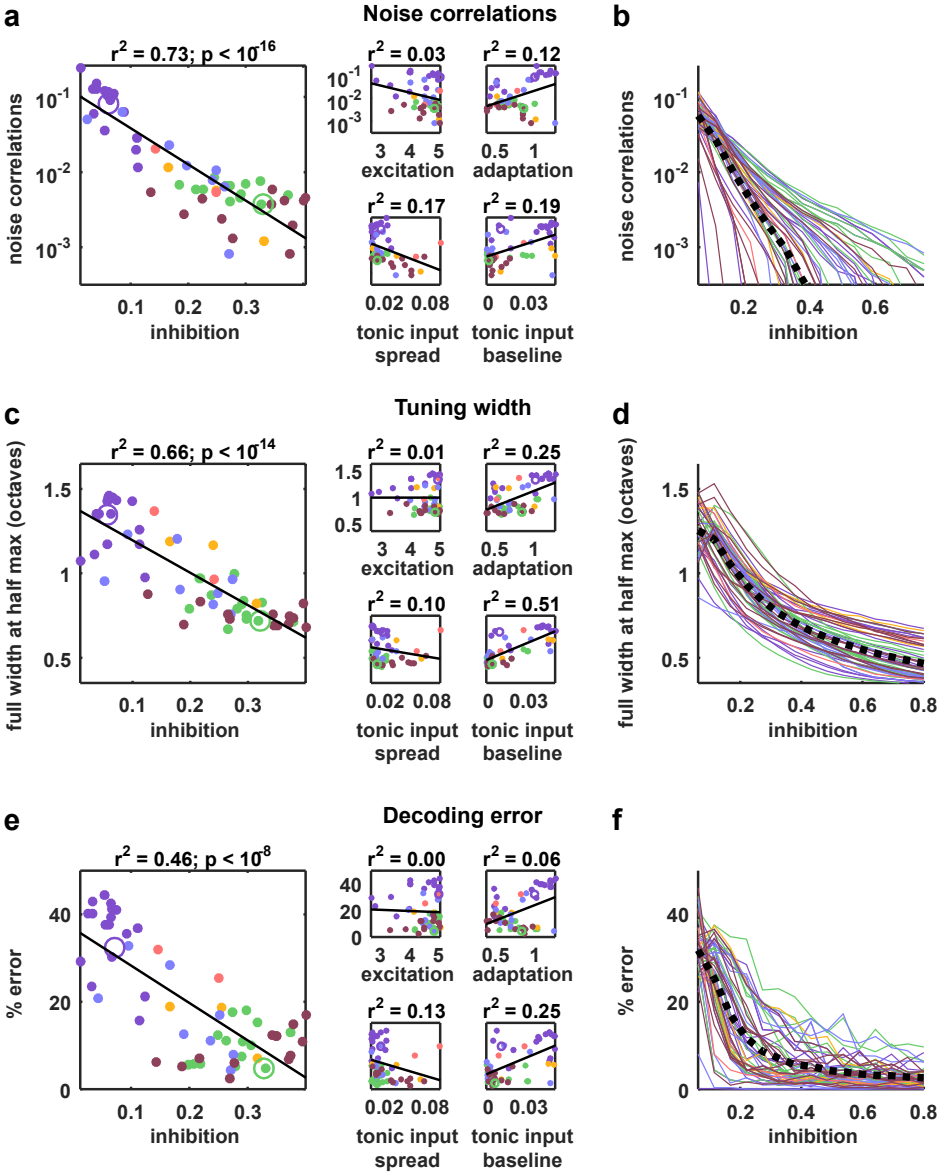

Figure 2.12: Strong inhibition suppresses noise correlations and enhances selectivity and decoding. **(a)** Scatter plots showing the mean pairwise noise correlations measured from simulations of the network model fit to each recording when driven by external input versus the value of the different model parameters. Colors correspond to recording types as in Figure 2.10. The recordings shown in Figure 2.10D are denoted by open circles. **(b)** The mean pairwise noise correlations measured from network simulations with different values of the inhibition parameter w_I. The values of all other parameters were held fixed at those fit to each recording. Each line corresponds to one recording. Colors correspond to recording types as in Figure 2.10. **(c,e)** Scatter plots showing tuning width and decoding error, plotted as in **a**. **(d,f)** The tuning width and decoding error measured from network simulations with different values of the inhibition parameter w_I, plotted as in **b**.

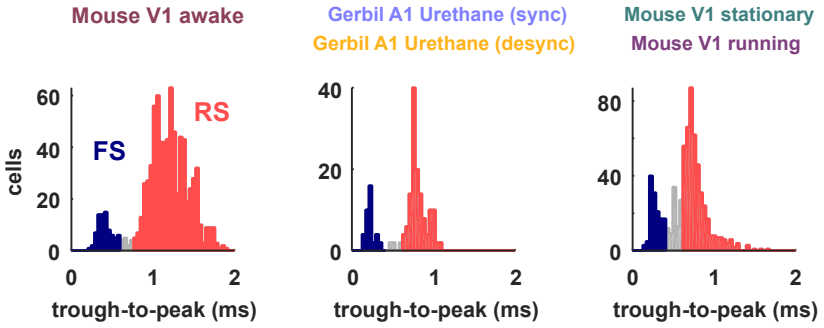

Figure 2.13: Classification of neuron types by spike width. The spike widths of the classified spike waveforms are plotted as histograms for the different recording types. The neurons classified as FS based on their width are colored in red and the RS in blue. These were the neurons used to compute FS and RS spike rates in Figures 2.14 and 2.15. Neurons that were not clearly part of either class, which were not included in the analyses in Figures 2.14 and 2.15, are shown in gray.

[Markram et al., 2004]), it is still effective for approximating the overall levels of inhibitory and excitatory activity in a population.

Given the results of our model-based analyses, we hypothesized that the overall level of activity of FS neurons should vary inversely with the strength of noise correlations. To identify sets of trials in each recording that were likely to have either strong or weak noise correlations, we measured the level of cortical synchronization. Previous studies have shown that noise correlations are strong when the cortex is in a synchronized state, where activity is dominated by concerted, large-scale fluctuations, and weak when the cortex is in a desynchronized state, where these fluctuations are suppressed [Pachitariu et al., 2015, Schölvinck et al., 2015].

We began by analyzing our recordings from V1 of awake mice. We classified the cortical state during each stimulus presentation based on the ratio of low-frequency LFP power to high-frequency LFP power [Sakata and Harris, 2012] and compared evoked responses across the most synchronized and desynchronized subsets of trials (Figure 2.14a). As expected, noise correlations were generally stronger during synchronized trials than during desynchronized trials, and this variation in noise correlations with cortical synchrony was evident both within individual recordings and across animals (Figure 2.14b-c). As predicted by our model-based analyses, the change in noise correlations with cortical synchrony was accompanied by a change in FS activity; there was a four-fold increase in the mean spike rate of FS neurons from the most synchronized trials to the most desynchronized trials, while RS activity

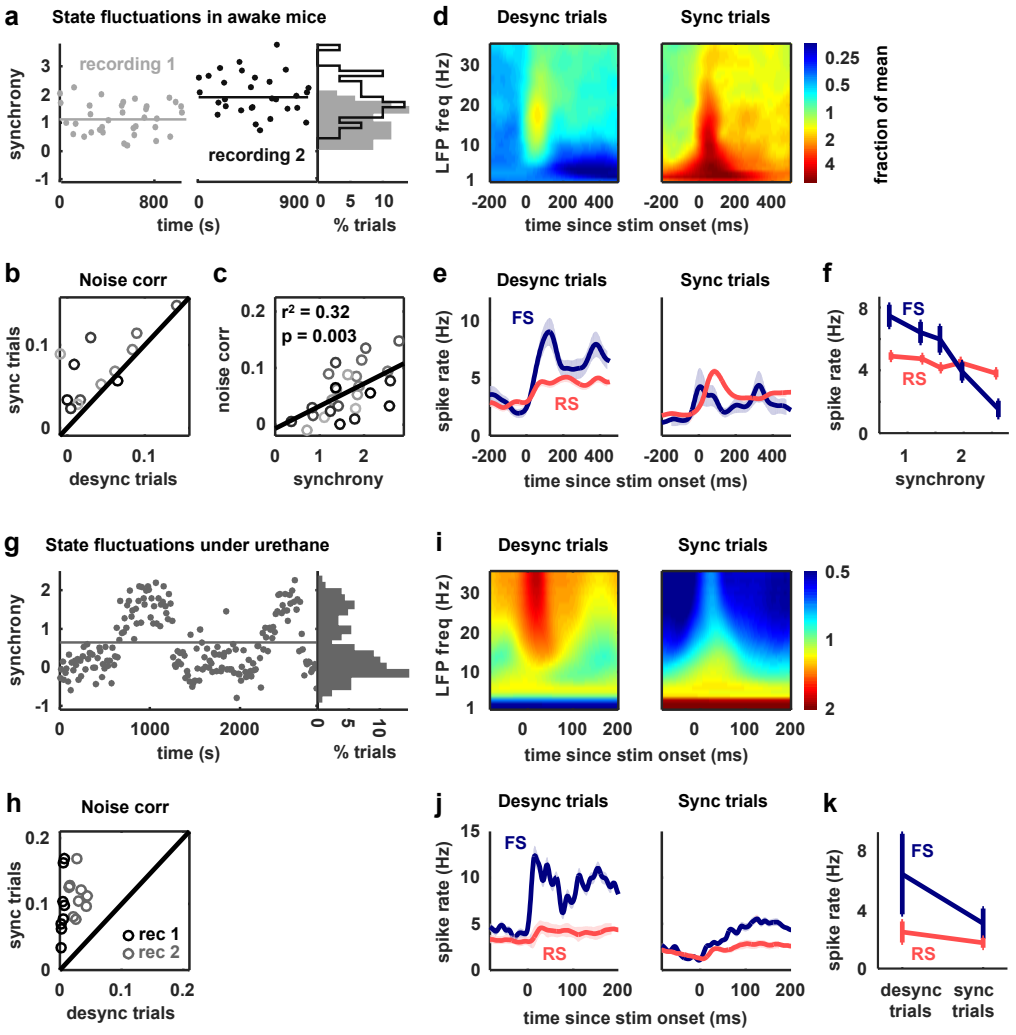

Figure 2.14: Fast-spiking neurons are more active during periods of cortical desynchronization with weak noise correlations.

remained constant (Figure 2.14d-f).

Figure 2.14: Fast-spiking neurons are more active during periods of cortical desynchronization with weak noise correlations. **(a)** The cortical synchrony at different points during two recordings from V1 of awake mice, measured as the log of the ratio of low-frequency (3 -10 Hz) LFP power to high-frequency (11 - 96 Hz). The distribution of synchrony values across each recording is also shown. The lines indicate the median of each distribution. **(b)** A scatter plot showing the noise correlations measured during trials in which the cortex was in either a relatively synchronized (sync) or desynchronized (desync) state for each recording. Each point indicates the mean pairwise correlations between the spiking activity of all pairs of neurons in one recording (after binning the activity in 15 ms bins). Trials with the highest 50% of synchrony values were classified as sync and trials with the lowest 50% of synchrony values were classified as desync. Values for 13 different recordings are shown. **(c)** A scatter plot showing noise correlations versus the mean synchrony for trials with the highest and lowest 50% of synchrony values for each recording. Colors indicate different recordings. **(d)** Spectrograms showing the average LFP power during trials with the highest (sync) and lowest (desync) 20% of synchrony values across all recordings. The values shown are the deviation from the average spectrogram computed over all trials. **(e)** The average PSTHs of FS and RS neurons measured from evoked responses during trials with the highest (sync) and lowest (desync) 20% of synchrony values across all recordings. The lines show the mean across all neurons, and the error bars indicate +/-1 SEM. **(f)** The average spike rate of FS and RS neurons during the period from 0 to 500 ms following stimulus onset, averaged across trials in each synchrony quintile. The lines show the mean across all neurons, and the error bars indicate +/-1 SEM. **(g)** The cortical synchrony at different points during a urethane recording, plotted as in A. The line indicates the value used to classify trials as synchronized (sync) or desynchronized (desync). **(h)** A scatter plot showing the noise correlations measured during trials in which the cortex was in either a synchronized (sync) or desynchronized (desync) state. Values for two different recordings are shown. Each point for each recording shows the noise correlations measured from responses to a different sound. **(i)** Spectrograms showing the average LFP power during synchronized and desynchronized trials, plotted as in **d**. **(j)** The average PSTHs of FS and RS neurons during synchronized and desynchronized trials, plotted as in **e**. **(k)** The average spike rate of FS and regular-spiking RS neurons during the period from 0 to 500 ms following stimulus onset during synchronized and desynchronized trials. The bars show the mean across all neurons, and the error bars indicate +/-1 SEM.

We next examined our recordings from gerbil A1 under urethane in which the cortex exhibited transitions between distinct, sustained synchronized and desynchronized states (Figure 2.14g). As in our awake recordings, cortical desynchronization under urethane was accompanied by a decrease in noise correlations and an increase in FS activity (Figures 2.14h-k). In fact, both FS and RS activity increased with cortical desynchronization under urethane, but the increase in FS activity was much larger (110% and 42%, respectively). The increase in RS activity suggests that cortical desynchronization under urethane may involve other mechanisms in addition to an increase in feedback inhibition (a comparison of the model parameters fit to desynchronized and synchronized urethane recordings (Figure 2.7) suggests that the average level of tonic input is significantly higher during desynchronization (desynchronized: 0.075+/-0.008, synchronized: 0.0195+/-0.0054, p =0.006)).

The change in cortical state that accompanies locomotion can be explained by an increase in feedback inhibition

Finally, we asked whether the same mechanisms might be used to control the changes in network dynamics that accompany transitions in behavioral state, such as those induced by locomotion. We recorded four separate populations of 100-200 neurons each, from two head-fixed mice that were allowed to run on a treadmill. We found that stationary periods were often accompanied by slow timescale population-wide fluctuations in firing (Figure 2.15a-b, top row). We fit the network model to these stationary periods, and verified that we could reproduce these dynamics (Figure 2.15a-b, top row, and statistics for all recordings and models in Figure 2.16). Running epochs were, by comparison, much more desynchronized (Figure 2.15a-b, bottom row), consistent with previous observations made with intracellular and LFP measurements [Vinck et al., 2015, Niell and Stryker, 2010, McGinley et al., 2015a, Polack et al., 2013, Bennett et al., 2013].

To determine which changes in our model best captured this state transition, we allowed either one or two parameters to change from the values fit to stationary periods. By changing two parameters, inhibition and adaptation, the model was able to reproduce the statistics of the neural population activity during running (Figure 2.15a-b, bottom row). Out of all the possible single-parameter changes, the best fits were achieved through changes in inhibition, while out of all the possible two-parameter changes, the best fits were achieved through changes in inhibition and adaptation (Figure 2.15c). In all four recordings, the model captured the change in dynamics associated with running through an increase in inhibition and a decrease in adaptation (Figure 2.15d). The changes in FS and RS activity in the recordings were consistent with such changes. Although both FS and RS populations increased their activity during running, the relative increase in FS activity was significantly larger (Figure 2.15e; on average, FS activity increased by 87% and RS activity increased by 28%). Our results suggest that the increase in RS activity during running despite increased FS activity is likely due to an accompanying decrease in adaptation.

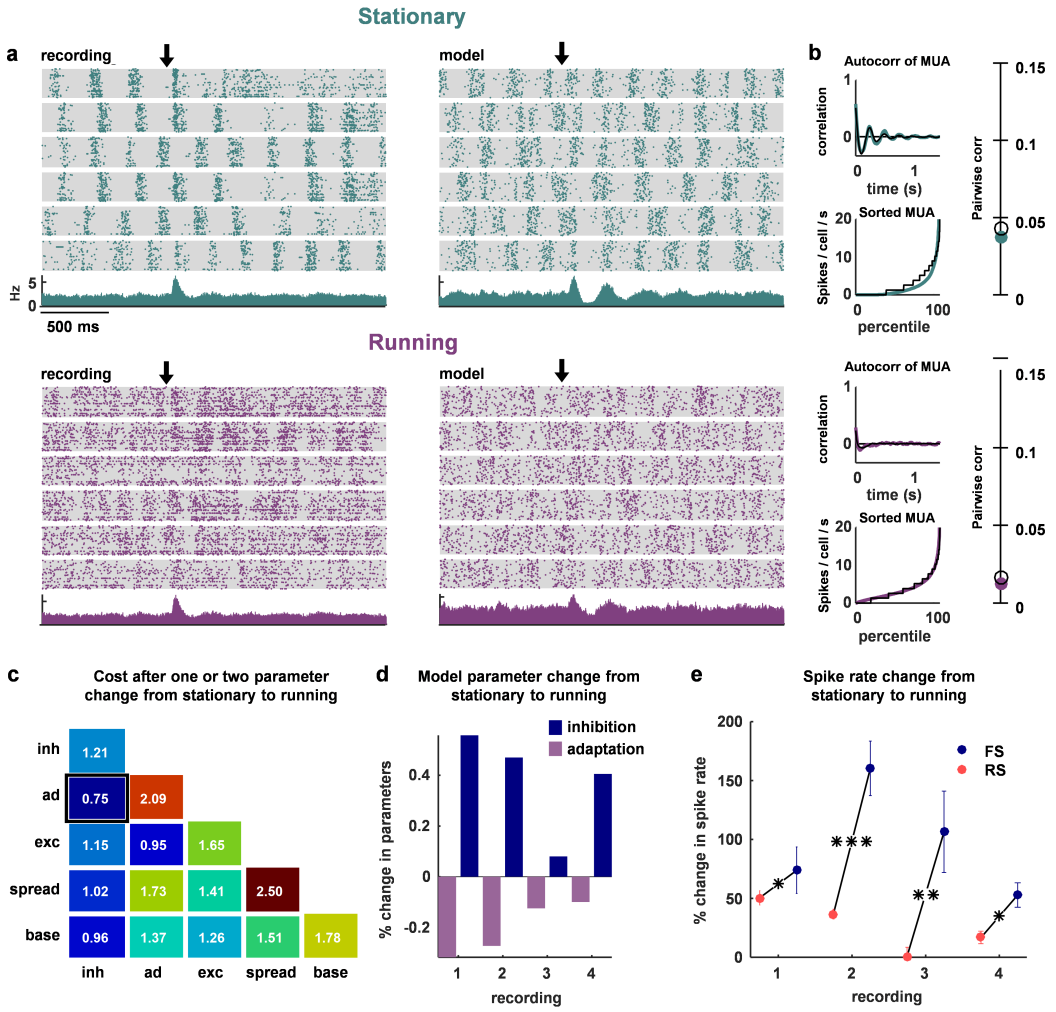

Figure 2.15: The change in dynamics during locomotion is best explained by an increase in inhibition and a reduction in adaptation. **(a)** We recorded populations of neurons in head-fixed mice that were allowed to run on a treadmill. We obtained four separate recordings from two mice, which we divided into running and stationary epochs. The raster plots and PSTHs show evoked responses recorded of one example population when the animal was stationary (top) or running (bottom), along with the activity generated by the network model fit to each set of epochs. The units for the vertical axis on the PSTH are spikes / cell / s. The arrow indicates stimulus onset. **(b)** Model and data summary statistics for stationary (top) and running (bottom) epochs for one example population, plotted as in Figure 3. The model fits shown for running epochs were achieved by allowing two parameters (inhibition and adaptation) to change from fits to stationary epochs. **(c)** We fit our network model to activity from stationary epochs and investigated which changes in either one or two parameters best captured the change in dynamics that followed the transition to running. The best achieved cost with changes in each parameter (values along diagonal), or pair of parameters (values off diagonal), is shown (lower is better). **(d)** For the pair of parameters that best described the change in dynamics that followed the transition to running, model inhibition increased and adaptation decreased for each recording. **(e)** The spike rates of both FS and RS neurons were increased by running, but the relative increase was significantly larger for FS neurons in all four recordings (Wilcoxon rank-sum test, $p = 0.043; p < 10^{-5}; p < 10^{-2}; p = 0.037$ respectively). Across all recorded neurons, FS activity increased by 87% and RS activity increased by 28% during running (Wilcoxon rank-sum test, $p < 10^{-6}$).

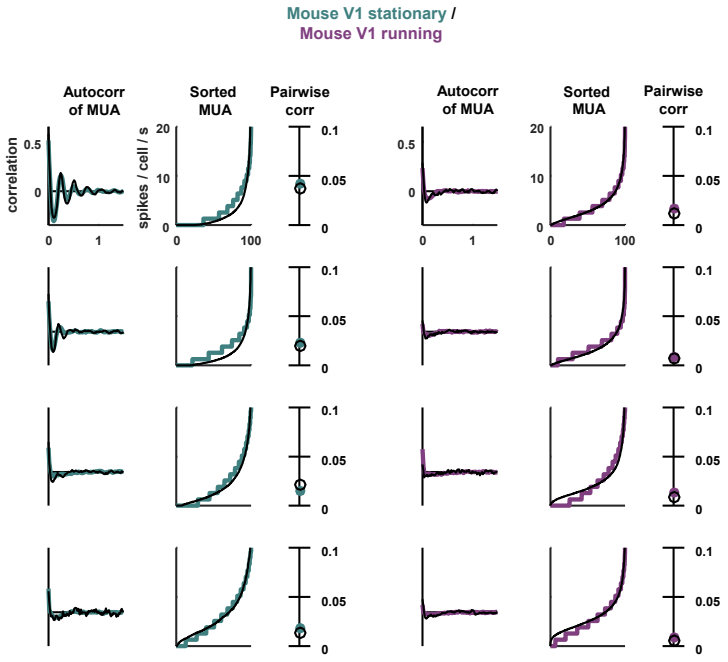

Figure 2.16: Statistics for all fits. The model was fit to stationary and running epochs for each of four separate recordings. The statistics of the activity measured from the recordings and generated by the model fit to the recordings are shown, plotted as in Figure 2.11.

Discussion

We have shown here that a deterministic spiking network model is capable of intrinsically generating population-wide fluctuations in neural activity, without requiring external modulating inputs. It has been observed in vitro that population-wide fluctuations in neural activity persist without external input [Sanchez-Vives et al., 2010, Sanchez-Vives and McCormick, 2000]. Such fluctuations also arise in vivo in localized cortical networks, in both awake and anesthetized animals, without feedforward inputs [Shapcott et al., 2016] or any external inputs [Malina et al., 2016]. However, no previous models have been able to reproduce such large-scale coordinated activity in a deterministic network of connected neurons; previous models only reproduced single-neuron variability [Vogels and Abbott, 2005, Litwin-Kumar and Doiron, 2012]. By fitting our spiking network model with adaptation currents directly to experimental recordings, we demonstrated that the model is able to reproduce the wide variety of multi-neuron cortical activity patterns observed in vivo without the need for external noise. Through chaotic amplification of small perturbations, the model generates activity with both trial-to-trial variability

in the spike times of individual neurons and coordinated, large-scale fluctuations of the entire network. These fluctuations continue in the presence of sensory stimulation, thus creating noise correlations in a deterministic neural network.

We developed a network model that can reproduce experimentally-observed activity patterns through intrinsic variability, instead of through external noise inputs [Doiron et al., 2016, de la Rocha et al., 2007, Renart et al., 2010, Ecker et al., 2014]. Networks in the classical balanced state produce variability in single neuron spiking without external noise, but this variability does not result in variable population-wide firing [Doiron et al., 2016, van Vreeswijk and Sompolinsky, 1996, Renart et al., 2010]. Thus, they are not suitable to describe the population-wide fluctuations that are observed in many brain states in vivo [Okun et al., 2015]. To obtain single-neuron rate fluctuations in balanced networks, structured connectivity has been used to create clustered networks [Doiron et al., 2016]. However, while clustered networks do produce activity with positive correlations between a small fraction of neuron pairs (less than 1 in 1000), the average noise correlations across all pairs are still near zero and, thus, these networks are still unable to generate population-wide fluctuations.

We were able to overcome the limitations of previous models and generate intrinsic large-scale variability that is quantitatively similar to that observed in vivo by using spike-frequency adaptation currents in excitatory neurons, which have been well-documented experimentally [Nowak et al., 2003, Compte et al., 2003]. The population-wide fluctuations generated by the interaction between recurrent excitation and adaptation were a robust feature of the network and persisted in more sophisticated networks that included multiple conductance timescales, many more neurons, spiking inhibitory neurons, structured connectivity, and kurtotic distributions of synaptic efficacies (see Figure 2.3).

Inhibition controls the strength of the large-scale fluctuations that drive noise correlations

Our results are consistent with experiments showing that one global dimension of variability largely explains both the pairwise correlations between neurons [Okun et al., 2015] and the time course of population activity [Ecker et al., 2014]. In our network model, the coordinated, large-scale fluctuations that underlie this global

dimension of variability are generated primarily by the interaction between recurrent excitation and adaptation. When inhibition is weak, small deviations from the mean spike rate can be amplified by strong, non-specific, recurrent excitation into population-wide events (up states). These events produce strong adaptation currents in each activated neuron, which, in turn, result in periods of reduced spiking (down states) [Latham et al., 2000, Destexhe, 2009, Curto et al., 2009, Mochol et al., 2015]. The alternations between up states and down states have an intrinsic periodicity given by the timescale of the adaptation currents, but the chaotic nature of the network adds an apparent randomness to the timing of individual events, thus creating intrinsic temporal variability. Several previous studies [Tsodyks et al., 1998, Loebel et al., 2007] have modelled alternations between up states and down states using synaptic depression rather than spike-frequency adaptation. However, to our knowledge, there is no experimental evidence for the involvement of synaptic depression in the control of cortical state.

The intrinsic temporal variability in the network imposes a history dependence on evoked responses; because of the build-up of adaptation currents during each spiking event, external inputs arriving shortly after an up state will generally result in many fewer spikes than those arriving during a down state [Curto et al., 2009]. This history dependence creates a trial-to-trial variability in the total number of stimulus-evoked spikes that is propagated and reinforced across consecutive stimulus presentations to create noise correlations. However, when the strength of the inhibition in the network is increased, the inhibitory feedback is able to suppress some of the amplification by the recurrent excitation, and the transitions between clear up and down states are replaced by weaker fluctuations of spike rate that vary more smoothly over time. If the strength of the inhibition is increased even further, such that it becomes sufficient to counteract the effects of the recurrent excitation entirely, then the large-scale fluctuations in the network disappear, weakening the history dependence of evoked responses and eliminating noise correlations.

Strong inhibition sharpens tuning curves and enables accurate decoding by stabilizing network dynamics

Numerous experiments have demonstrated that inhibition can shape the tuning curves of cortical neurons, with stronger inhibition generally resulting in sharper

tuning [Isaacson and Scanziani, 2011]. The mechanisms involved are still a subject of debate, but this sharpening is often thought to result from structured connectivity that produces differences in the tuning of the excitatory and inhibitory synaptic inputs to individual neurons; lateral inhibition, for example, can sharpen tuning when neurons with similar, but not identical, tuning properties inhibit each other. Our results, however, demonstrate that strong inhibition can sharpen tuning in a network without any structured connectivity simply by controlling its dynamics.

In our model, broad tuning curves result from the over-excitability of the network. When inhibition is weak, every external input will eventually excite every neuron in the network because those neurons that receive the input directly will relay indirect excitation to the rest of the network. When inhibition is strong, however, the indirect excitation is largely suppressed, allowing each neuron to respond selectively to only those external inputs that it receives either directly or from one of the few other neurons to which it is strongly coupled. Thus, when inhibition is weak and the network is unstable, different external inputs will trigger similar population-wide events [Bathellier et al., 2012], so the selectivity of the network in this regime is weak and its ability to encode differences between sensory stimuli is poor. In contrast, when inhibition is strong and the network is stable, different external inputs will reliably drive different subsets of neurons, and the activity patterns in the network will encode different stimuli with high selectivity and enable accurate decoding.

Two different dynamical regimes with weak noise correlations

A number of studies have observed that the noise correlations in cortical networks can be extremely weak under certain conditions [Ecker et al., 2010, Renart et al., 2010, Hansen et al., 2012, Pachitariu et al., 2015]. It was originally suggested that noise correlations were weak because the network was in an asynchronous state in which neurons are continuously depolarized with a resting potential close to the spiking threshold [Renart et al., 2010, van Vreeswijk and Sompolinsky, 1996]. Experimental support for this classical asynchronous state has been provided by intracellular recordings showing that the membrane potential of cortical neurons is increased during locomotion [McGinley et al., 2015a] and hyper-arousal [Constantinople and Bruno, 2011], resulting in tonic spiking. However, other experiments have shown that the membrane potential of cortical neurons in

behaving animals can also be strongly hyperpolarized with clear fluctuations between up and down states [Sachidhanandam et al., 2013, Tan et al., 2014, McGinley et al., 2015a, Polack et al., 2013].

Many forms of arousal tend to reduce the power of these low-frequency fluctuations in membrane potential [Sachidhanandam et al., 2013, Bennett et al., 2013, Polack et al., 2013, McGinley et al., 2015a, Crochet et al., 2011]; however, there is mounting evidence suggesting that different forms of arousal may have distinct effects on neural activity [McGinley et al., 2015b]. Locomotion in particular tends to depolarizate cortical neurons, and in some cases increases tonic spiking [Niell and Stryker, 2010]. In contrast, task-engagement in stationary animals has been associated with hyperpolarization and suppression of activity [McGinley et al., 2015a, Otazu et al., 2009, Buran et al., 2014] (but not all studies find a decrease in membrane potential during task engagement [Sachidhanandam et al., 2013]). The existence of two different dynamical regimes with weak noise correlations was also apparent in our recordings; while some recordings with weak noise correlations resembled the classical asynchronous state with spontaneous activity consisting of strong, tonic spiking (e.g. desynchronized urethane recordings and some awake recordings), other recordings with weak noise correlations exhibited a suppressed state with relatively low spontaneous activity that contained clear, albeit weak, up and down states (e.g. FMM recordings and other awake recordings). Our model was able to accurately reproduce spontaneous activity patterns and generate evoked responses with weak noise correlations in both of these distinct regimes.

In addition to strong inhibition, the classical asynchronous state with strong, tonic spiking appears to require a combination of weak adaptation and an increase in the number of neurons receiving strong tonic input (see parameter sweeps in Figures 2.2c-d and parameter values for awake mouse V1 recordings in 2.7). Since large-scale fluctuations arise from the synchronization of adaptation currents across the population, reducing the strength of adaptation diminishes the fluctuations [Destexhe, 2009, Curto et al., 2009, Mochol et al., 2015]. Increasing tonic input also diminishes large-scale fluctuations, but in a different way [Latham et al., 2000]; when a subset of neurons receive increased tonic input, their adaptation currents may no longer be sufficient to silence them for prolonged periods, and the activity of these neurons during what would otherwise be a down state prevents the entire

population from synchronizing. When the network in the asynchronous state is driven by an external input, it responds reliably and selectively to different inputs. Because the fluctuations in the network are suppressed and its overall level of activity remains relatively constant, every input arrives with the network in the same moderately-adapted state, so there is no history dependence to create noise correlations in evoked responses.

Unlike in the classical asynchronous state, networks in the suppressed state have slow fluctuations in their spontaneous activity, and the lack of noise correlations in their evoked responses is due to different mechanisms (see parameter values for gerbil A1 FMM recordings in Figure 2.7). The fluctuations in the hyperpolarized network are only suppressed when the network is driven by external input. In our model, this suppression of the correlated variability in evoked responses is caused by the supralinearity of the feedback inhibition [Rubin et al., 2015]. The level of spontaneous activity driven by the tonic input to each neuron results in feedback inhibition with a relatively low gain, which is insufficient to suppress the fluctuations created by the interaction between recurrent excitation and adaptation. However, when the network is strongly driven by external input, the increased activity results in feedback inhibition with a much higher gain, which stabilizes the network and allows it to respond reliably and selectively to different inputs. This increase in the inhibitory gain of the driven network provides a possible mechanistic explanation for the recent observation that the onset of a stimulus quenches variability [Churchland et al., 2010] and switches the cortex from a synchronized to a desynchronized state [Tan et al., 2014], as well as the suppression of responses to high-contrast stimuli in alert animals [Zhuang et al., 2014].

In the following chapter, we will investigate stimulus-driven activity in the visual cortex of awake mice. This activity is less dominated by one-dimensional fluctuations than activity during anesthetized states.

Materials and Methods

Electrophysiological recordings and processing

All of the recordings analyzed in this study have been described previously. Only a brief summary of the relevant experimental details are provided here. Each recording is considered as a single sample point to which we fit our model. Thus, our sample size is 59. This is justified as sufficient because our samples span multiple brain regions and multiple species, and may be considered as representative activity for a range of different brain states. Due to the sample size, we used the Spearman's (non-parametric) rank correlation in most of our analyses.

Mouse V1, awake passive

The experimental details for the mouse V1 awake passive recordings have been previously described [Okun et al., 2015]. The recordings were performed on male and female mice older than 6-7 weeks, of C57BL/6J strain. Mice were on 12 hours non-reversed light-dark regime. The mice were implanted with head plates under anaesthesia. After head-plate implant each mouse was housed individually. After a few days of recovery the mice were accustomed to having their head fixed while sitting or standing in a custom built tube. On the day of the recording, the mice were briefly anaesthetised with isoflurane, and a small craniectomy above V1 was made. Recordings were performed at least 1.5h after the animals recovered from the anaesthesia. Buzsaki32 or A4x8 silicon probes were used to record the spiking activity of populations of neurons in the infragranular layers of V1.

Visual stimuli were presented on two of the three available LCD monitors, positioned 25 cm from the animal and covering a field of view of 120 60, extending in front and to the right of the animal. Visual stimuli consisted of multiple presentations of natural movie video clips. For recordings of spontaneous activity, the monitors showed a uniform grey background.

Mouse V1, running

Two additional recordings were performed on two female mice, 14 and 20 weeks old. These mice expressed ChR2 in PV+ neurons (Pvalbtm1(cre)Arbr driver crossed with

Ai32 reporter). Mice were on 12 hour non-reversed light-dark regime. The mice were implanted with head plates under anesthesia. After head-plate implant each mouse was housed individually. After a few days of recovery the mice were accustomed to having their head fixed while standing or running on a styrofoam treadmill. On the day of the recording, the mice were briefly anesthetized with isoflurane and a small craniectomy was made above V1. Recordings were performed at least 1.5 hours after recovery from the anesthesia. Mice were head-fixed above the treadmill and allowed to run at will while multi-neuron recordings were made across all layers using silicon probes that were inserted roughly perpendicular to the cortical surface. Raw voltage signals were referenced against an Ag/AgCl wire in a saline bath above the craniectomy, amplified with analog Intan amplifiers, and digitized at 25 kHz with a WHISPER acquisition system. Visual stimuli were presented on three LCD monitors, positioned as three sides of a square, 20 cm from the animal and covering a field of view of approximately 270 x 70, centered on the direction of the mouses nose [Okun et al., 2015]. Visual stimuli consisted of drifting gratings of different sizes, approximately centered on the receptive field location of the recorded neurons. Stimuli were either 1 or 2 seconds long, and in the periods between stimuli (durations of 0.4 - 1 seconds), the monitors showed a uniform grey background.

Rat A1

The experimental procedures for the rat A1 recordings have been previously described [Luczak et al., 2009]. Briefly, head posts were implanted on the skull of male Sprague Dawley rats (300-500 g, normal light cycle, regular housing conditions) under ketamine-xylazine anesthesia, and a hole was drilled above the auditory cortex and covered with wax and dental acrylic. After recovery, each animal was trained for 6-8 d to remain motionless in the restraining apparatus for increasing periods (target, 1-2 h). On the day of the recording, each animal was briefly anesthetized with isoflurane and the dura resected; after a 1 h recovery period, recording began. The recordings were made from infragranular layers of auditory cortex with 32-channel silicon multi-tetrode arrays.

Sounds were delivered through a free-field speaker. As stimuli we used pure tones (3, 7, 12, 20, or 30 kHz at 60 dB). Each stimulus had duration of 1s followed by 1s of silence. All procedures conformed to the National Institutes of Health Guide for

the Care and Use of Laboratory Animals.

Gerbil A1

The gerbil A1 recordings have been described in detail previously [Pachitariu et al., 2015]. Briefly, adult male gerbils (70-90 g, P60-120, normal light-dark cycle, group housed) were anesthetized with one of three different anesthetics: ketamine/xylazine (KX), fentanyl/medetomidine/midazolam (FMM), or urethane. A small metal rod was mounted on the skull and used to secure the head of the animal in a stereotaxic device in a sound-attenuated chamber. A craniotomy was made over the primary auditory cortex, an incision was made in the dura mater, and a 32-channel silicon multi-tetrode array was inserted into the brain. Only recordings from A1 were analyzed. Recordings were made between 1 and 1.5 mm from the cortical surface (most likely in layer V). All gerbils recorded were used in this study, except for one gerbil under FMM which exhibited little to no neural activity during the recording period.

Sounds were delivered to speakers coupled to tubes inserted into both ear canals for diotic sound presentation along with microphones for calibration. Repeated presentations of a 2.5 s segment of human speech were presented at a peak intensity of 75 dB SPL. For analyses of responses to different speech tokens, seven 0.25 s segments were extracted from the responses to each 2.5 s segment.

Gerbil IC

The gerbil IC recordings have been described in detail previously [Garcia-Lazaro et al., 2013]. Recordings were made under ketamine/xylazine anesthesia using a multi-tetrode array placed in the low-frequency laminae of the central nucleus of the IC. Experimental details were otherwise identical to those for gerbil A1. In addition to the human speech presented during the A1 recordings, tones with a duration of 75 ms and frequencies between 256 Hz and 8192 Hz were presented at intensities between 55 and 85 dB SPL with a 75 ms pause between each presentation.

All relevant data are available from the authors upon request.

Spike sorting and filtering

Details of spike sorting for most recordings have been described in detail before in the respective original publications. Briefly, recordings from mouse V1 (awake passive) and rat A1 were spike sorted with KlustaKwik and further manually inspected in KlustaViewa. Recordings from gerbil A1 or IC were spike sorted with a custom-modified version of KlustaKwik. The unpublished recordings from mouse V1 (awake running) were spike sorted with Kilosort using the default settings [Pachitariu et al., 2016a] and inspected in Phy [Pachitariu et al., 2016a]. Only units with spike rates above 0.1 Hz were considered in the analysis. The spike waveforms considered in the FS/RS classification for all recordings were obtained from the mean, raw, unfiltered spike snippets.

Classifying FS and RS neurons

We classified fast-spiking and regular-spiking neurons based on their raw, unfiltered, spike shape [Okun et al., 2015]. We determined the trough-to-peak time of the mean spike waveform after smoothing with a gaussian kernel of $\sigma = 0.5$ samples. The distribution of the trough-to-peak time τ was clearly bimodal in all types of recordings. Following [Okun et al., 2015] we classified FS neurons in the awake data with $\tau < 0.6$ms and RS neurons with $\tau > 0.8$ms. The distributions of τ in the anesthetized data, although bimodal, did not have a clear separation point, so we conservatively required $\tau < 0.4$ms to classify an FS cell in these recordings and $\tau > 0.65$ms to classify RS neurons (see Figure 2.13). The rest of the neurons were not considered for the plots in Figures 2.14 and 2.15 and are shown in gray on the histogram in Figure 2.13.

Although one recent study has raised doubts on the accuracy of spike-width based classification [Moore and Wehr, 2013], a large number of other studies have shown 90-100% classification accuracy of FS neurons as PV+ interneurons [Nowak et al., 2003, Kawaguchi and Kubota, 1997, Barthó et al., 2004, Cho et al., 2010, Madisen et al., 2012, Stark et al., 2013, Cardin et al., 2009, Cohen and Mizrahi, 2015]. Even [Moore and Wehr, 2013] show that the classification is near-perfect using other features of the spike waveform; their finding that spike-width based classification was not accurate may be due to the filtering that they performed during

pre-processing. In our recordings, the distributions of the trough-to-peak duration of the raw waveform are highly bimodal in all cases (see Figure 2.13), unlike the distributions shown in [Moore and Wehr, 2013].

Local field potential

The low-frequency potential (LFP) was computed by low-pass filtering the raw signal with a cutoff of 300 Hz. Spectrograms with adaptive time-frequency resolution were obtained by filtering the LFP with Hamming-windowed sine and cosine waves and the spectral power was estimated as the sum of their squared amplitudes. The length of the Hamming-window was designed to include two full periods of the sine and cosine function at the respective frequency, except for frequencies of 1 Hz and above 30 Hz, where the window length was clipped to a single period of the sine function at 1 Hz and two periods of the sine function at 30 Hz respectively. The synchrony level was measured as the log of the ratio of the low to high frequency power (respective bands: 3-10 Hz and 11-96 Hz, excluding 45-55 Hz to avoid the line noise). We did not observe significant gamma power peaks except for the line noise, in either the awake or anesthetized recordings.

Spiking network model and fitting procedure

Spiking network model

We developed a network model using conductance-based quadratic integrate and fire neurons. There are three currents in the model: an excitatory, an inhibitory and an adaptation current. The subthreshold membrane potential for a single neuron i obeys the equation

$$\tau_m \frac{dV_i}{dt} = (V_i - E_L) * (V_i - V_{th}) - g_{E_i}(V_i - E_E) - g_{I_i}(V_i - E_I) - g_{A_i}(V_i - E_A).$$

When $V > V_{th}$, a spike is recorded in the neuron and the neuron's voltage is reset to $V_{reset} = 0.9V_{th}$. For simplicity, we set $V_{th} = 1$ and the leak voltage $E_L = 0$. The excitatory voltage $E_E = 2V_{th}$ and $E_I = E_A = -0.5V_{th}$. Each of the conductances has a representative differential equation which is dependent on the spiking of the neurons in the network

at the previous time step, $\mathbf{s}_{t-1}$. The excitatory conductance obeys

$$\tau_E \frac{\mathrm{d}\mathbf{g}_E}{\mathrm{d}t} = -\mathbf{g}_E + J\mathbf{s}_{t-1} + \mathbf{b}.$$

where J is the matrix of excitatory connectivity and $\mathbf{b}$ is the vector of tonic inputs to the neurons. The matrix of connectivity is random with a probability of 5% for the network of 512 neurons and their connectivities are randomly chosen from a uniform distribution between 0 and w_E. The tonic inputs $\mathbf{b}$ have a minimum value b_0, which we call the tonic input baseline added to a random draw from an exponential distribution with mean b_1, which we call the tonic input spread, such that for neuron i, $\mathbf{b}(i) = b_0 + \mathrm{exprnd}(b_1)$. The inhibitory conductance obeys

$$\tau_I \frac{\mathrm{d}g_I}{\mathrm{d}t} = -g_I + w_I * (\exp(\sum \mathbf{s}_{t-1} * c) - 1).$$

where c controls the gain of the inhibitory conductance. The inhibitory conductance is global, i.e. each neuron receives the same inhibitory feedback, and it obeys an exponential supralinearity [Rubin et al., 2015].

The adaptation conductance obeys

$$\tau_A \frac{\mathrm{d}\mathbf{g}_A}{\mathrm{d}t} = -\mathbf{g}_A + w_A\mathbf{s}_{t-1}.$$

The simulations are numerically computed using Eulers method with a time-step of 0.75 ms (this was the lock-out window used for spike-sorting the in vivo recordings and allowed for fast simulations). To avoid numerical instabilities at low voltages, we rectified the voltages at the activation potential of the inhibitory conductance. Each parameter set was simulated for 900 seconds. The timescales are set to $\tau_m = 20$ ms, $\tau_E = 5.10$ ms, $\tau_I = 3.75$ ms, $\tau_A = 375$ ms, and the inhibitory non-linearity controlled by $c = 0.25$. The remaining five parameters (w_I, w_A, w_E, b_1, and b_0) were fit to the spontaneous activity from multi-neuron recordings using the techniques described below. Their ranges were (0.01-0.4), (0.4-1.45), (2.50-5.00), (0.005-0.10), and (0.0001-0.05) respectively.

To illustrate the ability of the network to generate activity patterns with macroscopic variability, we simulated spontaneous activity with a parameter set that produces up and down state dynamics. Figure 2.2a shows the membrane potential of

a single neuron in this simulation and its conductances at each time step. Figure 2.2b shows the model run twice with the same set of initial conditions and parameters, but with an additional single spike inserted into the network on the second run (circled in green).

Parameter sweep analysis

Figure 2.2c and d summarize the effects of changing each parameter on the structure of the spontaneous activity patterns generated by the model. We held the values for all but one parameter fixed and swept the other parameter across a wide range of values. The fixed parameter values were set to approximately the median values obtained from fits to all in vivo recordings. Figure 2.12b, d, and f and Figure 2.11 show the results of similar parameter sweep analyses for stimulus-driven activity with the external input to the network derived from IC activity as described below. For these analyses, the values of the parameters that were not swept were fixed at those fit to each individual recording.

GPU implementation

We accelerated the network simulations by programming them on graphics processing units (GPUs) such that we were able to run them at 650x real time with 15 networks running concurrently on the same GPU. We were thus able to simulate $\sim$10000 seconds of simulation time in 1 second of real time. To achieve this acceleration, we took advantage of the large memory bandwidth of the GPUs. For networks of 512 neurons, the state of the network (spikes, conductances and membrane potentials) can be stored in the very fast shared memory available on each multiprocessor inside a GPU. A separate network was simulated on each of the 8 or 15 multiprocesssors available (video cards were GTX 690 or Titan Black). Low-level CUDA code was interfaced with Matlab via mex routines.

Summary statistics

Several statistics of spikes were used to summarize the activity patterns observed in the in vivo recordings and in the network simulations. Because there were on the order of 50 neurons in each recording, all of the statistics below were influenced by

small sample effects. To replicate this bias in the analysis of network simulations, we subsampled 50 neurons from the network randomly and computed the same statistics we computed from the in vivo recordings.

The noise correlations between each pair of neurons in each recording were measured from responses to speech. The response of each neuron to each trial was represented as a binary vector with 15 ms time bins. The total correlation for each pair of neurons was obtained by computing the correlation coefficient between the actual responses. The signal correlation was computed after shuffling the order of repeated trials for each time bin. The noise correlation was obtained by subtracting the signal correlation from the total correlation.

The multi-unit activity (MUA) was computed as the sum of spikes in all neurons in bins of 15 ms.

The autocorrelation function of the MUA at time-lag τ was computed from the formula

$$\mathrm{ACF}(\tau) = \frac{1}{N_{samples}} \sum \mathrm{MUA}(t) * \mathrm{MUA}(t+\tau)$$

In the awake recordings (mouse V1 passive and running, rat A1) we observed slow-timescale fluctuations on the order of tens of seconds, which significantly affected the autocorrelation function of the MUA at lags ¿ 1 s. We chose to ignore these fluctuations during model fitting by high-pass filtering the MUA at 1 Hz before computing the autocorrelation function.

To measure the autocorrelation timescale, we fit one side of the ACF with a parametric function

$$\mathrm{ACF}(\tau) \sim a \exp(-\tau/T) \cdot \cos(\tau/(2\pi t_{\mathrm{period}}))$$

where a is an overall amplitude, T is a decay timescale and t_{period} is the oscillation period of the autocorrelation function. There was not always a significant oscillatory component in the ACF, but the timescale of decay accurately captured the duration over which the MUA was significantly correlated.

Parameter searches

To find the best fit parameters for each individual recording, we tried to find the set of model parameters for which the in vivo activity and the network simulations

had the same statistics. We measured goodness of fit for each of the three statistics: pairwise correlations, the MUA distribution, the MUA ACF. Each statistic was normalized appropriately to order 1, and the three numbers obtained were averaged to obtain an overall goodness of fit.

The distance measure D_c between the mean correlations c_θ obtained from a set of parameters θ and the mean correlations c_n in recording n was simply the squared error $D_c(c_n, c_\theta) = (c_n - c_\theta)^2$. This was normalized by the variance of the mean correlations across recordings to obtain the normalized correlation cost Cost_c, where $\langle x_n \rangle_n$ is used to denote the average of a variable x over recordings indexed by n.

$$\text{Cost}_c = \frac{D_c(c_n, c_\theta)}{\langle D_c(c_n, \langle c_n \rangle) \rangle}$$

The distance measure D_m for the MUA distribution was the squared difference summed over the order rank bins k of the distribution $D_m(\text{MUA}_n, \text{MUA}_\theta) = \sum_k (\text{MUA}_n(k) - \text{MUA}_\theta(k))^2$. This was normalized by the distance between the data MUA and the mean data MUA. In other words, the cost measures how much closer the simulation is to the data distribution than the average of all data distributions.

$$\text{Cost}_m = \frac{D_m(\text{MUA}_n, \text{MUA}_\theta)}{D_m(\text{MUA}_n, \langle \text{MUA}_n \rangle)}$$

Finally, the distance measure D_a for the autocorrelation function of the MUA was the squared difference summed over time lag bins t of the distribution $D_a(\text{ACF}_n, \text{ACF}_\theta) = \sum_t (\text{ACF}_n(t) - \text{ACF}_\theta(t))^2$. This was normalized by the distance between the data ACF and the mean data ACF.

$$\text{Cost}_a = \frac{D_a(\text{ACF}_n, \text{ACF}_\theta)}{D_a(\text{ACF}_n, \langle \text{ACF}_n \rangle)}$$

The total cost of parameters θ on recording n is therefore $\text{Cost}(n, \theta) = \text{Cost}_c + \text{Cost}_m + \text{Cost}_a$. Approximately one million networks were simulated on a grid of parameters for 600 seconds each of spontaneous activity, and their summary statistics ($c_\theta, \text{MUA}_\theta$ and ACF_θ) were retained. The Cost was smoothed for each recording by averaging with the nearest 10 other simulations on the grid. This ensured that some of the sampling noise was removed and parameters

were estimated more robustly. The best fit set of parameters was chosen as the minimizer of this smoothed cost function, on a recording by recording basis.

Evaluation of the goodness-of-fit of the model

We computed the upper limit for the explained variance of the model based on the recordings. We split each neural recording into two halves (interleaved segments of 4s each) and computed the amount of variance in statistics from one half of the recording that is explained by the other half of the recording. We compared this to the amount of variance in the statistics of the full recording that was explained by the model.

Alternative Gibbs sampling parameter optimization

We also demonstrate an alternative approach to finding the best fitting parameters through a sampling-based optimization procedure (Figure 2.5). This reduces the necessary number of simulations from 1 million to 100,000. Future work might in principle devise even faster optimizations, thus allowing analysis on a bigger scale than presented here. Briefly, the sampling-based optimization is based on defining the energy landscape as the negative of the cost function, and thus defining a probability distribution over parameters $\mathrm{P}(\theta) = exp(-\mathrm{Cost}(\theta)/T)$, where T is the temperature. We use a proposal distribution that always proposes neighbors of the current sample on the grid on which we did the full parameter sweeps, and accept the proposals according to the balance equations of Markov Chain Monte Carlo sampling (MCMC):

$$\begin{aligned}\mathrm{prob}(\mathrm{accept}) &= \frac{\mathrm{P}(\theta_{\mathrm{new}})}{\mathrm{P}(\theta_{\mathrm{new}}) + \mathrm{P}(\theta_{\mathrm{old}})} \\ &= \frac{1}{1 + exp(-(\mathrm{Cost}(\theta_{\mathrm{new}}) - \mathrm{Cost}(\theta_{\mathrm{old}}))/T)}\end{aligned}$$

To avoid the MCMC chains getting stuck into low probability parts of the energy landscape, we restart the chain every 50 samples from the pool of already-sampled points, chosen with probability proportional to its $\mathrm{P}(\theta)$. Furthermore, we allow the chain used to optimize the model parameters for one recording to use information from the chains used for the other recordings by pooling together the already-sampled

points from all datasets and restarting chains based on all these points.

NMDA and GABA$_B$ conductance network

We added long timescale excitatory and inhibitory conductances to the model and simulated the model at multiple levels of inhibitory feedback strength. The strength of the NMDA conductance was 4% the strength of the AMPA conductance and $\tau_{NMDA} = 100$ ms (thus the integrated current was approximately the same as the AMPA integrated current injection). The strength of the GABA$_B$ conductance was 2% the strength of the GABA conductance and it had the same timescale as NMDA. The parameter set used for Figure 2.3a was θ = (0.51, w_I, 2.6, 0.008, 0.037), where w_I ranged from 0.02 to 0.25.

Clustered neuronal network with intrinsic variability and spiking inhibitory neurons

We also simulated a clustered architecture with variability and adaptation currents. This model consisted of 144 clusters, each with 32 neurons, 8 of which were inhibitory neurons and 24 of which were excitatory neurons. The probability of within cluster excitatory-excitatory (E-E) connectivity was 0.3, and within cluster inhibitory-excitatory (I-E) and excitatory-inhibitory (E-I) were 0.15 and 0.1 respectively. The probability of out of cluster E-E, I-E, and E-I connectivity were 0.012, 0.03, and 0.01 respectively. The inhibitory-inhibitory connectivity was unclustered. The probability of connection was 0.01 and its strength was 0.17. The average connection strengths for E-E and I-E were 0.024 and 0.016 respectively. The E-I strength in Figure 2.3b ranged from 0.025 to 0.057. The adaptation current had strength 0.45 and $\tau_A = 220$ ms. The membrane timescale for excitatory and inhibitory neurons were 25 ms and 5 ms respectively, and $\tau_E = 6$ ms and $\tau_I = 3$ ms.

Stimulus-driven activity

Once the simulated networks were fit to the spontaneous neuronal activity, we drove them with an external input to study their evoked responses. The stimulus was either human speech (as presented during our gerbil A1 recordings) or pure tones. The external input to the network was constructed using recordings from 563 neurons

from the inferior colliculus (IC). For all recordings in the IC the mean pairwise noise correlations were near-zero and the Fano Factors of individual neurons were close to 1 [Garcia-Lazaro et al., 2013], suggesting that responses of IC neurons on a trial-by-trial basis are fully determined by the stimulus alone, up to Poisson-like variability. Thus, we averaged the responses of IC neurons over trials and drove the cortical network with this trial-averaged IC activity. We binned IC neurons by their preferred frequency in response to pure tones, and drove each model cortical neuron with a randomly chosen subset of 10 neurons from the same preferred-frequency bin. We rescaled the IC activity so that the input to the network had a mean value of 0.06 and a maximum value of 0.32, which was three times greater than the average tonic input.

We kept the model parameters fixed at the values fit to spontaneous activity and drove the network with 330 repeated presentations of the stimulus. We then calculated the statistics of the evoked activity. Noise correlations were measured in 15-ms bins as the residual correlations left after subtracting the mean response of each neuron to the stimulus across trials:

$$c_{ij} = \frac{1}{N_{samples}} \sum_t (s_i(t) - < s_i(t) >)(s_j(t) - < s_j(t) >)$$

where $s_i(t)$ is the summed spikes of neuron i in a 15-ms bin and $< s_i(t) >$ is the mean response of neuron i to the stimulus. The noise correlation value given for each recording is the mean of c_{ij}.

Analysis of stimulus-driven activity

Tuning width

To determine tuning width to sound frequency, we used responses of IC neurons to single tones as inputs to the model network. The connections from IC to the network were the same as described in the previous section. Because the connectivity was tonotopic and IC responses are strongly frequency tuned, the neurons in the model network inherited the frequency tuning. We did not model the degree of tonotopic fan-out of connections from IC to cortex and, as a result, the tuning curves of the model neurons were narrow relative to those observed in cortical recordings [Pachitariu et al.,

2015]. We chose the full width of the tuning curve at half-max as a standard measure of tuning width.

Decoding tasks

We computed decoding error for a classification task in which the single-trial activity of all model neurons was used to infer which of seven different speech tokens was presented. The classifier was built on training data using a linear discriminant formulation in which the Gaussian noise term was replaced by Poisson likelihoods. Specifically, the activity of a neuron for each 15-ms bin during the response to each token was fit as a Poisson distribution with the empirically-observed mean. To decode the response to a test trial, the likelihood of each candidate token was computed and the token with the highest likelihood was assigned as the decoded class. This classifier was chosen because it is very fast and can be used to model Poisson-like variables, but we also verified that it produced decoding performance as good as or better than classical high-performance classifiers like support vector machines.

Awake behavioral state analysis

Dividing trials by synchrony

For the recordings from awake restrained mice, we computed a synchrony value for each trial in the 500 ms window following stimulus onset. The distribution of synchrony values was not clearly bimodal, but varied across a continuum of relatively synchronized and desynchronized states. To examine the effect of synchrony on noise correlations, we sorted all trials by their synchrony value, classified the 50% of trials with the lowest values as desynchronized and the 50% of trials with the highest values as synchronized, and computed the noise correlations for each set of trials for each recording. To examine the effect of synchrony on FS and RS activity, we pooled all trials from all recordings, divided them into quintiles by their synchrony value, and computed the average spike rates of FS and RS neurons for each set of trials. For Figure 2.14, noise correlations were computed aligned to the stimulus onsets in windows of 500 ms, to match the window used for measuring FS and RS activity as well as LFP power.

For urethane recordings, we computed the level of synchrony of the LFP (ratio of low frequency 1-10 Hz activity to high frequency 11-100 Hz) in sliding 10-second windows. The recordings were split into high and low synchrony based on the median level of synchrony, and 20 seconds around each transition point were discarded. We treated urethane recordings in synchronized and desynchronized states as separate recordings for the purposes of model fitting.

Dividing trials by behavioral state

We median-filtered the raw running speed of the treadmill with a window of 0.5 seconds. In order to discard extremely small speeds that may be noise, we discarded all points less than one hundredth of the standard deviation of the running speed. Using the processed running speed, we divided the data into periods of stationary and running behavior. We found periods of at least five seconds in which all bins were either zero or non-zero. We then excluded the first and last second of each of these segments from our computations, considering them to be periods of transition between stationary and running. We took all of the spiking in these segments and binned it into 15 ms bins for all further computations.

Modeling the transition from stationary to running

In order to fit the intrinsic variability in the population responses in the recordings that were made on the treadmill, we first removed the evoked responses from each recording. We computed the mean evoked response to each stimulus (10 stimuli total) for each neuron by computing the mean response across all trials and then subtracting the spontaneous spike rate. We then subtracted this mean evoked response from each neurons response on each trial. Because the spike rates of neurons varied from 0.1 Hz (our cutoff for inclusion) to 150 Hz, we divided each neurons binned spike rate by its average spike rate and then multiplied by the overall mean spike rate of the population. We then computed each of our statistics on these normalized population activity patterns. We fit the models to the statistics of the stationary periods in each of the recordings, and then changed one or two parameters of the stationary fits in order to best fit the running periods. We simulated evoked activity by driving the model with the mean evoked responses described above, scaled so that the overall spike rate

of the model responses matched the overall spike rate in each recording.

3

High-dimensional neural responses in visual cortex

The previous chapter shows that during wakefulness, cortical areas are able to encode stimuli with high fidelity. One popular hypothesis in neuroscience is that this encoding is nonetheless highly redundant: only a few dimensions of activity contain relevant information. Previous studies analyzing multi-neuron recordings have supported this interpretation, suggesting that the neural code is constrained to a low-dimensional manifold. However, the apparent manifold constraints may reflect low experimental complexity, rather than intrinsic limitations of the neural activity. To remove experimental limitations, we recorded populations of ∼10,000 neurons in response to 2,800 images, from the visual cortex of awake mice. We found that evoked neural activity was not constrained to any size of a lower dimensional subspace. Instead, the neural variance obeyed a $1/n$ power law distribution along the n-th dimension. The $1/n$ spectrum was not merely inherited from the $1/f$ spectrum of natural images, because it persisted under a new stimulus set, where we removed all but 8 dimensions of the presented images. Our results reveal a fundamental power-law scaling of activity along the neural dimensions, and suggest that multi-layer neural systems sequentially increase the dimensionality of their inputs at every layer through nonlinear computations.

Introduction

[1] The activity of single neurons in the brain is noisy, and has been described as Poisson-like in response to stimuli. This has led to the hypothesis that the fundamental unit of computation is not a single neuron, but a coordinated group of neurons that carry similar signals: an ensemble. The averaged activity of the ensemble would represent a reliable encoding of a stimulus on a single-trial basis. This framework has in turn led to the attractive hypothesis that large populations of neurons redundantly encode a small number of stimulus dimensions. In this view, the relevant encoding space is a low-dimensional stimulus embedding into the high-dimensional neural space. This hypothesis of low-dimensional representations implies that current generation neural recordings, of hundreds of neurons, are sufficient to capture the vast majority of information encoded by the local neural population. The low-dimensional hypothesis appears to even be validated by current generation multi-neuron recordings, which typically show that a few (3-10) dimensions of variation explain most of the neural activity related to a specific brain area, stimulus set or behavioral task [Gao et al., 2014, Gao and Ganguli, 2015].

However, recent theoretical analyses have questioned this interpretation [Gao et al., 2014, Gao and Ganguli, 2015]. Low estimates of dimensionality may instead reflect the low complexity of the experiments, not the intrinsic limitations of the neural activity. Indeed, most experiments are limited to at most dozens of stimuli or behavioral actions. Even then, these external covariates are highly correlated, for example adjacent orientations of a drifting grating stimulus, or adjacent reach directions for a motor task. In turn, if the stimuli are very similar, it should not be surprising that the neural responses to those stimuli is very similar. Another limitation pointed out by [Gao and Ganguli, 2015] relates to the number of neurons recorded: the estimated dimensionality can certainly not be higher than this number, and further limited by correlations between neurons. To remove limitations related to the low-complexity of previous experiments, we designed a protocol to record the activity of $\sim$10,000 neurons in response to 2,800 independent natural images. In addition, we developed a novel theoretical framework to robustly estimate the singular value spectrum of a matrix from noisy observations.

[1] The work described in this chapter was done in collaboration with Marius Pachitariu.

Results

Spatial receptive fields

We obtained $\sim$ 10,000 cells from each recording, using multi-plane two-photon calcium imaging and a high-yield, high-accuracy processing pipeline we developed called Suite2p (Figure 3.1a). The majority of recorded cells had significant stimulus-driven variance on a single-trial basis (Figure 3.1b), which we calculated from the correlation of responses to two stimulus repeats (see section 3.4.4.1). The distributions of stimulus-related variance were typically skewed, with a small fraction ($\sim$10%) of cells being highly-driven by the stimuli, but a majority of the population being just moderately well-driven.

One potential limitation on estimated dimensionality may be determined by the size and distribution of receptive fields (RFs) of the recorded population. Large, highly-overlapping receptive fields would impose dimensionality limitations for a simple-cell like bank of image filters. Small receptive fields, spread out over a large area would result in high dimensions even for a bank of simple cells. Thus, we wanted to estimate the degree of overlap of the receptive fields in single recorded population. We calculated the linear RFs for each cell, using a low-rank regularization method (Figure 3.1cd; for the fitting method, see also Chapter 5). The recorded populations had, in fact, large and highly-overlapping receptive fields, suggesting that their dimensionality might be low (Figure 3.1d). Nonetheless, there was a high diversity of receptive field sizes and shapes (Figure 3.2), and the linear receptive fields explained very little of the responses. Thus, we reasoned, it could still be possible that the nonlinear responses of the cells are high-dimensional.

High dimensionality of responses to natural images

We wanted to estimate the dimensionality of the responses independently of any encoding model, because the recorded cells might have response properties unaccounted for by simple or complex-cell like receptive fields. Suppose we had access to the trial-averaged responses of 10,000 neurons to 2,800 stimuli, such that all trial-to-trial fluctuations were averaged out. The singular value spectrum of this matrix would tell us how many significant dimensions of activity there are, and how

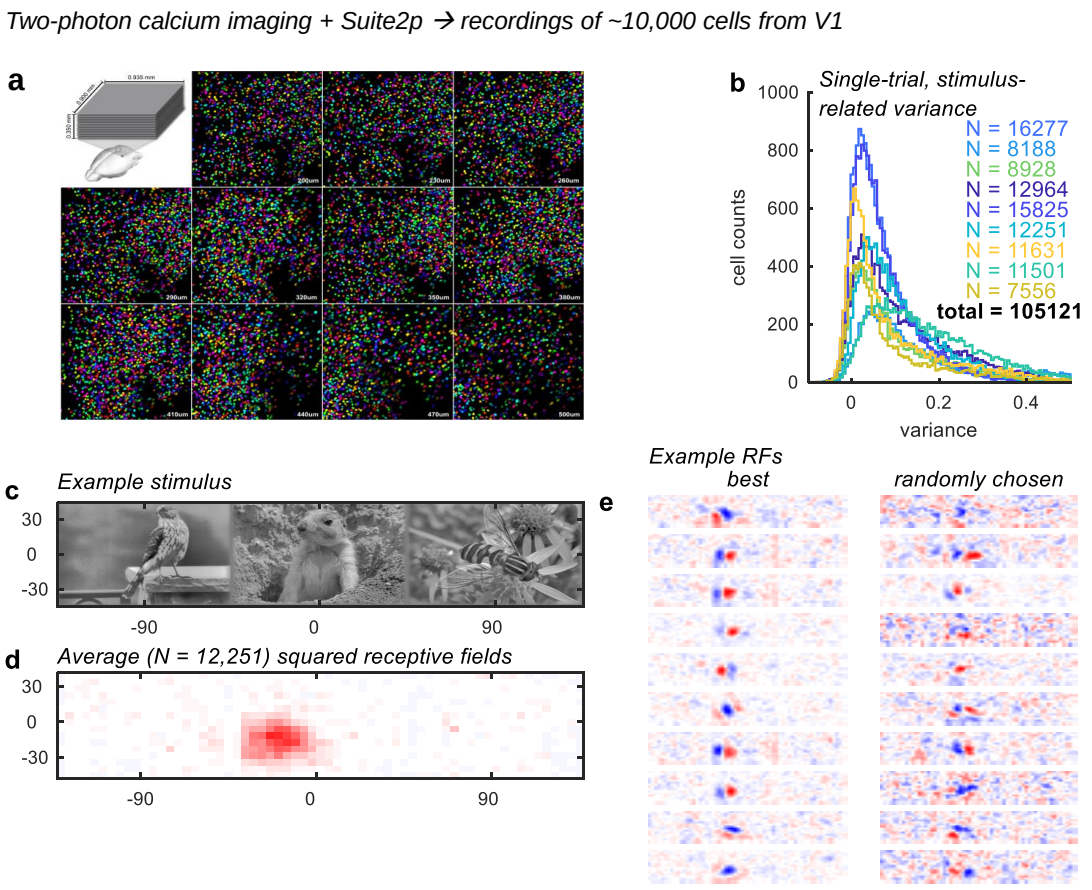

Figure 3.1: Recorded populations of ≈10,000 had highly overlapping receptive fields. **(a)** Schematic of multi-plane imaging method and cell masks obtained from Suite2p, an automated pipeline which we developed for processing these data. **(b)** Distribution of stimulus-related variance in single trials, for all cells recorded in 9 sessions from 9 mice. **(c)** Example single flashed stimulus consisting of three concatenated natural images, presented on the corresponding three monitors. **(d)** Average squared receptive fields obtained from one recording. **(e)** Example single cell receptive fields, estimated using a linear method. Most pairs of cells had highly overlapping receptive fields.

their variance is distributed across cells. In practice, we do not have access to the trial-averaged responses, because we can only collect two repeats of each stimulus. Instead, we need an estimation method of the underlying spectrum of a matrix, from two independent, noisy observations of it. We have developed such a method (see Appendix C), which we show can provide a tight lower bound for the spectrum of the underlying matrix.

We show in Appendix C that two presentations are in fact sufficient for the estimation, in the novel theoretical framework we developed. Briefly, this approach first calculates the percent of single-trial neural variance attributable to stimuli (∼10% in our case). This percentage was not substantially different for recordings at 30Hz

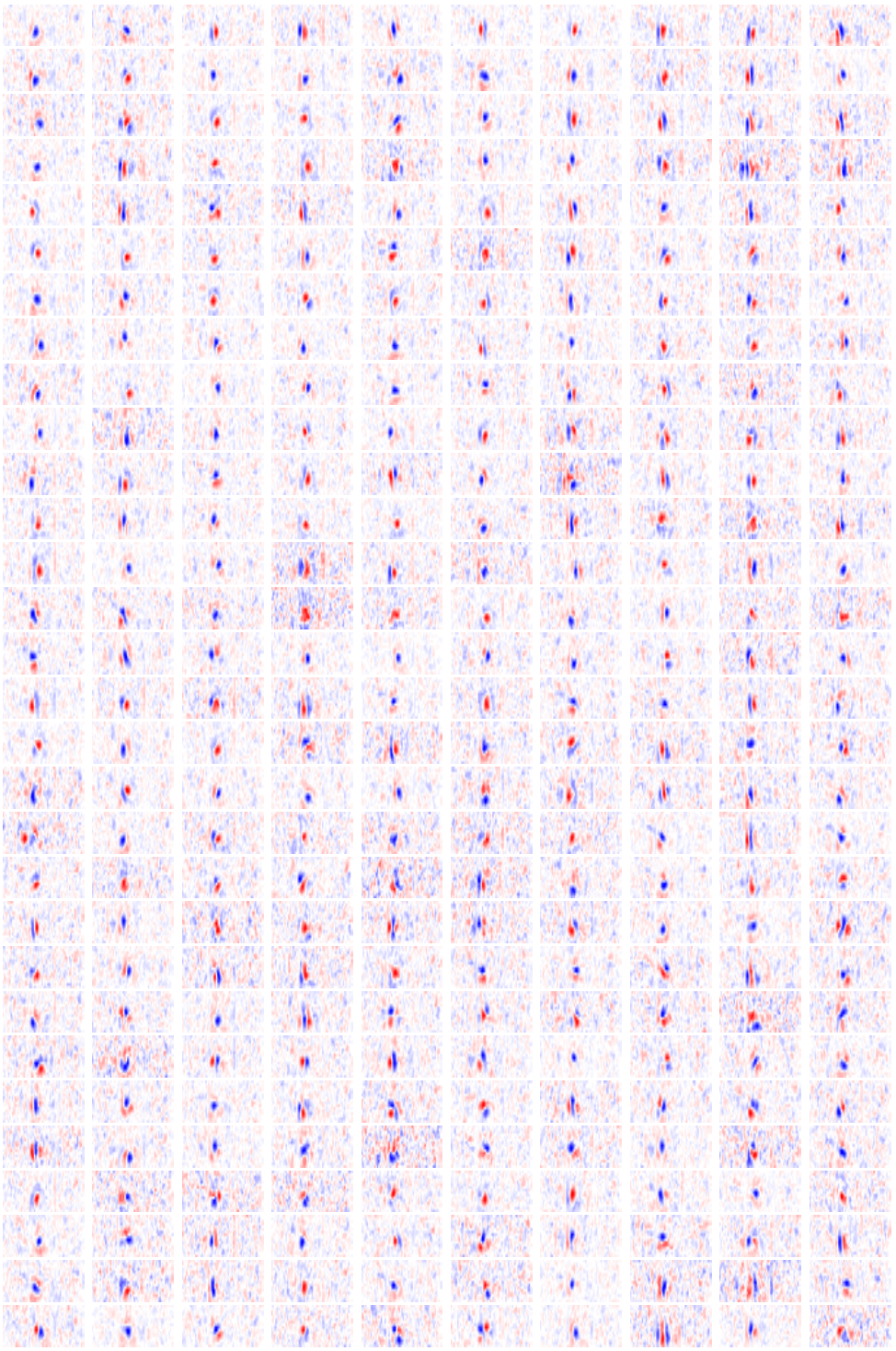

Figure 3.2: High diversity of receptive fields in a single session. Despite the high overlap between receptive fields of a single session, we observed a large diversity of receptive field shapes, orientations and sizes.

and recordings at 2.5Hz (data not shown), suggesting that our multi-plane imaging approach loses very little information. The percentage of stimulus-related variance was also not different in electrode recordings from visual cortex, while presenting the same stimulus set (data not shown). Similar to the estimation of single-neuron, single-trial variance, we can estimate variance along any neural dimension. When these neural dimensions are the singular vectors of the data matrix, we obtain a lower bound on the singular value spectrum of the trial-averaged, unobserved neural response matrix. Figure 3.3abc describes this procedure graphically. After obtaining the neural dimensions from the training data, we can project both training and testing data into these dimensions, and treat each dimension the same way we would treat individual neurons. Each dimensions has its own stimulus responses (Figure 3.3d), and the top dimensions have more robust responses to the two stimulus repeats (Figure 3.3e). The correlation of responses to two repeats (Figure 3.3f) is in fact equivalent to the percent stimulus-related variance on single trials (see section 3.4.4.1), i.e. the quantity we showed in Figure 3.1b for single neurons.

Using this method, our estimates of dimensionality were very high. Both the signal and the noise spectrum had a clear $1/n$ distribution (Figure 3.3g,h). Correspondingly, the cumulative signal spectrum had a logarithmic increase in explained variance with the number of dimensions (Figure 3.3i). Only 100 dimensions of neural activity were necessary to explain 50% of the neural response variance, but 1,000 dimensions were necessary to explain 95% of the variance.

Scaling of dimensionality with number of neurons and stimuli

As discussed in the introduction, any estimates of dimensionality are limited by the number of neurons recorded and stimuli shown. It is thus possible that our data is insufficient to estimate the true spectrum of the neural responses, especially given that most dimensions have relatively little variance (Figure 3.3h). If our data was insufficient, our estimates would be unstable: with more recorded neurons/stimuli, the estimates would shift. We can check directly if the estimates of dimensionality are stable, by studying the dependence of the spectra on the number of neurons and stimuli considered, as a subset of all recorded neurons/stimuli. We find that the spectra do converge for increasingly more neurons (Figure 3.4a). This shows that for a fixed number of stimuli (2,800), we recorded enough neurons to estimate the

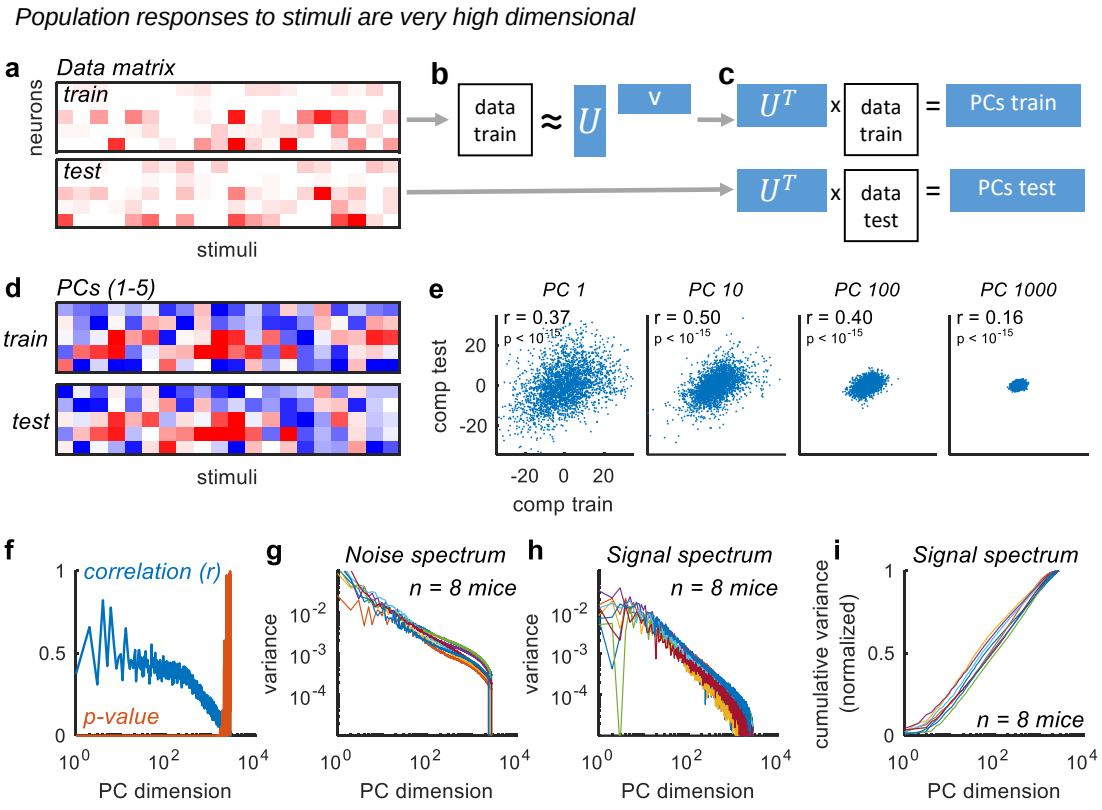

Figure 3.3: Responses in 10,000 cells are high-dimensional. **(a)** We obtained two data matrices on separate repeats of stimulus presentations. We call these the "train" and "test" matrices. **(b)** We factorized the training data into principal components U. **(c)** We projected both the training and test data into the principal components U, obtained from the training data. **(d)** Responses of the top principal components to the same stimuli as shown in (a). **(e)** Correlation of PC responses on the training and testing set, for the first, 10th, 100th and 1000th component. The correlation is also the percentage of stimulus-related variance along that principal component. **(f)** Correlation of PC responses to two repeats as a function of dimension. **(g)** The estimated noise spectrum as a function of dimension for all 9 recordings. **(h)** The estimated signal-variance for all 9 recordings. The signal-variance along dimension N corresponds to the squared estimate of the singular value corresponding to that dimension. **(i)** Same data as in (h), shown as a cumulative spectrum.

spectrum of the responses to these 2,800 stimuli. This conclusion holds if we consider the dimensionality required to account for 25%, 50%, 75% or even 95% of the variance (Figure 3.4b). In contrast, the spectra do not converge when we consider increasingly more stimuli (Figure 3.4cd). This implies that for a fixed number of neurons (∼10,000) we would have needed to present more stimuli than we did (2,800) in order to observe the true dimensionality of the population.

To get a more complete view of the dimensionality scaling with neurons and stimuli, we can also consider all combinations of subsets of neurons and stimuli. We summarize the inferred spectra with their dimensionality at 50% explained variance and 95% explained variance respectively (Figure 3.4ef).For any fixed number of stimuli, dimensionality converged close to the maximum possible (the number of stimuli), as we increased the number of neurons (Figure 3.4g). Similarly, for small, fixed numbers of neurons (up to $\approx 1,000$), dimensionality converged close to the maximum possible as we increased the number of stimuli (Figure 3.4h). However, for large numbers of neurons, we did not present enough stimuli to observe the convergence.

These observations are consistent with an underlying $1/n$ spectrum of signal dimensions, because for such spectra estimates of dimension should approach the unity line. In other words, the population responses should be full dimensional.

Low dimensionality of stimulus-triggered responses

We have shown that cortical populations of neurons span a full-dimensional space of responses to stimuli. Where are all the dimensions coming from? Are they inherited from the structure of the stimulus, or do they get added along the visual pathway through nonlinear computations? To rule out the first possibility, we constructed new stimulus sets and ran more analyses.

Natural images, such as the ones we presented during experiments, have the notable property that their 2D spectrum decays like $1/f$, as a function of spatial frequency f [Field, 1987]. This is thought to account for the multi-scale, fractal-like properties of natural images, and has lead to generative models of natural images with similar $1/f$ properties (for example, the "dead leaves" model [Lee et al., 2001]). It is thus possible that neural responses to natural images directly inherit their spectrum, either through linear or nonlinear representations. For this hypothesis to

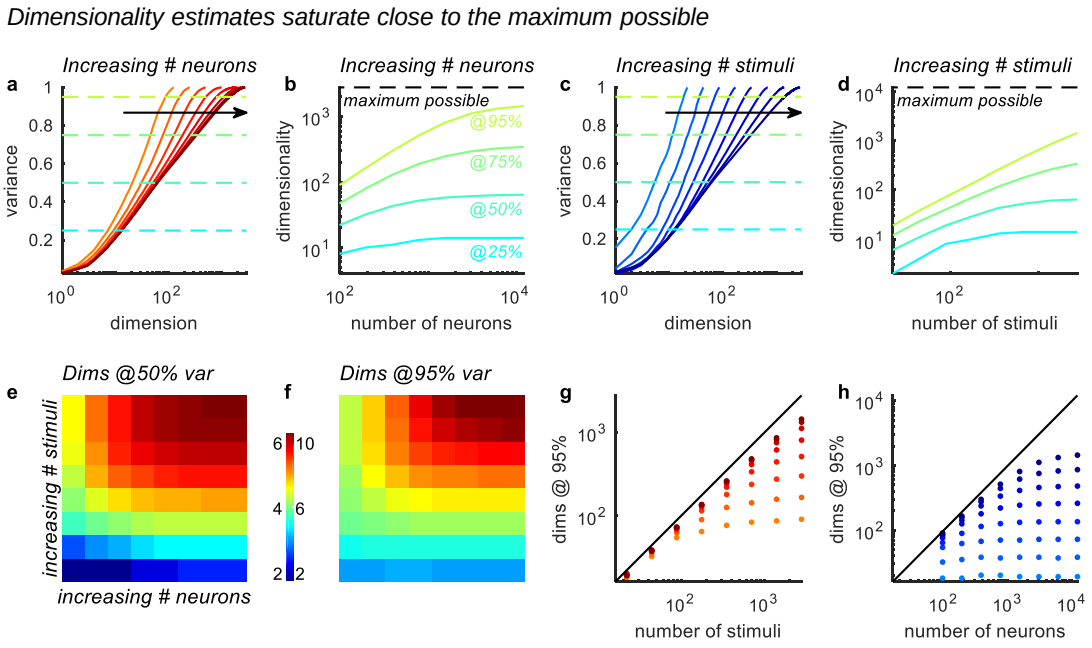

Figure 3.4: Scaling of dimensionality with neurons and stimuli. (a) We considered random subsets of all recorded neuron, spaced out logarithmically, i.e. a half, a quarter etc of all neurons. For each subset, we computed dimensionality curves as in Figure 3.3. **(b)** Based on these curves, we computed the number of dimensions required to reach 25%, 50%, 75% and 95% explained variance, for each subset of neurons. We found that these dimensionality estimates saturated with the number of neurons. **(c,d)** We repeated (ab) for random subsets of stimuli, rather than neurons. There were fewer stimuli than neurons to start with (2800 vs $\approx 10,000$), which explains why the dimensionality curves do not converge with the number of stimuli. **(ef)** For all combinations of subsets of neurons and stimuli, we estimated the dimensionality at 50% and at 95%. **gh** Same data as in (f) shown as scatter plots as a function of the number of stimuli (g) or neurons (h).

hold, the high spatial frequencies of the images should maintain a similarly low proportion of variance in the neural responses. To some extent, we know this hypothesis to be untrue: the retina is thought to be a bandpass filter that amplifies certain spatial frequencies and dampens others. In the mouse, it is thought that only low spatial frequencies pass through to V1, on the order of $0.05 - 0.2$ cpd, consistent with V1 response sensitivity curves [Niell and Stryker, 2010].

Nonetheless, it is possible that neural tuning to low spatial frequencies dominate response properties, while the tuning to high spatial frequencies resides in the small variance dimensions which we have estimated in the previous section. To show that this is not the case, we constructed an artificial stimulus set that only contained low spatial frequencies. In addition, each image was constructed from a set of 8 basis functions, chosen separately for each experiment as the spatial receptive fields of the

top PCs of the neural activity (Figure 3.5a). Specifically, we took each image in the original stimulus set, and projected it into these 8 spatial dimensions. The resulting images had little complexity left: they were 8-dimensional. We then repeated the same dimensionality estimation on neural responses to this stimulus set. We found that all previous results still held with this much more limited stimulus (Figure 3.5b-i). Neural responses were still very high-dimensional, and certainly much higher-dimensional than 8, the dimensionality of the images we presented.

We did observe an attenuation of the dimensionality curves beyond $\sim$50 dimensions. The power-law relation seemed to shift suddenly in favor of a larger power. We suspected this had to do with the localization of the images to a small area of the screen. The images presented lacked surrounds (Figure 3.5), which are known to significantly affect neural responses. To make a more direct comparison between spectra, we also showed another stimulus set where we localized the original natural images to the same area we localized the 8-dimensional images (Figure 3.6a). Not only did we observe similar scaling of estimated dimensionality across neurons and stimuli (Figure 3.6b-i), but quantitatively the spectra had similar decay over dimensions. We conclude that the surround provides an important component of the neural responses, particularly for the smaller variance dimensions, beyond the 50-th PC. This is consistent with previous studies in monkey visual cortex – responses to images with surrounds elicit sparser responses [Vinje and Gallant, 2002].

Usefulness of high-dimensional code for decoding

What might all the dimensions be useful for? One possibility is that they are used for object recognition, which would require enough information to be extracted from the stimulus. To determine how much stimulus-related information there is in the population, we performed a decoding analysis. We considered as training set the first repeat, and test set the second repeat, and asked how well we can guess the identity of the stimulus presented on the second repeat from the neural responses alone. Our decoder simply correlated the neural vector on the second repeat with all neural vectors on the first repeat, and guessed the maximum correlation as the most likely stimulus identity. Notice that the decoder is prone to overfitting, since we only have one repeat of each stimulus available to train it on. Nonetheless, we suspected that the size of the recorded population will somewhat make up for any overfitting.

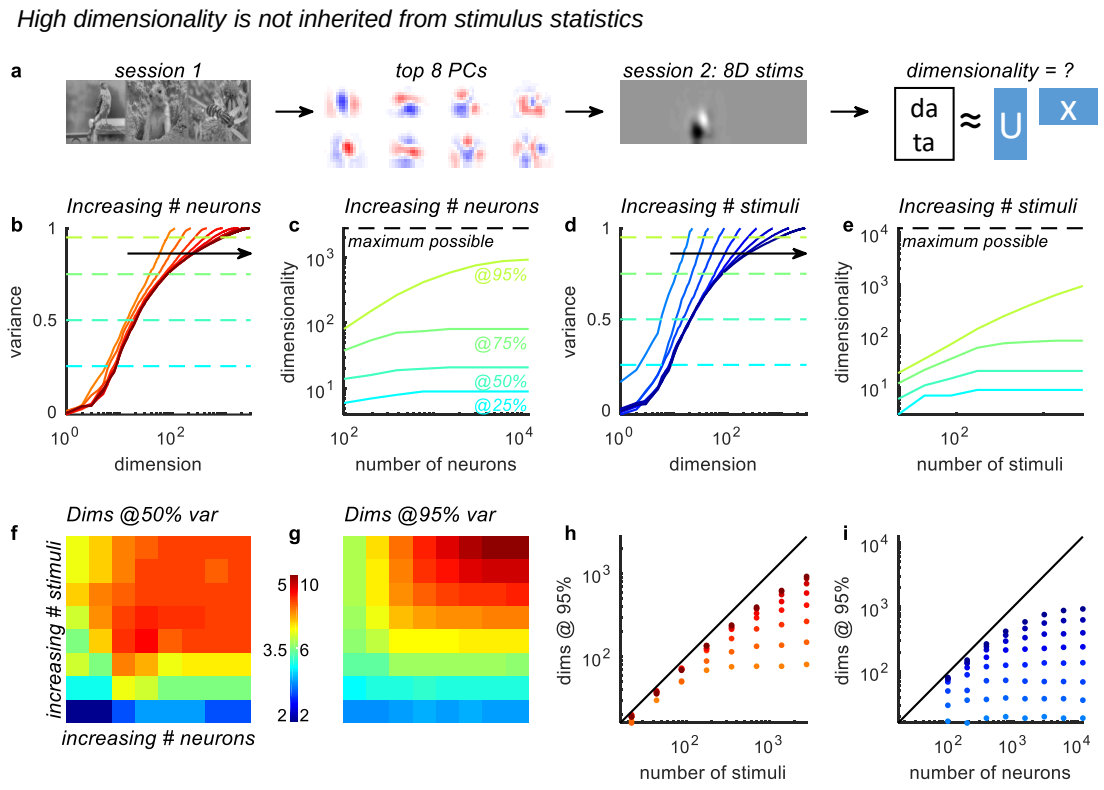

Figure 3.5: Dimensionality of 8-dimensional stimulus set. (a) We presented an artificial low-dimensional stimulus set, constructed from the original images, by projecting out all but 8 image dimensions. These 8 dimensions were chosen based on the neural responses in a first session, as the top PCs of the receptive fields of the population. (b-i) We then repeated all analyses from figure 3.4 on the neural responses to this constrained, 8-dimensional stimulus set.

Indeed, for all mice we were able to decode stimulus identity with more than 20% correct, reaching 80% correct for one mouse (Figure 3.7a and see mouse M2). Note that this classification task had a chance level of $\frac{1}{2800}$. It is thus remarkable that so much stimulus-related information is encoded by the neural population on a single-trial basis (≈10 bits).

In addition, we asked how many dimensions of the neural activity are needed to perform well in this task. A priori, one might expect that a dimensionality reduction approach would remove some of the noise, and thus allow better decoding with a small number of components. On the other hand, our analysis shows that the noise variance generally scales with the stimulus variance along any principal component (Figure 3.3g,h). Hence, even the smallest principal components might be able to contribute. The analysis reveals two surprising aspects: 1) performance is best when all principal components and all neurons are considered, and 2) for smaller numbers

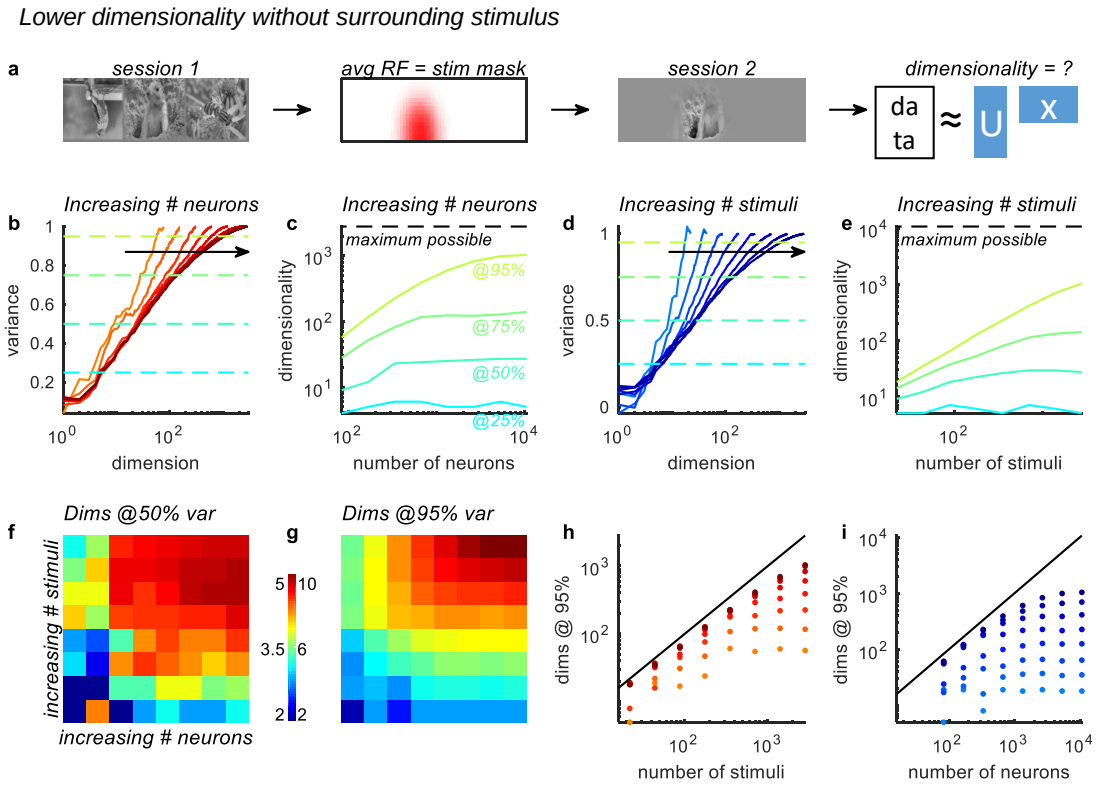

Figure 3.6: Dimensionality of spatially-localized stimulus set. (a) We presented an artificial stimulus set, in which the images were localized spatially. We constructed these images from the original images, by masking with the average receptive field of the recorded population. (b-i) We then repeated all analyses from figure 3.4 and 3.5 on the neural responses to this spatially-localized stimulus set.

of neurons, dimensionality reduction does have regularizing role and performance is best at an intermediate number of dimensions. This analysis shows that the full dimensionality of the neural responses helps in this decoding task, but this can only be appreciated when very large populations are recorded.

Discussion

We conclude that neural responses to stimuli are not limited to any lower dimensional subspace. The neural responses were distributed along dimensions in a $1/n^p$ fashion, where $p = 1$. Natural images also follow a power law spectrum (in the case of the natural images we presented, the power is $p = 1.6$). However, the neural circuit may not be inheriting the spectrum from the images. We observed a power law spectrum when we presented 8-dimensional images, which do not follow a power law spectrum. Because an input that lacked a $1/n$ spectrum resulted in neural

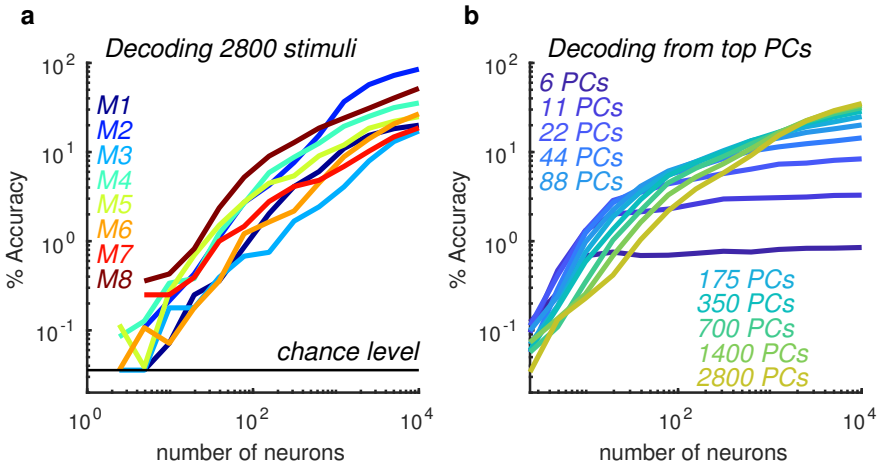

Figure 3.7: Decoding from 10,000 neurons. **(a)** For each of the 8 recorded populations, we computed decoding accuracy of the second repeat based on responses on the first repeat. We show the performance as a function of how many randomly selected neurons we consider out of the entire population. **(b)** For each population we project out all but the top N principal components from the first repeat, and redo the decoding analysis. Shown are averages over all mice.

activity which followed a power law, this suggests that this power law distribution may be produced by the architecture of the cortical circuit itself.

We observe $1/n$ scaling in simulations, when we consider a model with multiple rectifying neural layers, in which each layer is larger than the previous layer. This model expands the representation in each consecutive layer. Even with four-dimensional inputs, the model produces a power law scaling in its variance. Biophysical substrates for the stages of dimensionality expansion may be found in the retina, thalamus, or in the visual cortex itself. In visual cortex, multiple processing stages may proceed sequentially (i.e. starting in the input layers, then continuing in the superficial layers), or in parallel, via the recurrent dynamics of the local neural populations. For example, complex-cell computations could add dimensions of neural responses that are not linearly contained in the input dimensions.

Is there an advantage to having a power law scaling ($1/n^p$) of neural responses in which $p = 1$? Suppose the extreme case of $p = 0$. In this case, the neural stimulus representations would be orthogonal to each other. This would presume that there were no features in the natural images that were shared. However, some of these images are in fact similar to each other in pixel space. If each stimulus is orthogonal to each other in the neural space, then there would be no features in the stimuli that the neural activity could capture in a generalizable way.

If instead $p > 1$, then the first few dimensions will hold the majority of the variance. Because the activity is dominated by these first dimensions, the dimensionality of the responses will be lower. However, high-dimensional neural networks perform best on image classification tasks [Krizhevsky et al., 2012]. Projecting the input space into a high-dimensional space in order to perform feature computation and pattern separation is advantageous. Indeed the first layers of a deep network trained to perform image classification produce $1/n$ scaling in the variances of their principal components.

There are disadvantages to high-dimensional spaces. For instance, the representations may be over-fit to the inputs, making them less robust to perturbations. Small perturbations in single units may distort the representation sufficiently to decrease classification performance. One solution to this in the neural network literature is dropout: randomly removing units from the network during training to avoid over-fitting by forcing the network to produce multiple paths of information flow [Srivastava et al., 2014]. The brain also finds a robust solution to image classification: certain distortions of images are still recognizable to us. However, more research is needed to understand how the brain solves this task.

It is a topic of future research to uncover the precise functions computed throughout the visual pathway. Large-scale neural recordings promise to generate the required tuning data to enable the characterization of complex, distributed and high-dimensional systems like the visual cortex. The inferred functional tuning will provide valuable hypotheses to be tested with targeted connectomic mapping of synaptic connections.

Methods

Experimental

We recorded optically the neural activity of head-fixed awake mice implanted with 3-4mm cranial windows centered over visual cortex. We obtained $\sim 10,000$ neurons in all recordings. The recordings employed two-photon calcium imaging in combination with genetically encoded calcium indicators (GECIs, specifically GCaMP6s). The mice were free to run on an air-floating ball and were surrounded by

three computer monitors. We presented visual stimuli on these monitors arranged at 90^o angles to the left, front and right of the animal, so that the animal's head was approximately in the geometric center of the setup. Since our recording rates were relatively low (2.5-3Hz), we presented stimuli at relatively long intervals of 1-2 seconds, with the exact inter-stimulus interval randomized over a uniform range of values.

The recordings were performed using two photon calcium imaging, with multi-plane acquisition. The planes were spaced 30-35 μm apart in depth. 10-12 planes were acquired simultaneously at a scan rate of 2.5-3 Hz. We expressed GCaMP in large populations of neurons in one of two ways. Either we bred transgenic crosses of a GCaMP line and an excitatory cell driver (specifically EMX-CRE x Ai94 G6s, Rasgrf-CRE x Ai 94 G6s, or CamKII x tetO gcamp 6slow), or we injected non-specific AAV virus into transgenic lines expressing td-Tomato in GAD+ neurons. To obtain large fields of view for imaging 10,000 neurons, we typically performed 4-8 injections at nearby locations, sometimes at multiple depths ($\sim$500μm and $\sim$200μm). In this chapter, we only study the excitatory neurons from these recordings, but in Chapter 6 we analyze the inhibitory neurons as well. We implanted coverslips that were 3-4mm in diameter, allowing us, in the gcamp-transgenic animals, to select from several potential recording locations.

For each mouse imaged, we typically spent the first imaging day finding a suitable recording location, where the following three conditions held:

- the GCaMP signal was strong, in the sense that clear transients could be observed in large number of cells
- a large enough field of view could be obtained, to result in 10,000 neuron recordings,
- the receptive fields of the neuropil were localized inside the three monitors.

To determine receptive fields during imaging, we designed a stimulus that drives the neurons well, and allows for fast receptive field estimation using standard stimulus-triggered averages. This stimulus, called "sparse noise", contains a grid of large 5^o squares, which independently switch on for durations of 200ms. Each square only turns on every several seconds. Thus, at any one time, only a very small subset of squares are on. When a square is on, its brightness value is either white or black.

When a square is off, its brightness is gray. Thus, the image at all times is mostly gray, with some white and some black squares.

We found that it was sufficient to play this stimulus for 5 minutes, in order to get very accurate receptive fields by stimulus-triggered averaging (STA). To compute the STA, we divided out each frame into a grid of 3x3 or 4x4 regions, and averaged the brightness of all pixels in each region. Note that the dominant signal in most recordings was neuropil activity, not somatic spiking. We then estimated the increase in fluorescence caused by each square on the screen independently, over a baseline level preceding the stimulus, averaged over all presentations of that particular square. Since the stimulus was spatially white, i.e. uncorrelated, the STA produces a consistent estimate of the receptive fields.

In those animals where we had a choice over multiple valid recording locations (typically in the gcamp transgenic animals), we chose either: 1) a horizontally and vertically central retinotopic location or 2) a lateral retinotopic location, at 90^o from the center, but still centered vertically. This was in order to obtain enough populations in different mice at the same recording location, so we can combine the yields and further increase our number of neurons for some of the analyses below. We did not observe significant differences between recordings obtained from GCaMP transgenic animals, or from virus injections. Thus, we pooled data over all such recordings. We also did not observe differences related to retinotopic location (central or lateral), thus we pooled data across different recording locations as well.

Stimuli

We presented two types of stimuli: drifting gratings and flashed natural images.

First, we used drifting gratings with 96 different parameters to characterize the neurons' classical properties, like orientation tuning, spatial frequency tuning or temporal frequency tuning. For orientation tuning, we also separately presented drifting gratings narrowly spaced at 15^o, as we found that a large portion of the population had very narrow tuning curves, and larger spacing between grating orientation would risk missing completely a many neurons' preferred orientation. The characterization using drifting gratings primarily served to confirm the tuning properties of the neurons, and ensure that our recording method (10,000 neurons at 2.5Hz) was not too impoverished and noisy to record proper neural spiking

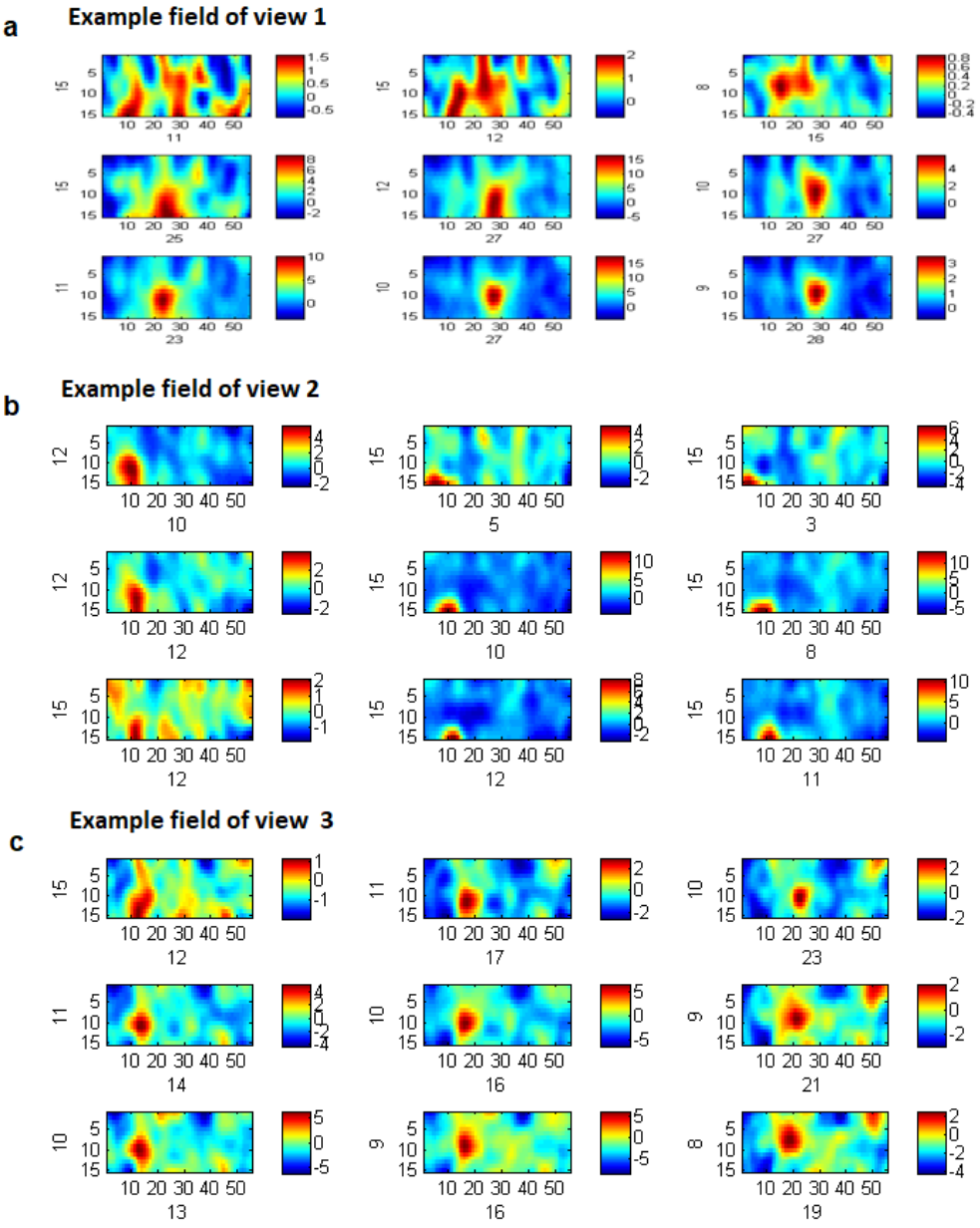

Figure 3.8: Online receptive fields. We presented a sparse noise stimulus for 5 minutes in each field of view, and estimated the average receptive fields of the neuropil online, during the recording session. This allowed us to center our recordings in fields of view with good stimulus responses, and centered on our preferred retinotopic locations. **(a)** In this experiment, the upper third of the field of view did not show good responses to stimuli, and thus produced poor receptive fields. **(b)** In this experiment, the field of view was at the bottom of the screen. We thus avoided recording here, to avoid potential artifacts due to the screen edges. **(c)** In this experiment, all quadrants of the image had good responses and localized receptive fields. The receptive field centers can be seen to shift consistently up and to the right (from the left to the right quadrants, and from the bottom to the top), consistent with known V1 topography.

responses. In fact, we found that we lost very little information by recording at 2.5Hz compared to $30Hz$, because the amount of stimulus-tuned variance in each neuron's responses was only slightly lower at 2.5 Hz.

Having confirmed good stimulus response properties, using classical characterizations, we presented up to 2,800 flashed natural images, covering all three screens. The images were manually selected from the ImageNet database, from ethologically-relevant categories: "birds", "cat", "flowers", "hamster", "holes", "insects", "mice", "mushrooms", "nests", "pellets", "snakes", "wildcat". We chose images where the subjects tended to fill out the image (less than 50% of the image was a uniform background), and if the images contained a good mixture of low an high spatial frequencies. We presented these images in one of two configurations. Either: 1) 2800 images presented twice, or 2) 32, 64, or 112 images presented between 32 and 128 times. The first configuration allows us to estimate the diversity of neural responses to a large range of images. However, it does not allow us to estimate precisely the trial-averaged responses. For this, we used the second configuration.

Data preprocessing

The pre-processing of all raw calcium movie data was done using a toolbox we developed called Suite2p, using the default settings [Pachitariu et al., 2016b]. Briefly, Suite2p aligns all frames of a calcium movie using 2D rigid registration based on a regularized form of phase correlation, subpixel interpolation and kriging (Appendix A). For all recordings we validated the inferred X and Y offset traces, to monitor any potential outlier frames that may have been incorrectly aligned. In a very small percentage of all recordings, frames that had artifacts were removed and replaced with NaNs. In all recordings, the registered movie appeared well-aligned by visual inspection. The recordings were corrected for motion artifacts in the depth of the imaged tissue volume (Appendix B).

Then, automated cell detection and neuropil correction was performed using Suite2p (Figure 3.9). To detect cells, Suite2p computes a low-dimensional decomposition of the data, and uses the decomposition to run a custom clustering algorithm that finds regions of interest (ROIs) based on the correlation of the pixels inside them. The extraction of ROIs stops when the potential ROIs drop below a certain correlation value, which is set as a fraction of the strongest available ROIs.

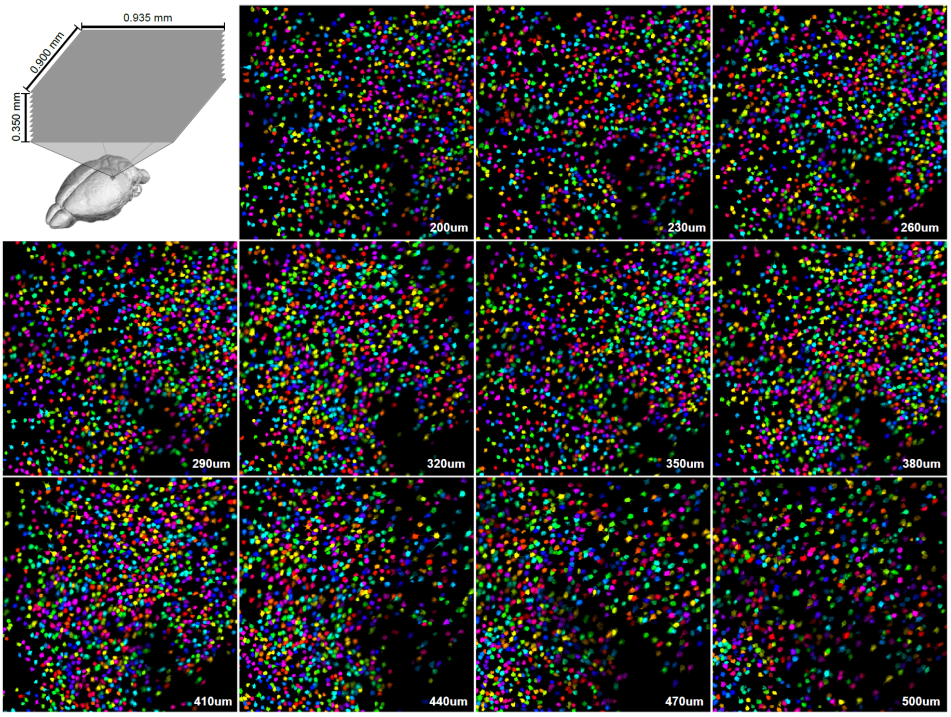

Figure 3.9: Output of Suite2P ROI detection: 13,451 simultaneously recorded cells. Imaging of an entire local cortical population, recorded from 11 planes in mouse visual cortex (GCaMP6s expressed in pyramidal cells of all layers using an Ai94; Camk2a-tTA; Emx1-Cre triple transgenic), using a standard Thorlabs B-scope with resonant scanning at 2.5Hz frame rate. The pseudocolor masks of all 13,451 detected cells are shown.

Thus, Suite2p does not require the number of clusters to be set a priori. Furthermore, in all cases that we inspected, Suite2p found all the large sources of activity in the field of view, whether they were somatic or dendritic in origin. A further step in the Suite2p GUI classifies these ROIs as somatic or not, partially based on user input, which is used to train a classifier. The classifier reaches 95% estimation of somatic/non-somatic signals on this data [Pachitariu et al., 2016b], thus allowing us to skip the manual step altogether for most recordings. We note that the 5% errors might either be attributable to human labelling error, or to dendritic signals, which would nonetheless most likely reflect backpropagating APs, which also directly measures the spiking signal for deeper cells. Thus, we were not worried about some of our ROIs potentially not being somatic. However, we were worried about the contamination with the surround neuropil signal, which has the potential to skew several of our analyses. This contamination is large in two-photon calcium imaging, and typically removed by subtracting out from the ROI signal a scaled-down version of the neuropil signal around the ROI. The precise scaling factor is set either to a

uniform value (0.7) or optimized to maximize the skewness, or the transient-characteristic of somatic spiking (the Suite2p approach). Importantly, for computing the neuropil signal, we excluded all pixels that Suite2p attributed to an ROI, whether this was a somatic or dendritic ROI.

Computational analyses

For the most part, we describe computational analyses in the specific section they are being used. One of the methods we use repeatedly here and in the next chapters is the unbiased estimation of signal-related variance on single-trials (like in Figure 3.1b for single neurons, and Figures 3.3e-i for principal components). Here we define this quantity formally and show that it can be estimated from two (or more) presentations of a stimulus set. The signal-related variance is at the core of the new dimensionality estimation method we introduce, which requires us being able to estimate the signal variance along any projection of the data.

Estimation of single-trial, signal-related variance

We consider the recorded neural response $f_k(n,i)$ of neuron n to stimulus i on repetition k. This can always be rewritten as

$$f_k(n,i) = \mu(n,i) + \varepsilon_k(n,i)$$

where $\mu(n,i)$ is the trial-averaged response of neuron n to stimulus i over an idealized, infinite number of repetitions of stimulus i, while $\varepsilon_k(n,i)$ is the trial-to-trial variability, or "noise". We would like to estimate the tuning of $\mu(n,i)$ across stimuli i, and summarize this tuning by a scalar quantity

$$V_n = \mathrm{E}_i\left[\left(\mu(n,i) - \mathrm{E}_i\left[\mu(n,i)\right]\right)^2\right],$$

which we call the stimulus-driven variance. We use the notation $\mathrm{E}_i[X]$ to denote the expectation of the quantity X over an infinite number of repeats i. By construction $\mathrm{E}_i\left[f_k(n,i)\right] = \mu(n,i)$ and $\mathrm{E}_i\left[\varepsilon_k(n,i)\right] = 0$.

We note that $\varepsilon_k(n,i)$ could depend on the neuron and stimulus. Our derivation makes no assumptions on these dependencies. However, we do assume that $\varepsilon_k(n,i)$

does not depend on the repetition number, i.e. for all k, $\varepsilon_k(n,i)$ are independent random samples from the same distribution. This last condition can be approximately achieved in practice by separating the presentation of the stimulus repeats by tens of minutes. This is necessary in order to avoid temporally correlated noise. Furthermore, we z-score across stimuli the responses of neuron n during repeat k, to remove the influence of temporally correlated noise with very long autocorrelation timescale, such as from global changes in state.

Under these assumptions, we show below that we can estimate the total stimulus-related variance as $V_n = Q_n - M_n^2$, by estimating the quantities $M_n = \mathrm{E}_i[\mu(n,i)]$ and $Q_n = \mathrm{E}_i\left[\mu(n,i)^2\right]$. It is clear that an unbiased estimate of M_n is given by $\frac{1}{N}\sum_i f_1(n,i)$. This can be formally shown by taking the expectation of the latter quantity with respect to the random variables $\varepsilon_1(n,i)$

$$\begin{aligned}\mathrm{E}_{\varepsilon_1}\left[\sum_i f_1(n,i)\right] &= \sum_i \mathrm{E}_{\varepsilon_1}\left[\mu(n,i)+\varepsilon_1(n,i)\right] \\ &= \sum_i \mu(n,i) + \mathrm{E}_{\varepsilon_1}\left[\varepsilon_1(n,i)\right] \\ &= N \cdot M_n\end{aligned}$$

The last equality follows because the random variables $\varepsilon_1(n,i)$ are coming from distributions with mean zero, so $\mathrm{E}_{\varepsilon_1}\left[\varepsilon_1(n,i)\right] = 0$ by construction. To improve the estimator, we can average it over all repeats k.

We now show that an unbiased estimate of Q_n is given by $\frac{1}{N}\sum_i f_1(n,i)\cdot f_2(n,i)$. The latter quantity equals Q_n in expectation over the random variables $\varepsilon_1, \varepsilon_2$:

$$\begin{aligned}&\mathrm{E}_{\varepsilon_1,\varepsilon_2}\left[\sum_i f_1(n,i)\cdot f_2(n,i)\right] = \\ &= \mathrm{E}_{\varepsilon_1,\varepsilon_2}\left[\sum_i (\mu(n,i)\cdot\mu(n,i) + \mu(n,i)\cdot(\varepsilon_1(n,i)+\varepsilon_2(n,i)) + \varepsilon_1(n,i)\cdot\varepsilon_2(n,i))\right] \\ &= \sum_i \mu(n,i)^2 + \sum_i \mu(n,i)\cdot(\mathrm{E}_{\varepsilon_1}[\varepsilon_1(n,i)] + \mathrm{E}_{\varepsilon_2}[\varepsilon_2(n,i)]) + \sum_i \mathrm{E}_{\varepsilon_1}\left[\varepsilon_1(n,i)\right]\cdot E_{\varepsilon_2}\left[\varepsilon_2(n,i)\right] \\ &= N\cdot Q_n.\end{aligned}$$

The last equality again follows because the random variables $\varepsilon_k(n,i)$ are coming from distributions with mean zero, so $\mathrm{E}_{\varepsilon_k}\left[\varepsilon_k(n,i)\right] = 0$ by construction.

Motivation for the new dimensionality reduction method

We devote an entire appendix to the derivation of the new dimensionality estimation method, that is less biased than standard approaches. This method allowed us to estimate the singular value spectrum of a matrix from two noisy observations of that matrix. In our case, the matrix is formed of the trial-averaged responses of $\sim 10,000$ neurons to $2,800$ stimuli. Each presentation of a stimulus adds independent Poisson noise to the matrix, correlated neural variability, as well as recording noise from the two-photon Calcium imaging approach. Thus, each repeat is noisy, with only $\sim 10\%$ of the recorded signal attributable to the stimulus identity (see Figure 1). To remove the bias from the noise, previous approaches typically average out responses over tens or hundreds of trials. However, since our total recording time is limited, this would only allow us to present 32-128 different images (which in some recordings we do). Thus, we would not be able to estimate and appreciate the large diversity of responses to different images, and specifically we could not estimate a dimensionality larger than the maximum number of presented stimuli. Thus, in the main recordings we analyze, we present each image only twice, allowing us to show 2800 different images. The new dimensionality estimation method is therefore crucial to a correct interpretation of the data.

4

Multi-dimensional spontaneous activity in awake mice

In Chapter 3, we observed high-dimensional stimulus-evoked activity in the visual cortex of awake mice. In these recordings, stimulus encoding did not appear to be impaired by one-dimensional spontaneous fluctuations of the sort we studied in non-aroused or anesthetized recordings in Chapter 2. Seeing that the one-dimensional population-wide fluctuations appeared to be quenched, we investigated whether other forms of structured activity may be present in the awake state. To quantify this structure, we recorded spontaneous activity over multiple hours, in the absence of visual stimuli using two-photon calcium imaging. We start by showing that the mean activity of the 10,000 cell populations was not dominated by widespread "up" and "down" states, and the mean pairwise correlations were near zero. Near-zero mean correlations may originate from a decorrelated neural state of independent neural firing (the desynchronized state), but they may also result from a balancing of large positive and large negative correlations. Our data clearly supported the second possibility: the correlation matrix was highly structured, and repeatable over separate halves of the data. Using a multi-dimensional model, we estimated that $\sim$100 components were significant. Of these, the top principal component was correlated with the running speed and the pupil area of the mouse. To visualize the structure of the population activity, we fit a non-negative matrix factorization model. The resulting clustering of neurons over components had a nearly random spatial distribution over the $\sim$0.3 mm^3 imaged volume, showing that the structured neural activity does not consist of spatial modes.

Introduction

In Chapter 2, we characterized the neural dynamics of populations of tens of neurons recorded simultaneously. These recordings were dominated by a single mode of variability which activated most of the neurons in the population. In awake brain states, we observed a decrease in the low-frequency fluctuations of cortical activity, consistent with previous studies. This decrease in low-frequency fluctuations was associated with an increase in inhibitory activity in our recordings and in our simulations, which consequently increased the fidelity of stimulus encoding. In awake, two-photon calcium imaging experiments, we observed reliable and flexible stimulus responses, with more than a thousand dimensions of stimulus-driven activity represented by populations of ∼10,000 neurons (Chapter 3).

In these awake states with diverse stimulus responses and near-zero mean pairwise correlations, is there any structure remaining that is not purely driven by external stimuli? To answer this question, we recorded the activity of ∼10,000 neurons in the visual cortex of awake mice in the absence of visual stimuli. We found that pairwise correlations were on average near zero, but there were significant positive and negative correlations that balanced on average. We developed a method to compute the dimensionality of such spontaneous neural activity, and used it to estimate that around ∼100 linear dimensions were explored by the neural population. These dimensions were not spatially organized in the imaged tissue.

Results

Pairwise correlations: near zero on average, but highly significant individually

In awake, engaged animals, it has been shown that neurons have low mean pairwise correlations relative to passive or anesthetized animals [Cohen and Maunsell, 2009, Ecker et al., 2010]. This does not necessarily imply that neurons are firing independently from each other. Low mean correlations can result from all pairwise correlations being low, or from a balancing of significant positive and negative pairwise correlations.

We investigated the structure of pairwise correlations in the visual cortex of

awake mice. The neural activity was binned in time bins of 1.2 seconds and the pairwise correlations among all neural pairs computed. We found that the mean pairwise correlations were low, but there was a wide distribution of pairwise correlations around their mean (Figure 4.1a). To show that this distribution is not due to noise, we split each neural recording into two halves (interleaved segments of 40 seconds each) and computed the pairwise correlation matrix of each half (Figure 4.1b). Let us call these matrices C^1 and C^2. The correlation between C^1 and C^2 was on average $r = 0.71$ (Figure 4.1c). Using the definition of "signal variance" from section 3.4.4.1, it follows that ~71% of the variance in the correlation matrix is "signal" and thus repeatable across non-overlapping segments of the recording. Thus, a majority of the correlations observed in Figure 4.1a are not due to noise. This shows that although the mean correlations were near zero, neural structure was not absent.

The top principal component of spontaneous activity is arousal

To quantify the structure of spontaneous activity, we started by obtaining its principal components (PCs). To visualize the activity, we used the weights of the top PC to sort the neurons in Figure 4.1a. The population could be clearly differentiated into neurons that were correlated and anti-correlated to this first principal component. As expected from the correlation matrix, there was a spread in the weights of these neurons' projections onto the first principal component. To get an intuition for this top PC across mice, we plotted the top PC for 12 recordings in Figure 4.2b.

We also extracted several behavioral variables, and plotted them together with the top PC in Figure 4.2b. The pupil area was estimated from a video of the mouse's face, using a robust, custom-made pupil detection method (see Chapter 5 for more information). The running speed of the mouse was extracted from optical mice that recorded the position of the air-floating ball, on which the mouse was free to run. Pupil area and running speed were correlated to each other in all recordings (average $r = 0.49$) (Figure 4.2b). These measures are often used to monitor animals' overall arousal [McGinley et al., 2015b, Vinck et al., 2015].

The arousal level of the mice was highly correlated with the principal component of the neural activity. Across 12 recordings in 9 mice, running speed was highly correlated with the first principal component, with the correlation variying from $r = 0.17$ to $r = 0.74$, with a median correlation of $r = 0.69$ (Figure 4.3a). The

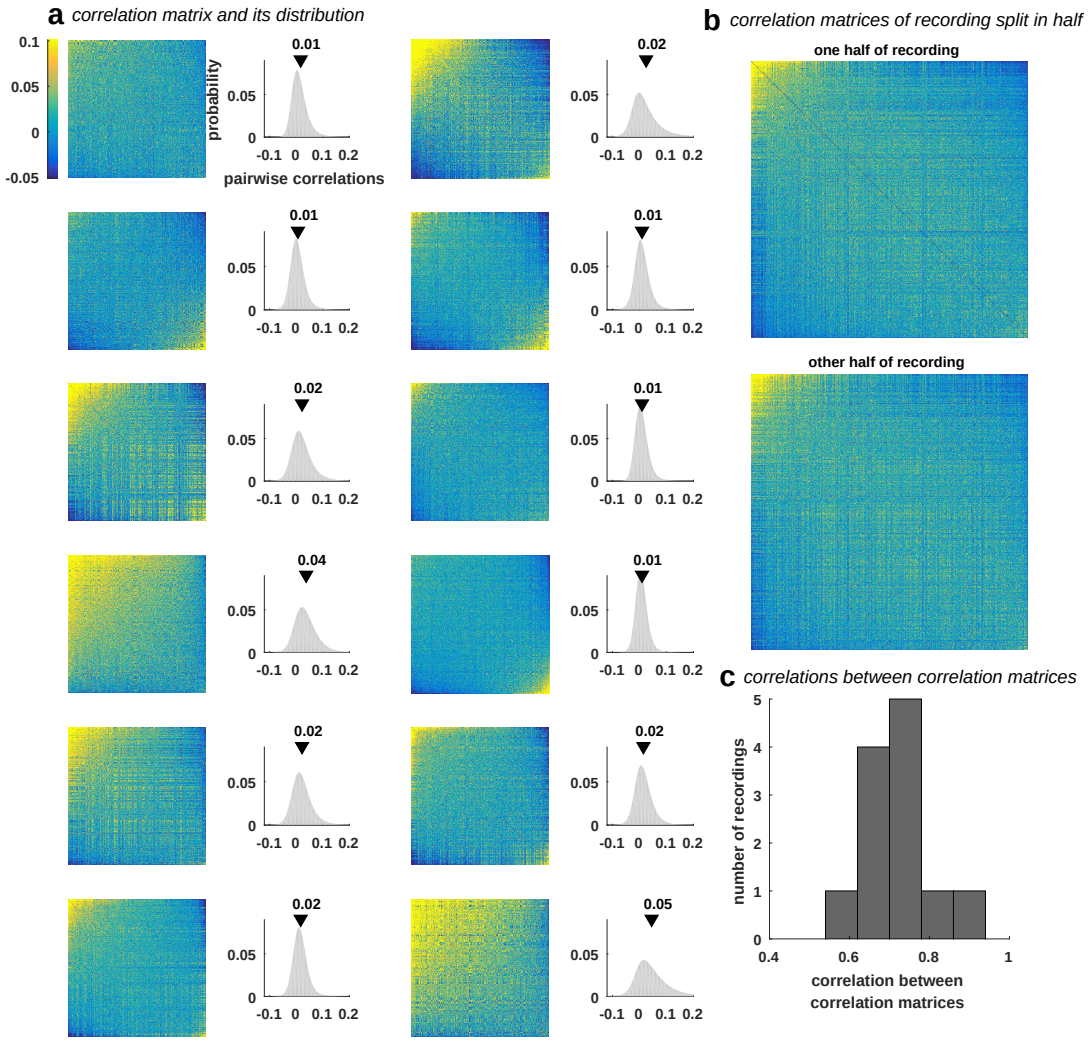

Figure 4.1: Correlation matrices of spontaneous activity in visual cortex. **(a)** The first and third columns are the matrices of correlations between pairs of neurons computed in bins of 1.2 seconds. Each matrix is from a different neural recording. The neurons in the matrix are sorted by their weight onto the first principal component of neural activity. The matrices' distributions of pairwise correlations are shown to their right. The mean of the correlation matrix is indicated by a black inverted triangle. **(b)** One neural recording was split in half in time (interleaved segments of 40 seconds each). The correlation matrix for each half is plotted. The correlation between these two matrices is $r = 0.68$. **(c)** We repeated this procedure for each recording. The correlation between correlation matrices computed on two separate halves of the data is plotted as a distribution across recordings. The mean of this distribution is $r = 0.71$.

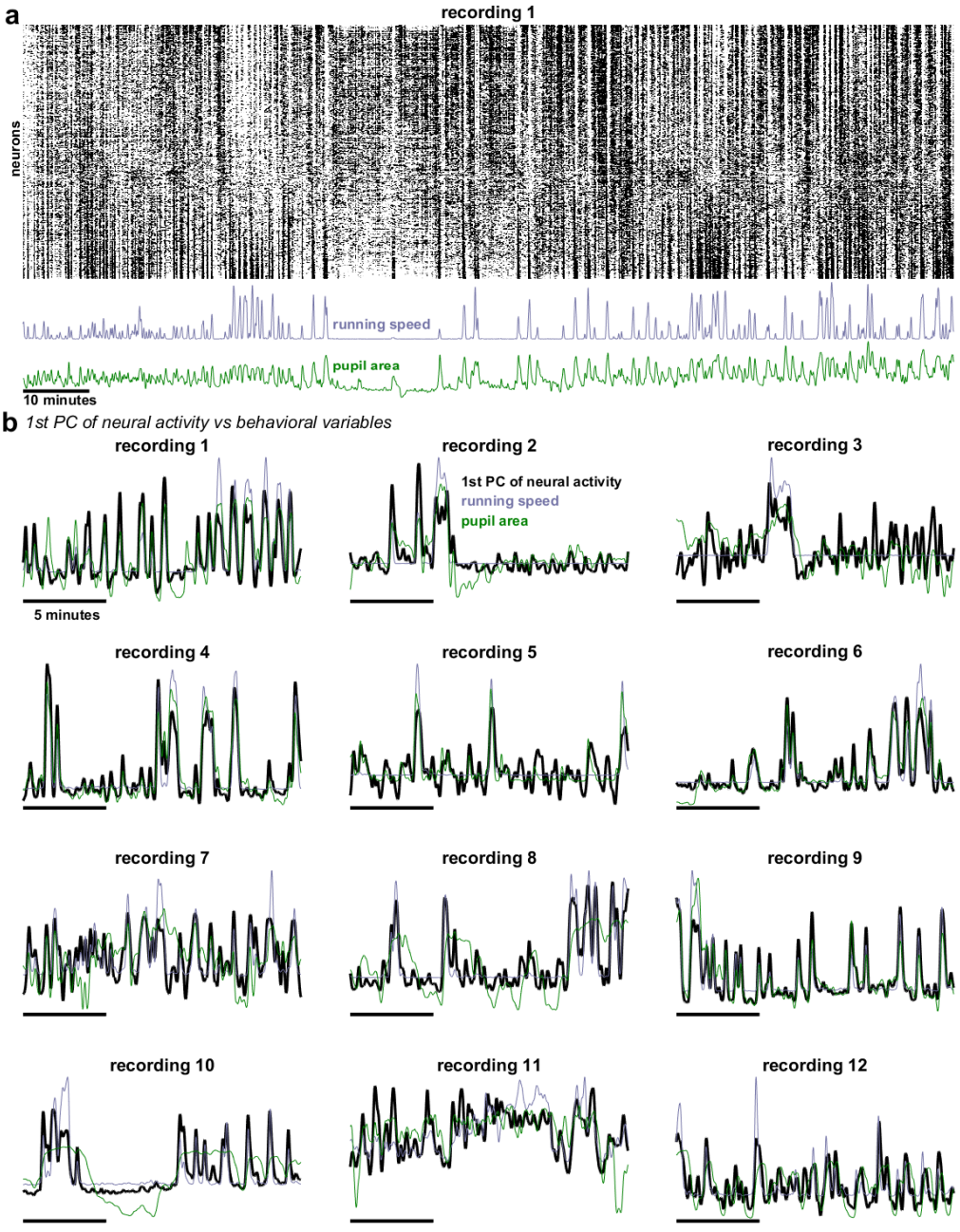

Figure 4.2: The first principal component of spontaneous neural activity and its relationship to behavioral factors. (a) Neural activity from a single recording. Neurons are sorted by their weights onto the first principal component. Their activity is z-scored and smoothed in time by a Gaussian with standard deviation of 40 seconds. The running speed, pupil area, and face motion are also smoothed in time by a Gaussian with standard deviation of 40 seconds. **(b)** The first PC of the neural activity, the running speed, the pupil area and the face motion are z-scored and smoothed by a Gaussian with a standard deviation of 3.6 seconds. The time segment plotted is 17 minutes long. Each plot is from a different recording.

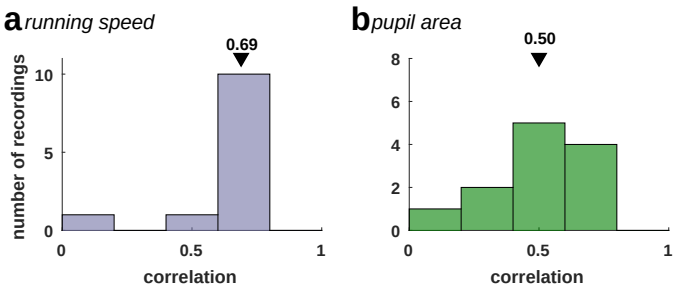

Figure 4.3: Behavioral factors relate to the first principal component of neural activity. (a) The correlation of running speed with the first principal component of neural activity in 1.2 second bins was computed for each recording. The distribution of correlation values is plotted. The median value is plotted as a black inverted triangle. **(b)** Same as **a**, but the correlation of pupil area with the neural activity. **(c)** Same as **a**, but the correlation of face motion with the neural activity.

pupil area was also highly correlated with the top PC, with the correlation varying from $r = 0.30$ to $r = 0.74$ with a median correlation of $r = 0.50$ (Figure 4.3b).

Although a strong predictor of the arousal level of the mouse, the top PC did not account for much of the neural activity: it only explained 2.5% of the variance at the level of single neurons (Figure 4.4). How many more relevant dimensions of variability there are in spontaneous activity?

Dimensionality of spontaneous activity

Spontaneous neural activity explores a space of 100 linear dimensions

In order to determine the dimensionality of spontaneous activity, we fit a multi-dimensional model and computed its average variance explained as we varied the number of dimensions. We modeled neural activity using peer prediction: we predicted the neural activity of one half of neurons from the other half of neurons. The prediction was based on neural projections from one half of the recording (split in time) and tested on the other half of the recording to produce a cross-validation score. We prevented the peer prediction model from overfitting by using the singular value decomposition to regularize the prediction (see section 4.4.2.1 and Figure 4.10 for more details).

We computed the peer prediction model for increasing numbers of dimensions of the singular value decomposition and predicted the neural activity (Figure 4.4). In

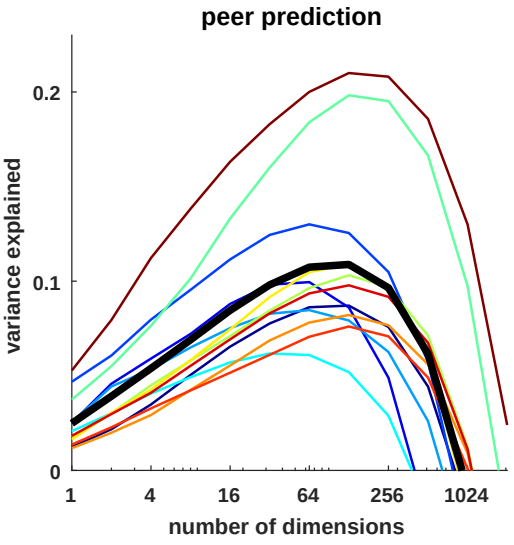

Figure 4.4: Cross-validated variance explained by peer prediction model. Each curve is the cross-validated variance explained as a function of the number of dimensions used in the peer prediction model for a single recording. The black curve is the average across all recordings. The one-dimensional model explained on average 2.5% of the variance. The peak of the curve was at 11% and it was achieved by a model with 128 dimensions.

the best recording, the peer prediction model accounted for 22% of the total neural activity variance in time bins of 1.2 seconds. Averaging across all recordings, the model peaked at 11% cross-validated variance explained. This peak was achieved by models with 128 dimensions. The variance explained decreased after 128 dimensions, suggesting that models with more than 128 dimensions were overfitting. We conclude that 128 dimensions of spontaneous neural activity are significant, and account for at least 11% of the recorded neural variance. The remaining variance may be unpredictable due to the Poisson-like neural spiking and the noise introduced by the neural recording method.

The correlation matrix also spans approx 100 dimensions

We next computed the dimensionality of the correlation matrix: we built a linear model of varying dimensions of the spontaneous activity correlation matrix and computed the variance explained by the model. We used the singular value decomposition of the correlation matrix as a linear model of the correlation matrix. We fit this model to one half of the recording, and then tested its performance on the

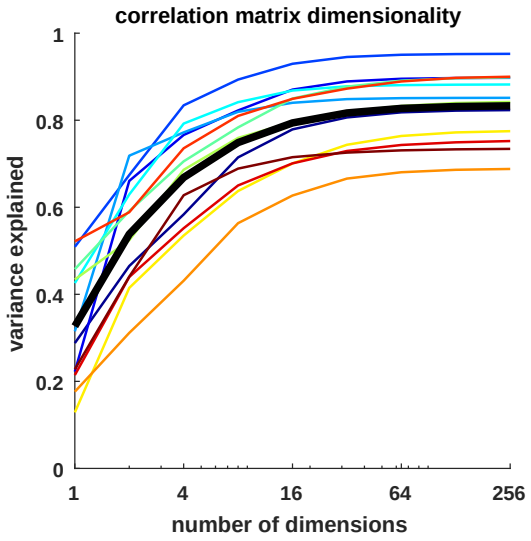

Figure 4.5: Cross-validated variance explained of the pairwise neural correlation matrix. Each curve is the variance explained of the correlation matrix as a function of the number of dimensions for a single recording. The variance explained is normalized by the fraction of explainable variance. The average variance explained is plotted in black. The average curve started at 33% variance explained and peaked at 83% variance explained. It reached 95% of this value (79%) with a model with 15 linear dimensions.

other half of the recording on half of the neurons, resulting in a cross-validated estimate of the variance explained of the model. To improve our model, we rescaled the singular values in accordance with the neural activity of the other half of the neurons (see section 4.4.2.2 for more details).

The one-dimensional model of the correlation matrix (the first principal component) explained a large fraction of the variance. The first principal component explained 18% to 52% of the variance, with an average variance explained of 33%. The average variance explained across recordings peaked at 83%. This suggests that a linear model such as singular value decomposition does a reasonable job of estimating the correlation matrix structure. It reached 95% of the maximum variance explained value (79%) with a model containing 15 linear dimensions, but needed at least 100 dimensions to saturate performance. The slow saturation in prediction is not surprising, as the singular values of the correlation matrix are approximately the squares of the singular values of the raw data, and thus decay more quickly.

Visualizing multiple dimensions of neural activity

The singular value decomposition allowed us to quantify the linear dimensionality of the spontaneous activity. However, singular vectors are allowed to have both positive and negative weights, thus they have a dense, unintuitive representation of the neural activity. The activity of single neurons is represented by a sum of many positively- and negatively-weighted singular vectors. For visualization, we wanted to obtain a sparse, multi-dimensional representation, in which each neuron is the sum of a few positively-weighted activity patterns. We were able to find such representations by running a non-negative matrix factorization (NMF) on the data [Lee and Seung, 1999].

NMF decomposes a matrix into two non-negative matrices of lower rank than the original matrix. In our case, we approximated the neural activity matrix F by two low rank matrices W and H

$$F \approx WH \quad \text{where } W \geq 0,\ H \geq 0,$$

using gradient descent to optimize W and H. We chose the rank of W and H to be 15 dimensions.

We fit W and H to one recording and plotted the neurons' activities sorted by their weights in the W matrix (Figure 4.6a). NMF found several dimensions of activity with different temporal dynamics. We computed the autocorrelation functions of each of these clusters in time (H), and found that their timescales varied from a few seconds to tens of seconds (Figure 4.6c). There were several distinct clusters of neurons correlated with running speed, and several anti-correlated. The structure uncovered by NMF was apparent in the pairwise correlation matrix among neurons (Figure 4.6b).

Figure 4.7 shows the neural activity sorted by non-negative matrix factorization in each mouse from which we recorded. There were similarities and differences between these patterns across mice. What might these patterns represent? One immediate possibility is that neural activity is clustered spatially, according to the position of each neuron in tissue. This hypothesis is supported by anatomical studies which suggest that neural connectivity decreases as a function of distance [Hellwig, 2000, Levy and Reyes, 2012].

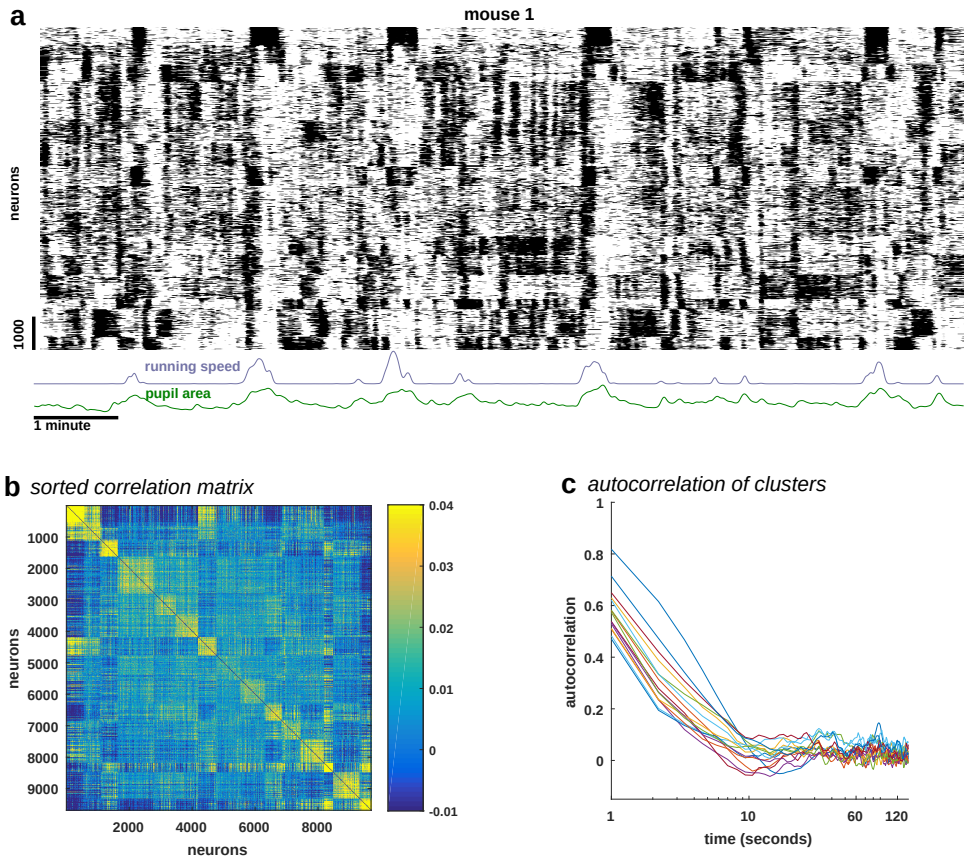

Figure 4.6: Spontaneous neural activity sorted using non-negative matrix factorization. (a) 15 minute segment of neural activity from a single recording. Neurons are sorted by their cluster identity as defined by NMF. Their activity is z-scored and smoothed in time by a Gaussian with standard deviation of 1.2 seconds. The running speed, pupil area, and face motion are also smoothed in time by a Gaussian with standard deviation of 1.2 seconds. **(b)** The pairwise correlation matrix among neurons sorted by NMF clustering as in **a**. The correlations are computed in bins of 1.2 seconds. **(c)** The autocorrelation functions of the time components extracted from NMF decomposition (H) matrix, binned in 1.2 second bins. In this example, the autocorrelation functions of the cluster activity decayed within 10-20 seconds.

Spontaneous neural activity is not spatially clustered

We tested whether individual NMF components were spatially localized to patches of neurons that activate together. We used the positive NMF weights to define a clustering of the neurons. Each neuron was assigned to a single cluster based on its maximal weight in W across components. An example of what this assignment produces is shown in Figure 4.8. Each neuron in the recording is pseudo-colored based on its cluster identity and plotted at its XY location. It is difficult to visually distinguish any clear spatial patterning of the cells according to their cluster identity.

We quantified the spatial organization of the clusters by computing the

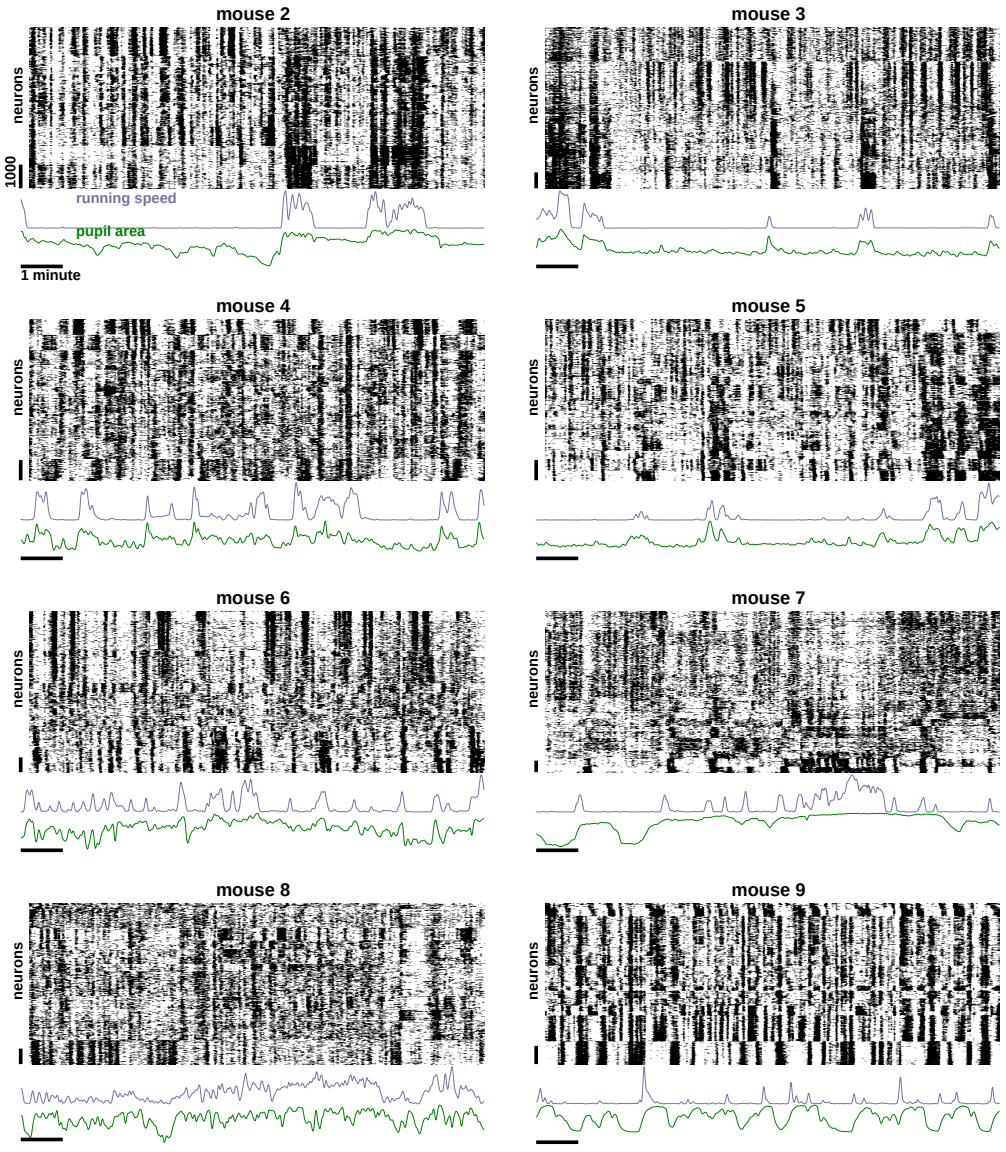

Figure 4.7: Spontaneous neural activity sorted using non-negative matrix factorization - all recordings. Each raster is a 13 minute segment of neural activity from a different mouse (the first mouse's neural activity is shown in Figure 4.6. Neurons are sorted by their cluster identity as defined by NMF. Their activity is z-scored and smoothed in time by a Gaussian with standard deviation of 1.2 seconds. The running speed, pupil area, and face motion are also smoothed in time by a Gaussian with standard deviation of 1.2 seconds.

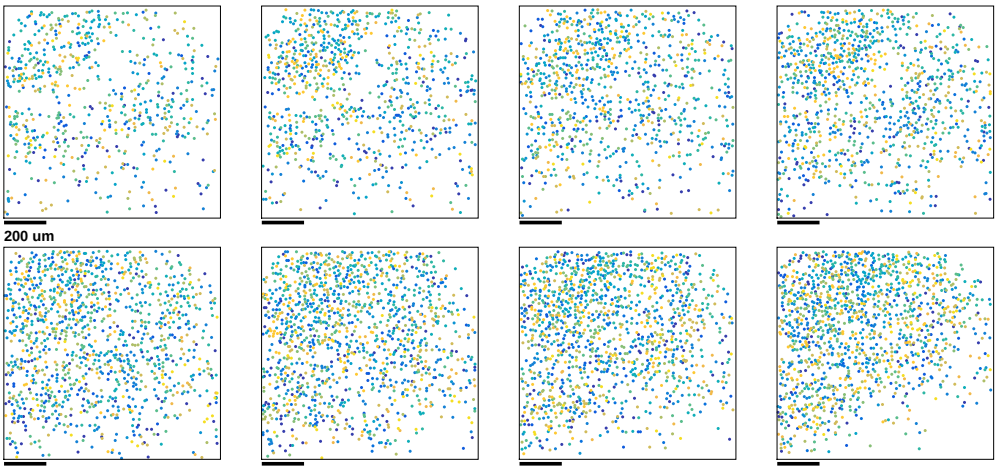

Figure 4.8: Example of cluster identities in one recording. Planes from the recording in Figure 4.6. Each box is a different plane in the recording, sorted in order of increasing depth. Each point is a neuron assigned to a single cluster based on its maximal weight in W across components, and pseudo-colored based on its cluster identity.

distributions of their locations and found that the clusters were not spatially localized in depth. All clusters spanned all planes of the recording (Figure 4.9a). The standard deviation of the depth distribution of the clusters varied from 75 μm to 89 μm, with an average standard deviation of 85 μm. The average standard deviation of the depth distribution of cells not within a specific cluster was 86 μm. The spread of the clusters in depth were not significantly different than the out-of-cluster spreads (Wilcoxon sign-rank test, p=0.81).

Were the clusters structured in their XY positions? We computed the pairwise distance among all cells from the same NMF cluster and call this the "in-cluster distance". We then computed the pairwise distance between cells in one cluster to all cells in other clusters and call this the "out-of-cluster distance". An example of the distribution of in-cluster distances and out-of-cluster distances for a neural activity cluster from the recording in Figure 4.6 is shown in Figure 4.9b. The distributions are similar but their means are slightly different. If the mean in-cluster distances and mean out-of-cluster distances are different, then this would signify non-random spatial organization of the neural activity clusters. We computed the mean in-cluster distance and mean out-of-cluster distance for all clusters in the recording from Figure 4.6. The median in-cluster distance was 477 μm and the average out-of-cluster distance was 482 μm. Each cluster was plotted as a point in Figure 4.9c. The majority

of clusters fell near the unity line. In this recording the difference between in-cluster distance and out-of-cluster distance was barely significant (Wilcoxon sign-rank test, $p = 0.041$). Across all recordings, the median in-cluster distance was lower than the median out-of-cluster distance (463 μm versus 498 μm). However, this difference in distances is arguably small. 36 μm is three to four times the diameter of a cell, and the recording field of view is 1 mm by 1 mm. Thus, 36 μm is only 3.6% of the distance across the recording field of view.

Discussion

In the absence of visual stimuli, the visual cortex of mice is nonetheless spontaneously active. This activity is structured across neurons. In Chapter 2, we found that the activity was dominated by a single one-dimensional mode of activity in recordings of 30-100 neurons in sensory cortical areas. However, in larger recordings, we observed neurons both positively and negatively correlated to the first principal component of neural activity. This first principal component was related to the overall arousal level of the mouse, as measured by running speed or pupil area, and thus there exist neurons both positively and negatively correlated with arousal [Pakan et al., 2016, Dipoppa et al., 2016]. Previous claims about independent neural activity, due to near zero average correlations need to be re-examined [Cohen and Maunsell, 2009, Ecker et al., 2010, Renart et al., 2010].

Instead, we observed multiple dimensions of structured activity in visual cortex. This activity explored a space of over 100 linear dimensions. We quantified the dimensionality using peer prediction: we estimated how many linear dimensions of activity from one half of the neurons were sufficient to explain the activity of the other half of the neurons in the recording. We may be able to explain more of the variance in the neural activity by using a more complicated model. For instance, we could introduce a non-linearity in the model, such as a threshold non-linearity to impose a positivity constraint on the prediction of the model. Other options could be to model the activity using a linear dynamical system in which the dimensions have their own time-varying components. These approaches will be explored in future work.

What are these different dimensions of spontaneous activity? One hypothesis is

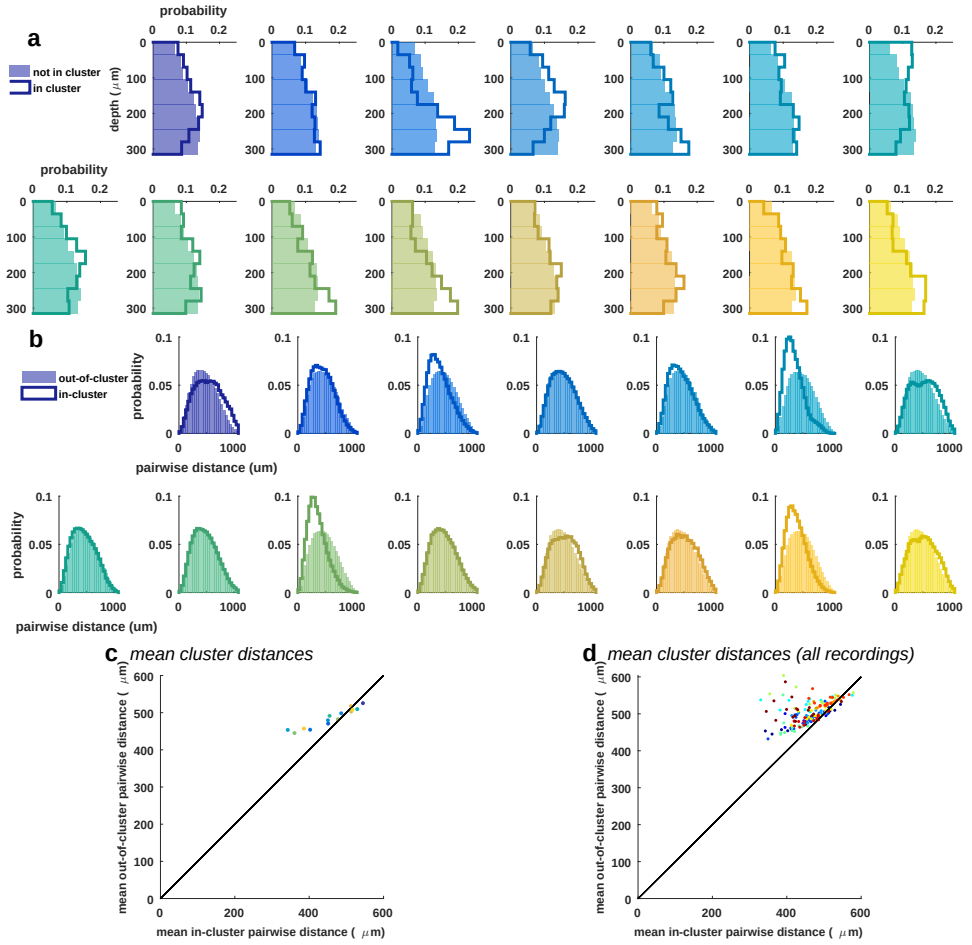

Figure 4.9: Spatial organization of spontaneous activity is near random. (a) Each subplot is a different cluster of activity as defined by NMF from the recording in Figure 4.6. Each cluster's distribution of cells in depth is plotted with a non-filled histogram. The distribution of the depths of cells which are not within that cluster is plotted with a filled histogram. The standard deviation of the depth distribution of the clusters varied from 75 μm to 89 μm, with an average standard deviation of 85 μm. The average standard deviation of the depth distribution of cells not within a specific cluster was 86 μm. The spread of the clusters in depth were not significantly different than the out-of-cluster spreads (Wilcoxon sign-rank test, p=0.81). **(b)** Each subplot is a different cluster from Figure 4.6. For each cluster, the distributions of pairwise distances among neurons within that cluster and those belonging to other clusters are plotted. The non-filled histogram is the distribution of pairwise distances among cells in the cluster. The filled histogram is the distribution of pairwise distances from cells in the cluster to cells outside that cluster. **(c)** Each cluster is summarized by a single point: its x-coordinate is the mean pairwise distance among neurons in-cluster and its y-coordinate the mean pairwise distance from neurons in the cluster to neurons out-of-cluster. The unity line is plotted in black. **(d)** The mean in-cluster pairwise distance and mean out-of-cluster pairwise distance are plotted for all clusters in each recording. The clusters from a single recording are plotted in the same color. The median in-cluster distance across all recordings is 463 μm and the median out-of-cluster distance is 498 μm. This difference in in-cluster distance and out-of-cluster distance was statistically significant (Wilcoxon sign-rank test, $p = 1.06 \times 10^{-26}$).

that these dimensions are different modes of activity that naturally arise through the structure of the connectivity in the network. Anatomical studies suggest that excitatory connectivity in cortex is organized spatially such that nearby neurons are more likely to be connected to each other [Hellwig, 2000, Levy and Reyes, 2012]. We investigated this hypothesis and found that the neural activity clusters were not spatially localized. They were distributed in depth throughout cortical layers 2, 3 and 4, and distributed throughout the field of view.

Alternatively, these multiple dimensions of spontaneous activity may be exploring the neural space of responses to visual stimuli [Berkes et al., 2011]. The mouse may be "hallucinating" visual stimuli in the absence visual stimulation. They suggested that this may be a mechanism for building a model of the external world. This hypothesis needs to be rigorously tested in future work.

Another hypothesis is that spontaneous neural activity is reflecting the complex internal state of the mouse, which might relate to the behavioral state of the mouse. For instance, some of these dimensions might reflect the exploratory drive of the mouse (which may be quantified by whisking) [Kurnikova et al., 2017]. The fear level of the mouse may be quantified through the pupil dilation. The mouse might have a drive to groom itself, which we can observe by monitoring the mouse's behavior. Other drives such as hunger may be harder to quantify using external measures.

In Chapter 5, we develop tools for extracting multi-dimensional behavioral state information from videos of the mouse, and relate this behavioral state information to neural activity.

Methods

We recorded optically the neural activity of head-fixed awake mice implanted with 3-4mm cranial windows centered over visual cortex. We obtained $\sim 10,000$ neurons in all recordings. The recordings employed two-photon calcium imaging in combination with genetically encoded calcium indicators (GECIs, specifically GCaMP6s). The mice were free to run on an air-floating ball (Figure 5.2). The position of the ball was monitored using two optical mice, allowing us to extract the running speed of the animals. For the analyses in this chapter, we only used spontaneous neural activity. For all experiments, the mice were surrounded by three computer screens. In order to collect spontaneous neural activity, the screens were either turned off, creating absolute darkness in the recording setup, or set to a constant gray background (as seen in Figure 5.2). We did not observe differences between spontaneous activity in complete darkness and during gray screen, and thus we pooled the data for all analyses. An infrared light was directed at the mouse's face, and the face of the mouse was recorded using an infrared camera (Figure 5.1). The face included the snout and the left whiskers and left eye.

The recordings were performed using two photon calcium imaging, with multi-plane acquisition. The planes were spaced 30-35 μm apart in depth. 10-12 planes were acquired simultaneously at a scan rate of 2.5-3 Hz. We expressed GCaMP in large populations of neurons in one of two ways. Either we bred transgenic crosses of a GCaMP line and an excitatory cell driver (specifically EMX-CRE x Ai94 G6s, Rasgrf-CRE x Ai 94 G6s, or CamKII x tetO gcamp 6slow), or we injected non-specific AAV virus into transgenic lines expressing td-Tomato in GAD+ neurons. To obtain large fields of view for imaging 10,000 neurons, we typically performed 4-8 injections at nearby locations, sometimes at multiple depths ($\sim$500μm and $\sim$200μm). In this chapter, we only study the excitatory neurons from these recordings, but in a later chapter we analyze the inhibitory neurons as well. We implanted coverslips that were 3-4mm in diameter, allowing us, in the gcamp-transgenic animals, to select from several potential recording locations.

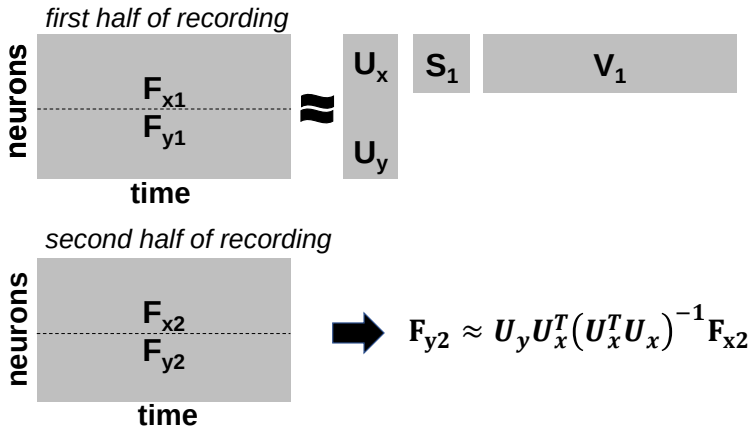

Figure 4.10: Cross-validated peer prediction schematic.

Data preprocessing

See the previous chapter for data preprocessing details. We used a toolbox that we developed called Suite2p, using the default settings [Pachitariu et al., 2016b].

Computational analyses

Single neuron dimensionality analysis using peer prediction

We binned the neural activity in time in bins of 1.2 seconds, and z-scored the activity: each neuron's mean activity was set to zero and its standard deviation to 1. Then we split the recording into two sets of neurons, x and y, so that we can predict the activity of neurons in y from the activity of neurons in x. We termed the activity of x neurons F_X and the activity of y neurons F_Y. Then we split F_x and F_y into two halves in time (interleaved segments of 40 seconds each), F_{X1} and F_{X2}, F_{Y1} and F_{Y2}. We could have directly predicted each neuron in x from each neuron in y by standard linear regression. However, each neuron's firing is noisy, so this method is prone to overfitting. Instead we reduced the dimensionality of the neural activity using singular value decomposition, and predicted the neural activity in the space of these singular vectors (see Figure 4.10 for the graphical representation of these matrices).

The singular value decomposition of the first half of the recording produces the approximation

$$\begin{bmatrix} F_{X1} \\ F_{Y1} \end{bmatrix} \approx \begin{bmatrix} U_X \\ U_Y \end{bmatrix} S_1 V_1^T.$$

The y neurons covary with the x neurons depending on their weights in the U matrix. We use U as our model to predict the activity of y neurons in the second half of the recording, F_{Y2}, from F_{X2}. We assume that the model for the second half of the recording maintains the same left singular vectors U, but requires different right singular vectors and singular values:

$$\begin{bmatrix} F_{X2} \\ F_{Y2} \end{bmatrix} \approx \begin{bmatrix} U_X \\ U_Y \end{bmatrix} S_2 V_2^T.$$

Under this model, we can thus derive $S_2V_2^T$ from U_X and F_{X2} using simple linear regression. It follows that:

$$\begin{aligned}(U_X(U_X^TU_X)^{-1})^TF_{X2} &= (U_X^TU_X)^{-1}U_x^TF_{X2} \\ &\approx (U_X^TU_X)^{-1}U_X^TU_XS_2V_2^T \\ &= S_2V_2^T.\end{aligned}$$

Notice that in general, $(U_X^TU_X)$ is very nearly a diagonal matrix with equal values on the diagonal, a consequence of the random choice of neurons in x out of the entire homogeneous population. Thus, we can formulate our prediction of the activity of Y neurons as

$$\begin{aligned}\hat{F}_{Y2} &\approx U_YS_2V_2^T \\ &\approx U_YU_X^T(U_X^TU_X)^{-1}F_{X2}.\end{aligned}$$

Note that this model uses as many dimensions of the neural activity as we specify in U_X. If the neural activity was N-dimensional for some N, we should not be able to increase our predictive power by keeping more than N singular vectors. Formally, we define the prediction using the top n singular vectors as:

$$\hat{F}_{Y2}^n \approx [u_Y^1\ u_Y^2\ \dots\ u_Y^n][u_X^1\ u_X^2\ \dots\ u_X^n]^T([u_X^1\ u_X^2\ \dots\ u_X^n]^T[u_X^1\ u_X^2\ \dots\ u_X^n])^{-1}F_{X2}$$

where u_X^k and u_Y^k are the columns of U_X and U_Y.

The variance explained by the model with n dimensions is

$$V_n = 1 - \frac{\text{var}\left[F_{Y2} - \hat{F}_{Y2}^n\right]}{\text{var}\left[F_{Y2}\right]}.$$

Correlation matrix dimensionality analysis

To compute the dimensionality of the correlation matrix, we again split the recording into two halves in time (in interleaved segments of 40 seconds) and computed the pairwise neural correlation matrices on each half the recording. These matrices are termed C^1 and C^2. Next, we fit a model with varying dimensionality to C_1. We chose the singular value decomposition as a linear model of the correlation matrix, and computed

$$[U\ S\ U^T] = \text{svd}(C^1)$$

An n-dimensional model M_n from the SVD can be formed from the top n components of the SVD is

$$M_n = [u^1\ \ u^2\ \ \ldots\ \ u^n][s^1\ \ s^2\ \ \ldots\ \ s^n][u^1\ \ u^2\ \ \ldots\ \ u^n]^T$$

where u^k and s^k are the columns of U and S.

As in the previous section, we split the prediction in half by neurons: one set of neurons x and the other y. U was also split in half according to these sets, producing two matrices U_X and U_Y. We normalized the columns of U_X and U_Y independently, so that they are orthonormal matrices. We recomputed the singular values S separately for the x neurons and the y neurons (the correlation matrix was also split by the sets of neurons, $C^1_{(X,X)}$ and $C^1_{(Y,Y)}$):

$$S_{X1} = U_X^T C^1_{(X,X)} U_X \quad \text{and}$$
$$S_{Y1} = U_Y^T C^1_{(Y,Y)} U_Y.$$

On the second half of the recording, the relative fractions of each SVD components might change, for example if the population activity explores one neural state disproportionately more. Thus, we re-estimated the singular values from the second half of the recording, using only half of the neurons x:

$$S_{X2} = U_X^T C^2_{(X,X)} U_X.$$

The predicted singular values for U_Y are then

$$S_{Y2} = \frac{S_{X2}}{S_{X1}} S_{Y1}.$$

The final model for the correlation matrix of the neurons y is

$$M = U_Y^T S_{Y2} U_Y,$$

and as a function of the number of dimensions n,

$$M_n = [u_Y^1 \;\; u_Y^2 \;\; \dots \;\; u_Y^n][s_{Y2}^1 \;\; s_{Y2}^2 \;\; \dots \;\; s_{Y2}^n][u_Y^1 \;\; u_Y^2 \;\; \dots \;\; u_Y^n]^T$$

where u_Y^k and s_{Y2}^k are the columns of U_Y and S_{Y2}. The variance explained is then:

$$V_n = 1 - \frac{\text{var}\left[C^2_{(Y,Y)} - M_n\right]}{\text{var}\left[C^2_{(Y,Y)}\right]}.$$

In all analyses shown, we scaled the variance explained by the fraction of explainable variance. The fraction of explainable variance is how much of the variance in the correlation matrix is consistent from one half of the recording to the other, rather than just noise. As we've shown in chapter 3 (Methods), the fraction of explainable variance is the correlation of the two correlation matrices from separate halves of the recordings (C_1 and C_2). Note that this is not quite the upper bound on the amount of variance we can predict in the correlation matrix structure (Figure 4.1), due to finite size effects: the rescaling of the singular values may help us predict more of the correlation matrix on the second half than we could have without any knowledge about these responses. Indeed we found this rescaling to be important for avoiding overfitting in the regime with > 32 dimensions. The final V_n is then

$$V_n = \frac{\left(1 - \frac{\text{var}\left[C^2_{(Y,Y)} - M_n\right]}{\text{var}\left[C^2_{(Y,Y)}\right]}\right)}{\text{corr}\left[C^1_{(Y,Y)}, C^2_{(Y,Y)}\right]}.$$

5

Multi-dimensional behavioral states and their neural correlates

In Chapter 4, we determined that spontaneous neural activity explores ∼100 linear dimensions. The top dimension was strongly correlated with the running speed and the pupil area of the mouse, suggesting it represents the animal's behavioral state. While it is well known that behavioral state influences neural activity [Niell and Stryker, 2010, Polack et al., 2013, McGinley et al., 2015a, Vinck et al., 2015, Pakan et al., 2016, Dipoppa et al., 2016], behavioral state is typically treated as a one-dimensional quantity in these studies. Recent studies ([Vinck et al., 2015, McGinley et al., 2015a, McGinley et al., 2015b]) have suggested there are at least two dimensions of behavioral state modulation in sensory cortices: those induced by running and those induced by increased pupil area. In addition, whisking has sometimes been studied as a proxy for brain state. We found that these three behavioral measures were correlated to brain activity, and explained a third of the correlation matrix variance. In addition, we found a high-dimensional influence of behavioral state on neural activity, when we quantified behavioral state according to the multi-dimensional orofacial behaviors of the mice. We found that orofacial behaviors accounted for more than half of the correlation matrix variance, but 10-20 dimensions of behavior were required for accurate prediction. We also developed a user-friendly, graphical interface to semi-automatically extract these orofacial behaviors from videos.

Introduction

Previous studies have found a strong relation between neural activity and behavioral measures like locomotion, whisking, and pupil area. For example, locomotion has been associated with changes in firing rates and the dynamics of the neural circuit [Niell and Stryker, 2010, Polack et al., 2013, McGinley et al., 2015a, Vinck et al., 2015, Pakan et al., 2016, Dipoppa et al., 2016]. Changes in the pupil area of the mouse have also been associated with firing rate changes in cortex, and with modified modulation by adrenergic and cholinergic inputs [McGinley et al., 2015a, Vinck et al., 2015, Reimer et al., 2016]. Finally, whisking has been shown to modulate the firing of neurons in barrel cortex [Gentet et al., 2010, Gentet et al., 2012, Peron et al., 2015].

Are all of these neural changes an indirect consequence of brain state modulation? Since all three behavioral measures (running, whisking, pupil area) are highly correlated to each other, their influence on neural activity may be through a one-dimensional underlying variable. On the other hand, these three behavioral measures are not always perfectly correlated. For example, while freely moving, mice perform diverse combinations of whisking, sniffing and locomotory patters [Wiltschko et al., 2015, Kurnikova et al., 2017], but it is unclear if head-fixed mice also engage in this diversity of behavioral sequences. If they did, we should be able to distinguish the separate influences of these three behavioral dimensions on neural activity. To quantify these behaviors in head-fixed mice, we developed a processing pipeline for video recordings of their faces. In addition to running, whisking and pupil area, we found multiple dimensions of behavioral variability which consisted of distinct movements of the whiskers, snout and other facial muscles. We collectively refer to these as orofacial behaviors. We distinguish orofacial behaviors from the three dimensions of global behavior (running, whisking, pupil area), and analyze the relation of all behaviors with high-dimensional neural activity.

Results

To measure spontaneous behaviors, we directed an infrared LED at the mouse, and recorded its full face using an infrared camera (Figure 5.1). The face included the snout, left whiskers, some of the right whiskers and the left eye (contralateral to recording location). The mice were free to run on an air-floating ball (Figure 5.2).

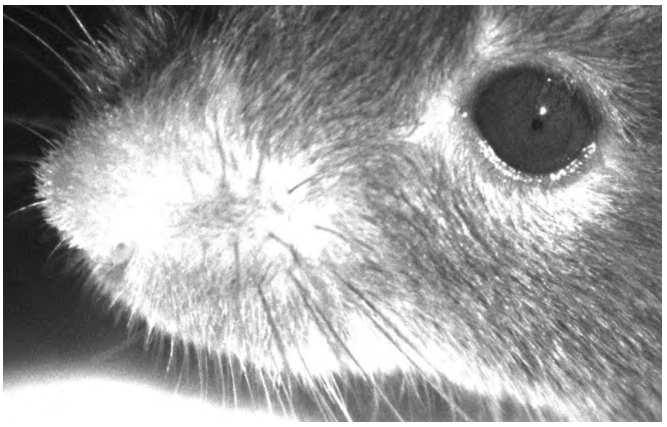

Figure 5.1: Example frame from infrared camera recording of mouse face.

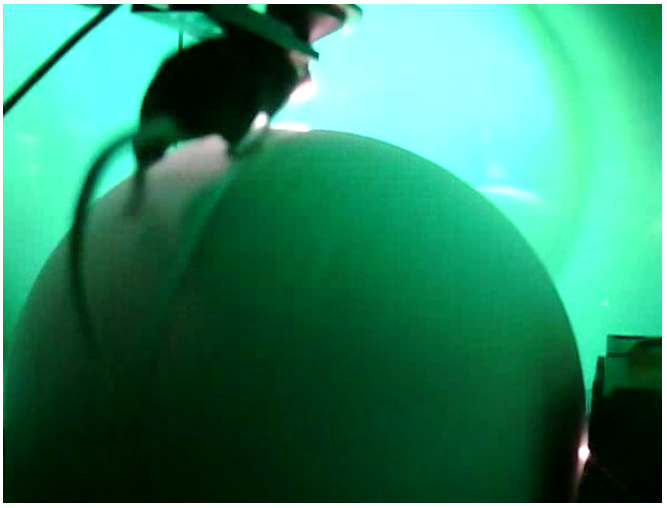

Figure 5.2: Head-fixed mouse running on air floating ball.

Predicting neural activity with one-dimensional measures of behavior

We first investigated how much neural activity could be explained by one-dimensional measures of the mouse's behavior such as running speed, pupil area and whisking. We used the first principal component of the whisker motion energy as a one-dimensional measure of whisking movements.

We wanted to disambiguate which of the three predictors (running, pupil area, whisking) contains unique information about the neural activity. For example, figure 5.3 shows that adding the pupil as a predictor or whisking as a predictor, on top of the running speed, improves prediction. Therefore, both the pupil and the whisking have independent information, not contained in the running alone. On the other hand, adding pupil as a predictor on top of whisking or running, only improved prediction in 3 of the 13 datasets, suggesting a more minor role for the pupil area in

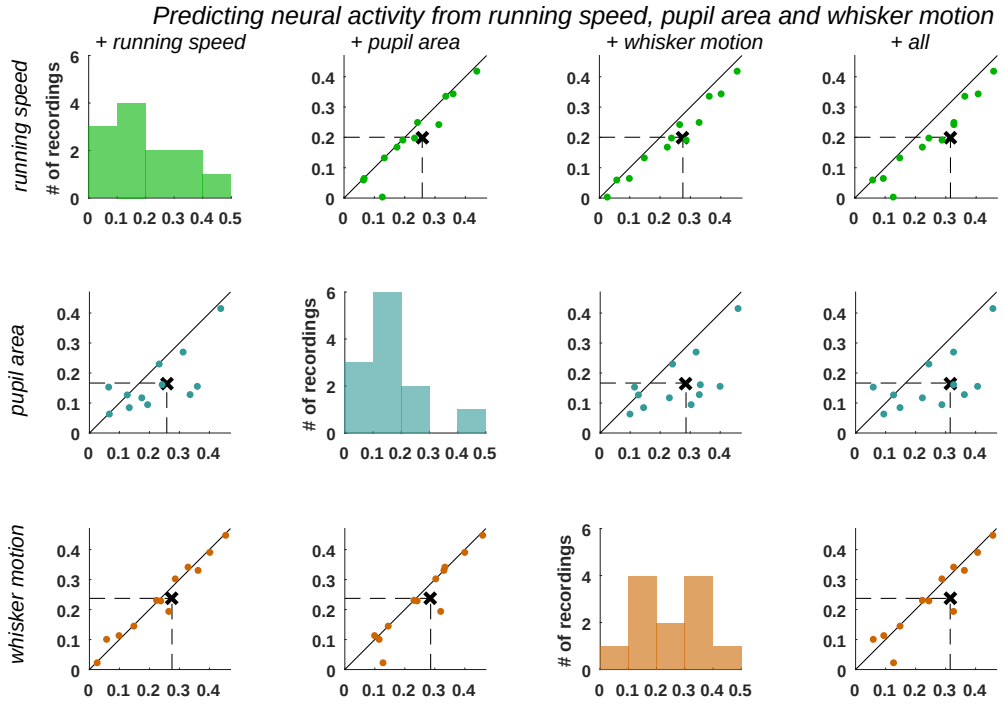

Figure 5.3: Predicting neural activity from one-dimensional behavioral variables. The neural activity (in bins of 1.2 seconds) was predicted by various behavioral variables on one half of the recording. This prediction was then tested on the other half of the recording. The variance explained was the amount of variance of the pairwise correlation matrix that was explained, normalized by the total explainable variance of the correlation matrix (see Figure 4.1). Each row is the variance explained of a different single-dimensional behavioral quantity. Each column is a different behavioral quantity added to the prediction of the single behavioral quantity in the column.

determining neural activity. Using all three predictors increased the variance explained to 32% on average, which was more than any single predictor alone (20%, 17% and 24% respectively, for running speed, pupil area and whisker motion), and more than any two predictors together (26%, 28% and 29% respectively for running + pupil area, running + whisking, pupil area + whisking). We conclude that all three behavioral variables have independent information not contained in the other ones.

Running speed, pupil area, and face or whisker motion are all associated with the overall arousal of the mouse. Around 32% of the variance in the correlation matrix among neurons is explained by these one dimensional measures of arousal. Can we explain more variability in the correlation matrix by using more dimensions of behavioral variability?

Spontaneous orofacial behaviors

We computed the principal components of the motion energy (squared difference of consecutive frames) from the faces of 9 different mice in 12 different recordings. Figure 5.4 shows the principal components from six of these mice. For one mouse, we also plotted the time traces for these components (W_{motion}, Figure 5.5). The components had similar features across mice. For instance, components 3, 4 and 4 of mouse 2, 3 and 4 respectively, were qualitatively similar. Several components across mice had strong motion in the whisker area of the face (Figure 5.4, and Figure 5.5 3, 4, and 5 in particular). Focusing on Figure 5.5, component 7 appeared to have strong motion energy in the nose area, suggesting it was a sniffing component. Component 8 had strong power in the eye area, suggesting that it was associated with blinking of the eye or wincing of the face. The timing of events in all traces was relatively similar, but different events had different activation patterns across all components.

In summary we found various components of motion, associated with a diverse repertoire of orofacial behaviors. Were these facial movements associated with neural activity?

Multiple dimensions of orofacial behaviors predict neural activity

We hypothesized that we might be better able to predict neural activity using these orofacial behaviors than using the more global behavioral variables (running, whisking, pupil area). To test this, we predicted the neural activity F from the face motion singular vectors' time components W_{motion} using reduced rank regression, and compared this with predictions using the global behavioral variables.

Reduced rank regression is a form of multivariate linear regression with the coefficient matrix restricted to a low rank (Figure 5.6) [Izenman, 1975]. Because the rank of the regression matrix is restricted, the model's dimensionality is lower, and it is thus less likely to overfit the data. To further avoid overfitting, we reduced the dimensionality of the neural activity F using the singular value decomposition of F to fit the reduced rank regression model. See section 5.4.3 for more details.

The cross-validated variance explained by the face motion components was computed for all recordings. This reached 52% variance explained with 16 dimensions of the reduced rank model, and saturated with more dimensions (Figure

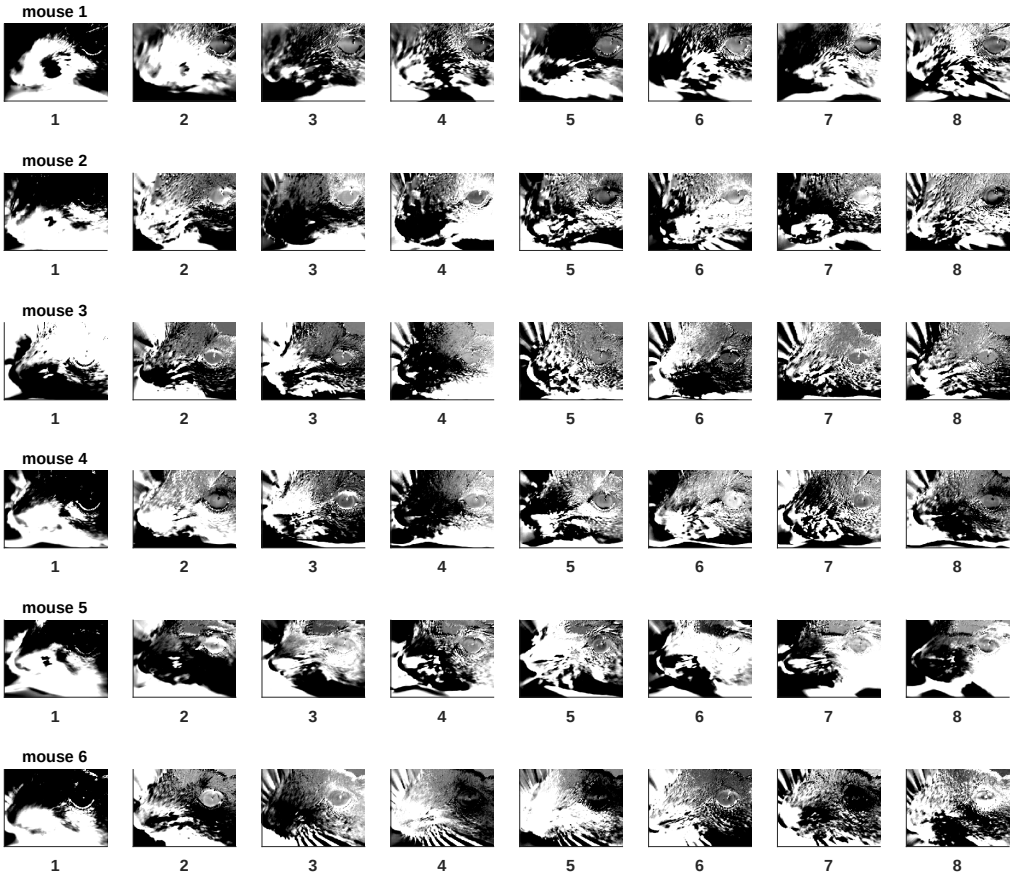

Figure 5.4: Principal components of face motion. Principal components of face motion (U_{motion}) in 6 mice, with white areas having the most motion. Each row is a different mouse, and each column is a different principal component. Column 1 explains the most face motion variance of all the faces in the row, and the subsequent columns show PCs of increasingly less variance.

5.7). Thus, on average, over half of the variance in correlation matrix structure was accounted for by the facial movements of the mouse. In contrast, a single dimension only explained about 26% of the variance (Figure 5.7b). This suggests that multiple dimensions of behavioral state influence neural activity in the visual cortex in mice.

Additionally, the information in the orofacial behaviors was redundant with the information from the running speed and the pupil area: adding these predictors only increased the explained variance by less than 1% (Figure 5.8b). Surprisingly, movements in the whisker area of the face explained almost as much variance as the movements in the full face (51% versus 52%) (Figure 5.8c). This may be due to the highly coherent movements of the face: any muscle twitch generates detectable movements all over the face.

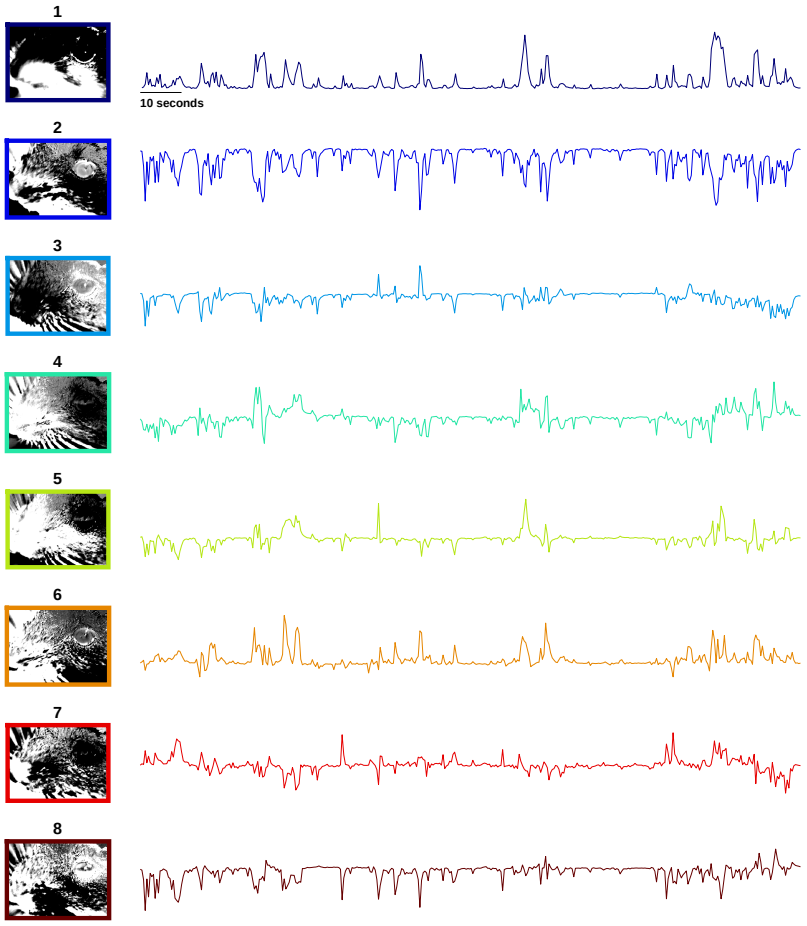

Figure 5.5: Principal components of face motion with time components. Temporal and spatial PCs of face motion (U_{motion}) for mouse 6. Each row is a different PC.

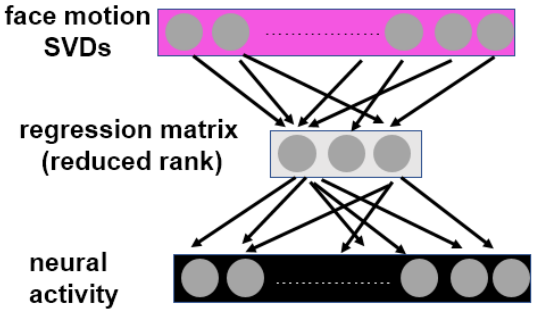

Figure 5.6: Reduced rank regression from face motion SVDs to neural activity.

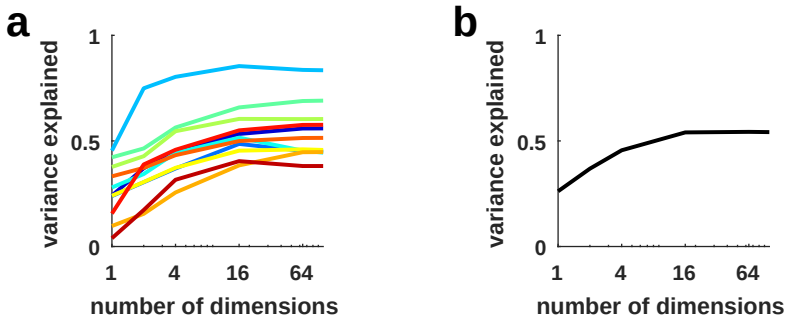

Figure 5.7: Cross-validated variance explained of the correlation matrix by face movements. The neural activity from half of the recording was predicted from the face motion SVDs using reduced rank regression with varying ranks. This prediction was then tested on the other half of the recording. The variance explained was the amount of variance of the pairwise correlation matrix that was explained, normalized by the total explainable variance of the correlation matrix. **(a)** Each line on the plot is a different recording. The number of dimensions denote the rank of the reduced rank regression prediction matrix from the face motion SVDs to the neural activity. **(b)** The average of the curves in **a**. The face motion SVDs explained on average 52% of the total explainable variance in the neural correlation matrix using a reduced rank regression model with rank 16. A rank one reduced rank regression model explained 26% of the total explainable variance.

Thus, 16 dimensions of face and whisker movements predicted two times more variance in neural activity than running speed and pupil area alone. Previous studies have treated behavioral modulation as a one-dimensional on neural activity. It was surprising that neurons in visual cortex should be modulated so strongly by running alone. It is even more surprising that they are modulated much more strongly by a combination of behavioral variables. The role of this modulation is an intriguing topic future research.

Interpretable features of the behavioral covariates of neural activity

Why might the neural activity be related to these orofacial features? Although we could not answer this question, we tried to make progress on it, by finding a more interpretable representation for the orofacial features. Reduced rank regression produces a dense transformation from the predictors to the neural activity, with equally many positive and negative weights. To obtain a more sparse and interpretable transformation from the face movement space to the neural space, we constrained our reduced rank regression matrix to be semi non-negative:

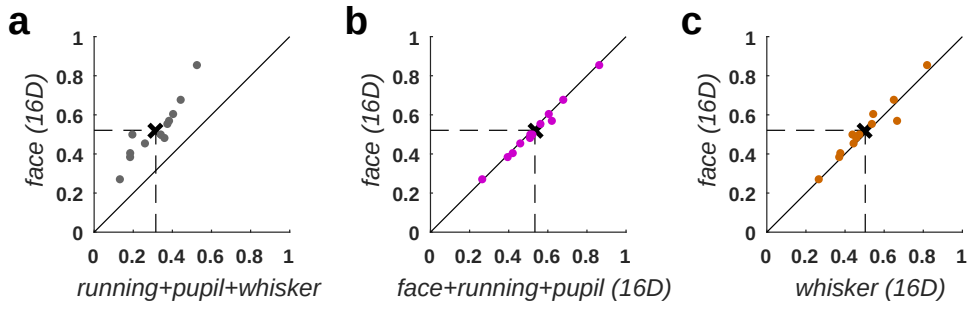

Figure 5.8: Neural activity explained by multi-dimensional behavioral variables. Comparison of the variance explained of the neural correlation matrix by various predictors. In each plot the variance explained by the rank 16 reduced rank regression model from the faces to the neural activity is the y-axis and is termed "face (16D)" **(a)** The x-axis is the variance explained by running speed, pupil area, and one-dimensional whisker motion. **(b)** The x-axis is the variance explained by the face SVDs combined with the running speed and pupil area (rank of model restricted to 16 dimensions). The average variance explained by the face, running speed and pupil area model was 53%. **(c)** The x-axis is the variance explained by the rank 16 reduced rank regression model from the whisker area SVDs to the neural activity ("whisker (16D)" - average variance explained 50%).

$$\min_{A,B} \left\| AB^T W_{motion} - F \right\|^2 \quad \text{where } A \geq 0.$$

This expression can be optimized using gradient descent (see Section 5.4.4 for details). We computed the optimal A and B with 8 dimensions and rotated W_{motion} and U_{motion} by the B vector: $W_{motion} \rightarrow B^T W_{motion}$ and $U_{motion} \rightarrow BU_{motion}$. We found distinctive patterns of facial movements and neurons correlated to the occurrence of these movements (Figure 5.9). For the top component, the first set of traces in Figure 5.9, captured the overall motion of the face. There were several neurons that were correlated with these movements. The next four components were more sparse in time than the first component, but still correlated with overall face movement. The face motion singular vectors $U_{motion}B$ associated with these components (in the right column) were distributed across the face – several parts of the face were active during these components, suggesting they were full face movements. When the mouse whisks its nose also moves for instance [Kurnikova et al., 2017]. The bottom three components appeared to be more specific. Components 6 and 7 are associated with the mouse blinking. Component 8 is associated with nose movement. In summary, these neurally-related components appear to be dominated by full-face movements, but there are some components related to more localized facial movements.

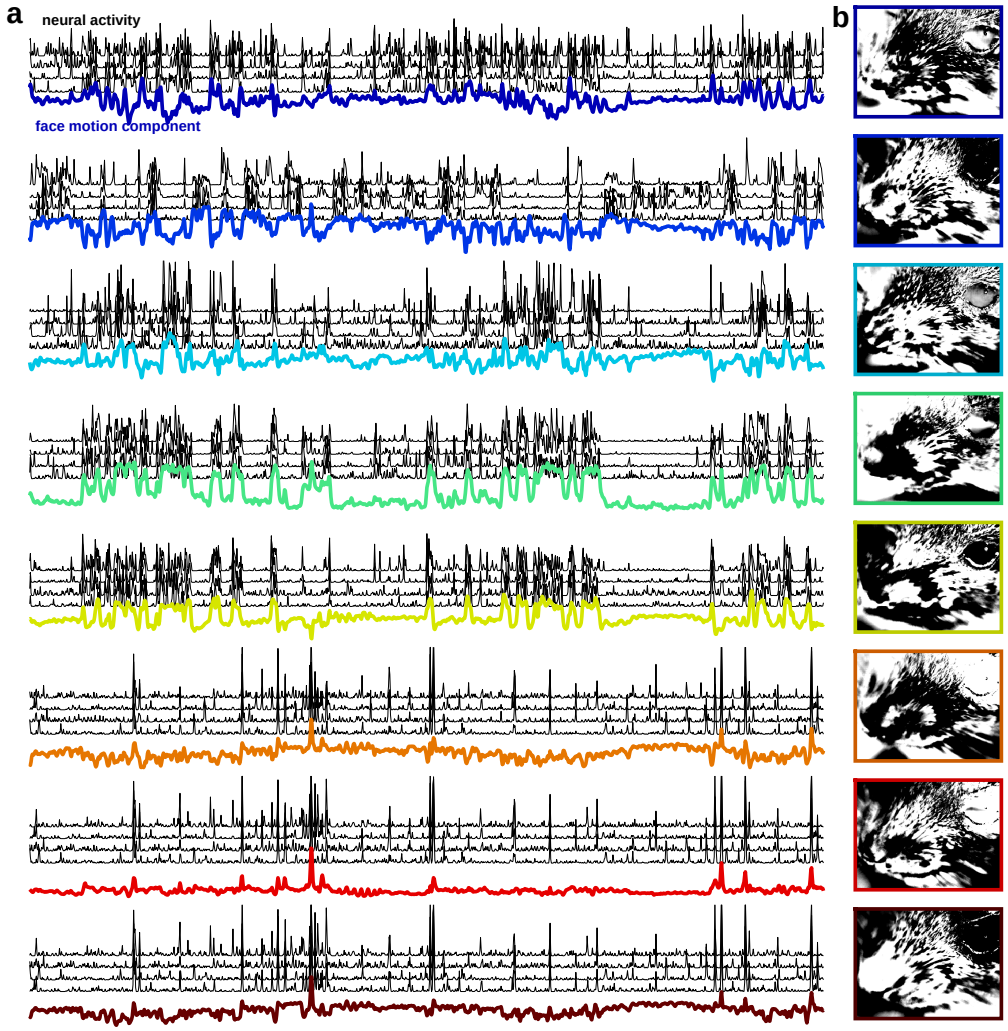

Figure 5.9: Semi non-negative reduced rank regression prediction of neural activity. (a) Activity of 4 neurons (in black) that are most correlated to the respective face motion components (in color, $B^T W_{motion}$). **(b)** The face motion components in semi-NMF space ($U_{motion} B$).

Discussion

A central question in neuroscience is the relationship between neural activity and behavior. When answering this question, neuroscientists often look to the motor cortex, particularly in monkeys [Afshar et al., 2011, Churchland et al., 2006]. In rodents, motor cortex has been shown to play a less pronounced role in motor output. For instance, motor cortical lesions do not interfere with the execution of well-learned motor skills [Kawai et al., 2015, Lopes et al., 2017]. Instead, sensory cortical areas may play a more significant role in informing the motor actions of the

animal [Mathis et al., 2017], because mice use many of their senses actively (i.e. active touch for whisking, sniffing for smell, active vision for for navigation).

In Chapter 4, we observed multiple dimensions of activity in visual cortex. The first principal component correlated with pupil area and running. How much of the total multi-dimensional neural variance can we explain from behavioral variables? Running speed, pupil area, and whisking explained 20%, 17% and 24% of neural variance respectively. Combining all these predictors explained more variance in visual cortex activity (32%) than any single predictor alone.

One possible hypothesis that can be drawn from this result is that using all three predictors provided the best estimate of the overall one-dimensional arousal of the mouse: for instance, the mouse could whisk in the absence of running and this corresponded to an aroused state that was not captured by running speed alone. An alternative hypothesis is that whisking, pupil area, and running speed correspond to different dimensions of neural activity: for instance, some neurons respond to whisking and not running, and vice versa. We extracted more dimensions of behavioral state from videos of the mouse's face, and investigated whether these behavioral state vectors are correlated with neural activity across multiple dimensions.

We observed activity in visual cortex coordinated with multiple dimensions of orofacial behaviors. We found that over half of the structure of the correlation matrix among neurons could be explained by the movements of the mouse's face. This may not be the upper bound on the amount of variance explained by the mouse's behavior, but only how much we were able to explain. We only recorded the face of the mouse. We might have found more neurally-related behavioral information if we had also monitored the paws, trunk and tail.

We might also be able to explain more neural variance if we construct a better model of the mouse's orofacial behavior. These movements were similar across mice (Figure 5.4). We could map these faces to each other using non-rigid registration techniques, and then use this large database of facial movements to build a model of the mouse's behavior. The whisker movements contained as much information about the neural activity as the full facial movements. Thus, building a model of whisker movements alone could improve prediction.

There are also different classes of behavioral models to consider. One class of

models, hidden Markov models (HMM), assumes that the behaviors are a sequence of discrete states [Wiltschko et al., 2015]. Another approach, linear dynamical systems, assumes a continuous state space in which the behaviors exist. Is the neural activity related more to discrete behavioral states, or is it reflecting the continuous dynamics of the facial movements? More research and model-fitting is required to distinguish between these two hypotheses.

Why are multi-dimensional representations of behavior present in visual cortex? This may not be surprising considering that mice use their facial muscles to enhance their exploration of the world. Rodents whisk in a rhythmic way that is phase-locked to sniffing [Deschênes et al., 2016, Kurnikova et al., 2017]. Their sniffs provide olfactory information and the whisks simultaneously provide touch information. In addition to this rhythmic modulation of the face muscles, mice move their heads in order to orient themselves to whisk relevant information or to better smell odors [Kurnikova et al., 2017, McElvain et al., 2017]. These orienting movements may be visually-guided, which may be one reason to represent the behavioral state in a multi-dimensional way in visual cortex.

Perhaps when rodents are required to perform a specific multi-sensory behavior, the orofacial information in the neural activity becomes relevant, and informs the rodent's visual perception. Further work will be required to test this hypothesis. In particular, large-scale recordings will need to be performed when the rodent is performing a task, and the rodent's behavior will need to be captured with multiple cameras.

We have observed multi-dimensional spontaneous activity which is correlated to behavioral variability. In the next chapter, we will build computational models of this multi-dimensional cortical activity and investigate the role of inhibitory activity in multi-dimensional cortical dynamics.

Methods

FaceMap: Automated classification of orofacial behaviors of mice

We developed an easy-to-use graphical user interface for processing the orofacial movements of mice that scales linearly with the number of frames, and runs 4x faster

Figure 5.10: FaceMap graphical interface.

than real-time on 30 Hz videos. We found similarities in orofacial behavioral motifs across mice.

Motion processing of regions of interest

The user can choose any region of the frame in which to compute the motion and the SVDs of the motion and the movie. These regions of interest are assigned different names, but the computations for each is the same. The motion ROIs are called "groom", "whisker", "snout", and "face", and can be manually moved to overlap any region of the movie.

We computed the motion frame G_T as the absolute value of the difference between frames F_{T+1} and F_T, for all T:

$$G_T = |F_{T+1} - F_T|.$$

The full matrix G is thus number of pixels by timepoints. For the behavioral analyses described here, we processed the whole frame and the whisker pads separately. G can also be thresholded using the saturation slider on the right, to compute a single motion trace for each region, in addition to the SVD components. The motion of the region is simply the sum of non-white pixels that appear in the

GUI panel on the right.

SVD processing of the movie and/or motion

In order to find the behavioral motifs in the data we reduced the dimensionality of the motion movies using SVD. The automated GUI allows for computation of the SVD of the movie F, the motion G, or both. The computation is identical for F and G. We outline the procedure for G. These matrices are too large to decompose in their raw form. We instead compute the SVD in two stages: first the SVD of temporal segments, then we concatenate the SVDs from different segments and recompute the SVD of these. Each segment of frames is a matrix G_i. Since the number of pixels is very large ($>$ 1 million), we avoid computing the SVD of this matrix directly, and instead compute the time by time covariance matrix $G_i^T G_i$. We keep the top 100 eigenvectors V_i of this matrix, which are also the top 100 right singular vectors of G_i. We now compute the spatial projections of these components $U_i = G_i V_i$. Notice that U_i now consists of the left singular vectors, scaled by the singular values. As such, the matrix U_i is a 100-dimensional summary of the original data.

We then concatenate U_i for all segments of the movie, and re-compute the SVD:

$$[U\ S\ V^T] = \text{svd}([\hat{U}_1 ... \hat{U}_n]).$$

We keep the top 100 components of this U matrix as the spatial components of the face motion. We then project the raw movies onto these spatial components, to obtain their temporal profiles:

$$W_{motion} = U^T G.$$

Pupil processing

In the FaceMap GUI, we also compute the area and position of the pupil in the region of interest that the user chooses. In this region of interest, a user-adjust threshold is used to keep only the darkest pixels, which correspond to the pupil. We compute the center of mass of these dark pixels as:

$$\mathbf{x}_{COM} = \frac{\mathbf{x}^T R(\mathbf{x})}{\sum_{\mathbf{x}} R(\mathbf{x})}$$

where $\mathbf{x}$ is the two-dimensional pixel location and R is the pixel's darkness level. We compute the covariance Σ of a 2D Gaussian fit to the region of interest:

$$\Sigma = (\mathbf{x} - \mathbf{x}_{COM})(\mathbf{x} - \mathbf{x}_{COM})^T.$$

We draw the outline of the pupil as the ellipse that is n standard deviations from the center of mass, where n can be set by the user (2 by default).

Blink processing

The user can also select a region in which to compute the degree of eye opening, which can be useful to detect blinks, and to detect when the animal is in a highly aroused state (very large eye opening). Again, a threshold is used to distingush between the brightness of the eye and that of the surround fur. The eye-opening area is the sum of non-white pixels displayed in the GUI.

Processing multiple videos

The user chooses a set of videos to process together using the folder options with the button "choose folder" (Figure 5.10). When a folder is chosen, the GUI will select all videos in that folder and one subfolder down and compile a list of movie files. The user can then choose a subset of these movie files for processing. All of the regions of interest selected in the GUI are processed across all the movies chosen in this list. The user can look at the raw movies by selecting the corresponding file in a drop down menu. Once the ROIs and settings are chosen, the user can save these settings for future use with the "Save ROIs and Settings" button or choose to process just this set of movies by selecting "Process ROIs". If the settings are saved, the user can then open another set of movies, draw the ROIs and save these settings as well. The GUI maintains a list of all the settings that have been saved during the current processing session. All of the sets of movies with user-defined ROIs can then be processed automatically by pressing "Batch Process ROIs".

Predicting shared neural activity from behavioral variables by standard linear regression

We wanted to know how much of the shared neural activity (the structure of the correlation matrix) can be predicted from behavioral variables. To answer this question, we computed a linear prediction on training data, used it to predict neural activity on test data. The correlation matrix of the prediction was then compared to the correlation matrix of the real data.

We first binned the neural activity and the behavior x in bins of 1.2 seconds, then z-scored the binned traces (each neuron's mean spiking rate was set to zero and its standard deviation set to 1). We split the neural activity F and the behavior x into two halves in time, F_1 and F_2, x_1 and x_2. The behavioral variables $x_{1,2}$ were either single traces (running, whisking, pupil area), pairwise combinations of these, or all three traces together. We predicted F_1 from x_1 by linear regression, obtaining the weights a:

$$a = \left(x_1 x_1^T\right)^{-1} \left(x_1 F_1^T\right).$$

We used a to obtain the prediction on the second half of the recording and computed the correlation matrix of this prediction:

$$\hat{F}_2 = a^T x_2$$
$$\hat{C}_2 = \mathrm{corr}\left(\hat{F}_2\right).$$

The similarity of $\hat{C}_2$ to C_2 determines how much variance the behavior explained of the correlation matrix:

$$V = 1 - \frac{\mathrm{var}\left(C_2 - \hat{C}_2\right)}{\mathrm{var}\left(C_2\right)}$$

The variance explained V was normalized by the fraction of explainable variance of the correlation matrix (see Chapter 4, Figure 4.1):

$$V_{norm} = \frac{V}{\mathrm{corr}\left[C_1, C_2\right]}.$$

Predicting shared neural activity from behavioral variables by reduced-rank regression

In order to obtain high-dimensional representations of the mouse's behavior, we recorded the full face of the mouse while imaging. We then computed the singular value decomposition of the motion energy of this movie (pixels by time) to reduce the movie to 100 dimensions by time (see section 5.4.1.2, this matrix is denoted as W_{motion}). We predicted the neural activity F from the face motion SVDs W_{motion} using reduced rank regression. Reduced rank regression is a form of regularized linear regression, with the prediction weights matrix restricted to a specific rank (Figure 5.6) [Izenman, 1975]. Because the rank of the regression matrix is restricted, the model's dimensionality is lower, making it less likely to overfit the data. To further avoid overfitting, we first reduced the dimensionality of the neural activity F using PCA, and fit the reduced rank regression model to the top PCs only:

$$[U\,SV^T] = \text{svd}[F].$$

We kept 100 singular vectors of the singular value decomposition of F, and set $V = SV^T$. We split the recordings in half in time (interleaved segments of 40 seconds), splitting F into F_1 and F_2, V into V_1 and V_2, and W_{motion} into W_1 and W_2. We computed the pairwise correlation matrices C_1 and C_2 of F_1 and F_2.

We computed the reduced rank regression matrices A_n and B_n with rank n that minimize the expression

$$\min_{A_n,B_n} \left|\left|A_n B_n^T W_1 - V_1\right|\right|^2.$$

This expression can be minimized analytically [Izenman, 1975]. We then reconstructed F_2 from the prediction $\hat{V}_2 = A_n B_n^T W_2$:

$$\begin{aligned}\hat{F}_2^n &= U\hat{V}_2 \\ &= UA_n B_n^T W_2.\end{aligned}$$

The predicted correlation matrix $\hat{C}_2^n$ was then computed as the correlation matrix of $\hat{F}_2^n$. The explained variance fraction is

$$V_n = 1 - \frac{\text{var}\left(C_2 - \hat{C}_2^n\right)}{\text{var}\left(C_2\right)}.$$

We normalized this explained variance by the explainable variance

$$V_{norm} = \frac{1 - \frac{\mathrm{var}\left(C_2 - \hat{C}_2^n\right)}{\mathrm{var}(C_2)}}{\mathrm{corr}\left[C_1, C_2\right]}.$$

We computed this value for increasing numbers of n, up to 100 dimensions (Figure 5.6).

Semi-non-negative reduced rank regression algorithm

To obtain a more sparse and interpretable transformation from the face movement space to the neural space, we constrained one of the matrices in the reduced rank regression (A) to be non-negative, and defined the constrained cost function

$$\min_{A,B} \left\| AB^T W_{motion} - F \right\|^2 \quad \text{where } A \geq 0.$$

We minimized this expression using gradient descent with a momentum term. Let $W = W_{motion}$. The expression we wish to minimize is

$$g = \left\| AB^T W - F \right\|^2 \quad \text{where } A \geq 0.$$

The gradient of this expression with respect to A is

$$\frac{\partial g}{\partial A} = (F - AB^T W)W^T B^T$$

and with respect to B is

$$\frac{\partial g}{\partial B} = W(F - AB^T W)^T A.$$

On each step of the optimization, any negative elements of A were set to zero. The optimization was performed with a step size of 1×10^{-6} and a momentum value of 0.9, and generally converged to good local minima.

6

Multi-dimensional inhibitory activity in cortical circuits

In order to model one-dimensional intrinsic fluctuations in neural networks, it was enough to have a single population of randomly-connected excitatory neurons (Chapter 2). However, in large-scale recordings of thousands of neurons, we observed multi-dimensional activity in response to stimuli and spontaneously. To model this activity in a network simulation, we would need either structured feedforward inputs, structured lateral connectivity, or a combination of both. To generate multi-dimensional activity intrinsically, we find that a network model with structured excitatory connectivity also requires multiple dimensions of inhibitory activity aligned to the excitatory activity dimensions. When the inhibition is instead global and unstructured, the excitatory modes compete with each other too aggressively to generate diverse activity. We predicted therefore that there would be multiple dimensions of inhibitory activity in visual cortex to stabilize the high-dimensional excitatory activity that we observed. As predicted, we observed these multi-dimensional inhibitory modes. We found inhibitory activity explored at least $\sim$64 dimensions during spontaneous activity and at least $\sim$391 dimensions when driven by natural images. This is counter to the hypothesis that inhibitory neurons indiscriminately pool the activity of local excitatory neurons [Kerlin et al., 2010, Fino and Yuste, 2011, Packer and Yuste, 2011]. Instead, our results suggest that inhibitory neurons receive specific excitatory inputs, resulting in high-dimensional activity.

Introduction

Networks of excitatory neurons in the brain can exhibit multiple modes of activity, in response to stimuli or spontaneously. Simulations of neural networks can also exhibit multiple modes when driving different subpopulations of excitatory neurons with feed-forward inputs, or by imposing specific connectivity patterns in the network of excitatory neurons. Here we primarily study the case where multi-dimensional activity is generated by the recurrent dynamics of a network of excitatory neurons with structured connectivity (i.e. clustered, or low-dimensional, see [Litwin-Kumar and Doiron, 2012]). We study the types of inhibitory connectivity patterns that may need to accompany such a model. We distinguish and study two extremes: (1) global, one-dimensional inhibition, and (2) structured, high-dimensional inhibition, for example paralleling the structure of the excitatory activity patterns. What are the functional consequences of global versus structured inhibition?

We first answered this question in network models of cortical activity. We found that a network with structured inhibition explored a much larger space of activity patterns than a network with global inhibition. This large space of patterns was consistent with the activity we observed in Chapters 3 and 4. Thus, the modelling work predicts that multi-dimensional excitatory activity is stabilized by multi-dimensional inhibitory activity. We tested this prediction in visual cortex and indeed found multi-dimensional inhibitory activity, both in response to stimuli and spontaneously.

Modelling of high-dimensional excitatory activity

Rate model of subnetworks

We built a model rate network with subnetworks of highly inter-connected excitatory neurons. We modeled the subnetworks as single dimensions of activity represented by the mean activity of the subpopulation. We grouped inhibitory neurons into subnetworks as well, assuming that the activity of each inhibitory neuron in the group is following the activity of its excitatory counterparts faithfully. Thus, the model contained $2n$ variables with n subnetworks of excitatory neurons and n subnetworks of inhibitory neurons. We termed the excitatory networks $\mathbf{x}$ and the

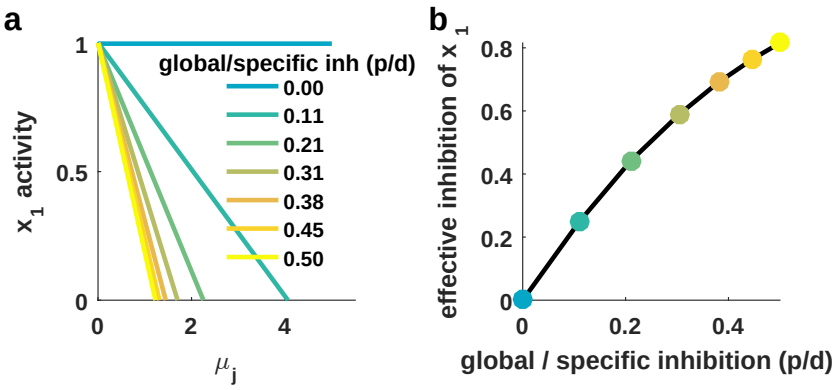

Figure 6.1: Excitatory subpopulations suppressed through global inhibition. (a) The activity of subpopulation x_1 as a function of the input to other excitatory subpopulations, μ_j. Each line represents a different value of p/d, the ratio between global and specific recurrent inhibition. As this ratio increases, the slope between x_1 and μ_j becomes more negative. **(b)** The negative slope between x_1 and μ_j in **a** is plotted here as a function of the ratio between global and specific inhibition. As the contribution of global inhibition increases, the slope becomes more negative. The effect of input to other excitatory subpopulations becomes more strong as global inhibition increases.

inhibitory networks $\mathbf{y}$. For simplicity, we assumed that the excitatory subnetworks were disconnected from each other, and we did not allow inhibitory-to-inhibitory interactions. Each subnetwork excited itself with a weight of g. We also assumed that the connection from excitatory subnetwork x_i to inhibitory subnetwork y_i was d, and the reverse connection from y_i to x_i was $-d$. The inter-connectivity between subnetwork x_i and y_j where $i \neq j$ were set to $\pm p$.

We investigated how varying the value of p affects the activity of excitatory subpopulation x_1, and we were able to derive this interaction as an analytical expression (see Methods). Increasing p is equivalent to increasing the global, non-specific inhibitory feedback onto x_1. We computed x_1^* as a function of μ_j for varying values of p (Figure 6.1a). As the global to specific inhibition ratio increased (p/d), the input to other excitatory subnetworks had a more strongly negative effect on the activity of subnetwork x_1. The slope of each line in Figure 6.1a was computed. The negative value of this slope is plotted in Figure 6.1b. This is the effective inhibition of x_1 by other excitatory populations. Increasing global inhibition increases the effective inhibition from other excitatory populations. When the ratio of global to specific inhibition is high, the activity of other excitatory subpopulations suppresses the activity of subpopulation x_1.

However, if there is a specific inhibitory subpopulation for each excitatory

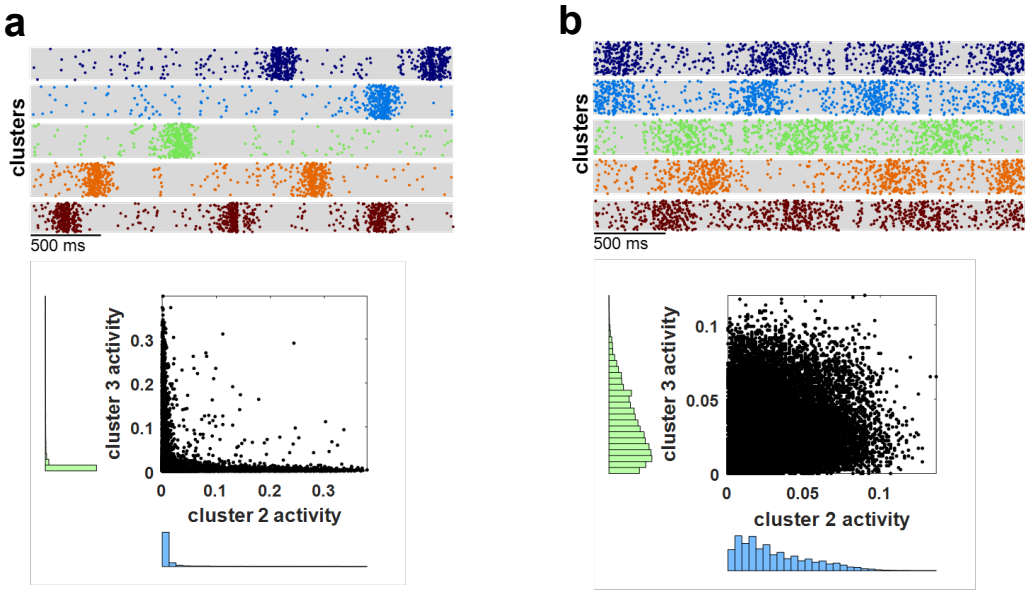

Figure 6.2: Network model with spiking excitatory and inhibitory neurons. Spiking network model in regime in which it spontaneously generates activity (no external input to the network). **(a)** Raster of activity of network. Each cluster is plotted in a separate color. The bottom plot shows the mean activity of cluster 2 vs the mean activity of cluster 3 computed in 100 ms bins. Out of cluster inhibitory connectivity is set to 10%. **(b)** Same as **a**, except out of cluster inhibitory connectivity is set to 3%.

subpopulation, then each excitatory subpopulation can in principle maintain activity when driven by inputs. Thus, multiple excitatory subpopulations can be active simultaneously. Global inhibition instead restricts the space of possible patterns that the network can represent: if the input is large to a single population, then only that population is active, and it suppresses the activity of other clusters. This limits the space of patterns to N possibilities, one for each subnetwork being active.

Spiking network model

The winner-take-all effect derived analytically in the rate network model can be observed in a spiking network of spiking excitatory and inhibitory neurons. We built a spiking network model with 5 clusters consisting of 4096 excitatory neurons and 1280 inhibitory neurons with strong but sparse synaptic connectivity. The probability of connectivity between excitatory neurons from the same cluster was 5%. The probability of connectivity to excitatory neurons from other clusters was 0.05%. The probability of connectivity of excitatory to inhibitory and from inhibitory to

excitatory from the same cluster was 3.5%. The probability of inhibitory to inhibitory neuron connectivity for inhibitory neurons from the same cluster was 3.5%.

We varied the connectivity from excitatory neurons to inhibitory neurons in other clusters and vice versa, and the connectivity from inhibitory neurons to inhibitory neurons. In Figure 6.2a, the excitatory to inhibitory out of cluster connectivity was 10% of the within cluster connectivity. In Figure 6.2b the out of cluster excitatory to inhibitory connectivity was 3% of the within cluster connectivity. When the out of cluster connectivity was lower, the clusters were more likely to be active simultaneously. The lower plots in Figure 6.2 represent the activity of cluster 2 and cluster 3. Each point is the mean activity of the cluster in 100 ms bins. Cluster 2 and 3 are more likely to be active simultaneously in the network with lower out of cluster connectivity, which is equivalent to less global inhibitory feedback.

Experimental investigation of inhibitory activity

Our theoretical work suggests that flexible and diverse pattern representation requires multi-dimensional inhibitory stabilization.

Multi-dimensional inhibitory spontaneous activity

In our recordings, we observed multi-dimensional spontaneous excitatory activity. The dimensions of activity did not display winner-take-all structure – multiple clusters could be active simultaneously (Figure 4.7). Thus, our modelling work predicts that the network will have multi-dimensional inhibitory activity that is aligned to the dimensions of excitatory variability.

We computed the dimensionality of inhibitory population activity in seven different recordings, by asking how many dimensions of the excitatory neurons are needed for best prediction (see Chapter 4). We regularized the prediction by taking the singular value decomposition of the neural activity first (same procedure that is described in section 4.4.2.1). A one-dimensional model of inhibitory activity explained 6.2% of the variance in the activity binned in time bins of 1.2 seconds. A model with 128 dimensions explained 23% of the variance in inhibitory neuron activity. Models with more than 128 dimensions performed worse, suggesting that they were overfitting. We concluded that the inhibitory population replicated at least

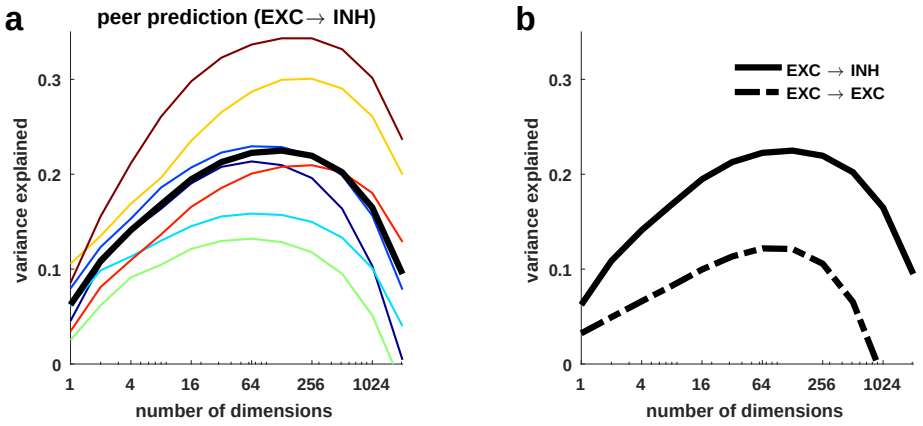

Figure 6.3: Predicting GAD+ inhibitory neurons from GAD- excitatory neurons. (a) Each curve is the cross-validated variance explained of the GAD+ inhibitory neurons as a function of the number of dimensions used in the peer prediction model (predicting inhibitory activity from excitatory activity). The black curve is the average across all recordings. The peak of this curve is 23% variance explained and it is achieved by a model with 128 linear dimensions. **(b)** The averaged cross-validated variance explained of the excitatory populations is plotted in a dashed line (from Figure 4.4).

128 dimensions of excitatory activity modes. This was similar to the total dimensionality of the excitatory populations (Figure 6.3b). In fact, the activity of inhibitory neurons during spontaneous activity was more predictable than that of excitatory neurons, perhaps because inhibitory neurons tend to be more active.

We also thought it would be instructive to analyze the residuals from this prediction. It is possible that inhibitory populations contain different modes of correlated activity, not included in the excitatory population. We explicitly asked if there was any residual inhibitory activity that could be predicted from the activity of other inhibitory neurons. We found that there was very little predictable activity left in the inhibitory population: the average of the maximum variance explained was 2.5% (Figure 6.4). This suggests that the majority of the predictable inhibitory activity was accounted for by the excitatory population activity.

High-dimensional inhibitory stimulus responses

Excitatory neurons in visual cortex also explored many dimensions of activity during stimulus-driven activity. Do inhibitory neurons have well-tuned responses to visual stimuli, and if so, does their activity inhabit a high-dimensional space?

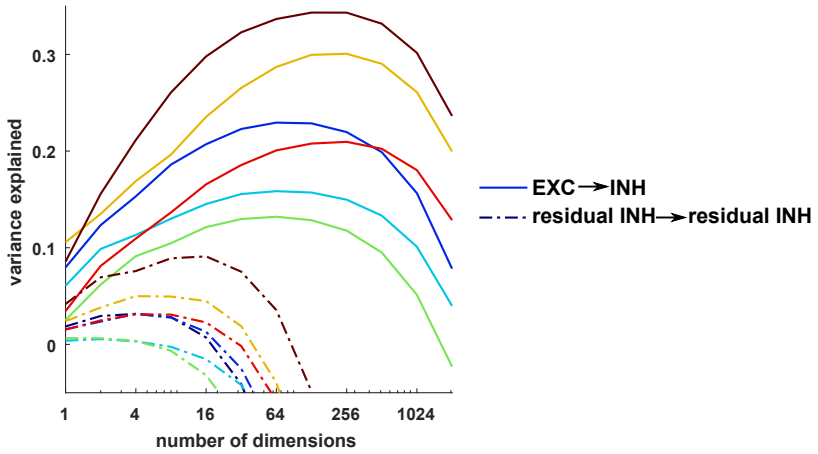

Figure 6.4: Cross-validated variance explained of residual GAD+ inhibitory activity. We subtracted the predicted inhibitory activity from the total inhibitory activity and attempted to predict this residual inhibitory activity using peer prediction. The dashed lines are the variance explained of the peer prediction model of the residual inhibitory activity. Each line is a different recording. The solid lines are matched in color to the dashed lines and represent the prediction of the total inhibitory activity in the recording by the excitatory population activity.

First, we investigated responses to drifting gratings, and found that the inhibitory neurons were well tuned to orientation (Figure 6.5). We computed cross-validated tuning curves for both excitatory and inhibitory neurons, by computing the preferred orientation on one half of the presentations, and aligning to this preferred orientation all tuning curves determined on the other half (which we also normalized to mean 1). We also computed a cross-validated OSI index in this manner. We averaged over all of these aligned and normalized tuning curves to produce Figure 6.5c. On average, excitatory and inhibitory neurons have similar orientation selectivity (an average OSI of 0.12 in inhibitory neurons, and an average OSI of 0.13 in excitatory neurons). However, the distribution of inhibitory neurons' OSIs is less skewed: inhibitory neurons are less likely to have high OSIs than excitatory neurons. Inhibitory neurons are also less likely to have high direction selectivity indices than inhibitory neurons. Notice that all these average OSI values are very small compared to other literature reports, which are typically ~ 0.4. This is simply due to cross-validation: when we did not cross-validate, we obtained similar distributions. In addition, small number of repeats typically performed in other studies result in noisy tuning curves, which, without cross-validation, highly bias the apparent OSI of the neuron's responses.

Although [Ma et al., 2010] showed high orientation selectivity indices for SOM+

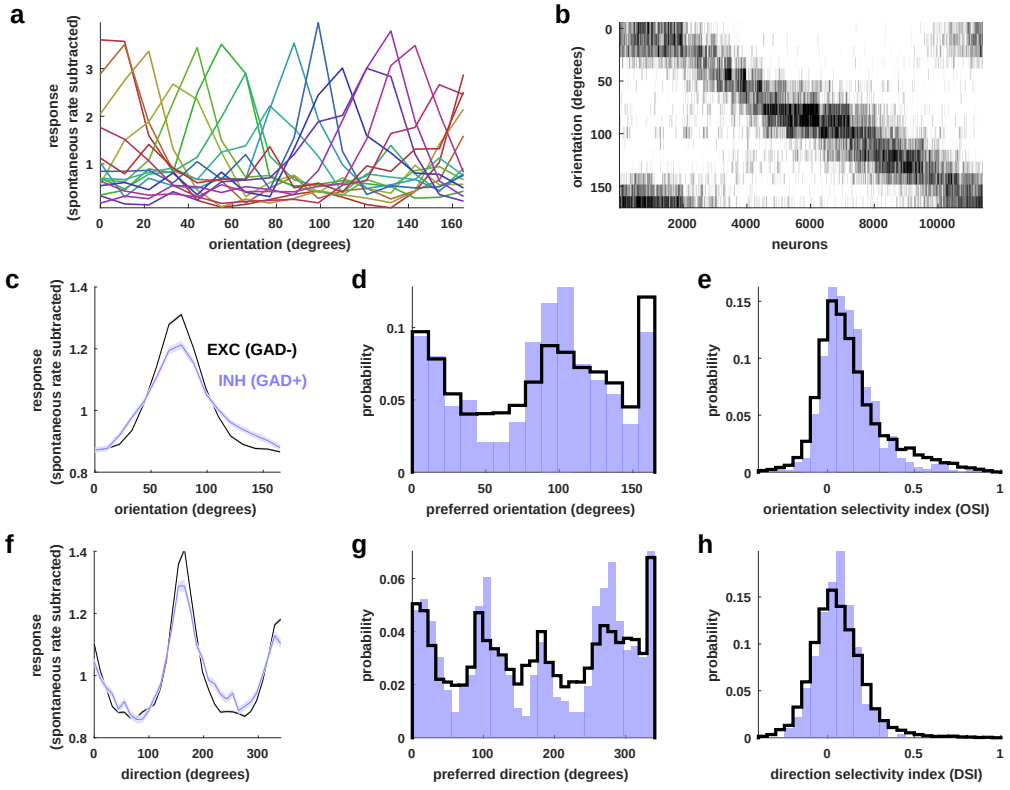

Figure 6.5: Responses of excitatory and inhibitory neurons to drifting gratings. We presented drifting gratings with a spacing of 12 degrees, for a total of 32 different directions (16 different orientations). **(a)** Example neurons with preferred orientations at each of the presented orientations. The neurons' responses were normalized by their mean response to all orientations. **(b)** All neurons' responses to the drifting gratings, sorted by their preferred orientations. The responses were normalized by the mean response over all gratings and then smoothed over neurons in the plot (with a Gaussian of width 10 neurons). **(c)** The mean tuning curves across excitatory neuron responses and inhibitory neuron responses. The responses of each neuron were normalized by the mean response across all orientations before the mean across neurons was computed. The error bars are the standard errors across neurons. **(d)** The distribution of preferred orientations across excitatory and inhibitory neurons. **(e)** The distribution of cross-validated orientation selectivity indices. The OSI of excitatory neurons was on average 0.13, the OSI of inhibitory neurons was on average 0.12. **(f)** Same as **c**, but the tuning across directions instead of orientations. **(g)** The distribution of preferred directions across excitatory and inhibitory neurons. **(h)** The distribution of cross-validated direction selectivity indices. The DSI of excitatory neurons was on average 0.056, the DSI of inhibitory neurons was on average 0.057. However the distribution of excitatory neuron DSI's was more skewed: 1.3% of the excitatory neurons had DSI's greater than 0.5, whereas only 0.4% of the inhibitory neurons had DSI's greater than 0.5.

inhibitory neurons, other studies have shown low orientation selectivity indices for inhibitory neurons compared to excitatory neuron responses [Liu et al., 2009, Kerlin et al., 2010, Atallah et al., 2012]. Previous studies however did not cross-validate the preferred orientation of the tuning curve. This means that low-firing, noisy neurons are more likely to have high orientation selectivity indices: if a neuron has very low firing throughout the recording, but responds strongly once to a single orientation presentation, then its non-cross-validated OSI would be close to one. However, if the OSI index is cross-validated, then the OSI for this neuron would be zero because in one half of the recording the neuron does not respond selectively to that orientation. Excitatory neurons have lower firing rates than PV+ inhibitory neurons, so they are more likely to have orientation selectivity indices biased in this way [Atallah et al., 2012]. Further, [Liu et al., 2009, Atallah et al., 2012] observed increases in baseline firing rates from excitatory to inhibitory neurons, but not large decreases orientation tuning width, suggesting that inhibitory neurons are tuned to different orientations, but have higher baseline responses to all stimuli. The reported difference in half width half height Gaussian tuning width for excitatory versus inhibitory neurons was 42 degrees versus 52 degrees in [Atallah et al., 2012].

We also investigated inhibitory neuron responses to 2,800 natural images. Inhibitory neuron responses were on average as reliable as excitatory neuron responses (Figure 6.6a). The average fraction of signal variance in inhibitory neurons was 0.19 and while it was 0.18 in excitatory neurons. Although their reliability was similar, inhibitory neurons responded more densely than excitatory neurons, as measured by their lower skewness (Figure 6.6b). We next computed the dimensionality of the inhibitory population. We found that inhibitory neurons required 391 dimensions to reach 95% of the stimulus variance (Figure 6.6c). The dimensionality of a subset of excitatory neurons of the same size as the inhibitory population was in fact very similar (dashed line in Figure 6.6c). We conclude that inhibitory neurons form a more dense code of natural images than excitatory neurons, but that the code is just as high dimensional.

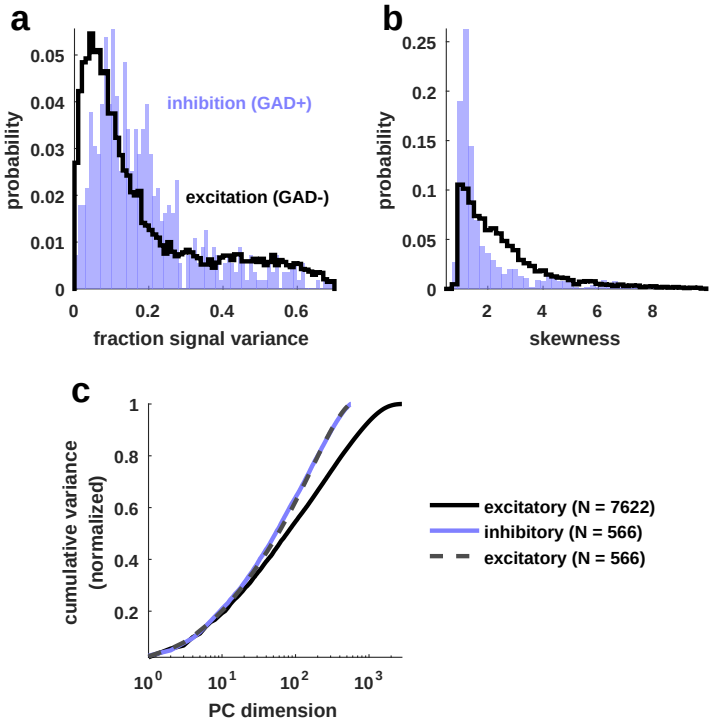

Figure 6.6: Inhibitory neuron responses to natural images. We recorded excitatory and inhibitory activity (GAD+) while presenting 2,800 natural images to the mouse. **(a)** The distribution of the fraction signal variance across neurons. The fraction signal variance is the fraction of the total variance of the neuron's activity that is related to the stimulus. Excitatory neurons have an average fraction signal variance of 0.18 and inhibitory neurons have an average fraction signal variance of 0.19. **(b)** The distribution of the skewness of natural image responses across neurons. Excitatory neurons have a response skewness of 2.60 on average. Inhibitory neurons have a response skewness of 1.91 on average. **(c)** The dimensionality of the excitatory and inhibitory population responses to natural images. The cumulative variance of each PC is computed in the same way as Figure 3.3i. Inhibitory neuron responses require fewer dimensions to reach 95% of the total variance than excitatory neuron responses. 95% of the variance in inhibitory neuron responses is accounted for by 391 dimensions, while 95% of the variance in excitatory neuron responses is accounted for by 1148 dimensions. However, this difference in dimensionality appears to be because fewer inhibitory neurons are recorded. The dimensionality of a subset of excitatory neurons of the same size as the inhibitory population is equivalent to the dimensionality of the inhibitory neuron population.

Discussion

We observed multi-dimensional excitatory activity in response to stimuli and during spontaneous activity. The diversity of patterns we observed could be recreated in a network model. However, the network model required multiple dimensions of inhibitory activity aligned to the excitatory activity dimensions to allow the network to have diverse, high-dimensional activity. We predicted therefore that activity of the recorded inhibitory neurons would also be high-dimensional.

We found ∼64 dimensions of spontaneous inhibitory activity, and this activity was predictable from local excitatory activity alone. Previous studies have suggested that local cortical circuits can spontaneously generate activity in the absence of input [Shapcott et al., 2016, Malina et al., 2016]. Thus, local excitatory neurons could drive multi-dimensional activity in inhibitory neurons. There is evidence to suggest that inhibitory neurons do receive local excitatory input [Fino and Yuste, 2011, Packer and Yuste, 2011, Kerlin et al., 2010]. These excitatory neurons may specifically target inhibitory neurons in order to produce multi-dimensional inhibitory activity. If excitatory neurons instead randomly targeted inhibitory neurons, then we would observe one-dimensional inhibitory activity which followed the mean activity of the excitatory population. Thus, one hypothesis for multi-dimensional inhibitory activity spontaneously is specific local excitatory input.

Alternatively, inhibitory neurons could receive the same input as the excitatory neurons, and this input could originate in other cortical areas or in subcortical regions. There is evidence to suggest that inhibitory neurons receive inputs from other cortical areas. For instance, inhibitory neurons in auditory cortex receive inputs from motor cortical neurons [Schneider et al., 2014], and inhibitory neurons in visual cortex receive inputs from auditory cortex [Ibrahim et al., 2016]. Thalamus may also modulate cortical inhibitory activity. Fast-spiking inhibitory neurons receive strong feed-forward input from thalamus in auditory and barrel cortex [Ji et al., 2015, Yu et al., 2016]. Inhibitory neurons in cortex also receive indirect thalamic input through layer 6 corticothalamic neurons [Bortone et al., 2014, Kim et al., 2014]. Other subcortical areas may indirectly modulate cortical interneurons by releasing neuromodulators into cortex. Some of these neuromodulators may selectively target certain types of inhibitory neurons [Chen et al., 2015, Sakata, 2016, Castro-Alamancos

and Gulati, 2014, Goard and Dan, 2009]. Further research is required to determine precisely the origin of multi-dimensional inhibitory activity in the absence of visual stimuli.

We also found ∼391 dimensions of stimulus-driven inhibitory activity. This is counter to the hypothesis that inhibitory neurons are untuned to visual stimuli. Instead, the inhibitory neurons appear to receive specific stimulus-evoked excitatory inputs, driving high-dimensional activity within the local inhibitory population. Some previous studies have reported orientation-tuned inhibitory neuron responses, but there is little previous work exploring the response properties of inhibitory neurons to natural images. A comparison of these responses to high-dimensional excitatory responses will be the subject of future work.

Our modelling work suggests that the purpose of this high-dimensional inhibitory activity is to stabilize high-dimensional excitatory activity. There is some experimental work that may support this claim. Intracellular recordings of excitatory neurons have shown that the ratio of excitatory and inhibitory conductances is conserved across excitatory neurons [Xue et al., 2014]. One way to produce this balance is to scale feedback from inhibitory neurons depending on how much feedforward excitation the cell receives. [Xue et al., 2014] suggest that their experimental findings support this hypothesis. This is equivalent to aligning inhibitory activity to excitatory activity, whether through shared feed-forward inputs or through feedback from specific excitatory neurons.

One method to examine the stabilizing role of inhibitory activity is to measure inhibitory synaptic conductances in excitatory neurons. [Monier et al., 2003] reports that in cat visual cortex, 60% of recorded neurons had inhibitory conductances tuned to their excitatory conductances in response to drifting gratings. However, [Haider et al., 2013] showed, on average, untuned inhibitory synaptic conductances in excitatory neurons in response to orientations. Further work is needed to fully characterize the role of inhibitory neurons in stabilizing excitatory neuron activity, and if this stabilization occurs across multiple dimensions.

Methods

Calculations for the rate model

Let

$$\mathbf{z} = [x_1, \ldots, x_n, y_1, \ldots, y_n] = \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}.$$

This results in the following connectivity matrix C from $\mathbf{z}$ to $\mathbf{z}$:

$$C = \left[\begin{array}{cccc|cccc} g & 0 & \ldots & 0 & -d & -p & \ldots & -p \\ 0 & g & \ldots & 0 & -p & -d & \ldots & -p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & -p \\ 0 & 0 & \ldots & g & -p & -p & \ldots & -d \\ \hline d & p & \ldots & p & 0 & \ldots & \ldots & 0 \\ p & d & \ldots & p & \vdots & \ldots & \ldots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ldots & \ldots & \vdots \\ p & p & \ldots & d & 0 & \ldots & \ldots & 0 \end{array}\right].$$

Rewriting C in terms of identity matrices I,

$$C = \left[\begin{array}{c|c} gI & (p-d)I - p\mathbb{1}\mathbb{1}^T \\ \hline (d-p)I + p\mathbb{1}\mathbb{1}^T & 0 \end{array}\right].$$

We assumed each subnetwork was leaky (the activity decayed to zero without input). The differential equation governing these excitatory and inhibitory subnetworks is

$$\tau\dot{\mathbf{z}} = C\mathbf{z} - \mathbf{z} + \mu$$

where

$$\mu = \begin{bmatrix} \mu_{\mathbf{x}} \\ \mu_{\mathbf{y}} \end{bmatrix}$$

are the external inputs to the excitatory subnetworks ($\mu_{\mathbf{x}}$) and inhibitory subnetworks ($\mu_{\mathbf{y}}$). τ is the timescale of the decay of the subnetwork activity. In steady-state, $\dot{\mathbf{z}} = 0$. Solving for $\mathbf{z}$ yields

$$\mathbf{z}^* = (I - C)^{-1}\mu$$

The matrix

$$I - C = \left[\begin{array}{c|c} (1-g)I & (1+d-p)I + p\mathbb{1}\mathbb{1}^T \\ \hline -(1+d-p)I - p\mathbb{1}\mathbb{1}^T & I \end{array}\right]$$

can be analytically inverted. The inverse of a block matrix is

$$\begin{bmatrix} D & E \\ F & G \end{bmatrix}^{-1} = \begin{bmatrix} (D - EG^{-1}F)^{-1} & -D^{-1}EH^{-1} \\ -H^{-1}FD^{-1} & H^{-1} \end{bmatrix}$$

where

$$H = G - FD^{-1}E$$

[Petersen and Pedersen, 2008].

In our case,

$$\begin{aligned} D &= (1-g)I \Rightarrow D^{-1} = 1/(1-g)I \\ E &= (d-p)I + p\mathbb{1}\mathbb{1}^T \\ F &= -(d-p)I - p\mathbb{1}\mathbb{1}^T = -E \\ G &= I \quad \Rightarrow \quad G^{-1} = I. \end{aligned}$$

Thus,

$$\begin{bmatrix} (1-g)I & E \\ -E & I \end{bmatrix}^{-1} = \begin{bmatrix} ((1-g)I + E^2)^{-1} & -\frac{1}{1-g}EH^{-1} \\ \frac{1}{1-g}H^{-1}E & H^{-1} \end{bmatrix}$$

where

$$H = I + \frac{1}{1-g}E^2.$$

The first block of this matrix is

$$\begin{aligned}((1-g)I+E^2)^{-1} &= [(1-g)I+((d-p)I+p\mathbb{1}\mathbb{1}^T)^2]^{-1}\\ &= [(1-g)I+((d-p)^2I+2p(d-p)\mathbb{1}\mathbb{1}^T+p^2\mathbb{1}\mathbb{1}^T\mathbb{1}\mathbb{1}^T)]^{-1}\\ &= [(1-g)I+((d-p)^2I+2p(d-p)\mathbb{1}\mathbb{1}^T+p^2\mathbb{1}N\mathbb{1}^T]^{-1}\\ &= [(1-g)I+(d-p)^2I+(2p(d-p)+p^2N)\mathbb{1}\mathbb{1}^T]^{-1}\\ &= [(1-g+(d-p)^2)I+(2p(d-p)+p^2N)\mathbb{1}\mathbb{1}^T]^{-1}.\end{aligned}$$

Let $\alpha = 1-g+(d-p)^2$ and $\beta = 2p(d-p)+p^2N$. Applying the Sherman-Morrison formula [Petersen and Pedersen, 2008],

$$\begin{aligned}[\alpha I+\beta\mathbb{1}\mathbb{1}^T]^{-1} &= \alpha^{-1}I-\frac{\alpha^{-1}I\beta\mathbb{1}\mathbb{1}^T\alpha^{-1}I}{1+\beta\mathbb{1}^T\alpha^{-1}I\mathbb{1}}\\ &= \alpha^{-1}I-\frac{\alpha^{-2}\beta\mathbb{1}\mathbb{1}^T}{1+\alpha^{-1}\beta N}\\ &= \alpha^{-1}\left(I-\frac{\beta}{\alpha+\beta N}\mathbb{1}\mathbb{1}^T\right).\end{aligned}$$

For the next block of the matrix, we must compute H^{-1}:

$$\begin{aligned}H^{-1} &= \left[I+\frac{1}{1-g}E^2\right]^{-1}\\ &= \left[I+\frac{1}{1-g}\left((d-p)^2I+(2p(d-p)+p^2N)\mathbb{1}\mathbb{1}^T\right)\right]^{-1}\\ &= \left[\frac{1-g+(d-p)^2}{1-g}I+\frac{2p(d-p)+p^2N}{1-g}\mathbb{1}\mathbb{1}^T\right]^{-1}\\ &= \left[\frac{\alpha I+\beta\mathbb{1}\mathbb{1}^T}{1-g}\right]^{-1}\\ &= (1-g)\left[\alpha I+\beta\mathbb{1}\mathbb{1}^T\right]^{-1}.\end{aligned}$$

From the derivation above,

$$\begin{aligned}H^{-1} &= (1-g)\left[\alpha I+\beta\mathbb{1}\mathbb{1}^T\right]^{-1}\\ &= \alpha^{-1}(1-g)\left(I-\frac{\beta}{\alpha+\beta N}\mathbb{1}\mathbb{1}^T\right).\end{aligned}$$

Then, the upper right block of the matrix is

$$\begin{aligned}
-\frac{1}{1-g}EH^{-1} &= -\alpha^{-1}\left((d-p)I + p\mathbb{1}\mathbb{1}^T\right)\left(I - \frac{\beta}{\alpha+\beta N}\mathbb{1}\mathbb{1}^T\right) \\
&= -\alpha^{-1}\left((d-p)I + p\mathbb{1}\mathbb{1}^T - \frac{(d-p)\beta}{\alpha+\beta N}\mathbb{1}\mathbb{1}^T - \frac{p\beta}{\alpha+\beta N}\mathbb{1}\mathbb{1}^T\mathbb{1}\mathbb{1}^T\right) \\
&= -\alpha^{-1}\left((d-p)I + \left(p - \frac{(d-p)\beta}{\alpha+\beta N} - \frac{p\beta N}{\alpha+\beta N}\right)\mathbb{1}\mathbb{1}^T\right) \\
&= -\alpha^{-1}\left((d-p)I + \left(\frac{p\alpha + p\beta N - (d-p)\beta - p\beta N}{\alpha+\beta N}\right)\mathbb{1}\mathbb{1}^T\right) \\
&= -\alpha^{-1}\left((d-p)I + \left(\frac{p(\alpha+\beta) - d\beta}{\alpha+\beta N}\right)\mathbb{1}\mathbb{1}^T\right).
\end{aligned}$$

The lower left block is $\frac{1}{1-g}H^{-1}E$. Because E, H^{-1}, and EH^{-1} are symmetric matrices, E and H^{-1} must commute. In other words, $H^{-1}E = EH^{-1}$. Therefore, the lower left block is the same as the upper right block with an inverse sign:

$$\frac{1}{1-g}EH^{-1} = \alpha^{-1}\left((d-p)I + \left(\frac{p(\alpha+\beta) - d\beta}{\alpha+\beta N}\right)\mathbb{1}\mathbb{1}^T\right).$$

The lower right block is

$$H^{-1} = \alpha^{-1}(1-g)\left(I - \frac{\beta}{\alpha+\beta N}\mathbb{1}\mathbb{1}^T\right).$$

Combining these blocks,

$$(I-C)^{-1} = \alpha^{-1}\begin{bmatrix} I - \frac{\beta}{\alpha+\beta N}\mathbb{1}\mathbb{1}^T & (d-p)I + \left(\frac{p(\alpha+\beta)-d\beta}{\alpha+\beta N}\right)\mathbb{1}\mathbb{1}^T \\ -(d-p)I - \left(\frac{p(\alpha+\beta)-d\beta}{\alpha+\beta N}\right)\mathbb{1}\mathbb{1}^T & (1-g)\left(I - \frac{\beta}{\alpha+\beta N}\mathbb{1}\mathbb{1}^T\right) \end{bmatrix}$$

The fixed point solution is

$$\mathbf{z}^* = \alpha^{-1}\begin{bmatrix} I - \frac{\beta}{\alpha+\beta N}\mathbb{1}\mathbb{1}^T & (d-p)I + \left(\frac{p(\alpha+\beta)-d\beta}{\alpha+\beta N}\right)\mathbb{1}\mathbb{1}^T \\ -(d-p)I - \left(\frac{p(\alpha+\beta)-d\beta}{\alpha+\beta N}\right)\mathbb{1}\mathbb{1}^T & (1-g)\left(I - \frac{\beta}{\alpha+\beta N}\mathbb{1}\mathbb{1}^T\right) \end{bmatrix}\begin{bmatrix} \mu_{\mathbf{E}} \\ \mu_{\mathbf{I}} \end{bmatrix}.$$

If we assume that the external input to the inhibitory subnetworks is zero ($\mu_{\mathbf{y}} = 0$), then

$$\mathbf{z}^* = \begin{bmatrix} \mathbf{x}^* \\ \mathbf{y}^* \end{bmatrix} = \alpha^{-1}\begin{bmatrix} \left(I - \frac{\beta}{\alpha+\beta N}\mathbb{1}\mathbb{1}^T\right)\mu_{\mathbf{x}} \\ -\left((d-p)I + \left(\frac{p(\alpha+\beta)-d\beta}{\alpha+\beta N}\right)\mathbb{1}\mathbb{1}^T\right)\mu_{\mathbf{x}} \end{bmatrix}.$$

Suppose the input to subnetwork x_1 is 1, and the input to all other excitatory

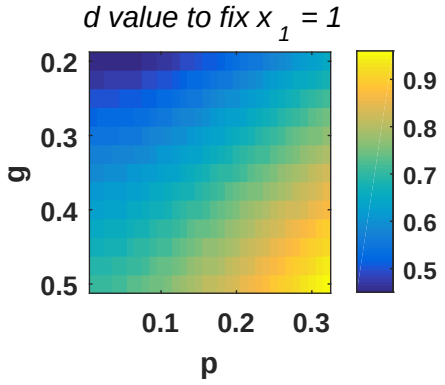

Figure 6.7: Constraining activity of excitatory population using excitatory-inhibitory connectivity. As g and p vary, the value of d required to fix $x_1 = 1$ at $\mu_{x_1} = 1$ varies. For each pair of g and p values, we fixed d such that $x_1 = 1$.

subnetworks (x_j, $(j \neq 1)$) is 0. Suppose d, the connectivity between subnetwork x_1 and y_1 is fixed. Then, as p increases (the connectivity to y_j, $(j \neq 1)$), the activity of x_1 decreases. We instead want the activity of x_1 to be constant across values of p. Therefore, d must be a function of p. We fixed the activity of subnetwork x_1 to 1 for an input μ_{x_1} of 1. We numerically solved for the value of d that satisfies this expression (Figure 6.7).

How is the activity of excitatory subnetwork x_1 influenced by input to other subnetworks $x_2, \ldots, x_n$? Suppose the input to subnetwork x_1 is 1, and the input to all other excitatory subnetworks μ_j, $(j \neq 1)$ is equal. Then the steady state activity of subnetwork x_1 is

$$\begin{aligned} x_1^* &= \alpha^{-1}\left(1 - \frac{\beta}{\alpha+\beta N} - \mu_j(N-1)\frac{\beta}{\alpha+\beta N}\right) \\ &= \alpha^{-1}\left(1 - \frac{(\mu_j(N-1)+1)\beta}{\alpha+\beta N}\right). \end{aligned}$$

Spiking network model

We built a clustered conductance-based spiking network model consisting of 4096 excitatory neurons and 1280 inhibitory neurons with strong but sparse synaptic connectivity. The membrane time constant for the excitatory neurons was 25 ms and for the inhibitory neurons it was 5 ms. The excitatory conductance time constant was 18 ms and the inhibitory conductance time constant was 10 ms. The single cell

adaptation current time constant was 200 ms. The connection probabilities among neurons was

$$p_{E_i \to E_i} = 5\%$$
$$p_{E_i \to E_j} = 0.05\%$$
$$p_{E_i \to I_i} = 3.5\%$$
$$p_{I_i \to E_i} = 3.5\%$$
$$p_{I_i \to I_i} = 3.5\%$$

where E means excitatory, I means inhibitory and the subscript denotes cluster identity.

We varied the connectivity from excitatory neurons to inhibitory neurons in other clusters and vice versa, and the connectivity from inhibitory neurons to inhibitory neurons. In Figure 6.2a, the excitatory to inhibitory out of cluster connectivity was 10% of the within cluster connectivity. In Figure 6.2b the out of cluster excitatory to inhibitory connectivity was 3% of the within cluster connectivity. No other parameters in the model changed.

The connection strengths for both models were

$$w_{E \to E} = 0.39$$
$$w_{E \to I} = 1.42$$
$$w_{I \to E} = 0.44$$
$$w_{I \to I} = 1.54$$

After each spike, the single cell adaptation conductance increased by 0.8. The threshold is set to a voltage of 1 and the reset voltage is 0.9.

7

Discussion

Network models reveal biophysical mechanisms underlying ongoing spontaneous fluctuations

Across brain states and sensory areas, we observe diverse cortical activity in the absence of sensory stimuli. In Chapter 2, we showed that a deterministic spiking network model is capable of intrinsically generating population-wide fluctuations in neural activity, without requiring external modulating inputs. When we fit the model to each of our individual recordings, we found that fluctuations were suppressed by inhibitory feedback. We also found that the activity of putative inhibitory neurons in our recordings was increased during periods of cortical desynchronization. The results of several previous experimental studies also support the idea that strong inhibition can stabilize cortical networks and enhance sensory coding.

This simple model provided a compact and intuitive description of the circuit mechanisms that are capable of coordinated dynamics in networks with intrinsic variability. We also extended the model using slower timescales of inhibition and excitation to represent $GABA_B$ and NMDA conductances respectively. The same mechanism of inhibition can control a more complicated model, but further work is needed to construct more components of the cortical circuit. For instance, we ignored different interneuron subtypes and the interactions among those subtypes [Pfeffer et al., 2013]. The dimensionality of the parameter space increases when more interneuron populations are considered, which means more data is needed to constrain the activity of these models. More studies of interneuron subtypes and

their functional roles are being performed [Pakan et al., 2016, Dipoppa et al., 2016, Muñoz et al., 2017]. Using this data, network models could be sufficiently constrained and then be used to better understand the dynamics of inhibitory neurons in cortical circuits. Several neuronal network architectures are being developed that could incorporate this physiological data [Izhikevich and Edelman, 2008, Markram et al., 2015].

High-dimensional stimulus encoding in visual cortex

Population-wide spontaneous fluctuations were suppressed in awake, aroused states, and cortical encoding of stimuli was reliable and well-tuned. The neural responses to natural images were distributed along dimensions in a $1/n$ fashion, with the largest dimensions contributing a substantial portion of the response (Chapter 3). In response to 8-dimensional stimuli, visual cortex responses were also distributed as $1/n$, suggesting that this power law scaling is a feature of the underlying circuit architecture and not inherited from the stimulus space.

We may be able to infer the structure of the circuit from the structure of the noise. The noise in the neural responses also scales with the signal of the responses, and thus decays as $1/n$. One mechanism to produce the noise scaling could be specific connectivity: neurons most aligned to a dimension of the stimulus response are also connected to each other. If one neuron in that population has increased activity on a stimulus presentation (perhaps due to neuromodulatory effects), then the other neurons in that dimension also increase their activity. The noise is thus proportional to the signal because the amplification factor of noise is proportional to the number of neurons in the sub-circuit.

The amplification factor of noise could also be proportional to the number of neurons activated by the stimulus through another mechanism. The neurons in visual cortex receive feedforward input from lateral geniculate nucleus (LGN). Suppose all the neurons in a given stimulus-driven dimension are activated by a single thalamic neuron. If the activity of that thalamic neuron is modulated on a stimulus presentation, then the activity of all of its feedforward targets will also be modulated, thus producing noise proportional to the number of neurons in this dimension. There is evidence for thalamic modulation across brain states and in

different behavioral contexts [Saalmann and Kastner, 2009, Lewis et al., 2015, Wimmer et al., 2015].

We present two hypotheses for the observed distributions of noise variance. These hypotheses could be tested by suppressing cortical spiking and recording EPSPs from thalamic inputs while presenting natural images, similar to the approach taken in [Malina et al., 2016]. It could also be tested using connectomic approaches such as dual patch clamp techniques [Cossell et al., 2015]. The coordination between thalamic firing and cortical activity could be investigated using simultaneous electrophysiological recordings from both visual cortex and LGN.

Multi-dimensional spontaneous activity and multi-dimensional behavioral variability

In Chapter 3, we observed high-dimensional stimulus-evoked activity in the visual cortex of awake mice. In these recordings, stimulus encoding did not appear to be impaired by one-dimensional spontaneous fluctuations of the sort we studied in non-aroused or anesthetized recordings in Chapter 2. However, we still hypothesized some structure in the neural activity may be present spontaneously. For example, when doing principal components analysis (PCA), we found neurons with large positive or negative weights onto the first PC, which defined two anti-correlated subpopulations of neurons. Due to the symmetry of the weights, the population had low average pairwise correlations (an ”asynchronous” state), but nonetheless the activity was highly structured, with at least 100 significant dimensions in the population activity (Chapter 4). In addition, this first principal component was correlated to the arousal state of the animal, as measured by the running speed or the pupil area, which motivated us to pursue an explanation of neural activity in the context of behavioral states.

If locomotion and pupil area are correlated with neural activity changes, then are other behaviors of the mouse also associated with neural activity? To answer this question, we developed a processing pipeline for classifying the orofacial behaviors of head-fixed mice (Chapter 5). After we classified their behaviors, we correlated them with the neural activity recorded. One dimension of facial movements explained 26% of the variance of the spontaneous activity correlation matrix. However, a model using

15 dimensions of facial movements explained much more: over 50%.

The activity of these movement-modulated neurons may aid rodents in multi-sensory decision-making and in generating complex sequences of actions that require sensory (or specifically visual) feedback for completion. Further work will be required to test this hypothesis.

Networks with multi-dimensional excitation and inhibition

In Chapter 2, we modelled spontaneous fluctuations which dominated the activity of the population. Thus, we modelled the activity with a single population of randomly-connected excitatory neurons, similar to other previous studies of population-wide fluctuations [Renart et al., 2010, Mochol et al., 2015, Kanashiro et al., 2017].

However, in large-scale recordings of thousands of neurons, we observed multi-dimensional excitatory activity in response to stimuli and during spontaneous activity. We could recreate multi-dimensional excitatory activity in a network model, either through specific feedforward inputs, specific connectivity, or a combination of both. But what structure of inhibition should the circuit have? Two possibilities arise: specific inhibition aligned to the dimensions of excitatory activity, or global inhibition which is the average of all excitatory activity. Through network simulations, we found that these two types of network architectures have different implications. A network with multiple populations of inhibitory neurons can explore a much larger space of activity patterns than a network with only global inhibitory feedback (Chapter 6). The modelling work predicted that inhibitory activity would also be multi-dimensional. This prediction would seem highly improbable, given the current view that inhibitory neurons pool over the local network activity, and thus have indiscriminate responses.

Nonetheless, we tested the prediction in recordings where we labelled all interneurons with tdTomato, so we could easily identify them during the analysis. As predicted, we found multi-dimensional inhibitory activity during both spontaneous and stimulus-driven activity in visual cortex. In fact, the inhibitory population appeared to be just as tuned to stimuli as the excitatory population, and in fact contained more information related to the spontaneous activity states. (Chapter 6). We predicted therefore that there would be multiple dimensions of inhibitory activity

in visual cortex to stabilize the high-dimensional excitatory activity that we observed. In neural recordings, we observed multiple dimensions of inhibitory activity in response to stimuli and during spontaneous activity.

Do inhibitory neurons actively stabilize these high-dimensional patterns of activity in visual cortex? One potential experiment to answer this question could be to slightly suppress the activity of inhibitory neurons and measure the activity of excitatory neurons in response to visual stimuli. Do the excitatory neuron patterns retain the high-dimensional structure observed in an unperturbed brain? If the patterns are changed by the perturbation, then the inhibitory neurons could be stabilizing these high-dimensional computations. Alternatively, it could mean that the inhibitory neurons are necessary for computing these high-dimensional representations of visual stimuli.

How these patterns of activity are formed is also an important question. Inhibitory plasticity could produce high-dimensional stabilization of excitatory populations [Vogels et al., 2011]. Inhibitory learning rules may also be sufficient to enable computation of diverse features of the images [Pachitariu and Sahani, 2012]. Further work is required to uncover the cortical circuit implementation of high-dimensional activity.

The role of multi-dimensional activity in cortex

Deciphering large scale recordings will require much more work. Fruitful approaches may utilize deep learning techniques [Pandarinath et al., 2017]. To solve image classification tasks, deep networks project the space of natural images into much higher-dimensional spaces which compute various features of the images. The brain may use a similar strategy. We found that the brain can expand an 8-dimensional image into a space of hundreds of dimensions (Chapter 3).

The role of spontaneous activity patterns is less clear. Perhaps, these patterns of behaviorally-related activity serve as reinforcement signals to train the network to perform various tasks. Perhaps, an approach using reinforcement learning principles will help us better understand the role of these signals in sensory cortices [Higgins et al., 2017]. To reveal the role of these patterns, we need to record 10,000 neuron populations under more variable behavioral scenarios.

Appendix A

Motion correction for calcium imaging

Two-photon microscopy has enabled unprecedented recordings from vast populations of neurons. However, this method is thought to have significant weaknesses, particularly in comparison to electrophysiological "gold standard" recordings. One of the major weaknesses relates to motion of the tissue relative to the imaging objective. Such motion is largest, both in two-photon imaging and electrophysiology, when an animal engages in overt behaviors such as running, which produce large changes at the location of the recording instrument. If we want to relate neural activity to such behaviors, we must first ensure that the recording is not corrupted by the motion artifacts. In this chapter, we describe correction methods for 2D motion artifacts, aligned to the imaging plane, and in the next chapter we describe correction methods for Z-drift motion, orthogonal to the imaging plane. Within 2D motion artifacts, we distinguish between rigid and non-rigid movements, with the latter primarily a consequence of sequential line scanning. We develop computational methods to correct for these types of motion at a subpixel level, and validate them in simulations (2D methods), or by visual inspection. The registration methods account for all major types of 2D motion artifacts in two-photon microscopy, allowing for robust monitoring of neural activity. In addition, we demonstrate a surprising capability of subpixel registration to be used for creating super-resolution images.

Introduction

Several of our analyses depend crucially on being able to acquire long recordings of neural activity. However, over the period of several hours, the imaged tissue may move relative to the objective. But are long recordings really necessary? We offer two examples of such recordings necessary for other chapters of this thesis: 1) to estimate the dimensionality of a neural recording we are limited not only by the number of neurons recorded but also by the recording duration [Gao and Ganguli, 2015]; hence, to estimate the true dimensionality, we must acquire long recordings; 2) quantifying the relation between spontaneous activity and spontaneous behaviors requires a large number of different behavioral states, which can only be acquired over long periods of time, because animals are likely to perform the same behaviors state for several seconds or even minutes. To enable these studies, we needed robust, long recordings, and thus we developed robust motion correction methods. In addition to enabling our own studies, this work establishes two-photon microscopy as a robust method for neural recordings, unaffected by XYZ motion of the brain relative to the recording instrument. Similar developments are yet to be achieved for electrophysiological recordings [Pachitariu et al., 2016a]. These improvements are important, because any behaviors the mouse engages in (running, whisking, grooming etc.) might be correlated with neural activity (see Chapter 5), but also may move the neural field of view, thus generating an artifactual relation to the recorded data.

This Appendix will describe how to identify and correct motion artifacts induced by movement parallel to the imaging plane. We developed novel rigid and non-rigid registration algorithms to identify motion artifacts [Greenberg and Kerr, 2009, Pnevmatikakis and Giovannucci, 2017], then shifted the images by the identified movements. We used GPU functions to increase the speed of the algorithms.

The algorithms described here are implemented as freely available software, as part of Suite2p, a full data processing pipeline for calcium imaging processing at www.github.com/cortex-lab/Suite2P.

Image registration via phase correlation

The first stage involves correcting for the effects of brain movement parallel to the recording plane, by registering all frames in the movie to each other.

Rigid registration

Common registration techniques used in two-photon microscopy rely on finding the cross-correlation peak between a frame and a target image [Poort et al., 2015]. This peak can be computed efficiently with the fast Fourier transform (FFT) and determines the XY offset (not necessarily an integer) by which the frame should be shifted. The frame is then shifted using FFT-based interpolation. A disadvantage of this approach is that it is driven by the low spatial frequencies that dominate images, to the expense of the high-frequency content that is essential for registration. At typical magnifications, this implies ignoring somata and other calcium-filled cellular compartments. To emphasize the high-frequency content, we used phase correlation, which applies spatial whitening to the images before computing the cross-correlation map [Alba et al., 2015, Foroosh et al., 2002]. We extended this method to detect sub-pixel shifts (down to 1/10 of a pixel), by interpolating the phase correlation map near its maximum with a squared-exponential kernel (kriging). We tested the resulting algorithm on simulated data with known translation and realistic noise and found that it outperformed the standard method of cross-correlation (Figure A.1). Our method was $>$ 15 times faster than upsampled cross-correlation (Figure A.1b), and at least as accurate (errors as low as 0.1 pixels, Figure A.1c) . Moreover, our algorithm could be further accelerated $\approx$ 4 times using GPU-based computations.

Non-rigid registration

As an optional step, Suite2p can correct for rotational and non-rigid brain movements [Greenberg and Kerr, 2009]. For this step, we divided the image in blocks, and used phase correlation to estimate XY offsets for each block. We then interpolated these XY offsets using Gaussian basis functions centered on the midpoints of each block, thus generating a globally non-rigid transformation.

We applied this non-rigid method successfully in experiments where movements were large relative to the size of the regions of interest such as cells, boutons or spines

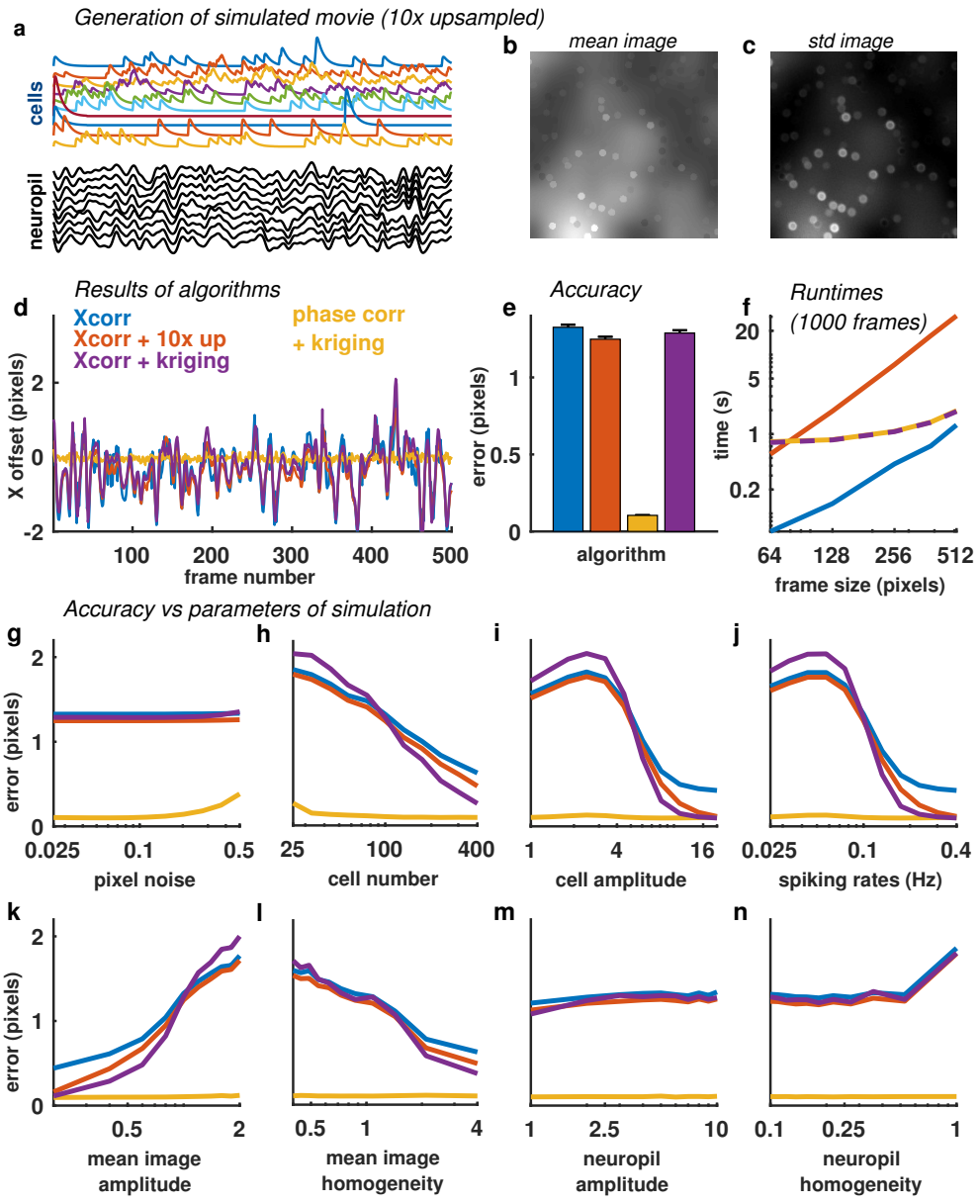

Figure A.1: Correcting rigid motion with subpixel phase correlation. **(a)** Example frame from a simulation of imaging data, subject to random rigid movement and activity-dependent changes in cellular and extracellular fluorescence. The cell masks were disks distributed randomly over the field of view. **(b,c)** Mean and standard deviation of fluorescence signal (without image registration). **(d)** Residual registration error (x offset) as a function of time, for four registration algorithms (standard cross-correlation, cross-correlation with Fourier upsampling, cross-correlation with kriging interpolation and phase correlation with kriging interpolation). Phase correlation with kriging provides best performance. **(e)** Mean residual error for these four algorithms (averaged over 10 randomized simulations). **(f)** Runtimes of registration for the four algorithms. Color codes as in **d**, dashed line indicates identical runtime. **(g-n)** Mean residual error as a function of eight simulation parameters: **(g)** pixel noise, **(h)** cell number, **(i)** amplitude of cell responses, **(j)** cell spiking rates, **(k)** amplitude of mean image, **(l)** spatial homogeneity of the mean image, **(m)** amplitude of the neuropil activity, and **(n)** homogeneity of the neuropil activity. Phase correlation with kriging provides better performance over a wide range of parameter values.

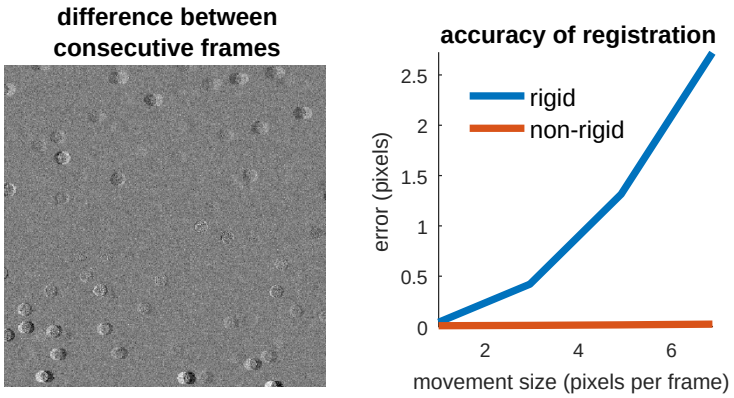

Figure A.2: Correcting nonrigid motion with block registration. (a) Difference of consecutive frames in a simulation where XY motion varies line-by-line. Because the time taken to scan all lines in a frame is comparable to the timescale of motion, cells near the top of the frame move in opposite direction compared to cells near the bottom. **(b)** Non-rigid block based registration is able to correct movements up to very large sizes, where the rigid method results in errors of several pixels.

(Movies 2-3 , Figure A.2). More evidence for the validity of the method comes from an implementation in another processing pipeline [Pnevmatikakis and Giovannucci, 2017]. These authors compared their implementation with our earlier one [Pachitariu et al., 2016b], and found that both implementations performed very well and much faster than previous methods. Some differences in performance between our implementation and theirs might be attributable to different parameter settings, such as the number of interpolation blocks.

Correcting movement

We acquired full-plane subpixel shifts from phase correlation with kriging. We shifted the images by subpixel amounts by performing the shifts in the FFT domain.

If non-rigid registration was performed, then the shifts varied across the pixels in the imaged plane. We discretized the shifts to integer values and applied the shift to the original image at the pixel level.

Appendix B

Drift correction for calcium imaging

As we have shown in the previous chapter, movements of the animal create large 2D motion aligned to the imaging plane. Naturally, these movements must also be generating vertical Z-motion, but such "Z-drift" is difficult to visually observe in the raw data. The effects of Z-drift are not well understood, and typically ignored during data processing, despite the potential confounds to behavioral experiments. Correcting for Z-drift is thought to require an auxiliary Z-stack recording, which is time-consuming to obtain, and may not accurately represent the conditions during the original recording. Here we demonstrate a novel algorithm for detecting and correcting Z-drift, which does not require an auxiliary Z-stack recording, and is thus convenient to use under typical recording conditions. Our method relies on the observation that fluorescence changes to pixels within a cell are either caused by spiking, and thus highly correlated across pixels, or caused by Z-drift, and thus different between pixels in the center of a cell compared to pixels in the cytoplasm. We validate this new method by comparing it to Z-drift inferred from recordings with auxiliary Z-stacks. We demonstrate even better performance when a second, non-functional channel is acquired during the recording. Finally, we show that not correcting for Z-drift causes major confounds, because it appears to create tuning of neural activity to behavioral variables.

Introduction

[1] Several of our analyses depend crucially on being able to acquire long, stable recordings of neural activity. However, over long periods of time, the imaged tissue may move relative to the objective. While 2D motion in the imaging plane is typically accounted for by most data processing pipelines, Z-drift is almost always not. This may, perhaps, not matter much, because the Z-drift may not produce sufficiently large fluorescence changes to be noticeable. However, over the course of hours, larger Z-drifts become likely, as we show below. In addition, large Z-drift may be generated on fast timescales, if an animal is engaging in behaviors, such as running or licking [Chen et al., 2013].

In Figure B.1, we show an example of a recording with Z-drift. Cell 1's activity is not dominated by drift. When the cell's activity is corrected by neuropil subtraction (subtraction of the averaged activity of surrounding pixels), the cell's activity lacks slow timescale changes in its fluorescence. Cell 2, however, still has slow timescale changes in its fluorescence after neuropil correction. The fluorescence of the non-functional imaging channel (the red channel) reflects the slow timescales of the cell. Any changes in a non-functional imaging channel cannot be due to the activity of the cell – instead they are caused by the movement of the tissue over the recording.

But are long recordings, on the order of three hours, really necessary? We offer two examples of such recordings necessary for other chapters of this thesis: 1) to estimate the dimensionality of a neural recording we are limited not only by the number of neurons recorded but also by the recording duration [Gao and Ganguli, 2015]; hence, to estimate the true dimensionality, we must acquire long recordings; 2) quantifying the relation between spontaneous activity and spontaneous behaviors requires a large number of different behavioral states, which can only be acquired over long periods of time, because animals are likely to perform the same behaviors for several seconds or even minutes. To enable these studies, we needed long recordings, and thus we developed robust motion correction methods.

In addition to enabling our own studies, this work establishes two-photon microscopy as a robust method for neural recording. When processed with our algorithms, the two-photon data is unaffected by XYZ motion of the brain relative to

[1] The work described in this chapter was done in collaboration with Marius Pachitariu.

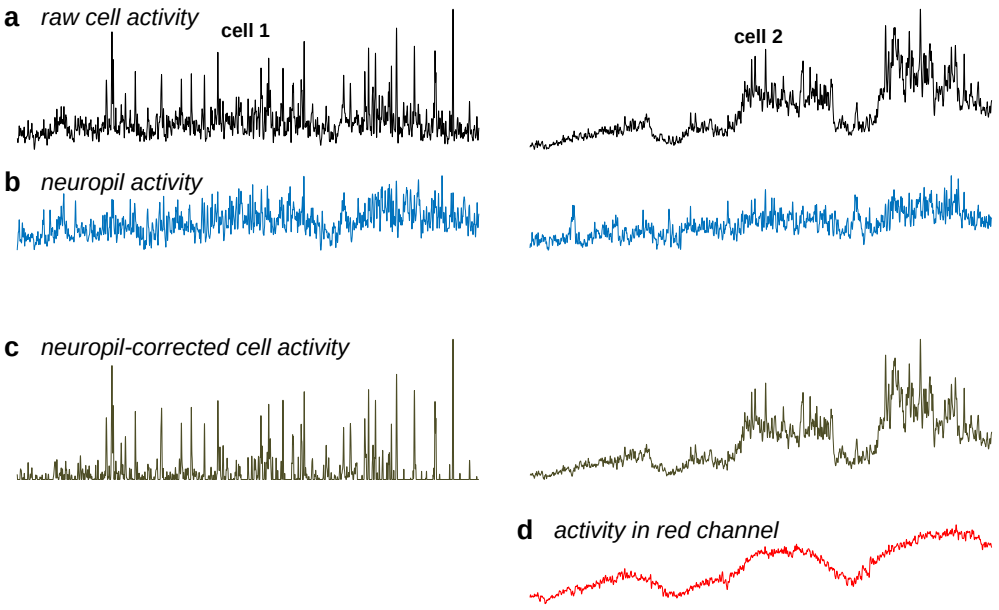

Figure B.1: Example cell activity in a recording with Z-drift. This is a two-photon calcium imaging recording that lasted 1.5 hours. **(a)** The raw cell fluorescence from GCaMP from two cells in the recording. **(b)** The neuropil activity surrounding each of these cells. The neuropil activity is defined as the sum of the fluorescence in pixels surrounding the cell. The surround is defined as pixels within 2.5 cell diameters of the cell, and excludes pixels within other detected cells. **(c)** The neuropil activity is subtracted from the cell fluorescence in **a**. **(d)** The fluorescence of the cell across the recording from a non-functional imaging channel. The imaging was performed in a GAD-cre x td-Tomato mouse, cell 2 is a GAD+ cell.

the recording instrument. Similar developments are yet to be achieved for electrophysiological recordings [Pachitariu et al., 2016a]. These improvements are important, because any behaviors the mouse engages in (running, whisking, grooming etc.) might be correlated with neural activity (see Chapter 5), but also may move the neural field of view, thus generating an artifactual relation to the recorded data.

We separate the Z-drift motion correction strategy into two parts: 1) detecting the amount of motion and 2) correcting the motion. We incrementally describe three strategies to detect the drift. The first strategy requires a dense volume recording ("Z-stack") in addition to the single-plane calcium movie. After correcting each recorded frame for XY movement, we compute the frame's correlation with all the frames of the Z-stack. We estimate Z-position by the maximal correlation, interpolated at a sub-plane level. In the second strategy we do not require a Z-stack,

but require simultaneous acquisition of a non-functional imaging channel, for example a red channel, which records td-Tomato in subsets of neurons, in our case interneurons. Even though the red channel does not have spiking-related activity, the Z-movements result in fluorescence changes which, we show, can be tracked, and processed to obtain an estimate of the Z-position. Finally, the third strategy does not require any auxiliary recordings, and only uses the recorded calcium movie. It takes advantage of the differential changes that Z-movements generate on pixels in the nucleus versus the cytoplasm of a cell. These differential changes allow us to distinguish fluorescence changes due to drift from those due to spiking. The latter generates fluorescence changes which are much more highly correlated across pixels.

In the second portion of this chapter, we develop methods to remove the effect of drift on neural recordings. We demonstrate that a running baseline estimate of the cell's fluorescence effectively removes the dependence of fluorescence on Z-position. This correction can also be confirmed on recordings without an auxiliary Z-stack, by using the third detection strategy described in the first part of this chapter.

The algorithms described here are implemented as freely available software, as part of Suite2p, a full data processing pipeline for calcium imaging processing at www.github.com/cortex-lab/Suite2P.

Calcium imaging baseline fluorescence

Detection of Z-movement in neural recordings

Z-position estimation using auxiliary Z-stacks

For this correction technique, a dense multi-plane imaging recording (a Z-stack) must be available (Figure B.2a,b).

Relative position of imaging planes in tissue

The piezo controller moves at a constant speed throughout the recording. It moves in the YZ axis. Thus, there is a change in the z position across the acquistion of a plane. This change in z (dz) is proportional to the total z distance traversed by the piezo z_{total}

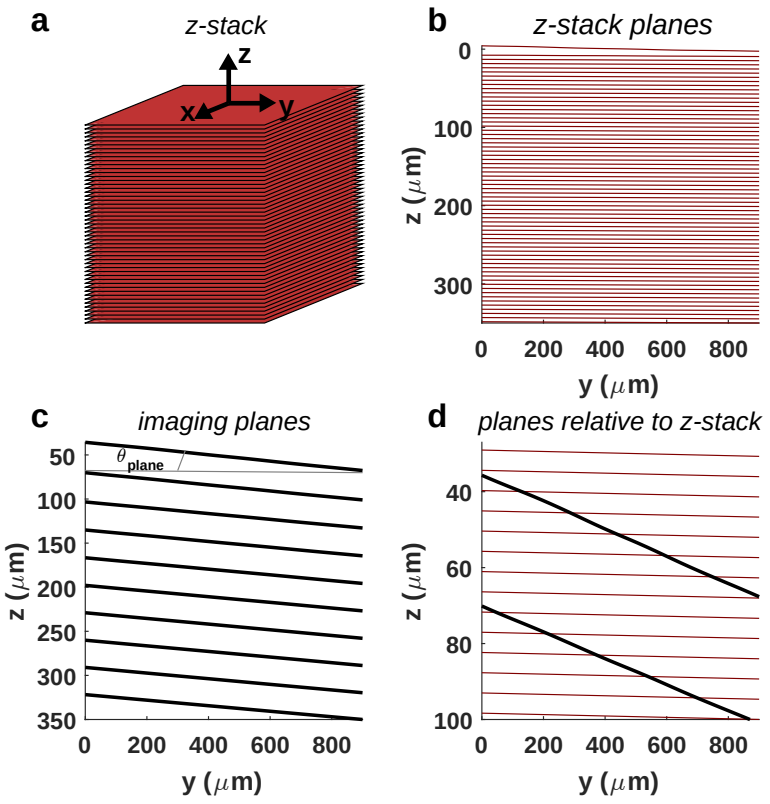

Figure B.2: Configuration of imaging planes. **(a)** A three-dimensional representation of a Z-stack. The Z-direction is perpendicular to the surface of the brain. Each plane is 2 μm apart **(b)** The position of each imaged z-stack plane in the YZ plane. Each line is a different imaging plane. **(c)** The position of the planes imaged in multi-plane configuration. Each line is a different imaging plane. These planes are acquired at an angle in the YZ plane because the piezo moves through the tissue at a constant speed (it does not stop when acquiring a plane). This angle is defined as θ_{plane}.

and the number of planes imaged n_{planes}:

$$dz = \frac{z_{total}}{n_{planes}}.$$

The angle of the plane with respect to the horizontal is then

$$\theta_{plane} = \tan^{-1}\left(\frac{dz}{L_y}\right)$$

where L_y is the length of the plane in the direction in which the piezo moves (the Y direction, see Figure B.2c). The Z-stack is also acquired at an angle with respect to the horizontal $\theta_{Z-stack}$. Thus, the approximate angle offset between a single imaged plane

and the Z-stack is

$$\theta \approx \theta_{plane} - \theta_{Z-stack}.$$

We will use this approximate angle to initialize an algorithm that more precisely finds the geometric transformation between the Z-stack and the imaged plane.

Aligning imaging planes to the Z-stack

We developed and implemented an algorithm to find the optimal mapping from the imaging plane to the Z-stack T_{aff}

$$\begin{bmatrix} y_{Z-stack} \\ x_{Z-stack} \\ z_{Z-stack} \end{bmatrix} = T_{aff} \begin{bmatrix} y_{plane} \\ x_{plane} \\ 1 \end{bmatrix}$$

and used this mapping to compute the plane's location in the Z-stack across time.

The optimization procedure is as follows:

1. First, we estimated the approximate location of the plane in the Z-stack. Due to the angled acquisition of the plane in multi-plane imaging, we estimated the angle between the planes and the Z-stack θ as outlined above. The Z-stack was then transformed by this angle in order to be parallel to the imaged planes. We next found the optimal Z-position of the plane within the stretched Z-stack by performing phase-correlation (see 2D registration chapter) between the imaged plane and each plane of the transformed Z-stack. The optimal Z-position was the plane with the highest phase-correlation with the imaged plane (z_{est}). The phase correlation also returned the XY offsets of the plane with respect to the Z-stack.

2. Next, we divided the plane into 64 patches in y and x with centers $[\mathbf{y}_{patch}\ \mathbf{x}_{patch}]$. We used the estimated position of the plane within the Z-stack to estimate the position of the patch in the Z-stack:

$$\begin{bmatrix} \mathbf{y}_{estZ} \\ \mathbf{x}_{estZ} \\ \mathbf{z}_{estZ} \end{bmatrix} \approx \begin{bmatrix} \mathbf{y}_{patch} + y_{offset} \\ \mathbf{x}_{patch} + x_{offset} \\ z_{est} - \mathbf{y}_{patch} \tan(\theta) \end{bmatrix}$$

3. We then created patches of the Z-stack surrounding these estimated locations with a spread in z of ± 10 planes. We computed the phase correlation of the plane patch with the Z-stack patch, and found the optimal position of the patch within this patch of the Z-stack

$$\begin{bmatrix} \mathbf{y}_{optZ} \\ \mathbf{x}_{optZ} \\ \mathbf{z}_{optZ} \end{bmatrix}.$$

An example of these Z-stack patches aligned to the imaged plane patches is shown in Figure B.3.

Using these paired values, we can estimate the transformation from the plane to the Z-stack by finding

$$T^*_{aff} = \min_{T_aff} \left(\left\| \begin{bmatrix} \mathbf{y}_{optZ} \\ \mathbf{x}_{optZ} \\ \mathbf{z}_{optZ} \end{bmatrix} - T_{aff} \begin{bmatrix} \mathbf{y}_{patch} \\ \mathbf{x}_{patch} \\ \mathbb{1} \end{bmatrix} \right\|^2 \right).$$

The least squares solution is

$$T^*_{aff} = \left(\begin{bmatrix} \mathbf{y}_{patch} \\ \mathbf{x}_{patch} \\ \mathbb{1} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{patch} \\ \mathbf{x}_{patch} \\ \mathbb{1} \end{bmatrix}^T \right)^{-1} \left(\begin{bmatrix} \mathbf{y}_{patch} \\ \mathbf{x}_{patch} \\ \mathbb{1} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{optimal} \\ \mathbf{x}_{optimal} \\ \mathbf{z}_{optimal} \end{bmatrix}^T \right)$$

We then re-initialized the estimated locations in the Z-stack

$$\begin{bmatrix} \mathbf{y}_{estZ} \\ \mathbf{x}_{estZ} \\ \mathbf{z}_{estZ} \end{bmatrix} = T^*_{aff} \begin{bmatrix} \mathbf{y}_{patch} \\ \mathbf{x}_{patch} \\ \mathbb{1} \end{bmatrix}$$

and used these locations to create the patches in the Z-stack with which to align the patches in the planes. Then we re-computed the optimal positions of the patches in the Z-stack and re-estimated T^*_{aff}. We repeated this step at least 5 times to ensure convergence of T^*_{aff}.

4. Once T^*_{aff} is estimated, we can align the Z-stack to the imaged plane. We

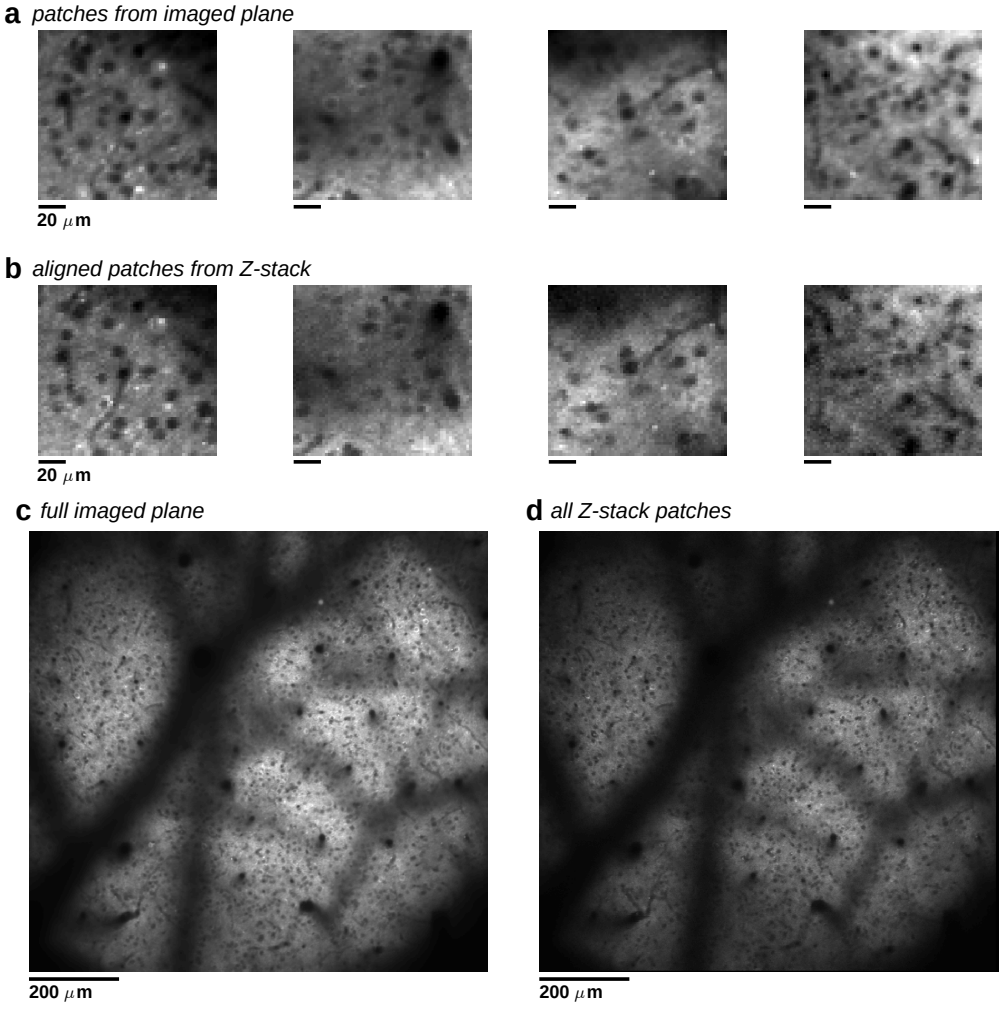

Figure B.3: Z-stack aligned to patches in imaged plane. (a) Mouse visual cortex was imaged for two hours with 12 plane imaging. One plane is represented in this figure. Each thumbnail is a patch from this plane. **(b)** The Z-stack was aligned to the patches in the plane using the coordinates $[\mathbf{y}_{optZ}\ \mathbf{x}_{optZ}\ \mathbf{z}_{optZ}]$. The intensity values at these coordinates for each of the patches in **a** is shown. **(c)** The full mean image of this plane. **(d)** All of the patches aligned to the imaged plane in **a, c** are stitched together into a single image.

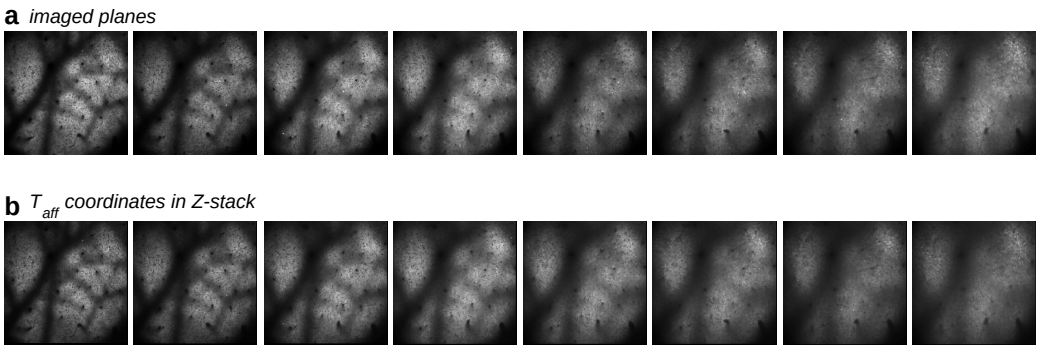

Figure B.4: Z-stack aligned to planes acquired in multi-plane imaging. (a) Mouse visual cortex was imaged for two hours with 12 plane imaging. Each thumbnail is the mean image from eight of the imaged planes - they are sorted in order of increasing depth from the surface of the brain. **(b)** For each plane, T^*_{aff} was estimated, and this transformation was applied to the coordinates in the planes in **a** in order to compute the coordinates of the planes within the Z-stack. Each thumbnail shows the pixel intensities in the Z-stack at the coordinates from the transformation T^*_{aff}.

multiplied T^*_{aff} by the y and x coordinates of the imaged planes to derive their coordinates in the Z-stack.

In a standard experiment, we imaged mouse visual cortex with 12 planes (Figure B.4a, 8 planes are shown). After imaging for two hours, we imaged with 200 planes (a Z-stack) for 30 minutes. For each plane, T^*_{aff} was estimated by the procedure outlined above. Figure B.4b shows the pixel intensities in the Z-stack at the coordinates from the transformation T^*_{aff}. The correlation between the pixel intensities of the imaged plane and the aligned Z-stack image was 0.98 ± 0.003 across planes in this experiment.

Computing Z-position across the recording

We applied this transformation T^*_{aff} to the Z-stack in order to create a Z-stack parallel to the plane and padded by ± 10 planes (equivalent to ± 20 μm). We whitened this aligned Z-stack in order to emphasize high-frequency features, such as cell edges. Next, we whitened the imaged plane at time t, and then correlated this whitened plane with the whitened Z-stack across all planes in the Z-stack. We performed spline interpolation on this correlation profile in Z to detect subpixel shifts. The maximum of the correlation profile in Z was inferred as the Z-position for the imaged plane at time t (Z-pos). This was repeated for all t, producing a Z-position for each time-point in the recording (Figure B.5a). For some of the analyses below, we

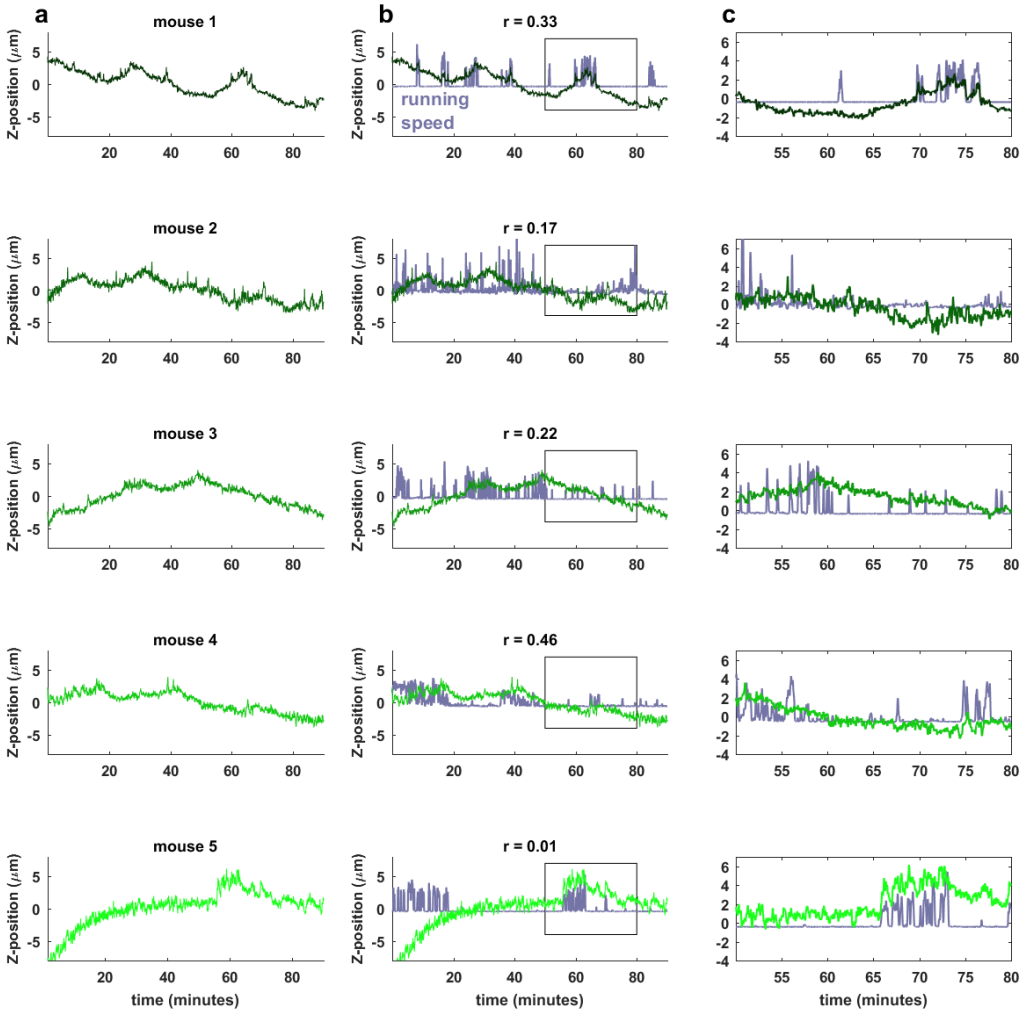

Figure B.5: Z-position of recording across time. (a) The Z-position of one plane over ninety minutes. The Z-position is smoothed by a Gaussian with a standard deviation of four time-points (1.6 seconds). Each plot is the Z-position of a plane in a different recording from a different mouse. The recordings move in a range of 10 μm in the tissue. **(b)** The running speed of the mouse in a thirty minute period. The speed is z-scored and offset by the mean of the Z-position of the plane in the window. Running bouts (on the order of minutes) can cause shifts in the Z-position of the plane. The correlation between the running speed and the Z-position in this time period is reported as an r^2 value at the top of the plot.

smoothed the Z-position estimate by a Gaussian with a standard deviation of four time-points (1.6 seconds).

In all recordings we analyzed, there was slow drift in the position of the plane. This slow drift covered a range of $\approx$10 μm in the tissue. There were also faster movements that occurred within a single minute. These faster movements were correlated with the running speed of the mouse (Figure B.5b,c). These data show that when mice are walking or running, their adjusted posture shifts the imaging plane by

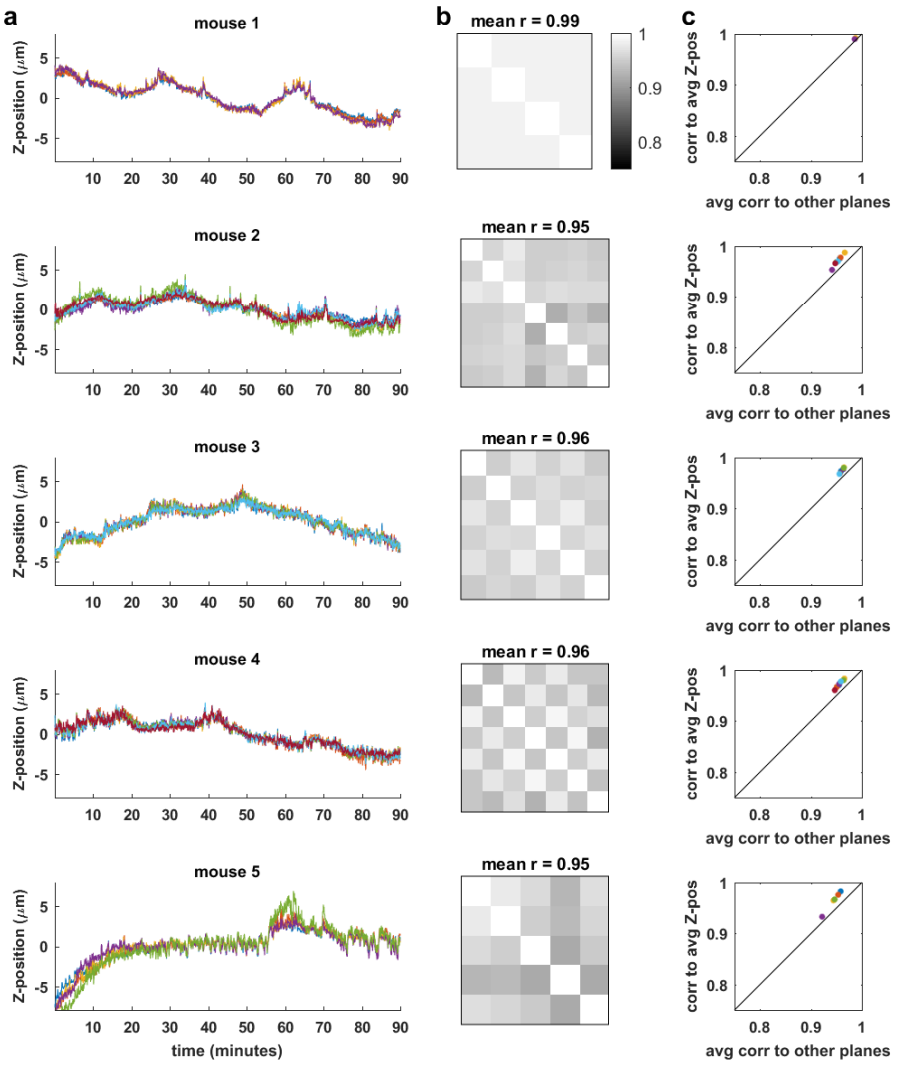

Figure B.6: Z-position of multiple planes in the recording. **(a)** The Z-position of multiple planes over 90 minutes. Each plane's Z-position is plotted as the Z-position subtracted by its mean position in the Z-stack. It is smoothed by two time-points. **(b)** The correlation between planes for each recording. **(c)** Each point represents a single plane. Its x-coordinate is the average correlation between its Z-position and the Z-position of each plane in the recording separately (the average of each column of the matrix in **b**, excluding the diagonal). Its y-coordinate is the correlation between its Z-position and the Z-position averaged across all other planes. Estimation of Z-position is improved by combining information from all other planes.

several microns. Similar shifts are likely when mice adjust their posture while stationary. This suggests that Z-movement may alter the relation of the recorded traces with running. Several studies have investigated the correlation between calcium recordings and running, but none have performed this correction [Dipoppa et al., 2016, Pakan et al., 2016, Fu et al., 2014].

We also found that the Z-positions for multiple planes in a recording were consistent (Figure B.6a,b). We correlated the position of one plane with the mean position across all other planes, and found that this correlation was higher than correlating a single plane position from another single plane position (Figure B.6c). In other words, combining information from all other planes improved the estimate of the Z-position for all planes in all recordings (Figure B.6c). These Z-position estimates were correlated with an $r > 0.96$ for all recordings. Fast movements were also highly consistent across planes.

Z-position estimation using a non-functional imaging channel

We developed a technique to detect the Z-position of an imaging plane if a non-functional imaging channel is simultaneously acquired. We applied this technique to a recording in which the non-functional channel was td-Tomato labelling in GAD+ inhibitory neurons. Thus, we termed this non-functional imaging channel the "red" channel. As the plane moves in the Z direction, the cells in the red channel become brighter or dimmer (Figure B.7a). We can plot the fluorescence traces of these cells in the red channel as a function of time (smoothed with a Gaussian of standard deviation of 1.6 seconds) (Figure B.7c). The fluorescence traces of these cells in the red channel are correlated with the Z-position of the plane as estimated from the Z-stack.

This suggests that a non-functional channel recording could produce an accurate estimate of the Z-position of the imaging plane. We found that the first principal component of non-functional, red channel imaging contained the Z-position information. For each cell, we computed its red channel fluorescence trace, and subtracted the mean of this trace from itself. This is a matrix of neurons by time, termed F_{red}. We then binned these traces in time bins of 30 seconds, in order to reduce the noise in the traces in time, and computed the singular value decomposition:

$$[U\ S\ V^T] = \text{svd}(F_{red\ (binned)}).$$

The weights of the cells onto the first principal component are given by the first column in matrix U, termed u_1. We sorted the cells by these weights and plotted a subset of them in Figure B.8a. In order to estimate the Z-position at each time-point,

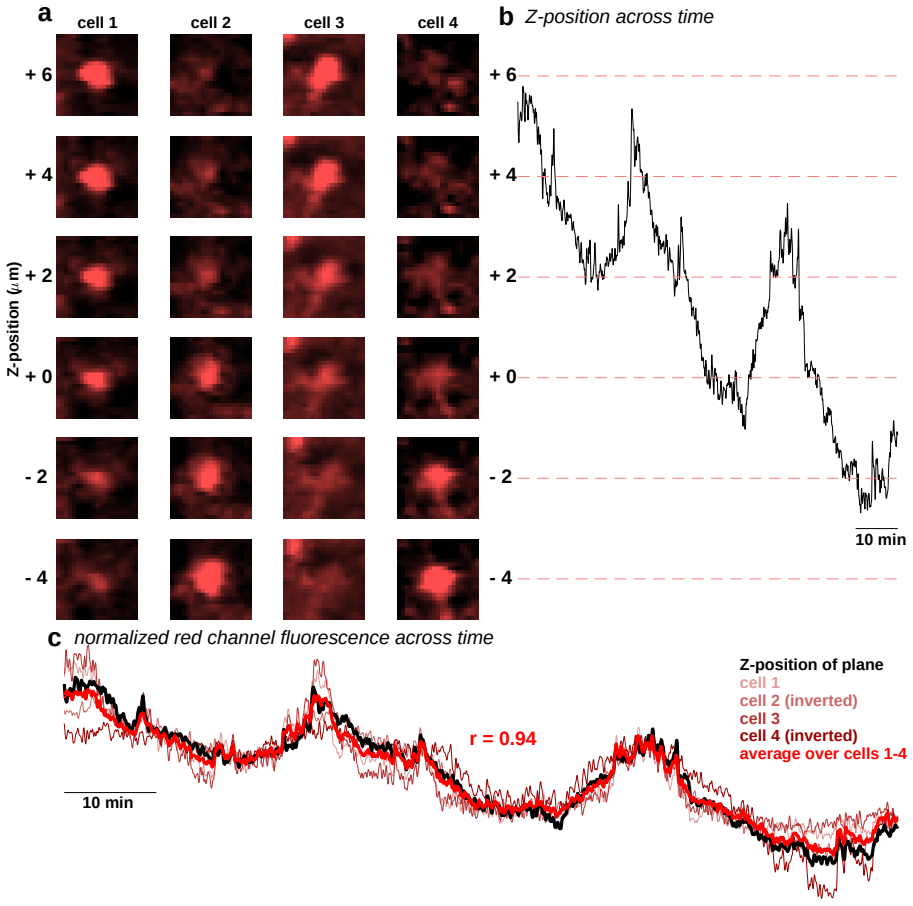

Figure B.7: Red channel fluorescence across depth. (a) Example cells in the red channel and their fluorescence profiles in XY across depth. Cells 1-4 appear to have monotonic fluorescence profiles as a function of depth. **(b)** The Z-position of a plane of the recording across time (90 minute period). **(c)** The red channel fluorescence of each cell in **a** as a function of time (90 minute period), smoothed with a Gaussian of standard deviation 1.6 seconds, and z-scored. The fluorescence of cells 2 and 4 are inverted in order to be positively correlated with the Z-position. The Z-position of the plane is z-scored and plotted in black. The average fluorescence of the four cells (with 2 and 4 inverted) is plotted in red. The average fluorescence of these 4 cells across time is correlated with the Z-position of the plane with a Spearman correlation of $r = 0.94$.

we projected in the space of the red cell fluorescence traces aligned to the first principal component:

$$Z_{red} = u_1^T F_{red}.$$

In this recording, the Z-position estimate from the red channel was correlated with the actual Z-position (as estimated from the Z-stack) with $r = 0.99$ (Figure B.8). This suggests that the Z-position of the plane can be estimated accurately from a non-functional channel. Taking a high-density Z-stack image of the imaged tissue may not be necessary for Z-position estimation.

Z-position estimation in single channel GCaMP recordings

Cells that express GCaMP primarily express the protein in their cell membranes. Thus, in the center of the cell where the nucleus is located, there is often lower fluorescence. However, the nucleus does not stretch the cell's full extent in Z. Imagine a sphere with a dark center and a bright shell. Central cross-sections of the sphere look like donuts. But cross-sections further from the center will contain less of the dark center and the bright shell will become more dominant. Examples of cells with these types of Z profiles are shown in Figure B.9. The dark centers of the cells are filled with fluorescence when moving in Z. This suggests that mean fluorescence in the cell may not work as a measure of depth like in the non-functional imaging channel (see B.3.2).

Instead, one could consider the ratio between the fluorescence of the interior of the cell and fluorescence of the exterior of the cell (Figure B.10). In the plane of the cell with the darkest center (the nucleus is largest), the ratio of the interior to the exterior will be very low. In the example in Figure B.10a, this would be at Z-positions from -2 to 0. Moving away from this plane in Z (moving away from the cell's nucleus), the interior of the cell fills with fluorescence, and the interior and exterior fluorescence of the cell become similar. The Z-profile of the interior fluorescence and the exterior fluorescence is shown in Figure B.10b. This cell's ratio varies across its Z-profile. At each time-point of the recording, we computed the logarithm of the ratio of the interior fluorescence to exterior fluorescence of the cell in bins of 30 seconds. We also computed the total fluorescence of the cell in 30 second bins (sum of interior and exterior fluorescence). The ratio of interior to exterior fluorescence was

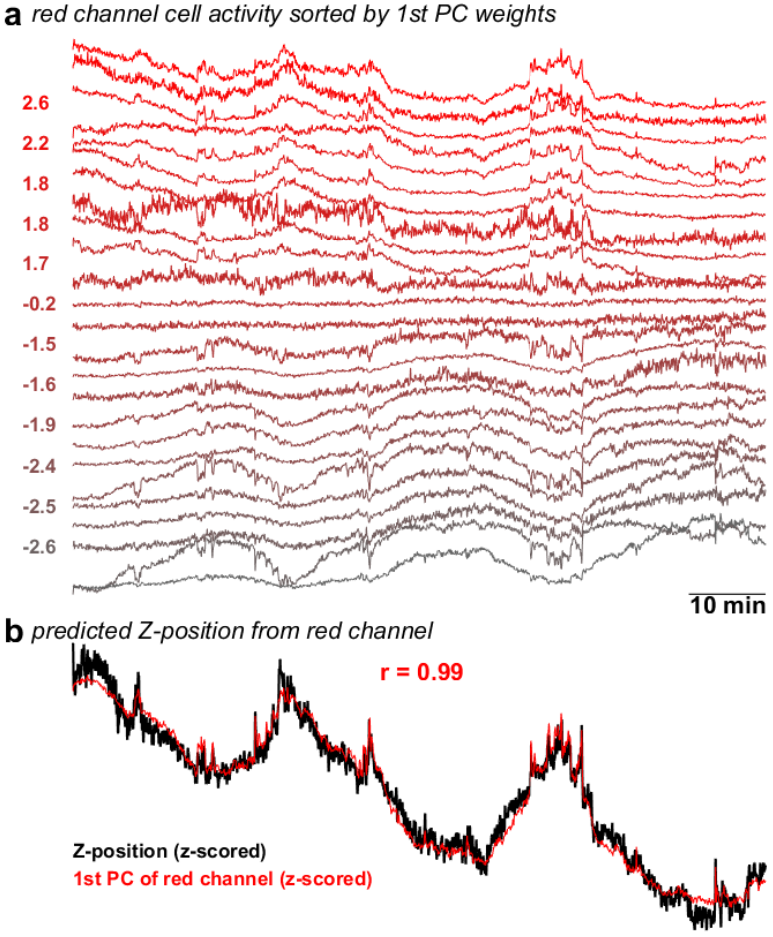

Figure B.8: Unsupervised estimate of Z-position from red channel. **(a)** The fluorescence traces of 25 cells in the red channel as a function of time (90 minute period). The cells are sorted by their weights onto the first principal component of all red cells' fluorescence traces. The first principal component is computed on the cells' fluorescence traces binned in time bins of 30 seconds. **(b)** The prediction Z-position of the plane based on the red channel first principal component. The Spearman correlation between the Z-position estimated from the Z-stack and the Z-position estimated from the red channel is $r = 0.99$.

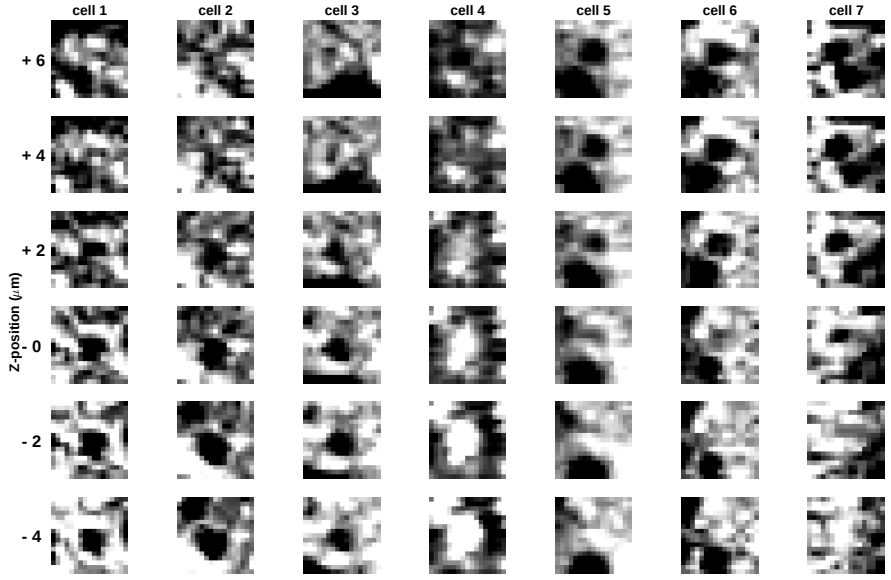

Figure B.9: Z-profile of cells expressing GCaMP6s. Each column is a different GCaMP6s-expressing cell in the Z-stack. Each row is a different depth in the Z-stack.

correlated to the Z-position of the plane with $r = 0.95$ while the total cell fluorescence was correlated to the Z-position of the plane with only $r = 0.76$ (Figure B.10).

Thus, we used the ratio of the interior to exterior fluorescence to estimate the Z-position instead of the total cell fluorescence . We computed the logarithm of the ratio of the interior to exterior fluorescence in the GCaMP channel for all cells ($F_{int/ext}$), and binned the traces in 30 second time bins. We then z-scored each cell's ratio and computed the singular value decomposition of this matrix. We then took the principal component u_1 of this decomposition and projected this back onto $F_{int/ext}$ to estimate the Z-position:

$$Z_{GCaMP} = u_1^T F_{int/ext}.$$

Correcting the drift

In this section, we develop strategies for diagnosing and attenuating or completely removing the effect of Z-drift on fluorescence. We use these strategies as controls and corrections, in order to study the relation between neural activity and multi-dimensional behaviors: whisking, running, arousal and other orofacial and locomotive behaviors. For a concrete example in this chapter, we analyze the

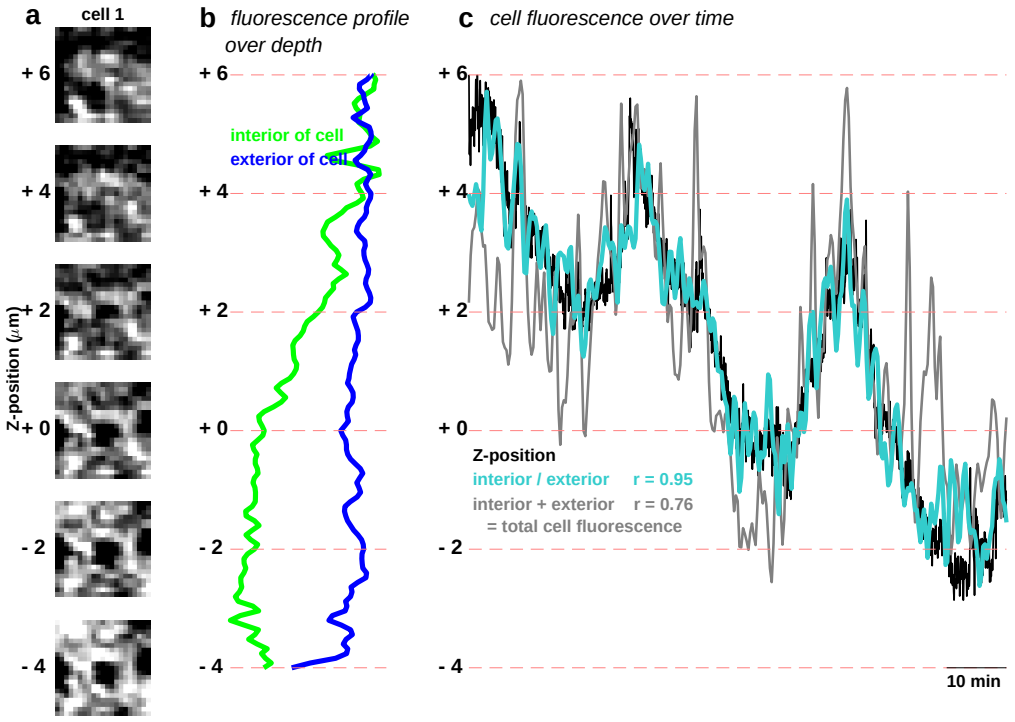

Figure B.10: Ratio of interior fluorescence to exterior fluorescence. correlation of blue = 0.93, correlation of gray = 0.73

influence of running speed on fluorescence.

We have shown in chapter 2 that running is able to modulate brain state effectively, and typically reduces the synchrony levels of neural populations. However, running may also modulate the firing rates of individual neurons [Niell and Stryker, 2010, Schneider et al., 2014, McGinley et al., 2015a], and potentially have differential effects on different neuron classes [Polack et al., 2013, Vinck et al., 2015, Pakan et al., 2016, Dipoppa et al., 2016]. We have seen in chapter 2 that during running, putative interneurons have large increases in firing rates, which are much larger than the increases in the rest of the neural population. Nonetheless, the precise relation between running and circuit dynamics is not well understood and actively debated [Pakan et al., 2016, Dipoppa et al., 2016].

The effect of Z-drift on neural recordings is also not well understood, and it may explain some of the differences in results between different labs or recording modalities. In our preparations for optical imaging, we often glue together multiple pieces of glass to make a coverslip. The role of the inner pieces, which go inside the skull, is to push down on the brain and hold it steady during subsequent

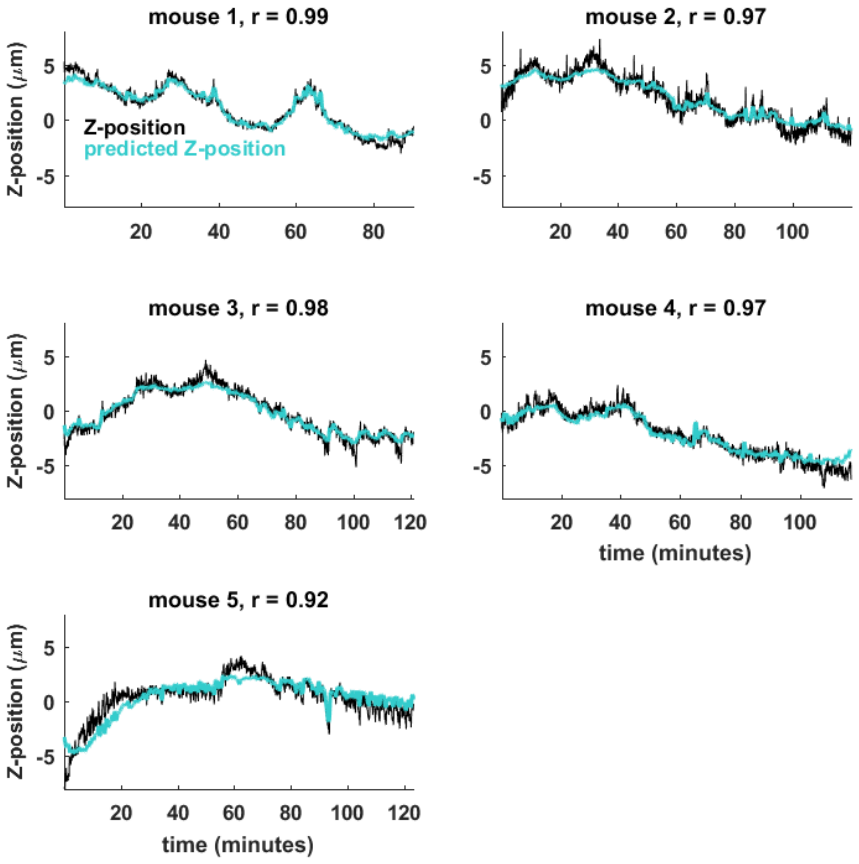

Figure B.11: Z-position of multiple planes in the recording.

head-fixation. Typical headfixed electrophysiology experiments do not have this additional stabilization. As such, drift may have an even larger effect on such recordings, and a larger effect still on non-headfixed recordings. Here we will only focus on the effect of drift on two-photon calcium imaging, and show that it can be diagnosed and removed effectively.

First, we show that at least some of the running-related fluorescence is artifactual due to Z-drift. To remove this artifact, we apply a running baseline procedure to correct the drifts in individual cells. We show that after correction, the artifactual relationship no longer holds, but the tuning of cells to running persists.

Using inferred drift to diagnose potential artifacts

Movement artifacts are a confound for almost all behavioral studies. For example, if an animal is trained to report a perceptual decision, the report is almost always a

motor action such as a lick, lever press or reach. If a population of neurons is found to be tuned for the perceptual decision, this may simply be due to the movement of the brain relative to the electrode or microscope, if the movement happens in stereotypical fashion on each trial. To determine if the effect is neural or artifactual, a powerful control could proceed as follows:

1. determine the effect of the movement artifact on neurons,
2. summarize this effect by a number indicating the strength and sign of the artefactual motor influence,
3. summarize the tuning of the neuron by a number,
4. correlate these two sets of numbers across the recorded population.

If the correlation is not significant, the neural tuning is likely not a consequence of movement artifacts. In our "tuning to running" example, having computed the z-drift in one of the three ways described above, we can compute for each cell its correlation with z-drift, as well as its correlation with running, and then correlate these two correlations across the population of neurons. A cell's baseline fluorescence will typically be positively/negatively correlated with z-drift if it sits above/below to the average imaging plane, respectively. If a cell sits roughly in the location of the average imaging plane, then z-drifts up or down will both decrease the cell's baseline fluorescence, resulting in an more nonlinear relation between fluorescence and drift. We only focus here on the linear, monotonic relationships, but note that the nonlinear relationsip can be treated similarly.

We find that the recorded fluorescence traces are indeed tuned significantly to running, with equal amounts of positive and negative correlations (Figure B.12a). Thus, we will measure the overall amount of running-related activity by the variance of this distribution, rather than by its mean. Similarly, the fluorescence traces were correlated with Z-position even more strongly, again with equal amounts of positive and negative correlations (Figure B.12b). Finally, the correlation with running and that with Z-drift were highly correlated with each other (Figure B.12c). This is worrying, because it is exactly the effect one would expect if Z-drift caused the fluorescence fluctuations during running. We repeated these analyses in 5 mice, where Z-stacks were available, and found similar effects: the correlations with running were centered on zero but had a large variance (Figure B.12d), the correlations with Z-position were also centered on zero and had an even larger

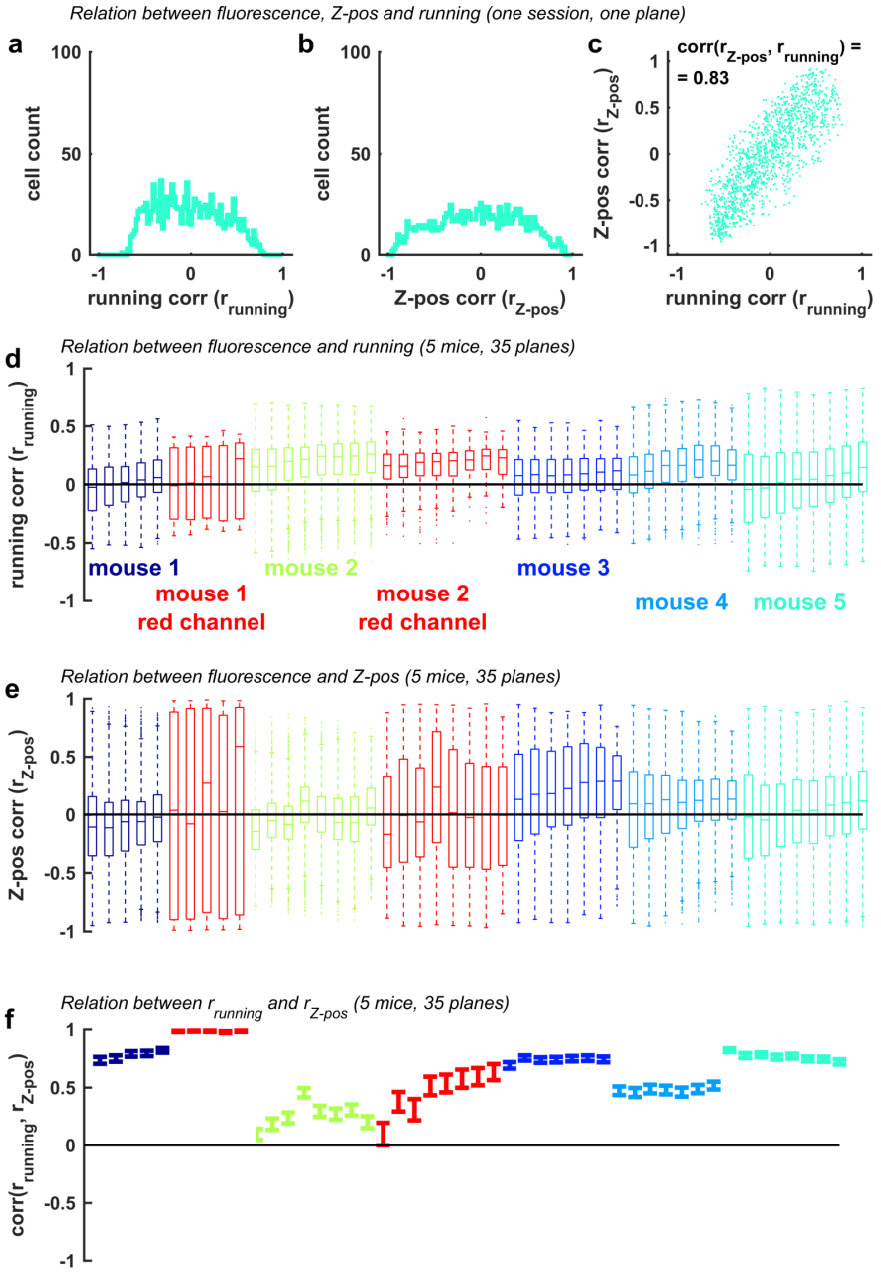

Figure B.12: Relation between fluorescence, running and Z-position. (a-c) For one recording session, in one plane, the distribution of fluorescence correlation with running (a), with Z-position (b), and the relation between these two correlations (c). **(d,e)** Summarized distributions of fluorescence correlation with running (r_{running}, d) and Z-position ($r_{\text{Z-pos}}$, e) across datasets. Boxes represent standard deviations, and the approximate range of values is indicated by the whiskers. **(f)** All correlation coefficients between r_{running} and $r_{\text{Z-pos}}$.

variance (Figure B.12e), showing that Z-drift is a better predictor of neural activity than running, and the correlation between these two sets of correlations was always positive, and often large (Figure B.12f).

This example has shown that there is relation between a cell's correlation with z-drift, and it's correlation with running. Since the z-drift itself is correlated with running, this relation might arise because either

1. cells' spiking is correlated with running, and so their fluorescence is indirectly correlated with z-drift.
2. cells' fluorescence is correlated with z-drift, and so it is indirectly correlated with running.
3. the fluorescence correlation with running comes from both sources: z-drift and real spiking.

We can show an example of the second possibility directly, by analyzing the fluorescence traces from the red channel (td-Tomato) of a recording, which does not contain spiking-related signals. We find that these traces are indeed significantly correlated with running (Figure B.12d), but much more correlated with Z-drift (Figure B.12e). In this case, the correlation with running must arise due to a correlation with z-drift, because there is no other source of time-dependent signal in these recordings (Figure B.12f).

It might still be the case that the tuning inherited from z-drift is much smaller (in the functional GCaMP channel) than the tuning of the respective neuron's spiking activity. To proceed further, we must remove the influence of z-drift on the recorded signals. If the correction is successful, the fluorescence will no longer be related to z-drift, and so any remaining tuning to running must be neural in origin. The next section describes this correction procedure, and the section after describes the results of the correction.

Baseline correction via morphological opening in time or in Z-position

We would ideally like to use a drift correction strategy that does not require an auxiliary recording of a z-stack. This will allow the correction to be employed in typical experiments, which usually lack a high-quality z-stack. One possibility would be to use our third strategy for detecting z-drift, and correct traces based on that.

While this strategy appears robust for our recordings, we advise users to first validate it on their own data, which might have different statistical properties. Another correction possibility that does not require a z-stack may be based on a running baseline correction. Since drift is typically slow-varying in time, the time-based correction should also remove the effect of drift. In addition, the time-based correction should also remove other slow-varying changes in baseline, such as those due to bleaching. However, the time-based correction will not be able to track well fast drifts, hence it may not be ideal for tasks where an animal's behavior is abrupt and brief, for example if cued with a stimulus.

Both drift and bleaching are thought to create a multiplicative change in a cell's fluorescence. In the case of drift, this happens due to varying size cross-sections of a cell at different imaging planes along its z-profile (Figure B.9). In the case of bleaching, the multiplicative fluorescence change is due to a reduction in the number of available GCaMP molecules [Donnert et al., 2007]. If we had access to the baseline fluorescence of a cell, termed $F_0(t)$, a function of time t, then we would typically correct the trace $F(t)$ as follows

$$F_{\text{corrected}}(t) = \frac{F(t) - F_0(t)}{max(c_0, c_0 + F_0(t))},$$

where c_0 is a small constant which ensures the denominator is large enough. For this transformation to accurately correct the fluorescence, $F_0(t)$ would need to be the true baseline of a cell, in the absence of spiking. This is almost impossible to get in practice, because most neurons fire relatively often. An approximation to this running baseline may nonetheless be sufficient. A typical choice is the running minimum in a medium-sized time window (minutes). To reduce the impact of photon shot noise on the running minimum, traces are typically smoothed first, over short time windows (seconds). Although this is a popular baseline choice [Mao et al., 2001, Vogelstein et al., 2010], the running minimum can produce artifacts when the window size is similar to the timecourse of true baseline changes, which is our case. We cannot use windows smaller than a few dozen seconds, because that would interact with the timescale of GCaMP, and have the net effect of high-pass filtering the true spiking-related signal.

A solution that does not require smaller window sizes is replacing the running minimum with the morphological opening, a strategy which we borrow from the

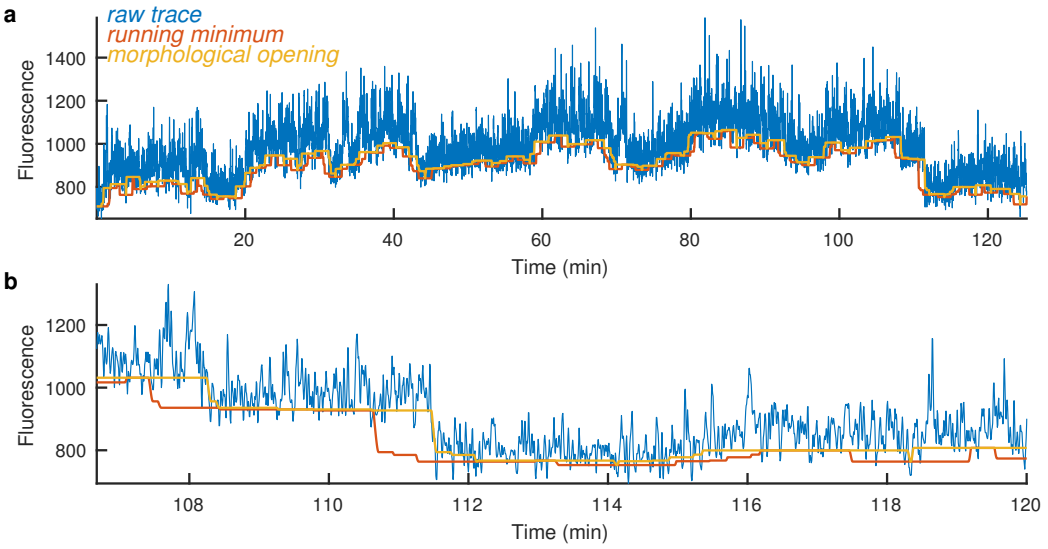

Figure B.13: Running minimum compared to morphological opening. (a) A 2 hour calcium trace recorded at 2.5Hz and smoothed with a 1-sample Gaussian filter. The running minimum (red) and the morphological opening (yellow) are obtained from the same trace smoothed with a 4-sample Gaussian filter. **(b)** Zoom in of (a).

image processing literature [Dougherty and Lotufo, 2003]. Morphological opening is a sequence of an ”erosion” and a ”dilation”. For a one-dimensional signal, ”erosion” is equivalent to a running minimum filter. Similarly, ”dilation” is equivalent to a running maximum. ”Erosion” removes from the trace all features smaller than a given size (in our case the GCaMP transients), while ”dilation” restores the shapes of the remaining, slower-timecourse elements (in our case the drift). These operations are illustrated on a real neuron trace in Figure B.13.

We can use a similar strategy to estimate a baseline that is Z-position dependent, rather than time dependent. One simple strategy is to sort all fluorescence values by estimated Z-drift, rather than by time, and run the same baseline correction procedure along the Z-drift axis, rather than along the time-axis. This strategy has the advantage that the baseline estimation will automatically be more precise for Z-positions that have more samples, and will automatically be smooth at the edges of the Z-position range, where few samples are available.

When run on data from the non-functional, red channel, both the temporal-, and Z- corrections are able to almost perfectly track the baseline (Figure B.14). The two correction methods compute similar baselines on the red channel, and also on the green channel of the first neuron shown in Figure B.14. However, they give different baselines on the green channel of the second neuron, where the temporal baseline

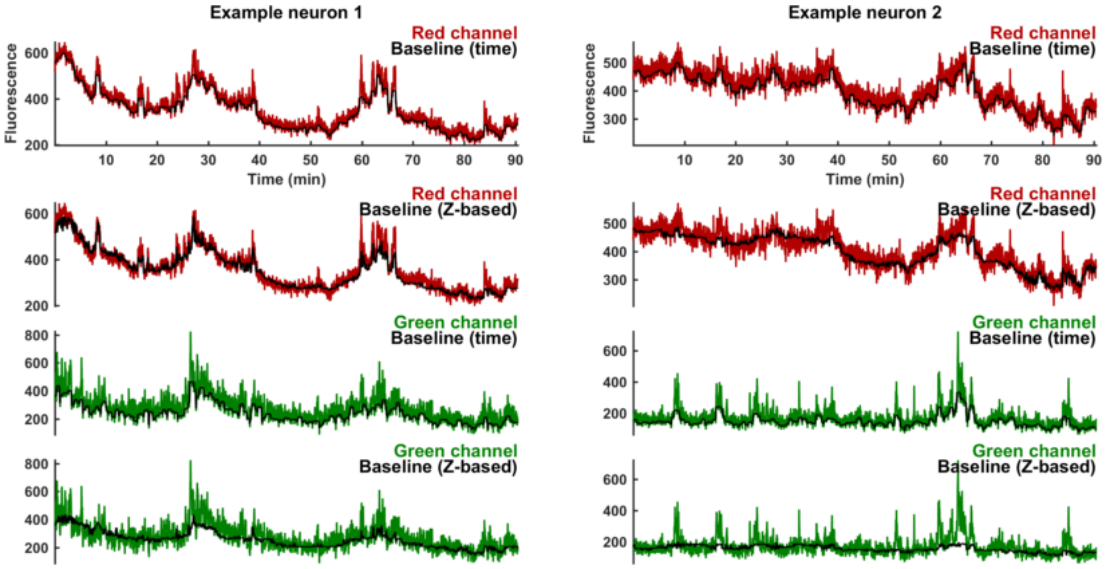

Figure B.14: Temporal and Z-position baselines. For two example neurons, we show their fluorescence traces from the red channel and from the green channel, together with the temporal baseline correction, and the Z-position-based baseline correction.

tracks some of the GCaMP responses, while the z-position method correctly avoids tracking them. As we show below, the temporal baselining strategy may aggressively over-correct traces for periods of continued spiking activity, where the fluorescence does not have time to return back to baseline.

Z-position, but not temporal-, baseline correction removes signal dependence on Z-drift

We applied both methods for drift correction (based on a temporal running baseline, or a Z-position baseline), and re-ran the analyses of Figure B.12. We find that the Z-position baselining is able to successfuly correct the traces, while the temporal baseline only partially corrects them.

First, for the Z-position baseline correction, we find that the dependence of the signals on running was significantly reduced, as shown by the width of the distribution of correlations (Figure B.15a,d). Interestingly, the distributions remained symmetrical around 0. On the other hand, the dependence on Z-position was almost completely removed, as illustrated by the very narrow distribution of correlation with Z-position (Figure B.15be). Finally, the relation between these two sets of correlations was nearly completely nullified (Figure B.15cf), suggesting that any remaining tuning to running must originate in the functional GCaMP activity, not in

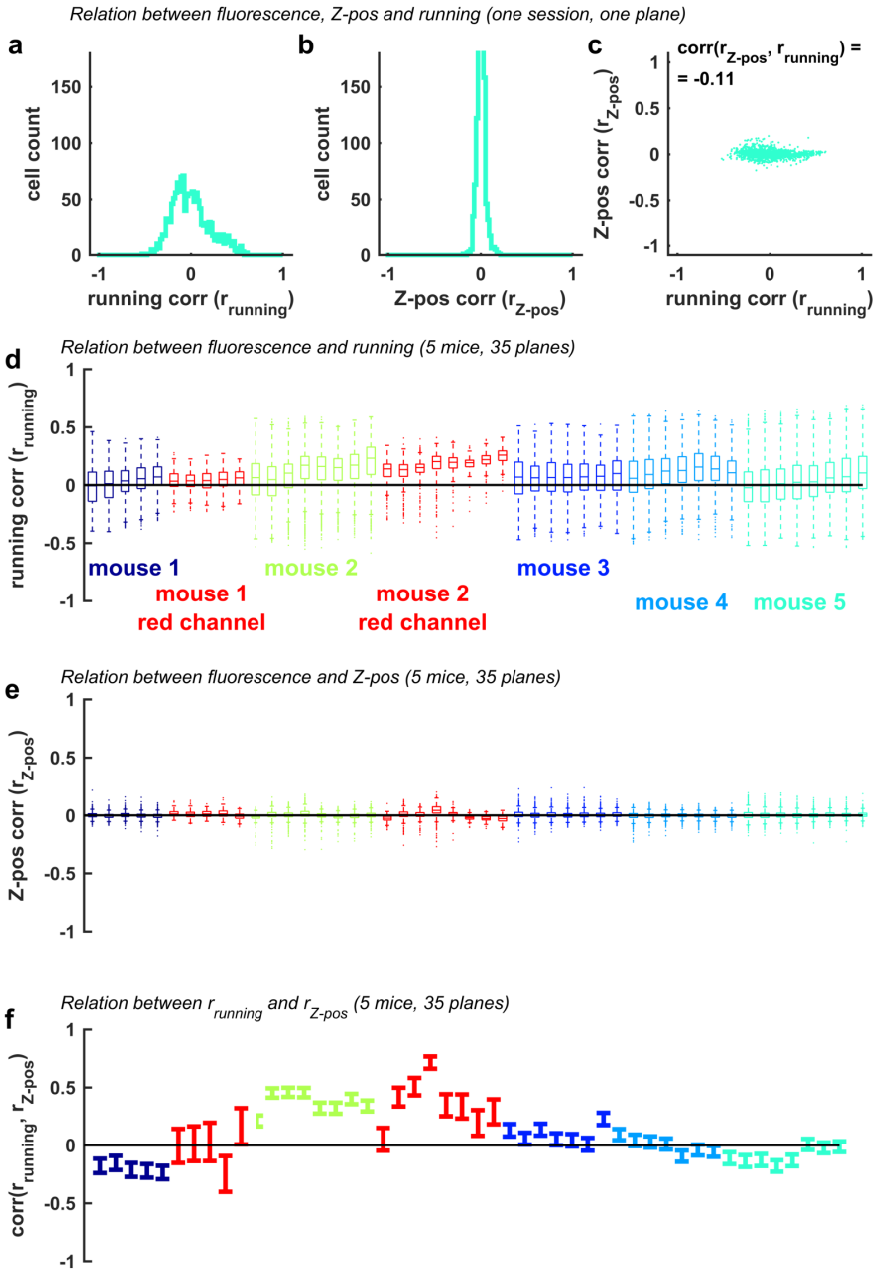

Figure B.15: The effect of Z-baseline correction. Same ordering as figure B.12. **(a-c)** For one recording session, in one plane, the distribution of Z-corrected fluorescence correlation with running (a), with Z-position (b), and the relation between these two correlations (c). Note the much narrow distributions in both cases compared to Figure B.15ab, and the lack of a relationship in (c). **(d,e)** Summarized distributions of Z-corrected fluorescence correlation with running (r_{running}, d) and Z-position ($r_{\text{Z-pos}}$, e) across datasets. Boxes represent standard deviations, and the approximate range of values is indicated by the whiskers. Compare to Figurez B.12de. **(f)** All correlation coefficients between r_{running} and $r_{\text{Z-pos}}$. There was (almost) no relation left after correction.

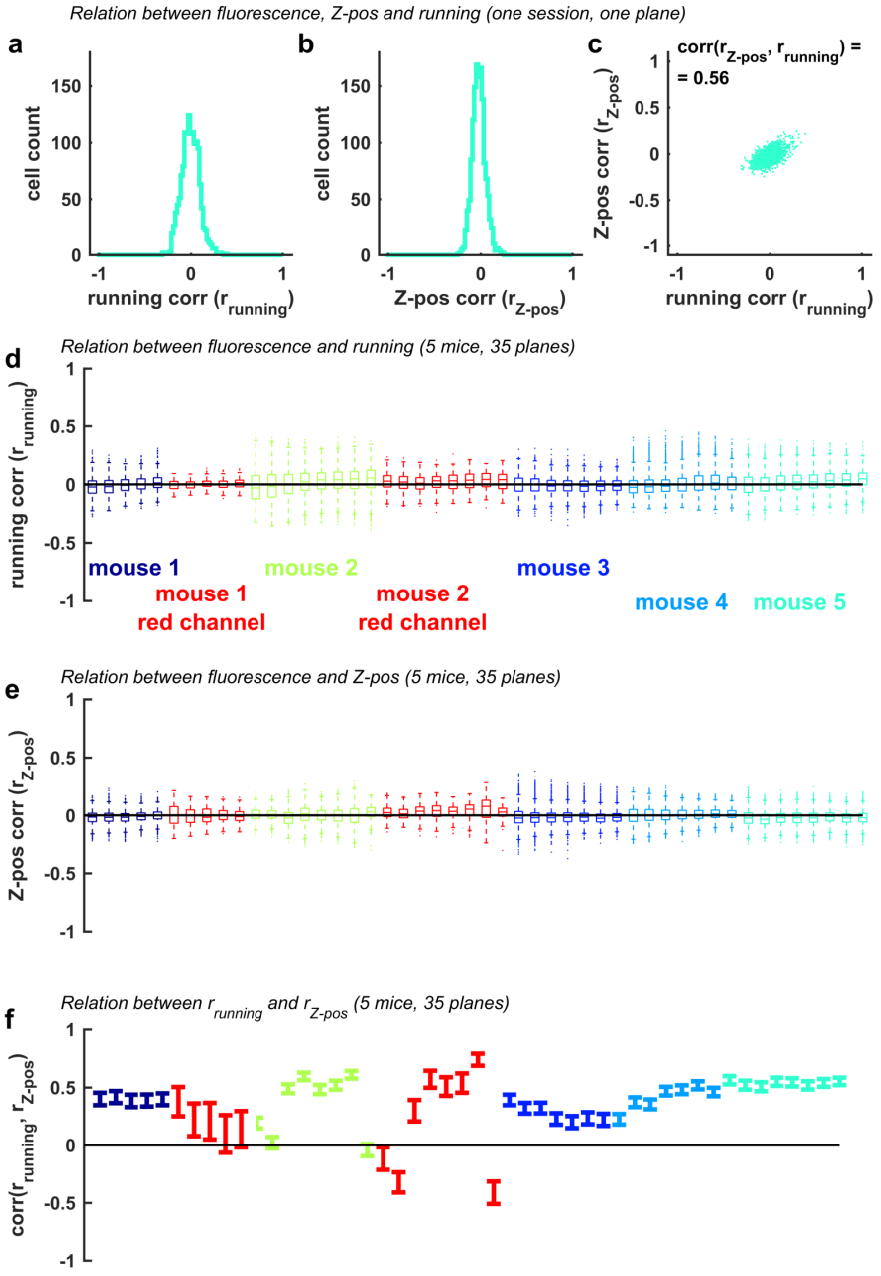

Figure B.16: The effect of temporal running baseline correction. Same ordering as figure B.12. **(a-c)** For one recording session, in one plane, the distribution of Z-corrected fluorescence correlation with running (a), with Z-position (b), and the relation between these two correlations (c). Note the much narrow distributions in both cases compared to Figure B.15ab, and the lack of a relationship in (c). **(d,e)** Summarized distributions of Z-corrected fluorescence correlation with running (r_{running}, d) and Z-position ($r_{\text{Z-pos}}$, e) across datasets. Boxes represent standard deviations, and the approximate range of values is indicated by the whiskers. Compare to Figurez B.12de. **(f)** All correlation coefficients between r_{running} and $r_{\text{Z-pos}}$. There was (almost) no relation left after correction.

the Z-drift changes.

In contrast, the temporal baselining was less effective in removing the artifact (Figure B.16). Not only did it not remove as well the dependence of the signals with Z-drift (Figure B.16be), but it removed much more of the signals' dependence with running (Figure B.16ad). It might be that this correction was warranted, because it is removing z-drift related-signals. However, the Z-baselining method is able to better remove dependence on Z-drift, while affecting dependence with running much less. In addition, for the temporal baselining method, the relation between running and z-position correlation was only partially reduced (Figure B.16cf), again suggesting imperfect removal of drift-related signals. We conclude that the running baseline correction overly subtracts running-related information from the traces, and does not subtract all drift-related information. This might be due to the aggressive tracking of the fluorescence baseline (Figure B.13b), which is not warranted during extended periods of running, when the cells' fluorescence may be elevated due to continuous spiking.

Discussion

We conclude that Z-drift is a real concern during calcium imaging, that can create artifactually high correlations between fluorescence and behavioral measures, like running. Available methods to correct these traces focus on removing a temporal, running baseline of the trace, which we find to not be very effective in removing the relation with Z-drift. In addition, the running baseline removal appears to overly correct the relation between fluorescence and running. This caveat has precluded the application of this method as a processing step in studies where the relation of neural activity and running is considered [Fu et al., 2014, Pakan et al., 2016, Dipoppa et al., 2016]. We developed a new Z-baselining procedure that does not have these caveats, and appears to effectively remove the dependence of the signals on Z-position. After the correction, there were still running-related signals, although they were smaller overall. Our method is thus much more effective than the running baseline correction, making it a practical choice for behavioral studies under head-fixation and two-photon calcium imaging.

Appendix C

Dimensionality estimation from noisy data

In order to characterize a physical system, we often start by asking how many degrees of freedom it possesses. To estimate the degrees of freedom, we would typically record the activity of the system under a diverse set of conditions, while monitoring many sub-parts. For example, in order to characterize the brain, we can start by recording many neurons and presenting many stimuli. This produces a large data matrix of responses, which is amenable to dimensionality reduction methods, like factor- and principal components- analysis. However, existing methods to estimate the number of significant principal components are based on cross-validation, and do not offer reliable estimates when the noise is realistically large, due to overfitting. We developed a more robust framework, in which the principal components are obtained from the training set, but the variances associated with these vectors are estimated using both the training and test data. This offers a good estimate of the singular value spectrum, and avoids the problem of overfitting that plagues standard cross-validation based approaches. In addition to allowing estimation of the dimensionality of a system, our method produces robust estimates of the full distribution of the data singular value spectrum, which are also informative for characterizing the system.

Introduction

[1] We would like to estimate the singular value spectrum of a data matrix A. Unfortunately, we can only observe noisy versions of A. Fortunately, we might have multiple independent observations A_k of A, such that $A_k = A + \varepsilon_k$, with $k = 1, 2, ..., K$, with different noises ε_k. Here, we study the case where $K = 2$, but note that extending to $K > 2$ is straightforward. The case of $K = 1$ can also be efficiently dealt with, but this is work in progress and not described in this thesis. The raw data matrices A_k have size N by T, where we generally refer to the N rows as "coordinates" and the T columns as timepoints, or data samples.

In the following, we will informally refer to the matrix A as "the signal", and to ε_k as "the noise", and assume that both are mean-zero throughout (otherwise we subtract out their means). If the noise is the same magnitude as the signal or larger, the situation we study here, then the spectrum of A_k changes significantly compared to that of A. The noise

1. adds significant power to all subspaces of A, whether or not they contain signal power,
2. changes the order of singular values in A_1 compared to A, because different amounts of noise are added to different subspaces,
3. prevents accurate estimation of eigenvectors associated with small eigenvalues, because other noise dimensions become larger

Given these varied effects of the noise, it is not trivial to determine the spectrum of A from the noisy observation(s) A_k. When at least two repeats are available, we can obtain a robust estimate of the spectrum by comparing the two repeats. We define $A_k = A + \varepsilon_k = USV^T + \varepsilon_k$, where ε_k is independent random noise, and USV^T is a singular value decomposition of A. Can we recover the spectrum S of A from the noisy observations A_1? We are not interested in accurately estimating the singular vectors U and V, but only in the distribution of singular values of A. This estimation may be used, for example, to characterize the dimensionality of the underlying (bio)physical system that has generated A. Is most of the variance in A concentrated in a few top dimensions, or is it spread out over all degrees of freedom of the underlying system?

[1] The work described in this chapter was done in collaboration with Marius Pachitariu.

Results

Our method relies on three observations: 1) we can estimate signal-related variance along any dimension, in an unbiased manner; 2) the signal-related variance along principal components of A is an estimate of the corresponding singular value; and 3) the summed signal variance in any N-dimensional basis is maximized by that basis consisting of the top N principal components of the signal matrix A. Together, these observations imply that the signal variance in any N-dimensional basis is lower bounding the summed squared singular values corresponding to the top N principal components. If we can tighten this bound, we will obtain a good estimate of the true underlying spectrum. In practice, we show that as our estimate of the top N dimensions improves (for example by acquiring more coordinates or timepoints), the cumulative signal variance spectrum converges.

Estimating signal-related variance along arbitrary dimensions

We have described a method to compute the single-trial signal variance, for single coordinates, in the Methods section of Chapter 3. The single-trial signal variance is defined as

$$V_n = \mathrm{E}_i \left[A(n,i) - \mathrm{E}_i[A(n,i)] \right)^2],$$

where

$$A(n,i) = A(n,i) + \varepsilon_k(n,i),$$

with A the trial-averaged response of neuron n to stimulus i and ε_k the noise on repeat k. Similarly, we can estimate the signal variance along any dimension u

$$V(u) = \mathrm{E}_i \left[u^T A - \mathrm{E}_i[u^T A] \right)^2],$$

Like in the case of the signal variance, the main difficulty is in estimating $\mathrm{E}_i(u^T A)_i^2$, which proceeds like before. This quantity is thus approximated, in an unbiased

manner, by

$$\mathrm{E}_i(u^T A)_i^2 \sim \sum_i (u^T A_1)_i \cdot (u^T A_2)_i = u^T (A_1^T A_2) u.$$

Signal-related variance along principal components of A

Assuming the data has been zero-centered, the signal variance along any principal component U_n is approximated by $\frac{1}{N}\sum_i (U_n^T A)_i^2$. This is equal to

$$\begin{aligned}
\mathrm{E}_i[u^T A]^2 &\sim \frac{1}{N}\sum_i (U_n^T A)_i^2 \\
&= \mathrm{Tr}(U_n^T A A^T U_n) \\
&= \mathrm{Tr}(U_n^T U S V^T V S U^T U_n) \\
&= \mathrm{Tr}(U_n^T U S^2 U^T U_n) \\
&= \mathrm{Tr}(\mathbf{e}_n^T S^2 \mathbf{e}_n) \\
&= S_n^2,
\end{aligned}$$

where $e_n(n) = 1$ and $e_n(k) = 0, \forall k \neq n$. Thus, the signal variance along the principal component U_n is approximating the squared n-th singular value of U_n.

Lower bounding the spectrum of A

By definition, the top n principal vectors $U_{1:n}$ of A capture the most variance of A possible of any n-dimensional orthonormal basis set, i.e they maximize over X the quantity $\sum_{k=1}^{n} \left\| X_k^T A \right\|^2$. We have just shown that we can obtain a reliable estimate for $\left\| X_k^T A \right\|^2$ from the noisy repeats A_1 and A_2. It follows that for any basis X, the empirical signal-related variance is a lower bound for the true singular value spectrum (squared). Notice that this bound holds only in expectation, and not for any particular instantiation of the noise ε_1 and ε_2. The bound is tight if and only if our estimated basis X consists of the true singular vectors U. In practice we estimate the basis X from one of the two repeats, as the top principal components of that repeat's responses. As we consider more neurons and stimuli, this estimate becomes more accurate.

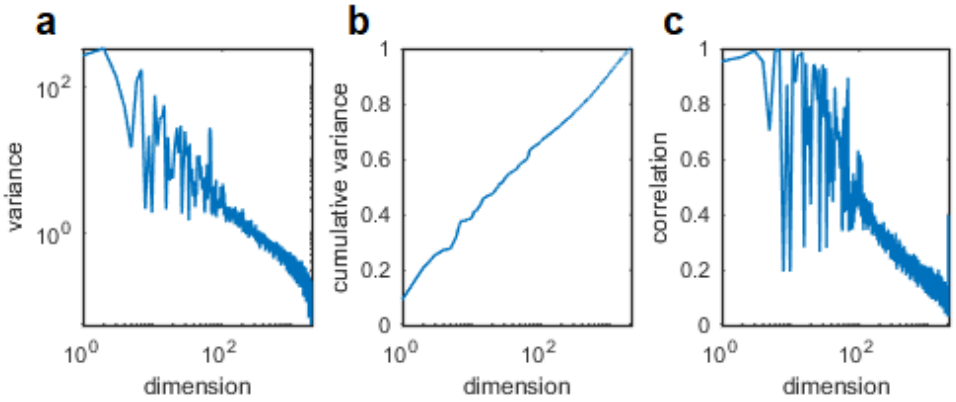

Figure C.1: Recovering the singular values from simulations. Compare to Figure 3.3. **(a)** Inferred signal variance along the top estimated dimensions. **(b)** Cumulative spectrum of the estimated singular values. **(c)** Correlation of responses to the first and second repeat, measured within each estimated dimension.

Simulations

To test the method, we performed simulations where A had a $1/n$ power-law distribution of singular values, and the noise scaled with the size of the signal along each dimension. We matched to the real neural data the average signal variance ($\sim 10\%$), the number of neurons and the number of stimuli. We found that the estimated spectrum of A indeed represented a noisy version of the true spectrum of A.

Discussion

We have demonstrated a new method for estimating the singular value spectrum of a matrix from two noisy observations of that matrix. It is beyond the scope of this thesis, but a topic of future research, to estimate the limits of this method, and compare it with more standard alternative, like standard cross-validation of model-based estimates.

Bibliography

Abbott, L. F. and Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Computation*, 11(1):91–101. (Cited on page 23.)

Afshar, A., Santhanam, G., Byron, M. Y., Ryu, S. I., Sahani, M., and Shenoy, K. V. (2011). Single-trial neural correlates of arm movement preparation. *Neuron*, 71(3):555–564. (Cited on page 124.)

Alba, A., Vigueras-Gomez, J. F., Arce-Santana, E. R., and Aguilar-Ponce, R. M. (2015). Phase correlation with sub-pixel accuracy: a comparative study in 1d and 2d. *Computer Vision and Image Understanding*, 137:76–87. (Cited on page 158.)

Amit, D. J. and Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex*, 7(3):237–252. (Cited on pages 19 and 25.)

Atallah, B. V., Bruns, W., Carandini, M., and Scanziani, M. (2012). Parvalbumin-expressing interneurons linearly transform cortical responses to visual stimuli. *Neuron*, 73(1):159–170. (Cited on page 141.)

Averbeck, B. B., Latham, P. E., and Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, 7(5):358–366. (Cited on page 23.)

Barthó, P., Hirase, H., Monconduit, L., Zugaro, M., Harris, K. D., and Buzsáki, G. (2004). Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. *Journal of Neurophysiology*, 92(1):600–608. (Cited on pages 45 and 61.)

Bathellier, B., Ushakova, L., and Rumpel, S. (2012). Discrete neocortical dynamics predict behavioral categorization of sounds. *Neuron*, 76(2):435–449. (Cited on page 55.)

Bennett, C., Arroyo, S., and Hestrin, S. (2013). Subthreshold mechanisms underlying state-dependent modulation of visual responses. *Neuron*, 80(2):350–357. (Cited on pages 14, 50, and 56.)

Berkes, P., Orbán, G., Lengyel, M., and Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, 331(6013):83–87. (Cited on pages 12, 18, and 109.)

Berridge, C. W. and Waterhouse, B. D. (2003). The locus coeruleus-noradrenergic system: Modulation of behavioral state and state-dependent cognitive processes. *Brain Research Reviews*, 42(1):33–84. (Cited on page 17.)

Bortone, D. S., Olsen, S. R., and Scanziani, M. (2014). Translaminar inhibitory cells recruited by layer 6 corticothalamic neurons suppress visual cortex. *Neuron*, 82(2):474 – 485. (Cited on page 143.)

Buran, B. N., von Trapp, G., and Sanes, D. H. (2014). Behaviorally gated reduction of spontaneous discharge can improve detection thresholds in auditory cortex. *The Journal of Neuroscience*, 34(11):4076–4081. (Cited on pages 15, 23, and 56.)

Cardin, J. A., Carlén, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., Tsai, L.-H., and Moore, C. I. (2009). Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature*, 459(7247):663–667. (Cited on pages 45 and 61.)

Castro-Alamancos, M. A. and Gulati, T. (2014). Neuromodulators produce distinct activated states in neocortex. *The Journal of Neuroscience*, 34(37):12353–12367. (Cited on pages 17, 18, and 144.)

Chen, J. L., Pfäffli, O. A., Voigt, F. F., Margolis, D. J., and Helmchen, F. (2013). Online correction of licking-induced brain motion during two-photon imaging with a tunable lens. *The Journal of physiology*, 591(19):4689–4698. (Cited on page 162.)

Chen, N., Sugihara, H., and Sur, M. (2015). An acetylcholine-activated microcircuit drives temporal dynamics of cortical activity. *Nature Neuroscience*, 18(6):892–902. (Cited on pages 16, 17, 18, and 144.)

Cho, K.-H., Jang, J. H., Jang, H.-J., Kim, M.-J., Yoon, S. H., Fukuda, T., Tennigkeit, F., Singer, W., and Rhie, D.-J. (2010). Subtype-specific dendritic ca2+ dynamics of inhibitory interneurons in the rat visual cortex. *Journal of Neurophysiology*, 104(2):840–853. (Cited on pages 45 and 61.)

Churchland, M. M. and Abbott, L. F. (2012). Two layers of neural variability. *Nature Neuroscience*, 15(11):1472–1474. (Cited on page 25.)

Churchland, M. M., Afshar, A., and Shenoy, K. V. (2006). A central source of movement variability. *Neuron*, 52(6):1085–1096. (Cited on page 124.)

Churchland, M. M., Yu, B. M., Cunningham, J. P., Sugrue, L. P., Cohen, M. R., Corrado, G. S., Newsome, W. T., Clark, A. M., Hosseini, P., Scott, B. B., Bradley, D. C., Smith, M. a., Kohn, A., Movshon, J. A., Armstrong, K. M., Moore, T., Chang, S. W., Snyder, L. H., Lisberger, S. G., Priebe, N. J., Finn, I. M., Ferster, D., Ryu, S. I., Santhanam, G., Sahani, M., and Shenoy, K. V. (2010). Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nature Neuroscience*, 13(3):369–378. (Cited on page 57.)

Cohen, L. and Mizrahi, A. (2015). Plasticity during motherhood: Changes in excitatory and inhibitory layer 2/3 neurons in auditory cortex. *Journal of Neuroscience*, 35(4):1806–1815. (Cited on pages 45 and 61.)

Cohen, M. R. and Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience*, 14(7):811–819. (Cited on page 23.)

Cohen, M. R. and Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience*, 12(12):1594–1600. (Cited on pages 15, 23, 96, and 107.)

Cohen-Kashi Malina, K., Mohar, B., Rappaport, A. N., and Lampl, I. (2016). Local and thalamic origins of ongoing and sensory evoked cortical correlations. *bioRxiv*. (Cited on page 19.)

Collura, T. F. (1993). History and evolution of electroencephalographic instruments and techniques. *Journal of clinical neurophysiology*, 10(4):476–504. (Cited on page 11.)

Compte, A., Sanchez-Vives, M. V., McCormick, D. A., and Wang, X.-J. (2003). Cellular and network mechanisms of slow oscillatory activity (¡ 1 hz) and wave propagations

in a cortical network model. *Journal of Neurophysiology*, 89(5):2707–2725. (Cited on page 53.)

Constantinople, C. M. and Bruno, R. M. (2011). Effects and mechanisms of wakefulness on local cortical networks. *Neuron*, 69(6):1061–1068. (Cited on pages 12, 18, 23, and 55.)

Cossell, L., Iacaruso, M. F., Muir, D. R., Houlton, R., Sader, E. N., Ko, H., Hofer, S. B., and Mrsic-Flogel, T. D. (2015). Functional organization of excitatory synaptic strength in primary visual cortex. *Nature*. (Cited on page 153.)

Crochet, S., Poulet, J. F. a., Kremer, Y., and Petersen, C. C. H. (2011). Synaptic mechanisms underlying sparse coding of active touch. *Neuron*, 69(6):1160–1175. (Cited on page 56.)

Curto, C., Sakata, S., Marguet, S., Itskov, V., and Harris, K. D. (2009). A simple model of cortical dynamics explains variability and state dependence of sensory responses in urethane-anesthetized auditory cortex. *The Journal of Neuroscience*, 29(34):10600–10612. (Cited on pages 18, 24, 54, and 56.)

Dadarlat, M. C. and Stryker, M. P. (2017). Locomotion enhances neural encoding of visual stimuli in mouse v1. *Journal of Neuroscience*, 37(14):3764–3775. (Cited on page 14.)

de Kock, C. P. J. and Sakmann, B. (2009). Spiking in primary somatosensory cortex during natural whisking in awake head-restrained rats is cell-type specific. *Proceedings of the National Academy of Sciences*, 106(38):16446–16450. (Cited on page 15.)

de la Rocha, J., Doiron, B., Shea-Brown, E., Josić, K., and Reyes, A. (2007). Correlation between neural spike trains increases with firing rate. *Nature*, 448(7155):802–806. (Cited on pages 23, 24, and 53.)

Deschênes, M., Takatoh, J., Kurnikova, A., Moore, J. D., Demers, M., Elbaz, M., Furuta, T., Wang, F., and Kleinfeld, D. (2016). Inhibition, not excitation, drives rhythmic whisking. *Neuron*, 90(2):374–387. (Cited on page 126.)

Destexhe, A. (2009). Self-sustained asynchronous irregular states and Up-Down states in thalamic, cortical and thalamocortical networks of nonlinear integrate-and-

fire neurons. *Journal of Computational Neuroscience*, 27(3):493–506. (Cited on pages 29, 54, and 56.)

Dipoppa, M., Ranson, A., Krumin, M., Pachitariu, M., Carandini, M., and Harris, K. D. (2016). Vision and locomotion shape the interactions between neuron types in mouse visual cortex. *bioRxiv*, page 058396. (Cited on pages 107, 115, 116, 152, 171, 177, and 187.)

Doiron, B., Litwin-Kumar, A., Rosenbaum, R., Ocker, G. K., and Josić, K. (2016). The mechanics of state-dependent neural correlations. *Nature Neuroscience*, 19(3):383–393. (Cited on pages 23, 24, 25, 26, and 53.)

Donnert, G., Eggeling, C., and Hell, S. W. (2007). Major signal increase in fluorescence microscopy through dark-state relaxation. *Nature methods*, 4(1):81. (Cited on page 182.)

Dougherty, E. R. and Lotufo, R. A. (2003). *Hands-on morphological image processing*, volume 59. SPIE press. (Cited on page 183.)

Ecker, A. S., Berens, P., Cotton, R. J., Subramaniyan, M., Denfield, G. H., Cadwell, C. R., Smirnakis, S. M., Bethge, M., and Tolias, A. S. (2014). State dependence of noise correlations in macaque primary visual cortex. *Neuron*, 82(1):235–248. (Cited on pages 24 and 53.)

Ecker, A. S., Berens, P., Keliris, G. a., Bethge, M., Logothetis, N. K., and Tolias, A. S. (2010). Decorrelated neuronal firing in cortical microcircuits. *Science*, 327(5965):584–587. (Cited on pages 23, 55, 96, and 107.)

Ecker, A. S., Berens, P., Tolias, A. S., and Bethge, M. (2011). The effect of noise correlations in populations of diversely tuned neurons. *The Journal of Neuroscience*, 31(40):14272–14283. (Cited on page 23.)

Engel, T. A., Steinmetz, N. A., Gieselmann, M. A., Thiele, A., Moore, T., and Boahen, K. (2016). Selective modulation of cortical state during spatial attention. *Science*, 354(6316):1140–1144. (Cited on pages 13 and 15.)

Erisken, S., Vaiceliunaite, A., Jurjut, O., Fiorini, M., Katzner, S., Busse, L., Reichardt, W., and Neuroscience, I. (2014). Article Effects of Locomotion Extend throughout the Mouse Early Visual System. *Current Biology*, 24(24):1–9. (Cited on pages 14 and 23.)

Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Josa a*, 4(12):2379–2394. (Cited on page 80.)

Fino, E. and Yuste, R. (2011). Dense inhibitory connectivity in neocortex. *Neuron*, 69(6):1188–1203. (Cited on pages 29, 133, and 143.)

Foroosh, H., Zerubia, J. B., and Berthod, M. (2002). Extension of phase correlation to subpixel registration. *Image Processing, IEEE Transactions on*, 11(3):188–200. (Cited on page 158.)

Fu, Y., Tucciarone, J. M., Espinosa, J. S., Sheng, N., Darcy, D. P., Nicoll, R. A., Huang, Z. J., and Stryker, M. P. (2014). A cortical circuit for gain control by behavioral state. *Cell*, 156(6):1139–1152. (Cited on pages 171 and 187.)

Gao, P. and Ganguli, S. (2015). On simplicity and complexity in the brave new world of large-scale neuroscience. *Current opinion in neurobiology*, 32:148–155. (Cited on pages 74, 157, and 162.)

Gao, P., Trautmann, E., Yu, B., Santhanam, G., Ryu, S., Shenoy, K., and Ganguli, S. (2014). A theory of neural dimensionality and measurement. *Cosyne abstract*. (Cited on page 74.)

Garcia-Lazaro, J. A., Belliveau, L. A., and Lesica, N. A. (2013). Independent population coding of speech with sub-millisecond precision. *The Journal of Neuroscience*, 33(49):19362–19372. (Cited on pages 42, 60, and 69.)

Gentet, L. J., Avermann, M., Matyas, F., Staiger, J. F., and Petersen, C. C. (2010). Membrane potential dynamics of gabaergic neurons in the barrel cortex of behaving mice. *Neuron*, 65(3):422–435. (Cited on pages 15 and 116.)

Gentet, L. J., Kremer, Y., Taniguchi, H., Huang, Z. J., Staiger, J. F., and Petersen, C. C. (2012). Unique functional properties of somatostatin-expressing gabaergic neurons in mouse barrel cortex. *Nature Neuroscience*, 15(4):607–612. (Cited on pages 15 and 116.)

Goard, M. and Dan, Y. (2009). Basal forebrain activation enhances cortical coding of natural scenes. *Nature Neuroscience*, 12(11):1444–1449. (Cited on pages 17, 18, and 144.)

Goris, R. L. T., Movshon, J. A., and Simoncelli, E. P. (2014). Partitioning neuronal variability. *Nature Neuroscience*, 17(6):858–65. (Cited on page 24.)

Greenberg, D. S. and Kerr, J. N. (2009). Automated correction of fast motion artifacts for two-photon imaging of awake animals. *Journal of neuroscience methods*, 176(1):1–15. (Cited on pages 157 and 158.)

Gutierrez, G. J., O'Leary, T., and Marder, E. (2013). Multiple mechanisms switch an electrically coupled, synaptically inhibited neuron between competing rhythmic oscillators. *Neuron*, 77(5):845–858. (Cited on page 36.)

Haider, B., Duque, A., Hasenstaub, A. R., and McCormick, D. A. (2006). Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *Journal of Neuroscience*, 26(17):4535–4545. (Cited on page 12.)

Haider, B., Häusser, M., and Carandini, M. (2013). Inhibition dominates sensory responses in the awake cortex. *Nature*, 493(7430):97–100. (Cited on pages 15, 16, 29, and 144.)

Hansen, B. J., Chelaru, M. I., and Dragoi, V. (2012). Correlated variability in laminar cortical circuits. *Neuron*, 76(3):590–602. (Cited on page 55.)

Harris, K. D. and Thiele, A. (2011). Cortical state and attention. *Nature Reviews Neuroscience*, 12(9):509–523. (Cited on pages 23 and 28.)

Hattox, A. M. and Nelson, S. B. (2007). Layer v neurons in mouse cortex projecting to different targets have distinct physiological properties. *Journal of Neurophysiology*, 98(6):3330–3340. (Cited on page 11.)

Hellwig, B. (2000). A quantitative analysis of the local connectivity between pyramidal neurons in layers 2/3 of the rat visual cortex. *Biological cybernetics*, 82(2):111–121. (Cited on pages 103 and 109.)

Higgins, I., Pal, A., Rusu, A. A., Matthey, L., Burgess, C. P., Pritzel, A., Botvinick, M., Blundell, C., and Lerchner, A. (2017). Darla: Improving zero-shot transfer in reinforcement learning. *arXiv preprint arXiv:1707.08475*. (Cited on page 155.)

Hofer, S. B., Ko, H., Pichler, B., Vogelstein, J., Ros, H., Zeng, H., Lein, E., Lesica, N. A., and Mrsic-Flogel, T. D. (2011). Differential connectivity and response dynamics of

excitatory and inhibitory neurons in visual cortex. *Nature Neuroscience*, 14(8):1045–1052. (Cited on page 29.)

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106–154. (Cited on page 19.)

Ibrahim, L. A., Mesik, L., ying Ji, X., Fang, Q., fu Li, H., tang Li, Y., Zingg, B., Zhang, L. I., and Tao, H. W. (2016). Cross-modality sharpening of visual cortical processing through layer-1-mediated inhibition and disinhibition. *Neuron*, 89(5):1031 – 1045. (Cited on page 143.)

Isaacson, J. S. and Scanziani, M. (2011). How inhibition shapes cortical activity. *Neuron*, 72(2):231–243. (Cited on page 55.)

Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of multivariate analysis*, 5(2):248–264. (Cited on pages 119 and 131.)

Izhikevich, E. M. and Edelman, G. M. (2008). Large-scale model of mammalian thalamocortical systems. *Proceedings of the national academy of sciences*, 105(9):3593–3598. (Cited on page 152.)

Ji, X.-y., Zingg, B., Mesik, L., Xiao, Z., Zhang, L. I., and Tao, H. W. (2015). Thalamocortical innervation pattern in mouse auditory and visual cortex: laminar and cell-type specificity. *Cerebral Cortex*, 26(6):2612–2625. (Cited on page 143.)

Jones, B. E. (2008). Modulation of cortical activation and behavioral arousal by cholinergic and orexinergic systems. *Annals of the New York Academy of Sciences*, 1129:26–34. (Cited on page 17.)

Kanashiro, T., Ocker, G. K., Cohen, M. R., and Doiron, B. (2017). Attentional modulation of neuronal variability in circuit models of cortex. *eLife*, 6. (Cited on page 154.)

Kato, H. K., Gillet, S. N., Peters, A. J., Isaacson, J. S., and Komiyama, T. (2013). Parvalbumin-expressing interneurons linearly control olfactory bulb output. *Neuron*, 80(5):1218–1231. (Cited on page 16.)

Kawaguchi, Y. and Kubota, Y. (1997). Gabaergic cell subtypes and their synaptic connections in rat frontal cortex. *Cerebral Cortex*, 7(6):476–486. (Cited on pages 45 and 61.)

Kawai, R., Markman, T., Poddar, R., Ko, R., Fantana, A. L., Dhawale, A. K., Kampff, A. R., and Ölveczky, B. P. (2015). Motor cortex is required for learning but not for executing a motor skill. *Neuron*, 86:800–812. Kawai, RisaMarkman, TimothyPoddar, RajeshKo, RaymondFantana, Antoniu LDhawale, Ashesh KKampff, Adam ROlveczky, Bence PENG2015/04/22 06:00Neuron. 2015 May 6;86(3):800-812. doi: 10.1016/j.neuron.2015.03.024. Epub 2015 Apr 16. (Cited on page 124.)

Kerlin, A. M., Andermann, M. L., Berezovskii, V. K., and Reid, R. C. (2010). Broadly tuned response properties of diverse inhibitory neuron subtypes in mouse visual cortex. *Neuron*, 67(5):858–871. (Cited on pages 133, 141, and 143.)

Kim, J., Matney, C. J., Blankenship, A., Hestrin, S., and Brown, S. P. (2014). Layer 6 corticothalamic neurons activate a cortical output layer, layer 5a. *Journal of Neuroscience*, 34(29):9656–9664. (Cited on page 143.)

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105. (Cited on page 86.)

Kuchibhotla, K. V., Gill, J. V., Lindsay, G. W., Papadoyannis, E. S., Field, R. E., Sten, T., Miller, K. D., and Froemke, R. C. (2017). Parallel processing by cortical inhibition enables context-dependent behavior. *Nature neuroscience*, 20(1):62–71. (Cited on page 16.)

Kurnikova, A., Moore, J. D., Liao, S.-M., Deschłnes, M., and Kleinfeld, D. (2017). Coordination of orofacial motor actions into exploratory behavior by rat. *Current Biology*, 27(5):688 – 696. (Cited on pages 14, 109, 116, 123, and 126.)

Latham, P. E., Richmond, B. J., Nelson, P. G., and Nirenberg, S. (2000). Intrinsic dynamics in neuronal networks. I. Theory. *Journal of Neurophysiology*, 83(2):808–827. (Cited on pages 29, 54, and 56.)

Lee, A. B., Mumford, D., and Huang, J. (2001). Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. *International Journal of Computer Vision*, 41(1):35–59. (Cited on page 80.)

Lee, D. D. and Seung, H. S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788. (Cited on page 103.)

Lee, S.-H., Kwan, A. C., Zhang, S., Phoumthipphavong, V., Flannery, J. G., Masmanidis, S. C., Taniguchi, H., Huang, Z. J., Zhang, F., Boyden, E. S., et al. (2012). Activation of specific interneurons improves v1 feature selectivity and visual perception. *Nature*, 488(7411):379–383. (Cited on pages 16, 18, and 43.)

Lee, W.-C. A., Bonin, V., Reed, M., Graham, B. J., Hood, G., Glattfelder, K., and Reid, R. C. (2016). Anatomy and function of an excitatory network in the visual cortex. *Nature*, 532(7599):370–374. (Not cited.)

Levy, R. B. and Reyes, A. D. (2012). Spatial profile of excitatory and inhibitory synaptic connectivity in mouse primary auditory cortex. *Journal of Neuroscience*, 32(16):5609–5619. (Cited on pages 103 and 109.)

Lewis, L. D., Voigts, J., Flores, F. J., Schmitt, L. I., Wilson, M. A., Halassa, M. M., and Brown, E. N. (2015). Thalamic reticular nucleus induces fast and local modulation of arousal state. *Elife*, 4:e08760. (Cited on page 153.)

Litwin-Kumar, A. and Doiron, B. (2012). Slow dynamics and high variability in balanced cortical networks with clustered connections. *Nature Neuroscience*, 15(11):1498–1505. (Cited on pages 19, 20, 25, 52, and 134.)

Liu, B.-h., Li, P., Li, Y.-t., Sun, Y. J., Yanagawa, Y., Obata, K., Zhang, L. I., and Tao, H. W. (2009). Visual receptive field structure of cortical inhibitory neurons revealed by two-photon imaging guided recording. *Journal of Neuroscience*, 29(34):10520–10532. (Cited on page 141.)

Loebel, A., Nelken, I., and Tsodyks, M. (2007). Processing of sounds by population spikes in a model of primary auditory cortex. *Frontiers in neuroscience*, 1:15. (Cited on page 54.)

London, M., Roth, A., Beeren, L., Häusser, M., and Latham, P. E. (2010). Sensitivity to perturbations in vivo implies high noise and suggests rate coding in cortex. *Nature*, 466(7302):123–127. (Cited on page 32.)

Lopes, G., Nogueira, J., Dimitriadis, G., Menendez, J. A., Paton, J. J., and Kampff, A. R. (2017). A robust role for motor cortex. *bioRxiv*, page 058917. (Cited on page 124.)

Luczak, A., Barthó, P., and Harris, K. D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron*, 62(3):413–425. (Cited on pages 12, 18, and 59.)

Lyamzin, D. R., Barnes, S. J., Donato, R., Garcia-Lazaro, J. A., Keck, T., and Lesica, N. A. (2015). Nonlinear transfer of signal and noise correlations in cortical networks. *The Journal of Neuroscience*, 35(21):8065–8080. (Cited on page 23.)

Ma, W.-p., Liu, B.-h., Li, Y.-t., Josh Huang, Z., Zhang, L. I., and Tao, H. W. (2010). Visual representations by cortical somatostatin inhibitory neurons—selective but with weak and delayed responses. *Journal of Neuroscience*, 30(43):14371–14379. (Cited on page 139.)

Macke, J. H., Buesing, L., Cunningham, J. P., Byron, M. Y., Shenoy, K. V., and Sahani, M. (2011). Empirical models of spiking in neural populations. *Advances in Neural Information Processing Systems*, pages 1350–1358. (Cited on pages 24 and 26.)

Madisen, L., Mao, T., Koch, H., Zhuo, J.-m., Berenyi, A., Fujisawa, S., Hsu, Y.-W. A., Garcia III, A. J., Gu, X., Zanella, S., et al. (2012). A toolbox of cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nature Neuroscience*, 15(5):793–802. (Cited on pages 45 and 61.)

Malina, K. C.-K., Mohar, B., Rappaport, A. N., and Lampl, I. (2016). Local and thalamic origins of correlated ongoing and sensory-evoked cortical activities. *Nature communications*, 7. (Cited on pages 13, 23, 24, 52, 143, and 153.)

Mao, B.-Q., Hamzei-Sichani, F., Aronov, D., Froemke, R. C., and Yuste, R. (2001). Dynamics of spontaneous activity in neocortical slices. *Neuron*, 32(5):883 – 898. (Cited on page 182.)

Marder, E., Goeritz, M. L., and Otopalik, A. G. (2015). Robust circuit rhythms in small circuits arise from variable circuit components and mechanisms. *Current Opinion in Neurobiology*, 31:156–163. (Cited on page 36.)

Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., et al. (2015). Reconstruction and simulation of neocortical microcircuitry. *Cell*, 163(2):456–492. (Cited on page 152.)

Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., and Wu, C. (2004). Interneurons of the neocortical inhibitory system. *Nature Reviews Neuroscience*, 5(10):793–807. (Cited on page 47.)

Martinez, L. M. and Alonso, J.-M. (2001). Construction of complex receptive fields in cat primary visual cortex. *Neuron*, 32(3):515–525. (Cited on page 19.)

Mathis, M. W., Mathis, A., and Uchida, N. (2017). Somatosensory cortex plays an essential role in forelimb motor adaptation in mice. *Neuron*, 93(6):1493–1503. (Cited on page 125.)

McElvain, L. E., Friedman, B., Karten, H. J., Svoboda, K., Wang, F., Deschłnes, M., and Kleinfeld, D. (2017). Circuits in the rodent brainstem that control whisking in concert with other orofacial motor actions. *Neuroscience*, pages –. (Cited on page 126.)

McGinley, M. J., David, S. V., and McCormick, D. A. (2015a). Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. *Neuron*, 87(1):179–192. (Cited on pages 13, 14, 15, 18, 23, 50, 55, 56, 115, 116, and 177.)

McGinley, M. J., Vinck, M., Reimer, J., Batista-Brito, R., Zagha, E., Cadwell, C. R., Tolias, A. S., Cardin, J. A., and McCormick, D. A. (2015b). Waking state: rapid variations modulate neural and behavioral responses. *Neuron*, 87(6):1143–1161. (Cited on pages 56, 97, and 115.)

Mitchell, J. F., Sundberg, K. a., and Reynolds, J. H. (2009). Spatial Attention Decorrelates Intrinsic Activity Fluctuations in Macaque Area V4. *Neuron*, 63(6):879–888. (Cited on pages 15 and 23.)

Mochol, G., Hermoso-Mendizabal, A., Sakata, S., Harris, K. D., and de la Rocha, J. (2015). Stochastic transitions into silence cause noise correlations in cortical circuits.

Proceedings of the National Academy of Sciences, 112(11):201410509. (Cited on pages 20, 54, 56, and 154.)

Monier, C., Chavane, F., Baudot, P., Graham, L. J., and Frgnac, Y. (2003). Orientation and direction selectivity of synaptic inputs in visual cortical neurons: A diversity of combinations produces spike tuning. *Neuron*, 37(4):663 – 680. (Cited on page 144.)

Moore, A. K. and Wehr, M. (2013). Parvalbumin-expressing inhibitory interneurons in auditory cortex are well-tuned for frequency. *The Journal of Neuroscience*, 33(34):13713–13723. (Cited on pages 61 and 62.)

Moreno-Bote, R., Beck, J., Kanitscheider, I., Pitkow, X., Latham, P., and Pouget, A. (2014). Information-limiting correlations. *Nature Neuroscience*, 17(10):1410–1417. (Cited on page 23.)

Muñoz, W., Tremblay, R., Levenstein, D., and Rudy, B. (2017). Layer-specific modulation of neocortical dendritic inhibition during active wakefulness. *Science*, 355(6328):954–959. (Cited on page 152.)

Niell, C. M. and Stryker, M. P. (2010). Modulation of Visual Responses by Behavioral State in Mouse Visual Cortex. *Neuron*, 65(4):472–479. (Cited on pages 14, 16, 50, 56, 81, 115, 116, and 177.)

Nowak, L. G., Azouz, R., Sanchez-Vives, M. V., Gray, C. M., and McCormick, D. A. (2003). Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. *Journal of Neurophysiology*, 89(3):1541–1566. (Cited on pages 11, 45, 53, and 61.)

Okun, M., Steinmetz, N. A., Cossell, L., Iacaruso, M. F., Ko, H., Barthó, P., Moore, T., Hofer, S. B., Mrsic-Flogel, T. D., Carandini, M., et al. (2015). Diverse coupling of neurons to populations in sensory cortex. *Nature*, 521(7553):511–515. (Cited on pages 12, 23, 53, 58, 59, and 61.)

Otazu, G. H., Tai, L.-H., Yang, Y., and Zador, A. M. (2009). Engaging in an auditory task suppresses responses in auditory cortex. *Nature Neuroscience*, 12(5):646–654. (Cited on pages 15 and 56.)

Pachitariu, M., Lyamzin, D. R., Sahani, M., and Lesica, N. A. (2015). State-dependent population coding in primary auditory cortex. *The Journal of Neuroscience*, 35(5):2058–2073. (Cited on pages 12, 13, 18, 23, 47, 55, 60, and 70.)

Pachitariu, M., Petreska, B., and Sahani, M. (2013). Recurrent linear models of simultaneously-recorded neural populations. *Advances in Neural Information Processing Systems*, pages 3138–3146. (Cited on pages 24 and 26.)

Pachitariu, M. and Sahani, M. (2012). Learning visual motion in recurrent neural networks. In *Advances in Neural Information Processing Systems*, pages 1322–1330. (Cited on page 155.)

Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., and Harris, K. D. (2016a). Kilosort: realtime spike-sorting for extracellular electrophysiology with hundreds of channels. *bioRxiv*, page 061481. (Cited on pages 61, 157, and 163.)

Pachitariu, M., Stringer, C., Schröder, S., Dipoppa, M., Rossi, L. F., Carandini, M., and Harris, K. D. (2016b). Suite2p: beyond 10,000 neurons with standard two-photon microscopy. *bioRxiv*, page 061507. (Cited on pages 20, 90, 91, 111, and 160.)

Packer, A. M. and Yuste, R. (2011). Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: a canonical microcircuit for inhibition? *The Journal of Neuroscience*, 31(37):13260–13271. (Cited on pages 29, 133, and 143.)

Pakan, J. M., Lowe, S. C., Dylda, E., Keemink, S. W., Currie, S. P., Coutts, C. A., and Rochefort, N. L. (2016). Behavioral-state modulation of inhibition is context-dependent and cell type specific in mouse visual cortex. *eLife*, 5:e14985. (Cited on pages 16, 107, 115, 116, 152, 171, 177, and 187.)

Pandarinath, C., O'Shea, D. J., Collins, J., Jozefowicz, R., Stavisky, S. D., Kao, J. C., Trautmann, E. M., Kaufman, M. T., Ryu, S. I., Hochberg, L. R., Henderson, J. M., Shenoy, K. V., Abbott, L. F., and Sussillo, D. (2017). Inferring single-trial neural population dynamics using sequential auto-encoders. *bioRxiv*. (Cited on page 155.)

Parga, N. and Abbott, L. F. (2007). Network model of spontaneous activity exhibiting synchronous transitions between up and down states. *Frontiers in neuroscience*, 1:4. (Cited on page 25.)

Peron, S. P., Freeman, J., Iyer, V., Guo, C., and Svoboda, K. (2015). A cellular resolution map of barrel cortex activity during tactile behavior. *Neuron*, 86(3):783 – 799. (Cited on pages 15 and 116.)

Petersen, K. B. and Pedersen, M. S. (2008). The matrix cookbook. (Cited on pages 146 and 147.)

Pfeffer, C. K., Xue, M., He, M., Huang, Z. J., and Scanziani, M. (2013). Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nature Neuroscience*, 16(8):1068–1076. (Cited on page 151.)

Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999. (Cited on pages 23 and 24.)

Pnevmatikakis, E. A. and Giovannucci, A. (2017). NoRMCorre: An online algorithm for piecewise rigid motion correction of calcium imaging data. *bioRxiv*, page 108514. (Cited on pages 157 and 160.)

Polack, P.-O., Friedman, J., and Golshani, P. (2013). Cellular mechanisms of brain state-dependent gain modulation in visual cortex. *Nature Neuroscience*, 16(9):1331–9. (Cited on pages 14, 16, 50, 56, 115, 116, and 177.)

Poort, J., Khan, A. G., Pachitariu, M., Nemri, A., Orsolic, I., Krupic, J., Bauza, M., Sahani, M., Keller, G. B., Mrsic-Flogel, T. D., and others (2015). Learning enhances sensory and multiple non-sensory representations in primary visual cortex. *Neuron*, 86(6):1478–1490. (Cited on page 158.)

Reimer, J., McGinley, M. J., Liu, Y., Rodenkirch, C., Wang, Q., McCormick, D. A., and Tolias, A. S. (2016). Pupil fluctuations track rapid changes in adrenergic and cholinergic activity in cortex. *Nature communications*, 7:13289. (Cited on page 116.)

Renart, A., de la Rocha, J., Bartho, P., Hollender, L., Parga, N., Reyes, A., and Harris, K. D. (2010). The asynchronous state in cortical circuits. *Science*, 327(5965):587–590. (Cited on pages 19, 25, 32, 53, 55, 107, and 154.)

Roth, M. M., Dahmen, J. C., Muir, D. R., Imhof, F., Martini, F. J., and Hofer, S. B. (2016). Thalamic nuclei convey diverse contextual information to layer 1 of visual cortex. *Nature neuroscience*, 19(2):299. (Cited on page 17.)

Rubin, D. B., Van Hooser, S. D., and Miller, K. D. (2015). The stabilized supralinear network: A unifying circuit motif underlying multi-input integration in sensory cortex. *Neuron*, 85(2):402–417. (Cited on pages 29, 57, and 63.)

Saalmann, Y. B. and Kastner, S. (2009). Gain control in the visual thalamus during perception and cognition. *Current opinion in neurobiology*, 19(4):408–414. (Cited on page 153.)

Sachidhanandam, S., Sreenivasan, V., Kyriakatos, A., Kremer, Y., and Petersen, C. C. (2013). Membrane potential correlates of sensory perception in mouse barrel cortex. *Nature Neuroscience*, 16(11):1671–1677. (Cited on pages 15 and 56.)

Sakata, S. (2016). State-dependent and cell type-specific temporal processing in auditory thalamocortical circuit. *Scientific Reports*, 6. (Cited on pages 17, 18, and 144.)

Sakata, S. and Harris, K. D. (2009). Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. *Neuron*, 64(3):404–418. (Cited on pages 12 and 23.)

Sakata, S. and Harris, K. D. (2012). Laminar-dependent effects of cortical state on auditory cortical spontaneous activity. *Frontiers in Neural Circuits*, 6(109):6. (Cited on page 47.)

Sanchez-Vives, M. V., Mattia, M., Compte, A., Perez-Zabalza, M., Winograd, M., Descalzo, V. F., and Reig, R. (2010). Inhibitory modulation of cortical up states. *Journal of Neurophysiology*, 104(3):1314–1324. (Cited on pages 24 and 52.)

Sanchez-Vives, M. V. and McCormick, D. A. (2000). Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nature Neuroscience*, 3(10):1027–1034. (Cited on pages 11, 24, and 52.)

Schmid, M. C., Schmiedt, J. T., Peters, A. J., Saunders, R. C., Maier, A., and Leopold, D. A. (2013). Motion-sensitive responses in visual area v4 in the absence of primary visual cortex. *Journal of Neuroscience*, 33(48):18740–18745. (Cited on page 13.)

Schneider, D. M., Nelson, A., and Mooney, R. (2014). A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature*, 513(7517):189–194. (Cited on pages 14, 16, 17, 23, 143, and 177.)

Schölvinck, M. L., Saleem, A. B., Benucci, A., Harris, K. D., and Carandini, M. (2015). Cortical state determines global variability and correlations in visual cortex. *The Journal of Neuroscience*, 35(1):170–178. (Cited on pages 12, 23, and 47.)

Shadlen, M. N., Britten, K. H., Newsome, W. T., and Movshon, J. a. (1996). A computational analysis of the relationship between neuronal and behavioral responses to visual motion. *Journal of Neuroscience*, 16(4):1486–1510. (Cited on page 23.)

Shapcott, K. A., Schmiedt, J. T., Saunders, R. C., Maier, A., Leopold, D. A., and Schmid, M. C. (2016). Correlated activity of cortical neurons survives extensive removal of feedforward sensory input. *Scientific Reports*, 6. (Cited on pages 13, 19, 52, and 143.)

Sippy, T. and Yuste, R. (2013). Decorrelating action of inhibition in neocortical networks. *The Journal of Neuroscience*, 33(23):9813–9830. (Cited on page 16.)

Snyder, A. C., Morais, M. J., and Smith, M. A. (2016). Dynamics of excitatory and inhibitory networks are differentially altered by selective attention. *Journal of Neurophysiology*, 116(4):1807–1820. (Cited on page 16.)

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *Journal of machine learning research*, 15(1):1929–1958. (Cited on page 86.)

Stark, E., Eichler, R., Roux, L., Fujisawa, S., Rotstein, H. G., and Buzsáki, G. (2013). Inhibition-Induced theta resonance in cortical circuits. *Neuron*, 80(5):1263–1276. (Cited on pages 45 and 61.)

Steriade, M., Timofeev, I., and Grenier, F. (2001). Natural waking and sleep states: a view from inside neocortical neurons. *Journal of neurophysiology*, 85(5):1969–1985. (Cited on page 12.)

Stringer, C., Pachitariu, M., Steinmetz, N. A., Okun, M., Bartho, P., Harris, K. D., Sahani, M., and Lesica, N. A. (2016). Inhibitory control of correlated intrinsic variability in cortical networks. *Elife*, 5:e19695. (Cited on page 23.)

Tan, A. Y. Y., Chen, Y., Scholl, B., Seidemann, E., and Priebe, N. J. (2014). Sensory stimulation shifts visual cortex from synchronous to asynchronous states. *Nature*, 509(7499):226–9. (Cited on pages 13, 56, and 57.)

Tasic, B., Menon, V., Nguyen, T. N., Kim, T. K., Jarsky, T., Yao, Z., Levi, B., Gray, L. T., Sorensen, S. A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature Neuroscience*. (Not cited.)

Tsodyks, M., Pawelzik, K., and Markram, H. (1998). Neural networks with dynamic synapses. *Neural computation*, 10(4):821–835. (Cited on page 54.)

Urbain, N., Salin, P. A., Libourel, P.-A., Comte, J.-C., Gentet, L. J., and Petersen, C. C. (2015). Whisking-related changes in neuronal firing and membrane potential dynamics in the somatosensory thalamus of awake mice. *Cell Reports*, 13(4):647 – 656. (Cited on page 17.)

van Vreeswijk, C. and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293):1724–1726. (Cited on pages 19, 25, 53, and 55.)

Vinck, M., Batista-Brito, R., Knoblich, U., and Cardin, J. A. (2015). Arousal and Locomotion Make Distinct Contributions to Cortical Activity Patterns and Visual Encoding. *Neuron*, 86(3):740–754. (Cited on pages 13, 14, 16, 18, 23, 50, 97, 115, 116, and 177.)

Vinje, W. E. and Gallant, J. L. (2002). Natural stimulation of the nonclassical receptive field increases information transmission efficiency in v1. *Journal of Neuroscience*, 22(7):2904–2915. (Cited on pages 19 and 82.)

Vogels, T. P. and Abbott, L. F. (2005). Signal propagation and logic gating in networks of integrate-and-fire neurons. *The Journal of neuroscience*, 25(46):10786–10795. (Cited on pages 25 and 52.)

Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C., and Gerstner, W. (2011). Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science*, 334(6062):1569–1573. (Cited on page 155.)

Vogelstein, J. T., Packer, A. M., Machado, T. A., Sippy, T., Babadi, B., Yuste, R., and Paninski, L. (2010). Fast nonnegative deconvolution for spike train inference from

population calcium imaging. *Journal of neurophysiology*, 104(6):3691–3704. (Cited on page 182.)

Wehr, M. and Zador, A. M. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature*, 426(6965):442–446. (Cited on page 16.)

Wertz, A., Trenholm, S., Yonehara, K., Hillier, D., Raics, Z., Leinweber, M., Szalay, G., Ghanem, A., Keller, G., Rózsa, B., et al. (2015). Single-cell–initiated monosynaptic tracing reveals layer-specific cortical network modules. *Science*, 349(6243):70–74. (Not cited.)

Wilson, N. R., Runyan, C. A., Wang, F. L., and Sur, M. (2012). Division and subtraction by distinct cortical inhibitory networks in vivo. *Nature*, 488(7411):343–348. (Cited on pages 16 and 18.)

Wiltschko, A. B., Johnson, M. J., Iurilli, G., Peterson, R. E., Katon, J. M., Pashkovski, S. L., Abraira, V. E., Adams, R. P., and Datta, S. R. (2015). Mapping sub-second structure in mouse behavior. *Neuron*, 88(6):1121 – 1135. (Cited on pages 116 and 126.)

Wimmer, R. D., Schmitt, L. I., Davidson, T. J., Nakajima, M., Deisseroth, K., and Halassa, M. M. (2015). Thalamic control of sensory selection in divided attention. *Nature*, 526(7575):705. (Cited on page 153.)

Wolf, F., Engelken, R., Puelma-Touzel, M., Weidinger, J. D. F., and Neef, A. (2014). Dynamical models of cortical circuits. *Current Opinion in Neurobiology*, 25:228–236. (Cited on pages 19 and 25.)

Xue, M., Atallah, B. V., and Scanziani, M. (2014). Equalizing excitation-inhibition ratios across visual cortical neurons. *Nature*, 511(7511):596. (Cited on page 144.)

Yu, J., Gutnisky, D. A., Hires, S. A., and Svoboda, K. (2016). Layer 4 fast-spiking interneurons filter thalamocortical signals during active somatosensation. *Nature neuroscience*, 19(12):1647–1657. (Cited on pages 17 and 143.)

Yu, X., Ye, Z., Houston, C. M., Zecharia, A. Y., Ma, Y., Zhang, Z., Uygun, D. S., Parker, S., Vyssotski, A. L., Yustos, R., et al. (2015). Wakefulness is governed by gaba and histamine cotransmission. *Neuron*, 87(1):164–178. (Cited on page 17.)

Zhu, Y., Qiao, W., Liu, K., Zhong, H., and Yao, H. (2015). Control of response reliability by parvalbumin-expressing interneurons in visual cortex. *Nature Communications*, 6. (Cited on pages 16 and 18.)

Zhuang, J., Bereshpolova, Y., Stoelzel, C. R., Huff, J. M., Hei, X., Alonso, J.-M., and Swadlow, H. A. (2014). Brain State Effects on Layer 4 of the Awake Visual Cortex. *The Journal of Neuroscience*, 34(11):3888–3900. (Cited on page 57.)