Title
The laminar organization of V1 neural activity in response to dynamic natural scenes

Author
Khosrowshahi, Amir

Publication Date
2011

Peer reviewed|Thesis/dissertation

The laminar organization of V1 neural activity in response to dynamic natural scenes

by

Amir Khosrowshahi

A dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Vision Science

and the Designated Emphasis

in

Computational Science and Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Bruno A. Olshausen, Chair
Professor Yang Dan
Professor Michael R. DeWeese
Professor Charles M. Gray

Fall 2011

The laminar organization of V1 neural activity in response to dynamic natural scenes

by

Amir Khosrowshahi

Abstract

The laminar organization of V1 neural activity in response to dynamic natural scenes

by

Amir Khosrowshahi

Doctor of Philosophy in Vision Science

and the Designated Emphasis in

Computational Science and Engineering

University of California, Berkeley

Professor Bruno A. Olshausen, Chair

Despite remarkable progress in understanding the neurophysiology of cortex, fundamental questions regarding its function remain unanswered. The present study explores how a local population of neurons along a cortical column in visual cortex responds to dynamic natural stimuli. I show how a novel application of sparse coding to neural recordings obtained with high-density laminar probes can separate the complex statistical structure of this data into biophysically interpretable, underlying causes. The resulting data representation is instrumental in characterizing both spiking and local field potential activity across lamina in response to natural movies. Finally, I present a framework for understanding laminar population activity in terms of a statistical model that accounts for network interactions as well as the driving influence of the stimulus.

Acknowledgements

I am deeply indebted to my advisor Bruno Olshausen for his unfailingly generous support and for sharing his many exciting ideas, some of which are explored in this thesis. I would like to especially thank Urs Köster for a fun and productive collaboration and Charles Gray for providing a great environment to experience experimental neuroscience. I would like to thank Jonathan Baker for his help in the early parts of this work. I would like to acknowledge the members of the Redwood Center for the many valuable interactions from which I have greatly benefited, and Stan Klein for steering me towards the Vision Science program. Additionally, I am grateful to Gaute Einevoll for his hospitality in two separate visits to his lab. I would also like to thank my friend Casimir Wierzynski for his help and advice over many years.

This work was made possible by funding through a CRCNS grant funded from NIH R01EY019965. Computations performed on the Stallo supercluster at the University of Tromsø, NOTUR project number n4661k, were supported by Grant 178892/V30 eNeuro as part of the eScience program of the Research Council of Norway.

To Line.

Contents

Chapter 1

Introduction

The goal of this work is to characterize how a local population of neurons along the length of a laminar probe in primary visual cortex responds to full-field, dynamic natural movies. Though local cortical physiology has been intensively studied through a variety of means, little work has been done to characterize it in this setting. As a result, this work will necessarily be exploratory in nature. However, it builds upon results along several lines of research which are highlighted in this section.

1.1 Historical background

A great deal of our current knowledge of cortical processing has been accumulated over decades of single-unit recordings using controlled test stimuli. These stimuli typically have a low dimensional parametrization that can be spanned during the short duration available in a recording session. They are designed to have desirable statistical properties to simplify off-line analysis and reduce bias in parameter estimation. But the cortex is a highly interconnected, nonlinear dynamical system and it is not clear how far this knowledge generalizes to populations of neurons acting in concert to complex, ecologically relevant stimuli. Despite remarkable recent advances in experimental methods[12,15,37,65], allowing us to record from larger populations of neurons and to manipulate them chemically and optically in intricate ways, our understanding of the early visual system and how it processes input is still largely based on knowledge gained from early experiments[55,56]. The predominant experimental paradigm remains to drive cortex with simple stimuli eliciting responses that are easily interpretable in terms of an established and highly oversimplified prescription for the role of early visual processing. This entrenched view has had wide-spread influence, permeating into other fields such as computer vision and psychology, where it is rarely questioned.

Many lines of evidence from a variety of recent studies of the early visual system reveal a messy picture full of incompatible and inconsistent results that suggest our whole framework

for understanding early vision and primary visual cortex in particular requires a complete revision[85]. The current models of response properties of individual neurons, despite their theoretical elegance and simplicity, have poor predictive power particularly when applied to natural stimuli[22,27,85]. Despite impressive recent advances, computer vision and robotic systems exploiting our knowledge of the biology of visual processing still fail to perform basic tasks such as inferring the presence of surfaces, occlusion, and clutter in a scene, grasping and manipulating simple objects, and navigating a landscape while avoiding obstacles. This thesis focuses on a primary concern of Olshausen and Field[85], that the dominant paradigm of characterizing cortical processing a single neuron at a time with simple, controlled stimuli is flawed. Here, this concern is addressed directly by simultaneously recording from a large population of neurons with stimuli that include long duration, full-field natural movies. Despite encouraging success in this promising, largely unexplored direction, it is important to remain mindful of shortcomings that will need to be addressed in future studies, such as the use of anesthesia, sampling biases, and using impoverished models of the stimulus to characterize neural responses.

1.2 Natural scene statistics and cortical responses

A key component in the evolution and development of sensory systems is their need to adapt to the statistics of the environment, suggesting that one goal of such a system could be to produce an efficient internal representation of sensory statistics in the responses of populations of early sensory neurons[4,9,38,100]. To explore this efficient coding hypothesis, researchers have characterized the probability distributions of natural image patches[11,54,82,83], natural movies[20], as well as auditory stimuli[68,102], in terms of optimality principles such as sparseness of representation[82] and coding robustness[29] that would be ecologically advantageous. Despite capturing only a fraction of the complexity of the distributions, these simple functional models have nonetheless shown considerable explanatory power in their correspondence with physiology[61,82,102]. Recently, researchers have increasingly used naturalistic stimuli in probing the early visual system and have found differences that do not fit into the tidy picture built upon results using artificial stimuli[34], such as sparser response[7,106,107] and greater sensitivity to spectral phase structure than to power[35] in V1. These results were derived primarily from experiments where only a single neuron was recorded at a time.

More recently, Yen *et al.*[111] used silicon tetrode probes to simultaneously record from several nearby neurons in visual cortex in response to natural movies. In agreement with previous findings with natural stimuli, neurons had sparse and transient responses that were precise in time and repeatable from trial-to-trial (Fig. 1.1). More significantly, they showed that the response of nearby neurons exhibited an unexpected degree of heterogeneity. Fig. 1.2 depicts the response of several simultaneously recorded neurons located in close proximity from a single tetrode of a muli-site silicon probe for three separate short natural movie

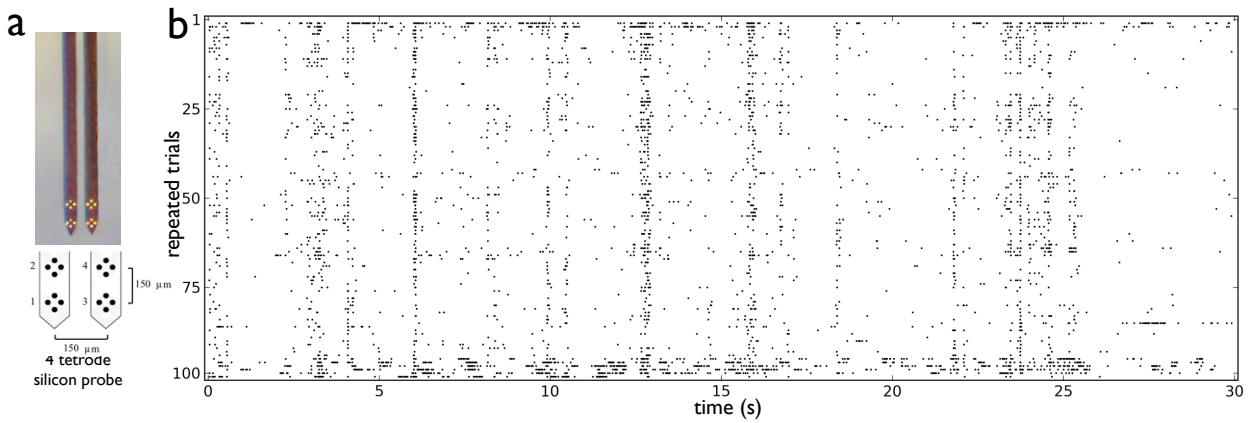

Figure 1.1: **Sample neural response to repeated natural movie. (a)** Silicon tetrode probe with 4 sites on two separate shanks. The tetrode arrangement improves single unit isolation and facilitates recording from several neurons simultaneously at each site. **(b)** Spike rasters of a single neuron in response to 100 repeats of a 30 second natural movie taken from a nature show documentary[111]. The raster displays temporal sparsity, selective and punctate response to certain features in the stimulus, as well as a high degree of trial-to-trial precision, properties shared by many such neurons. The standard functional model for response of these neurons is a set of linear receptive filters with outputs nonlinearly combined to estimate a firing rate[98]. However, these models fail to capture many of the response properties shown in this example[85,86].

repeats. Despite these neurons having similar classical properties such as orientation tuning to gratings and receptive fields as derived from binary noise stimuli[111], their responses show a remarkable degree of difference to natural movies. This result is in sharp contrast to a commonly accepted principle of cortical organization, that neural populations in a single column of cortex provide a redundant representation of local stimulus features[57], and that neuronal response variability could be reduced by pooling responses over nearby neurons with similar feature selectivity[74]. It suggests, rather, that a diverse degree of selectivity exists to stimulus features in a single cortical column, and that neurons in a columnar microcircuit can play distinct roles in the computations necessary in processing and representing the complex statistics of dynamic natural stimuli.

1.3 Large-scale neural recordings using silicon polytrodes

Technological advances have driven much of the progress in experimental neuroscience in recent years. One advance relevant to this work is the development of high-density laminar probes built out of a silicon substrate[14]. These polytrodes come in various recording site configurations, geometries, and number of shanks and can now be commercially purchased.

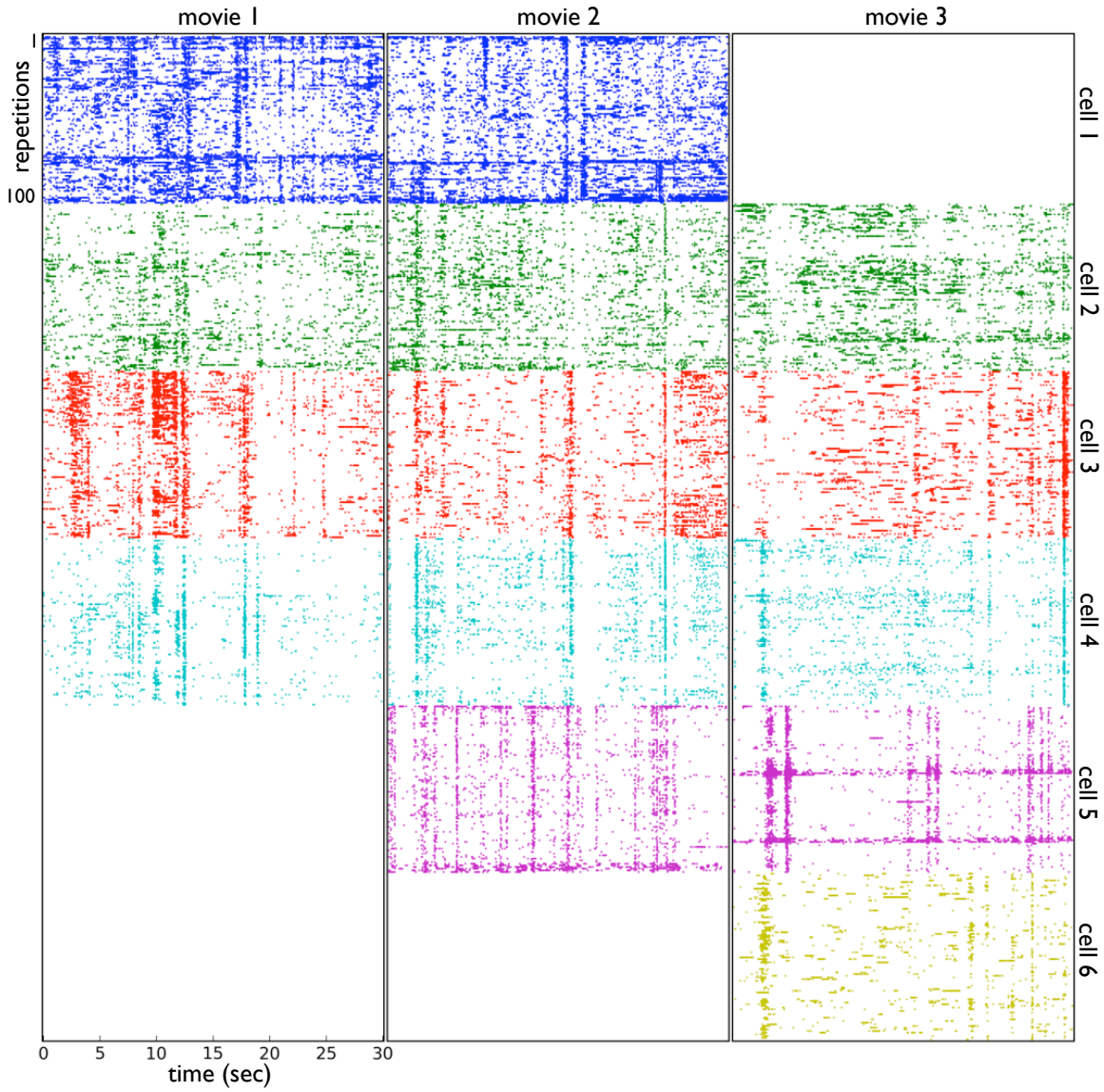

Figure 1.2: **Heterogenous response of nearby neurons to natural stimuli.** Spike rasters of a collection of neurons recorded on a single tetrode from a four tetrode probe (Fig. 1.1a) in response to 30 second repeats of three different natural movies[111]. Each neuron is depicted in a different color. The three columns correspond each to a different natural movie. Missing rasters indicate that the neuron was either lost during the recording or that it was not sufficiently isolated for spike sorting. The rasters display a remarkable diversity of response patterns, differing in overall firing rate, specificity in response, sensitivity to particular frames in each movie, and the nature of trial-by-trial variability, all despite the neurons being in close physical proximity. Chap. 4 presents a statistical model to try to account for this variability.

Several polytrode configurations including the one used for recordings in this thesis are displayed in Fig. 1.3. The primary advantage of polytrodes over methods such as 2-photon Ca^{2+} imaging is the ability to record from a full cortical column at high temporal resolution.

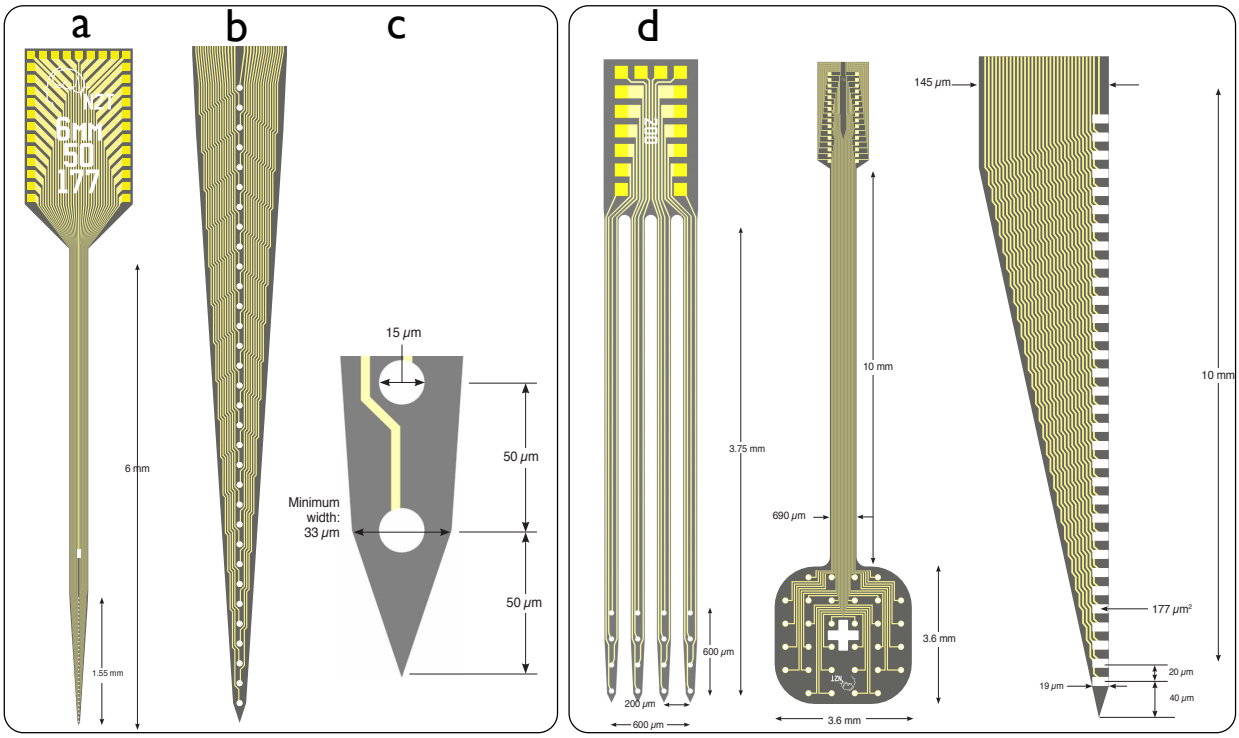

Figure 1.3: **Silicon polytrodes.** Silicon polytrodes[14] are a recent development in experimental neuroscience. Iridium recording sites of different sizes are patterned with varying density on a silicon substrate. These devices allow recording from tens of neurons simultaneously. The top of each probe is bonded to a head-stage interface board that carries the signal to an acquisition system. **(a)** The probe configuration used in most of the recordings in this work. **(b)** A closeup of **(a)** with channels arranged in a 1.55 mm linear array, sufficient in length to span a full cortical column. **(c)** Site geometry and spacings for the 32-channel probe. Site sizes can have a significant effect on the nature of the recording, such as the level of noise and how single unit waveforms are distributed on neighboring channels. **(d)** Various commercially available probe configurations[78] offer a large degree of flexibility in targeting different brain areas in a wide range of organisms.

Polytrodes with high electrode density capture the extracellular waveforms of single neurons on multiple channels, facilitating spike sorting. Sample spike traces of a recording from cat visual cortex using a 32-channel laminar probe are shown at different temporal scales in Figs. 1.4-1.6. The recordings show the statistically rich patterns of population firing across all lamina of a cortical column. Bursts of responses from groups of cells interrupt periods of

little to no activity. Extracellular waveforms of individual neurons are clearly visible when isolated but are more difficult to disambiguate when other neurons in close proximity or on the same channel are also firing. Waveforms often span several channels, though they can also be local to a single channel.

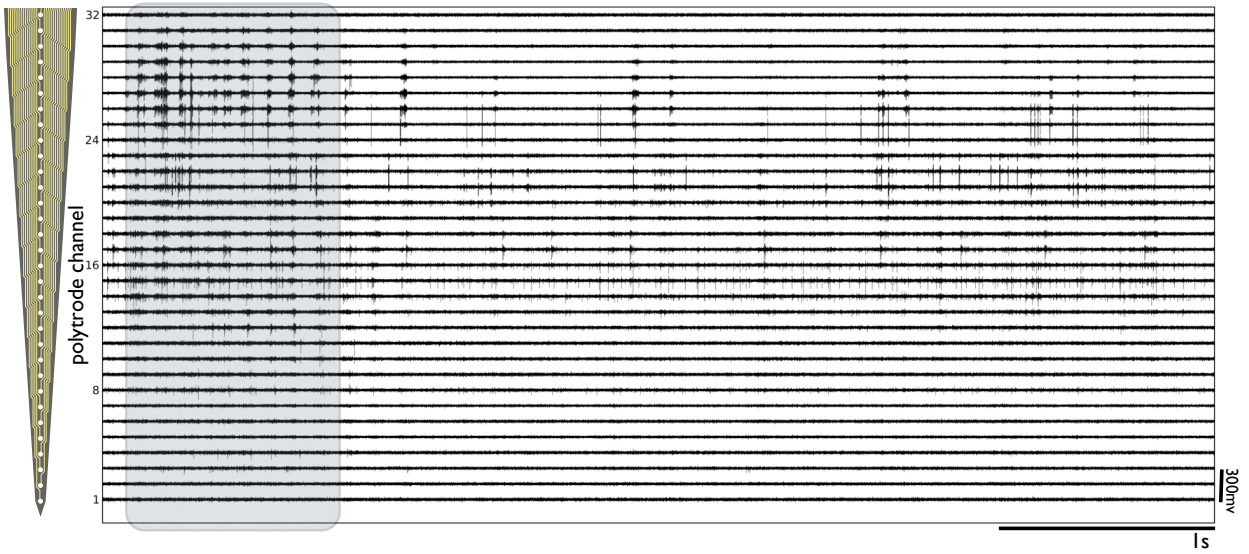

Figure 1.4: **Sample high-pass filtered polytrode traces at several time scales.** Recordings from cat visual cortex high-pass filtered between 0.5-10 kHz. After a burst of activity across channels due to the onset of a natural movie stimulus, activity becomes sparse, interrupted occasionally with short bursts of firing of groups of cells distributed across channels.

Fig. 1.7 shows a sample trace of the local field potential (LFP), the low-frequency portion of the recorded polytrode signal. This activity is thought to consist mainly of dendritic processing of synaptic inputs of populations of neurons[40] but remains poorly understood. Polytrode LFP signals are typically coherent over a large number of channels, but relative phases contain a rich structure. Oscillations in the LFP have faced several decades of intense scrutiny and theories abound for their role in cortical function, including a mechanism for coordinating between spatially separated regions of cortex[44,45,45,46,101]. Confirming the role of these oscillations has remained elusive. Large-scale recordings along a cortical column can help shed light on the biophysical mechanisms which generate the LFP[32,93] as well as the oscillations at characteristic frequencies[79].

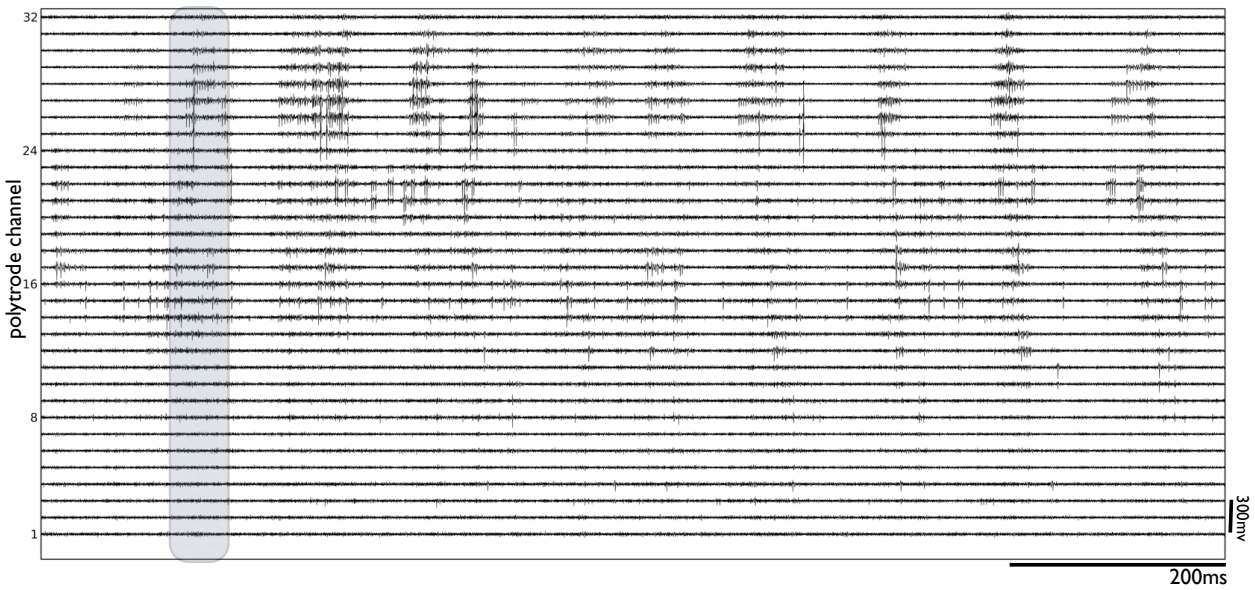

Figure 1.5: **Sample high-pass filtered polytrode traces at several time scales.** The greyed portion of Fig. 1.4 is shown at a finer temporal scale. Individual cells are visible, with waveforms distributed across several channels. Many of the cells fire in bursts of several action potentials.

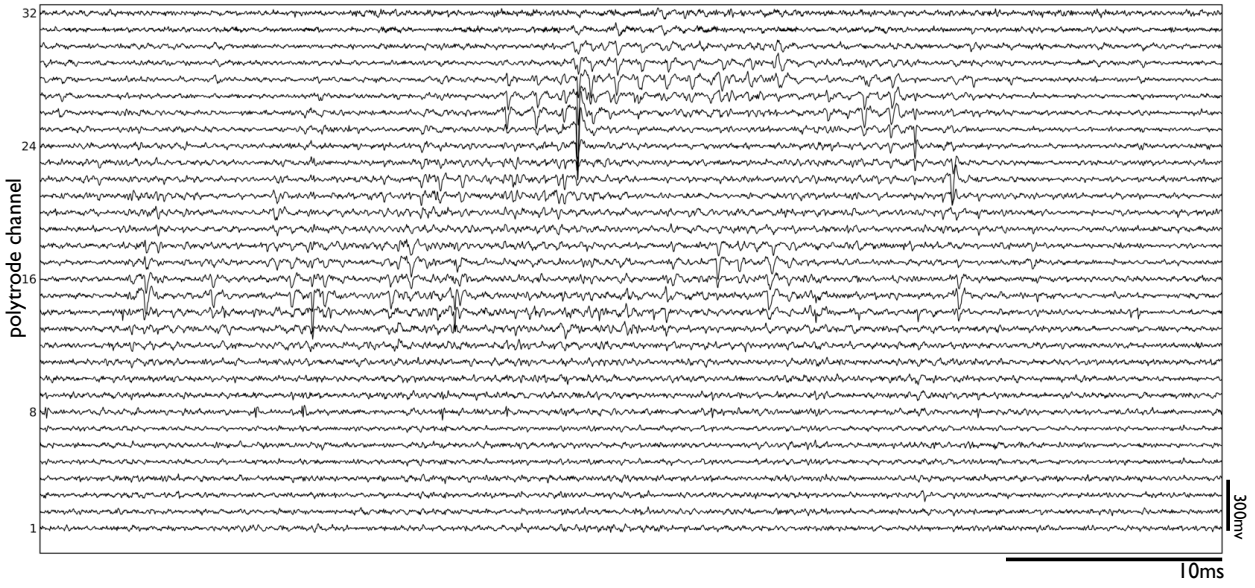

Figure 1.6: **Sample high-pass filtered polytrode traces at several time scales.** The greyed portion of Fig. 1.5 is shown at a finer temporal scale. Individual spike waveforms of different magnitudes and shapes are clearly visible.

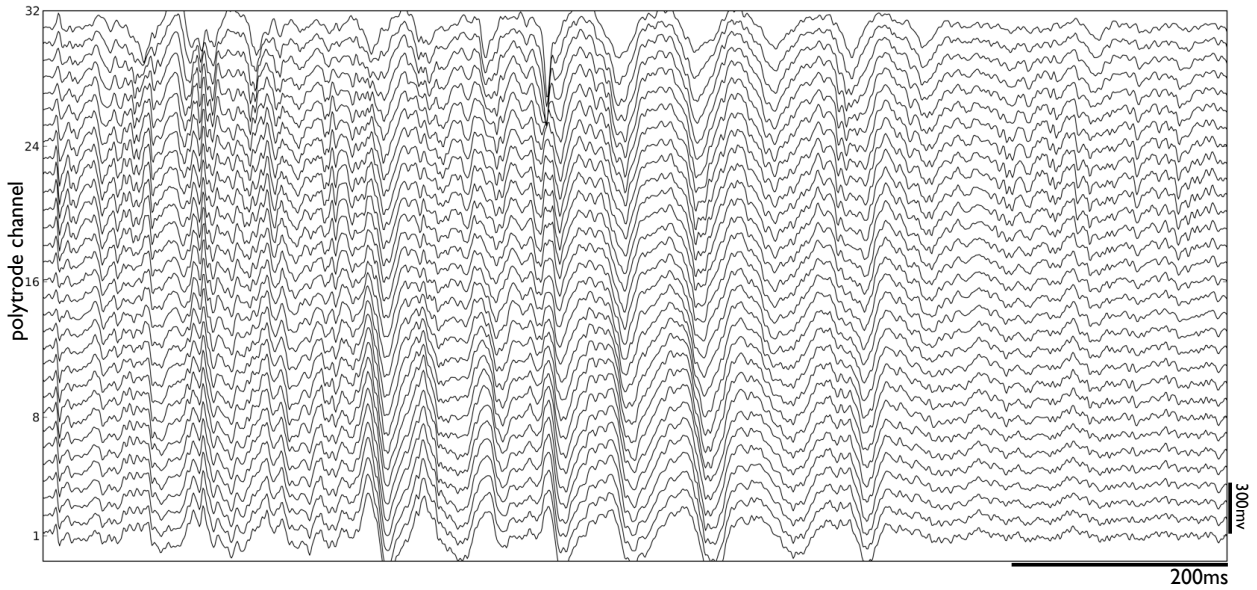

Figure 1.7: **Sample low-pass filtered polytrode trace.** Polytrode recordings from cat visual cortex filtered between 1-150 Hz. Some common features of laminar LFP data are oscillations at characteristic frequencies and complex patterns of phase relationships between channels.

1.4 Cortical microcircuits and models of population response

The layered structure of neocortex and the consistent pattern of synaptic connections between cell types both within and between different lamina suggest that this structure is to a large degree canonical and is replicated across cortex[31]. Though theories have been proposed about its computational role[31,42], insufficient data has been available to test their validity and significance. Recently, a class of parametric models, termed Generalized Linear Models (GLM's), have been introduced that attempt to capture the statistics of the spiking responses of populations of individual neurons in terms of self-inhibition, inter-neuronal coupling, the stimulus, and other factors such as the LFP[88,105]. These GLM's have been applied successfully to account for population responses obtained from large-scale, simultaneous recordings from the retina[95]. They are able to capture a significant portion of the joint statistical structure of the firing of a population of ganglion cells. However, few studies have attempted a similar approach to cortical neurons. Such models would provide a principled way to describe the statistics of cortical electrophysiology data and would provide a means of testing the mechanistic components of the canonical microcircuit hypothesis. Chap 4 presents an application of this approach to polytrode recordings.

1.5 Convex optimization and parallel computation

The computational methods underlying many machine learning algorithms and fitting of parametric statistical models are often convex optimizations[17]. Many of these methods are well established, sometimes over a century old. Despite this fact, the field of convex optimization has experienced explosive growth in the last two decades, driven by new theoretical results in several directions aided to a large extent by the dramatic increase in available computational power. One key success, for example, is the development of interior-point algorithms[62,77] which have made solving linear programs of 1 million or more variables routine on a typical workstation. This is an astonishing accomplishment given the usually exponential worst-case complexity of any strategy for picking an optimal vertex of a high-dimensional convex polytope. A number of new results have demonstrated that several classes of combinatorially intractable problems can be approximately and quickly solved when reformulated as relaxations of convex optimizations. Examples are the max-cut algorithm[43], kissing numbers for sphere packings[5], and lower bounds for chromatic numbers of finite graphs[6]. The methods developed in this thesis are formulated as high-dimensional convex optimizations as well as convex relaxations of combinatorially intractable problems. Their practical implementation relies heavily on recent developments in these fields and will be discussed in detail in Chap. 2 and 4.

Another recent development is the wide-spread availability of parallel computing architec-

tures, such as multicore processors, large clusters, and a new class of graphics cards (GPU's) that have been designed to perform a wide class of computations. The size and complexity of the data analyzed in this thesis required all key algorithms to be implemented in parallel. Furthermore, even routine tasks such as filtering of data and processing of stimuli could be done far more quickly exploiting the GPU's inherent parallelism in performing fourier transforms orders of magnitude more quickly than on a conventional multicore processor.

1.6 Sparse latent variable models

Latent variable statistical models are widely used in machine learning applications[13,108], including in several areas of biology, such as genomics. However, they have not found as much use in the description of electrophysiology data. The standard framework of studying recorded data is by linear filtering and modeling the magnitude of the output. Most fourier-based methods fall into this category. Latent variable models have a much greater degree of explanatory power as they they are able to 'explain away'[13] to some extent to infer true underlying causes of data. This inference is usually highly non-linear in contrast to traditional filtering methods, which only amount to projecting onto a carefully picked basis that may or may not be suitable for data with poorly understood structure.

A class of latent variable models that are central to this work are sparse coding models[39,82], in which variables that characterize the underlying causal structure of the data in some natural representation are sparsely active. The sparse representation and latent parameters for these models can be estimated in a variational framework through convex optimization. The sparse coding non-linearity has been shown to be a convex relaxation of a combinatorially hard problem[104] and has been studied in thousands of recent papers in several contexts, including compressed sensing[21]. Sparse coding models were used initially to describe the statistics of natural images[82,83] and auditory stimuli[68,102]. Increasing evidence from recordings of populations of neurons in different modalities and organisms suggests that sparse coding is a fundamental neurological principle, conferring advantages such as energy efficiency, increased storage capacity, and ease of read-out[84].

Chap. 2 applies sparse coding to polytrode data at two time scale regimes where neural activity is assumed to be sparse both in time as well as in space along the length of a polytrode. The crux of this thesis is that even the simplest sparse linear generative model of polytrode data as used here succeeds at characterizing the statistics of a rich, high-dimensional, poorly understood dataset in ways that traditional analysis methods would fail.

1.7 Outline of thesis

This thesis is organized into three parts. Chap. 2 develops a novel application of sparse coding that is applied in turn to high-pass and low-pass filtered polytrode data recorded from anesthetized cat visual cortex. This unsupervised algorithm is able to learn the statistically recurring components of spike and LFP datasets. Having transformed the data into this new representation, Chap. 3 characterizes the statistics of the components across lamina for different stimulus classes. Lastly, Chap. 4 combines the components into a GLM model for describing the joint statistics of the spiking and LFP to characterize the full statistics of the data in a unified manner. Chap. 5 concludes with a summary of results from this analysis and suggestions for future work.

Chapter 2

The statistical structure of laminar recordings

2.1 Introduction

Silicon polytrodes have become increasingly commonplace in recordings from a variety of brain areas. Built out of a silicon substrate with a high-density of recording sites, they are available in a variety of site arrangements and physical configurations (Fig. 1.3). They enable recording the extracellular waveforms of populations of neurons at a high spatial and temporal resolution. However, they share many of the challenges of other large-scale recording methods. Traditional spike sorting algorithms, designed for single electrodes or tetrodes, are in general not scalable to the increased number of channels and often cannot incorporate and exploit knowledge of the physical geometry of the probe. Polytrodes may have poorer single unit isolation than single electrodes as their position cannot be finely adjusted to move sites near cell bodies. Super-imposed spike waveforms on multiple channels cannot be handled by conventional thresholding and clustering spike sorting methods. Traditional signal processing algorithms such as spectral density estimation and linear filtering methods treat channels independently. It is not clear how to incorporate the complex phase relationships between polytrode channels in fourier-based analysis methods such as spike-field coherence[16]. These methods make implicit assumptions about the nature of recordings that may not hold, such as ascribing special significance to activity in narrow frequency bands. The statistics of the recorded data may be non-stationary, inconsistent across repeated experiments and containing experimental artifacts that are difficult to account for and remove.

In this Chapter, I present a novel application of sparse coding[82] (Sec. 1.6) to polytrode recordings from visual cortex. In this case, the underlying data is modeled as an approximately linear superposition of electric potentials from nearby biophysical events. Sparse coding assumes these events occur sparsely in time and in space along the polytrode. It is able to learn a natural representation for these events such the causal structure of the

data is explicit. The polytrode devices in this case are single shaft probes with 32 or 54 channels that span the extent of a cortical column. The goal is to record from a population of 30-40 neurons across all cortical lamina to study the spiking and LFP response properties of the population in response to dynamic, natural stimuli. Sparse coding models were first applied to natural image and audio statistics[61,82,102] and have recently been used in a wider range of applications. The simplest version of a sparse coding model, which is used here, is a linear generative model with a sparse prior on the latent variables. The basis functions learned provide a decomposition of the signal into dominant statistically recurring components. Their characteristics are revealing and interpretable and have the potential to shed light on structure that would be difficult to discern through other methods such as principal components analysis (PCA). An application is demonstrated in Chap. 3. Additionally, the inferred latent variables can be used as regressors in subsequent analyses in the place of the original data (Chap. 4).

The data was divide into high and low frequency, or 'spike' and 'LFP' datasets, respectively, and a sparse coding basis was learned per recording penetration. The learned spike bases represent components of extracellular potentials of individual neurons along the length of polytrode. Their relationship to spike sorted data using more conventional methods is demonstrated and a number of applications are presented. The learned LFP components reveal a rich laminar structure that is separated into distinct frequency bands. This novel representation of the LFP provides a promising new approach to understanding the biophysical cause and significance of these signals.

2.2 Methods

2.2.1 Surgery and preparation

Acute experiments were performed on three female (2.8-3.5 kg) and one male (3.5 kg) anesthetized and paralyzed cats. Prior to surgery, the animals were anesthetized with an intramuscular injection of ketamine (12 mg/kg) and acepromazine (0.3 mg/kg) and given atropine (0.025 mg/kg) subcutaneously to reduce salivation. Anesthesia was maintained using halothane (0.6-2%) in a mixture of nitrous oxide and oxygen (2:1) using a Surgivet Model 100 halothane vaporizer while the animals were actively ventilated using a Harvard Apparatus respirator pump. The cephalic vein was cannulated and a continuous infusion of Ringer's solution containing 2.5% dextrose was given throughout the experiment at a rate of 4 ml/kg/h. The antibiotic Cephazolin was administered intramuscularly (30 mg/kg) every eight hours to reduce the risk of infections. Vital signs including electrocardiogram, heart rate (140-180 bps), rectal body temperature ($37 - 39^\circ$ C), end-tidal CO_2 (3.5-4.5%), and SpO_2 (99-100%), were continuously monitored and maintained within the normal range using a Cardell 9500 HD veterinary life signs monitor and a Gaymar TP650 thermal water

blanket.

The animals were mounted in a stereotaxic frame and a stainless steel head post was affixed anterior to the craniotomy location using 6-8 stainless steel screws and covered with dental acrylic. This allowed us to retract the ear and eye bars to remove potentially painful pressure points and also served to electrically isolate the animal from the stereotaxic table. The stereotaxic frame was used to locate the primary visual cortex at coordinates 67 mm dorsally and ±3 mm laterally relative to AP0. After removal of the eye and ear bars, the head remained suspended from the head post for the remainder of the experiment.

An approximately 7 mm × 7 mm craniotomy was made in one hemisphere over area 17. After recording from several penetrations in this hemisphere, a second craniotomy was made in the opposite hemisphere for further penetrations. Paralysis was then induced with vercuronium at 1 mg/kg and maintained at 1 mg/kg/h intravenously.

A local application of Neosynephrine was made to retract nictitating membranes as well as an application of atropine to dilate the pupils. The eyes were focused on a computer monitor screen at a distance of 57 cm using a pair of gas permeable contact lenses chosen using the tapetal reflection technique[94], where an opthalmoscope light introduced into the eye produced an in-focus, reflected fundus image on a white screen facing the animal. At this distance, 1° of visual angle corresponds to 1 cm on the computer monitor.

To reduce pulsations from heartbeat and relieve intracranial pressure during the experiment, a catheter was inserted into the cisterna magna to allow for the partial drainage of cerebrospinal fluid. A small portion of the dura mater and, as necessary, the arachnoid layer, were then reflected using a dural knife custom made from a surgical needle. The head stage holding a silicon polytrode was carefully positioned just above the cortical surface using Kopf Instruments micromanipulators. A 4% mixture of agar in artificial cerebrospinal fluid (aCSF) was applied to protect the cortical surface and reduce pulsations. The polytrode was slowly lowered perpendicularly into cortex using a hydraulic microdrive Narishige Model MHO 110. The craniotomy and a portion of the polytrode assembly where then covered with a layer of bone wax to prevent the agar from drying out.

The protocol used in these experiments was approved by the Institutional Animal Care and Use Committee at Montana State University.

2.2.2 Recording procedure

All recordings were performed using single shank, planar silicon polytrodes purchased from NeuroNexus Technologies. In the first experiment, a 54-channel probe was used with contacts staggered in two columns, spaced 65 μm vertically, 56 μm horizontally, with an overall shaft width of 206 μm (54μmap2b in Blanche *et al.*[14]). In the remaining three experiments, 32-channel probes were used with sites arranged in a single column with contact diameters of 15 μm, 23 μm, and 30 μm arrayed over a length of 1.55 mm with full-shaft lengths of 6 mm or 10 mm and width tapering from 250 μm to a point over the active length of the probe

(Fig. 1.3a).

The polytrode was mounted onto a Multichannel Systems ADPT-NN-32 Probe adapter, which was connected to the hydraulic drive by a custom adapter. Signals were buffered with a head-stage amplifier with gain of 10 in the first three experiments, modified to a gain of 5 in the last, using Multichannel Systems model MPA32I. The signals were further amplified with a Multichannel Systems model FA64I amplifier with a gain of 500 and cutoffs at 1 Hz and 10 kHz. This signal was digitized with 14-bit precision and 30 kHz sampling rate with a PC-based data acquisition system containing a United Electronic Industries PowerDAQ PD2-MF-64-2M-14H. The broadband signals for all channels were recorded to hard disk for offline processing. For the last experiment, electrolytic lesions were made at the end of each recording session by injecting 2 μA of current for a duration of 12 s into the top and bottom recording sites according to the probe manufacturer's specification. However, lesions were not visible in the histology and did not aid in recovering polytrode position.

The raw data was divided into low-pass and high-pass datasets. For the high-pass dataset, a 4-th order Butterworth filter with passband between 0.5 kHz and 10 kHz was applied forward then backward on the full raw dataset for a net zero-phase filtering and the filtered data saved to disk without resampling at 30 kHz. For the low-pass dataset, the data was down-sampled from 30 kHz to 1 kHz by a succession of filtering and sub-sampling steps of 5, 3, and 2 times, with FIR filters designed using the window method and filter lengths of 256, 128, and 128, respectively. This data was then filtered with a zero-phase Butterworth filter with passbands of 1 to 150 Hz and saved to disk at 1 kHz sample rate. These two datasets are referred to as the spike and local field potential (LFP) datasets. All analysis software was custom written using a collection of open-source `python` tools, including `numpy`, `scipy`[81], `ipython`[89], `h5py`[24], and `matplotlib`[58].

2.2.3 Histological procedures

At the end of each experiment, the animal was euthanized with an intravenous injection of pentobarbital and perfused with a 4% formaldehyde solution in phosphate buffered saline. The brain was then removed and brain regions containing the recording locations were blocked and fixed in sucrose solution. Histological Nissl-stained slices of 60 μm thickness were made (FD Neurotechnologies) in a coronal direction at 120 μm intervals. For all but one penetration, the polytrode track was clearly visible in the slices. An example is shown in Fig. 2.1. The tracks were used to confirm recording depth, angle of penetration, and whether the recording spanned an entire single or multiple cortical columns.

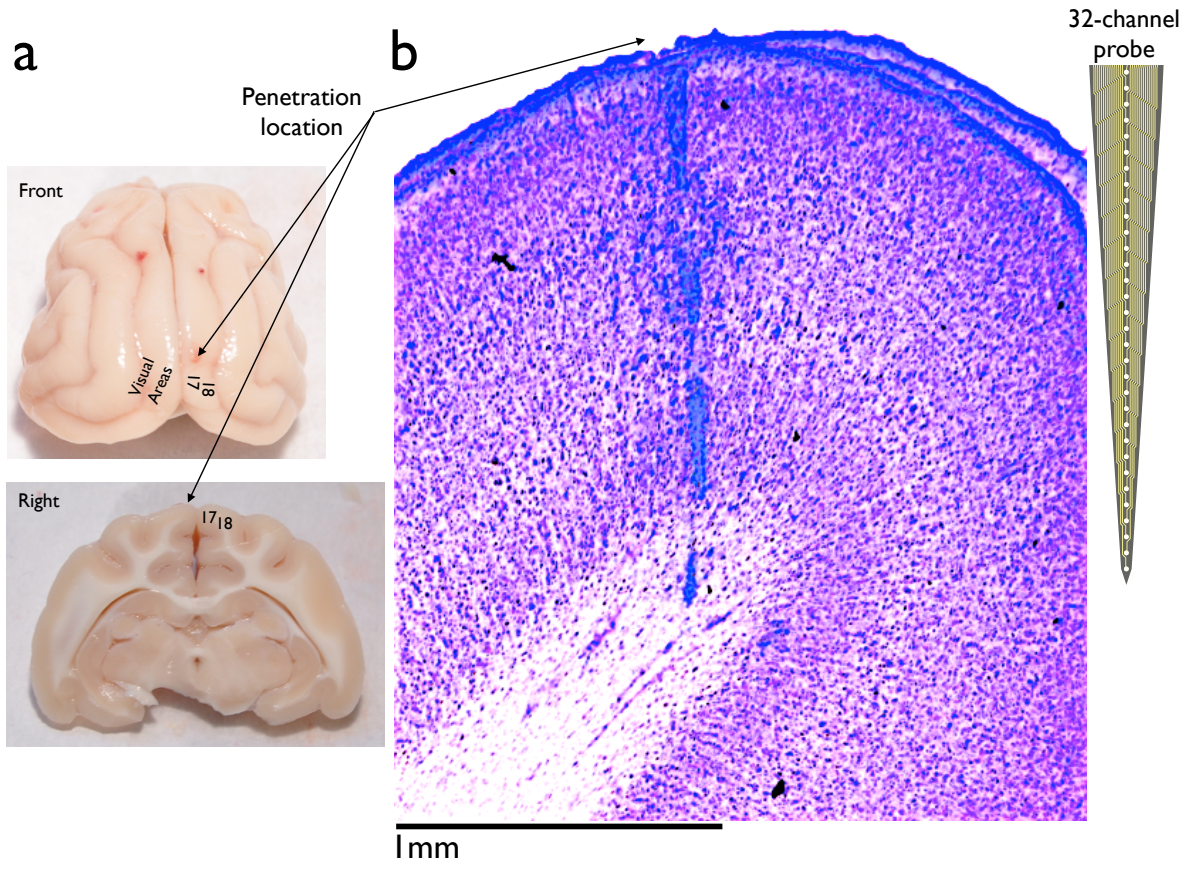

Figure 2.1: **Histological recording signature. (a)** Full brain from one experiment showing recording site and location of visual areas 17 and 18 as well as a coronal block in preparation for histology. **(b)** Reconstruction of one penetration from four superimposed Nissl-stained slices. In each of the four slices, some tissue damage due to the electrode passage is visible. The penetration spans all cortical layers and is aligned parallel to the striations of cortical mini-columns. Due to tissue shrinkage from the perfusion and fixation process, as well as the penetration spanning multiple slices, the scale bar does not exactly correspond to the span of recording locations of the polytrode.

2.2.4 Stimuli

A variety of visual stimuli were used in each experiment, including long and repeated short natural movies, 'Hilbert' movies, drifting gratings, binary white noise, and full-field black and white stimuli. The combined duration of all movies played for each penetration was approximately four hours. The stimuli and presentation protocol are presented in detail in Sec. 3.2.2.

2.2.5 Sparse coding model

Sparse coding[82,87,102] is a latent variable model that attempts to describe data in terms of a small number of additive components, or basis functions, selected out of a large dictionary. Let $y_i(t)$, the data on channel i at time t, be written as a temporal convolution of a set of basis functions $\phi_{ij}(t)$, with i and j denoting channel and basis function, respectively,

$$y_i(t) = \sum_j \phi_{ij}(t) * x_j(t) + \epsilon_i(t) \tag{2.1}$$

with $\epsilon_i(t) \sim \mathcal{N}(0, \sigma_n)$ small, uncorrelated gaussian noise on each channel. This model is illustrated in Fig. 2.2. To estimate model parameters, the data is assumed to be an identically distributed, independent ensemble of length T time samples with C channels $\mathbf{Y} = \{\mathbf{y}^{(i)}\}_{i=1\ldots D}$ with $\mathbf{y}^{(i)} \in \mathbb{R}^{C\times T}$. The log-likelihood of the model is,

$$\begin{aligned}
\mathcal{L}(\mathbf{\Phi}, \sigma_n) &= \log p(\mathbf{Y}|\mathbf{\Phi}, \sigma_n, \lambda) \\
&= \sum_{i=1}^{D} \log p(\mathbf{y}^{(i)}|\mathbf{\Phi}, \sigma_n, \lambda) \\
&= \sum_{i=1}^{D} \log \int \mathrm{d}\mathbf{x}\, p(\mathbf{y}^{(i)}|\mathbf{x}, \mathbf{\Phi}, \sigma_n)\, p(\mathbf{x}|\lambda)
\end{aligned}$$

where,

$$p(\mathbf{y}|\mathbf{x}, \mathbf{\Phi}, \sigma_n) \propto \exp\left(-\frac{1}{2\sigma_n^2}\sum_{t=1}^{T}\|\mathbf{y}_t - \sum_\tau \mathbf{\Phi}_\tau \mathbf{x}_{t-\tau}\|^2\right)$$

with $\mathbf{\Phi}_\tau \in \mathbb{R}^{C\times N}$, $\mathbf{x}_\tau \in \mathbb{R}^N$ with N the number of basis functions. The sparse prior on the coefficients $\mathbf{x}$ is parametrized with λ. The goal of learning is to maximize the likelihood

$\mathcal{L}(\mathbf{\Phi}, \sigma_n)$. The derivative of $\mathcal{L}(\mathbf{\Phi}, \sigma_n)$ is taken with respect to the model parameter $\mathbf{\Phi}_\alpha$,

$$\frac{\partial \mathcal{L}}{\partial \mathbf{\Phi}_\alpha} \propto -\sum_{i=1}^{D} \int \mathrm{d}\mathbf{x}\, p(\mathbf{x}|\mathbf{y}^{(i)}, \mathbf{\Phi}, \sigma_n, \lambda) \sum_{t=1}^{T} (\mathbf{y}_t - \sum_{\tau} \mathbf{\Phi}_\tau \mathbf{x}_{t-\tau})\, \mathbf{x}_{t-\alpha}^T$$

A number of ways exist to estimate the intractable integral in this expression, including a Laplace approximation[66] and Hamiltonian Monte Carlo sampling[26]. The simplest approach, taken here, is to assume the posterior distribution $p(\mathbf{x}|\mathbf{y}^{(i)}, \mathbf{\Phi}, \sigma_n, \lambda)$ is sufficiently peaked and to take one sample at its mode[83], that is,

$$\begin{aligned} \mathbf{x}^{(i)} &= \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}^{(i)}, \mathbf{\Phi}, \sigma_n, \lambda) \\ &= \arg\max_{\mathbf{x}} \log p(\mathbf{y}^{(i)}, \mathbf{x}|\mathbf{\Phi}, \sigma_n, \lambda) \\ &= \arg\min_{\mathbf{x}} \left(\frac{1}{2\sigma_n^2} \sum_{t=1}^{T} \|\mathbf{y}_t - \sum_{\tau} \mathbf{\Phi}_\tau \mathbf{x}_{t-\tau}\|^2 - \log p(\mathbf{x}|\lambda) \right) \end{aligned} \tag{2.2}$$

Model likelihood was maximized using an alternating scheme where $\mathbf{x}^{(i)}$ were inferred and then used to update $\mathcal{L}(\mathbf{\Phi}, \sigma_n)$[83]. An additional simplifying assumption was made to take the parameter of the gaussian noise σ_n as given, though a prior could be imposed on it and estimated along with $\mathbf{\Phi}$. In practice, an appropriate σ_n is approximated from descriptive statistics of the data. As the data is practically infinite, only a small sample of $\mathbf{y}^{(i)}$ was chosen in each step. Additionally, the model has a degeneracy due to the sparse prior $p(\mathbf{x}|\lambda)$ shrinking coefficients to zero, causing the norm of $\mathbf{\Phi}$ to grow without bound. Therefore a convex constraint was imposed such that,

$$\sum_{i=1}^{C} \sum_{\tau=1}^{P} \phi_{ij\tau}^2 \leq 1$$

where P are the number of time taps in the basis functions. The constraint makes the learning update of $\mathbf{\Phi}$ a quadratically constrained quadratic program (QCQP)[17]. In practice, however, this problem was solved by making a small stochastic gradient learning step and renormalizing the basis functions on each iteration. A full QCQP solver was implemented using a convolutional adaptation of the method proposed in Mairal *et al.*[72], but found that for neural datasets, the algorithm was prone to get stuck in local minima. During learning, the basis functions were recentered gradually over many iterations. This reflected an implicit prior that the basis functions should be temporally localized.

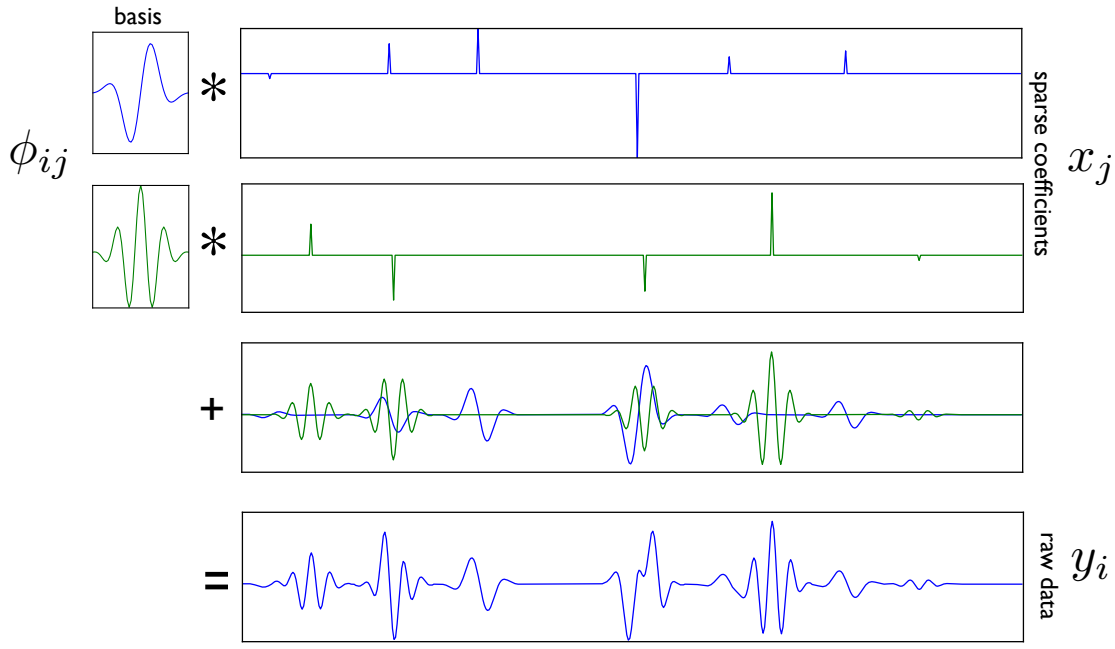

Figure 2.2: **Convolutional sparse coding.** A convolutional sparse coding model is a linear generative model where data is represented as a sum of a convolution of a set of bases with coefficients which are expected to be sparse and independent in both time and channel. This schematic shows how to reconstruct a portion of one channel of data at bottom using two basis elements at left convolved with corresponding sparse coefficients at right and summed. Estimation of basis functions ϕ_{ij} and coefficients x_j (Eq. 2.1) are called 'learning' and 'inference', respectively. Bases ϕ_{ij} are learned for spiking and LFP datasets separately for each recording penetration and coefficients are inferred for the entire recordings. This new representation has many desirable properties.

2.2.5.1 Matching pursuit inference

The nature of the spiking and LFP datasets required the inference problem 2.2 to be solved the using a different strategy for each case. For the spike dataset, which is made up primarily of highly sparse spike activity separated in time as well as in space among the channels and mixed with approximately gaussian noise, a greedy algorithm, matching pursuit [73,87,102], was chosen. This algorithm is efficient for a high degree of assumed sparsity when the basis functions can be assumed to be relatively incoherent. Its goal is to represent the data with

at most k basis functions,

$$\mathbf{x}^* = \underset{\|\mathbf{x}\|_0 \leq k}{\arg\min} \frac{1}{2\sigma_n^2} \sum_{t=1}^{T} \|\mathbf{y}_t - \sum_{\tau} \mathbf{\Phi}_\tau \mathbf{x}_{t-\tau}\|^2$$

where $\|\mathbf{x}\|_0$ denotes the number of non-zero elements of $\mathbf{x}$. For arbitrary $\mathbf{\Phi}$, this problem is NP-hard[28]. Matching pursuit is a greedy strategy that finds an approximate solution and overcomes this combinatorial complexity. Additionally, to make the learned basis functions and coefficients more physiologically interpretable, the coefficients were forced to be non-negative.

2.2.5.2 L_1-regularized quasi-Newton inference

For the LFP dataset, an L_1-regularized method was used to induce sparsity on the coefficients. The LFP dataset is sampled at a lower rate and is hence considerably smaller than the spike dataset, reducing requirements for computational efficiency by a large factor. An L_1 method performed better at learning in a less sparse regime where basis functions were more coherent and 'explaining away' was more critical. Briefly, the prior on coefficients was assumed to be exponentially distributed,

$$p(\mathbf{x}|\lambda) \propto e^{-\lambda\|\mathbf{x}\|_1}$$

giving the following convex minimization problem for inference,

$$\mathbf{x}^* = \underset{\mathbf{x}>0}{\arg\min} \frac{1}{2\sigma_n^2} \sum_{t=1}^{T} \|\mathbf{y}_t - \sum_{\tau} \mathbf{\Phi}_\tau \mathbf{x}_{t-\tau}\|_2^2 + \lambda\|\mathbf{x}\|_1$$

This objective is closely related to the Lasso, which has been intensively studied[23,103] and for which an abundance of methods exist. Despite its widespread use as a feature selection method, it is important to note that an L_1 regularizer has a deficiency. If the data is indeed generated by a Laplacian distribution, its order statistics are not sufficiently sparse to guarantee recovery[8]. Additionally, both inference methods used are not causal, and coefficients inferred for a given time can be affected by data in the future. In contrast, state-space models such as Kalman filters and hidden Markov models are causal by design, though operations such as smoothing are inherently acausal. Creating a causal inference algorithm in this setting is an open problem that is the subject of future work.

2.2.6 Implementation

For convolutional matching pursuit, the algorithm was implemented in `python` using the `numpy`[81] library. For L_1-regularized inference, a method[1] based on the widely used l-BFGS quasi-Newton algorithm[70] was chosen, which uses a particular choice of sub-gradient whenever the optimization attempts to cross from one octant to another. The advantages of this algorithm over the many others is that it does not require computing the Gram matrix, only requires an objective and gradient to be defined, is numerically stable for large numbers of parameters, converges quickly to an approximate solution, and can be efficiently run in parallel on multiple cores of a processor as it uses only BLAS level 1 operations.

Three versions of this algorithm were implemented. The first was a `cython`[10,99] wrapper of the `C++` library `liblbfgs`[80], which explicitly implements all linear algebra with SSE2 instructions. Next, a version in `cython` was implemented that could handle a vector of L_1 regularization parameters λ, a non-negative constraint on the coefficients, and would run more efficiently on multicore architectures through its use of vendor linear algebra libraries. Lastly, a version in `cython` and `pycuda`[64] was implemented to run fully on the NVIDA GPU avoiding all host to GPU transfers during the optimization. In this implementation, the convolutions in the objective were implemented both as a bank of 2-D FFTs as well as a single 3-D FFT. The 3-D FFT method, despite its theoretical inefficiency in this case, gave an approximately 30x speed-up versus computing the convolutions in the time domain on the CPU. The 2-D FFT bank gave only a 6x speed-up, most likely due to the GPU being constrained to computing one 2-D FFT at a time. Array slicing was implemented with 1-D texture maps and all norms, projections, and reductions were custom written to take advantage of the parallelism in the GPU.

The full learning algorithm was parallelized using the Message Passing Interface (MPI) in `python` using `mpi4py`[99]. On each learning iteration, the root node sampled the data from disk and distributed this data amongst the nodes using an MPI `Scatter`. All nodes then performed inference using one of the algorithms described above and the results were reduced on the root node with an MPI `Gather`. A learning step was taken and the basis was then MPI `Broadcast` to all nodes. This framework allowed the learning algorithm to exploit parallelism on a single multicore CPU with or without a GPU, a cluster of multicore CPU's, as well as a hybrid cluster of CPU's and GPU's.

2.2.6.1 Parallel inference of coefficients for a full dataset

After a basis was learned for a spike or LFP dataset, coefficients were inferred for the whole dataset in a chunk-wise parallel fashion (Fig. 2.3). Given N parallel computational nodes, the data was divided into N large chunks. Within each chunk, inference was performed in sequence on blocks of time length T with $T >> P$, where P is the number of time taps in the learned basis $\mathbf{\Phi}$. Blocking was used at it is computationally impractical to perform inference on arbitrarily large time segments. A block of length T yielded coefficients $\mathbf{x}$ of

time length $T + P - 1$. After one block was completed, all except a $P - 1$ length of its tail was written to disk. The $2P - 2$ tail portion of this block was used for initializing the next block. For the new block, the first $P - 1$ coefficients were held fixed whereas the next $P - 1$ coefficients were used to warm start the inference in the case of L_1 or were set to zero for matching pursuit.

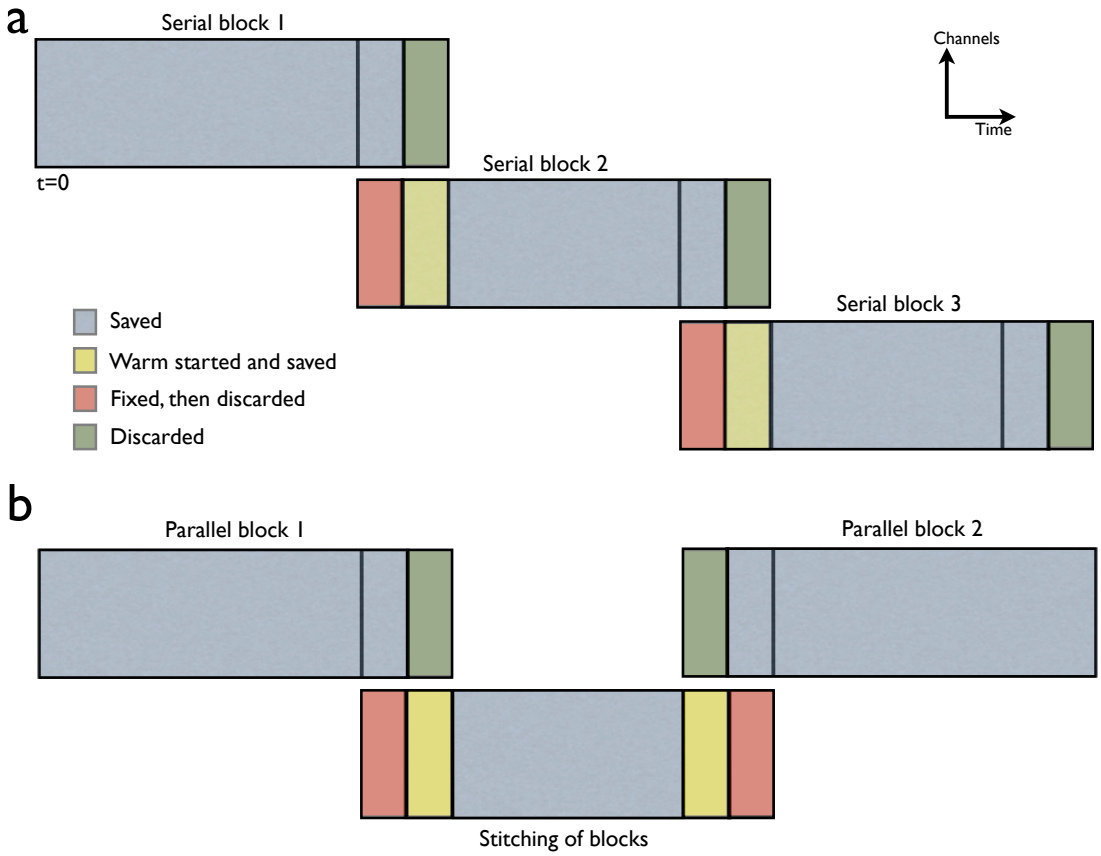

Figure 2.3: **Inferring coefficients over a full dataset. (a)** Each parallel chunk of data is blocked into tractable portions and processed serially by using tail portions of previous blocks to warm-start the following block. **(b)** When each parallel chunk is completed, the regions between chunks are processed to stitch chunks into one long set of coefficients. The stitched portion is warm-started from it's adjoining blocks.

The non-linear nature of inference raised the possibility that the implementation would have blocking artifacts. However, inference was tested by blocking over a several block region as well as inferring coefficients on the whole region with results in good agreement. This is due to the high degree of sparsity used as well as a side-effect of the convolutional formulation of the objective, that coefficients on the $P - 1$ borders received less derivative information

and were more likely to remain at zero.

Parallel blocks were stitched by performing inference on adjoining regions of length $3P-2$, with only the inner $P-2$ portion being free to be modified by the optimization. The parallel blocking allowed the inference to scale almost linearly and to compute coefficients for a recording session with approximately the same order of time as the experiment itself. The datasets with metadata were written to disk using a generic gzip level 4 compressed HDF5 format which afforded a 20-100 fold compression over the original dataset, depending primarily on the level of coefficient sparsity.

2.2.7 Conventional spike sorting

The dataset from one penetration was spike sorted using conventional methods for evaluation of the spike coefficients inferred using sparse coding. The high-pass dataset was divided into 8 groups of 4 non-overlapping channels, serving as virtual tetrodes. Putative spike waveforms were extracted using a threshold crossing criterion of 5 standard deviations after removing data outliers. A custom set of 11 features were computed for each extracted spike. The features were input into KlustaKwik[48], which uses a modified mixture-of-gaussians clustering algorithm, and the result was loaded into MClust[97]. Clusters were cleaned up by manually deleting and merging waveforms as necessary and saved as single units. Special care was given to identifying bursting cells, which have waveforms that change in amplitude and shape with each subsequent firing. After sorting for all virtual tetrodes were completed, cells with waveforms in adjacent tetrodes were identified and merged into single cells.

2.2.8 Model data

To create realistic model spike data[33] for testing the learning of spike waveforms and comparing model spike times with inferred sparse coefficients, the modeling software LFPy[69] was used. LFPy allows the specification of compartmental models of a population of neurons with given morphology, simulates them using Neuron[52], and records the extracellular potential at a set of electrodes using a line-source approximation[53]. To generate data, an example morphology provided by LFPy of a layer 5 pyramidal neuron[71] was used, with three sets of active conductances[50]. A set of 20 neurons were placed randomly with uniform probability along the length of a 32-channel model polytrode with distance from the probe distributed according to a gaussian distribution with a standard deviation of 30 μm. All neurons were oriented identically, roughly parallel to the polytrode. The model was run for 10 seconds at 31.25 kHz with each neuron's ion channels driven by constant magnitude synaptic inputs with gamma distribution arrival times. This data was then high-pass filtered between 0.5-10 kHz and a spike basis was learned. Sparse coefficients were inferred to compare with model spike times. To make learning and inference more challenging, varying degrees of noise was added to each channel of the filtered model data. This framework will be used

in future work to study the extracellular waveforms of different types of model neurons, to be able to assign cell types to recorded waveforms, to understand and identify other phenomenon such as back-propagating action potentials, to better understand the properties of the recording device itself, and to study the aggregate activity of a population of neurons in a model cortical column. Soma traces, synaptic currents, and recorded extracellular activity for one neuron are shown in Fig. 2.4.

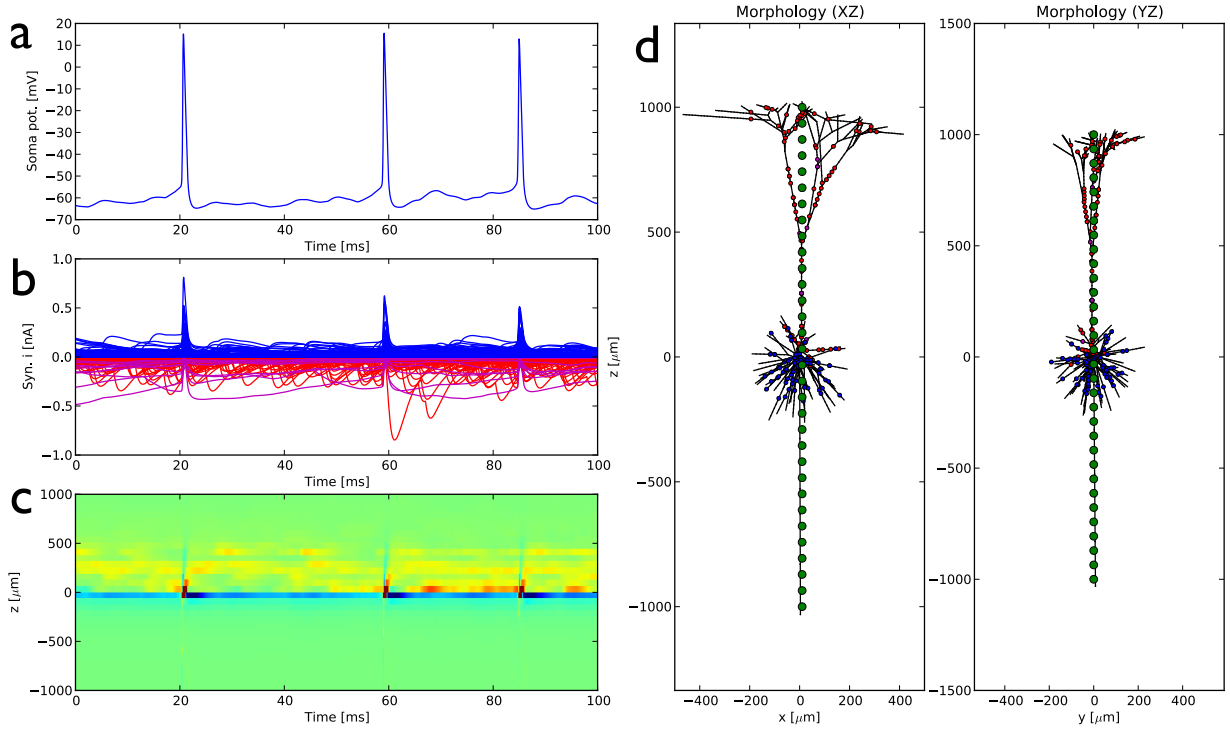

Figure 2.4: **Generating model spike data.** An LFPy[69] model layer 5 pyramidal cell with AMPA (red), NMDA (magenta), $GABA_A$ (blue) active channels distributed along the its dentritic arbors. **(a)** Soma trace with three action potentials. **(b)** Synaptic current traces in all channels. Input arrival times are gamma distributed. **(c)** Recorded extracellular activity at the 32-channel contact positions with time. **(d)** Two views of the geometry of the neuron, distribution of channels, and probe contacts (green).

To create model data for testing the learning of LFP basis functions, a set of spatiotemporal Gabor functions were distributed across channels, with coefficients sampled from Bernoulli or Laplace distributions, and reconstructed raw data from them. Varying degrees of noise were added including gaussian noise, sinusoidal line noise, and sinusoidal line noise mapped through a nonlinearity. The resulting model data shared many of the properties of the real LFP, including activity across lamina with both coherent and non-trivial phase relationships.

2.3 Results

2.3.1 Learned bases

2.3.1.1 Spike basis functions

The linear generative convolutional sparse coding model (2.2.5) was applied to a spike dataset consisting of high-pass filtered data from a single, 4 hour penetration. A linear generative model is particularly well-suited to this problem as the biophysical system in which the polytrode is embedded can be approximated as an isotropic, resistive medium satisfying the quasi-static approximation of Maxwell's equations. The signal recorded at the electrode contacts can be interpreted as linearly added potentials of extracellular activity surrounding the contacts[93].

The learned basis, sorted by lamina, is shown in Fig. 2.5. The basis elements divide into a number of distinct classes. Many basis elements are localized in time and channels, with shapes resembling spike waveforms. The relationship of these elements to spiking of individual neurons is discussed in 2.3.2.1. A number of basis functions with coherent activity across channels are harmonics of 60 Hz line noise bleeding into the high-pass filtered signal. Other basis functions represent unresolved multi-unit activity as well as activity entrained in noise. These latter basis functions appear at much lower voltage amplitudes in the data than the spiking activity. The basis was learned from sample patches of data from the entire recording penetration. A criterion that at least one channel exceed a standard deviation threshold was used to discard patches with no spiking activity to keep the model from trying to explain the structure in the noise. However, since one goal of this approach was to learn all the interesting structure in the data beyond well understood features such as spiking activity, the sparsity enforced was lower, reconstruction error smaller, and over-completeness of basis larger than one would use if the goal was only to only learn spike waveforms of the best isolated neurons. This approach introduces some confounds in interpreting coefficients as spike times and is addressed using model data in Sec. 2.3.3. Coupled with biophysically realistic modeling studies, this approach will allow identification of other interesting phenomena in this frequency domain such as back-propagating action potentials[19], action potentials along LGN axonal afferents, as well as allow assigning of putative cell types to different classes of waveforms.

Inference was performed with matching pursuit on blocks of data that were 4 times longer than the number of time taps in the kernels and at most 0.5% of coefficients were active for a given block. The number of basis functions was chosen to be at least twice the number of estimated single units present from visual inspection of the data so that the model would have enough representational power to capture low-amplitude and low-frequency spike waveforms. As a result, some basis functions did not learn any structure and were not activated in inference. A number of checks were performed to ensure the basis actually represented real structure in the data. This included re-running learning with different initial conditions,

on subsets of the recording, as well as randomly perturbing the basis during learning, in each case checking to see to what degree the bases changed. In general, the spike-like basis functions were highly consistent in each case with small variations in the shapes of the waveforms. There was a greater degree of variability in the remaining basis functions, but, overall, the qualitative differences were small. These variations are potentially useful for detecting non-stationarity in the data, such as would be caused, for example, by movement of the electrode.

The learned basis was then used to infer coefficients for the entire recording session. A example of inference is shown in Fig. 2.6. Many neurons, particularly in the superficial layers, fired in bursts, with waveforms that changed with each subsequent action potential. This is captured in the sparse coefficients as a series of activations of a single basis function in quick succession with some change in the coefficient amplitude. Often, many nearby neurons were active at the same time, with waveforms overlapping across channels and time. The inferred coefficients are able to explain away this confound to represent the data as a sum of activations of several basis functions with overlapping waveforms. Traditional cluster-based spike sorting methods[67] lack this flexibility and consequently have difficulty separating such coincident activity. Though much of the structure of the data is preserved in the reconstruction (Fig 2.6b), the sparsity of the representation helped in removing some of the background noise. It is important to remember that in contrast to traditional spike sorting algorithms, the sparse coding model attempts to represent all of the data and not just of waveforms extracted according to threshold criteria.

2.3.1.2 LFP basis functions

The sparse coding model was also applied to the LFP. An example of a learned basis for the same penetration as described in Sec. 2.3.1.1 is shown in Fig. 2.7. The basis functions, spanning 128 ms, separate into distinct classes. Many of the basis functions are localized in time with spatial structure that separates across lamina. Others are localized in time but are largely coherent in space and account for relatively more of the variance of the data than the other classes. The remaining basis functions have low-frequency temporal structure that is partly entrained in 60 Hz line noise. The overall sparsity of the representation was an order of magnitude less than with the spike bases, with approximately 10% of the coefficients active at any time.

To learn this basis, samples were selected at random from a full recording penetration. The sparsity penalty was initially set to a low value to let all basis elements, which were initialized randomly, to learn structure. The sparsity was then gradually increased from this level during learning until the inferred coefficients had punctate activations in time. If sparsity was too low, the basis functions, particularly the lower frequency ones, were smeared in time making the structure in the basis functions difficult to interpret. The 60 Hz line noise was particularly troublesome and did not separate clearly from low-frequency structure.

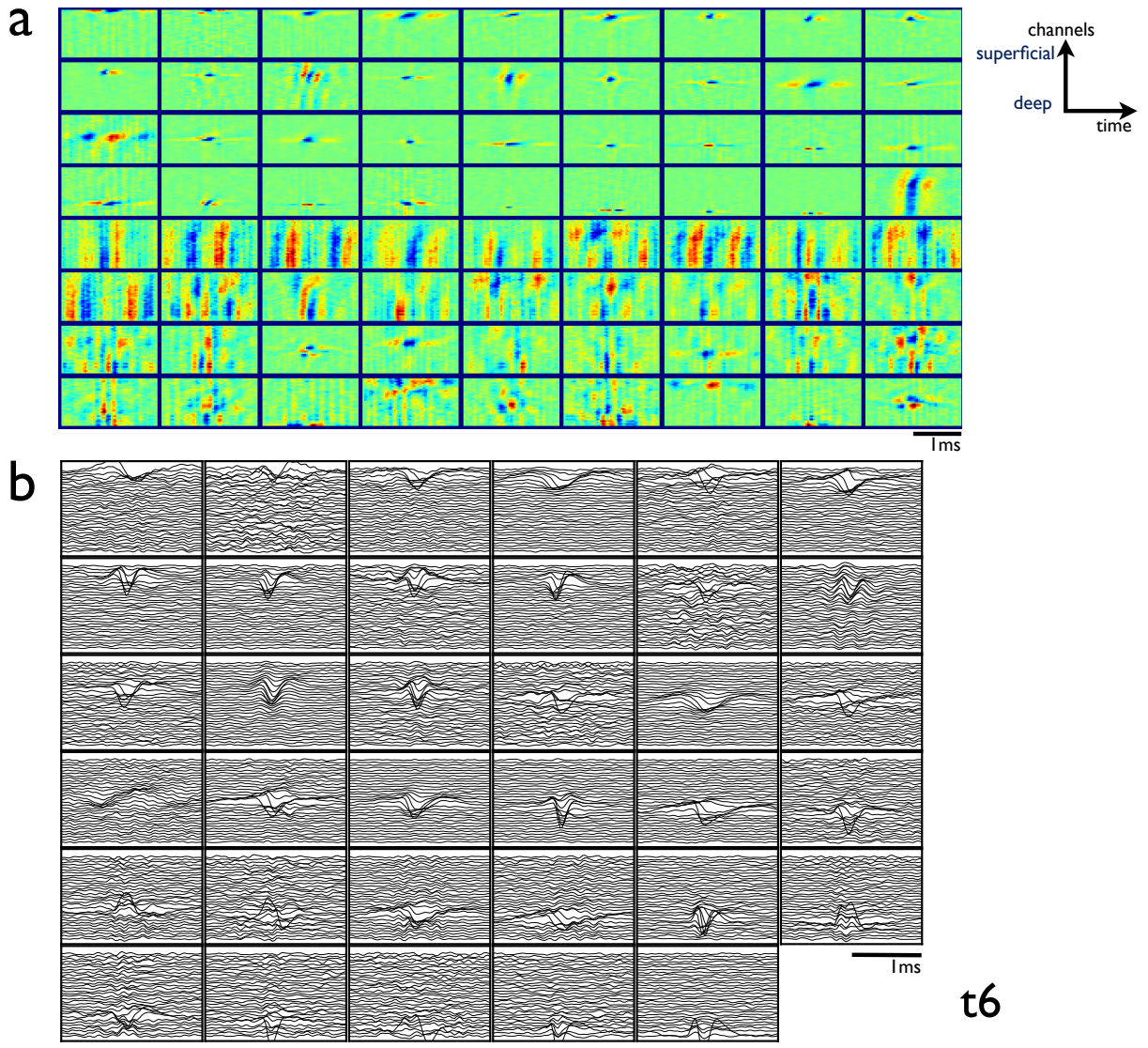

Figure 2.5: **Learned spike basis. (a)** A convolutional sparse coding basis learned from high-pass filtered data from one recording penetration. Each subplot consists of a single basis function with polytrode channels on the vertical axis oriented with superficial layers at top and 2 ms of time on the horizontal axis. The basis functions are plotted as correlation kernels. Plots are in normalized units, with reds indicating positive voltages and blues negative voltages. Several classes of basis functions are readily discernible, including spike-like waveforms, noise basis functions correlated across channels, as well as unresolved multi-unit activity. **(b)** A subset of basis functions in **(a)** that are localized in time and across channels redrawn as line plots for visualization, ordered from superficial to deep. These basis elements correlated well with spike events in the raw data.

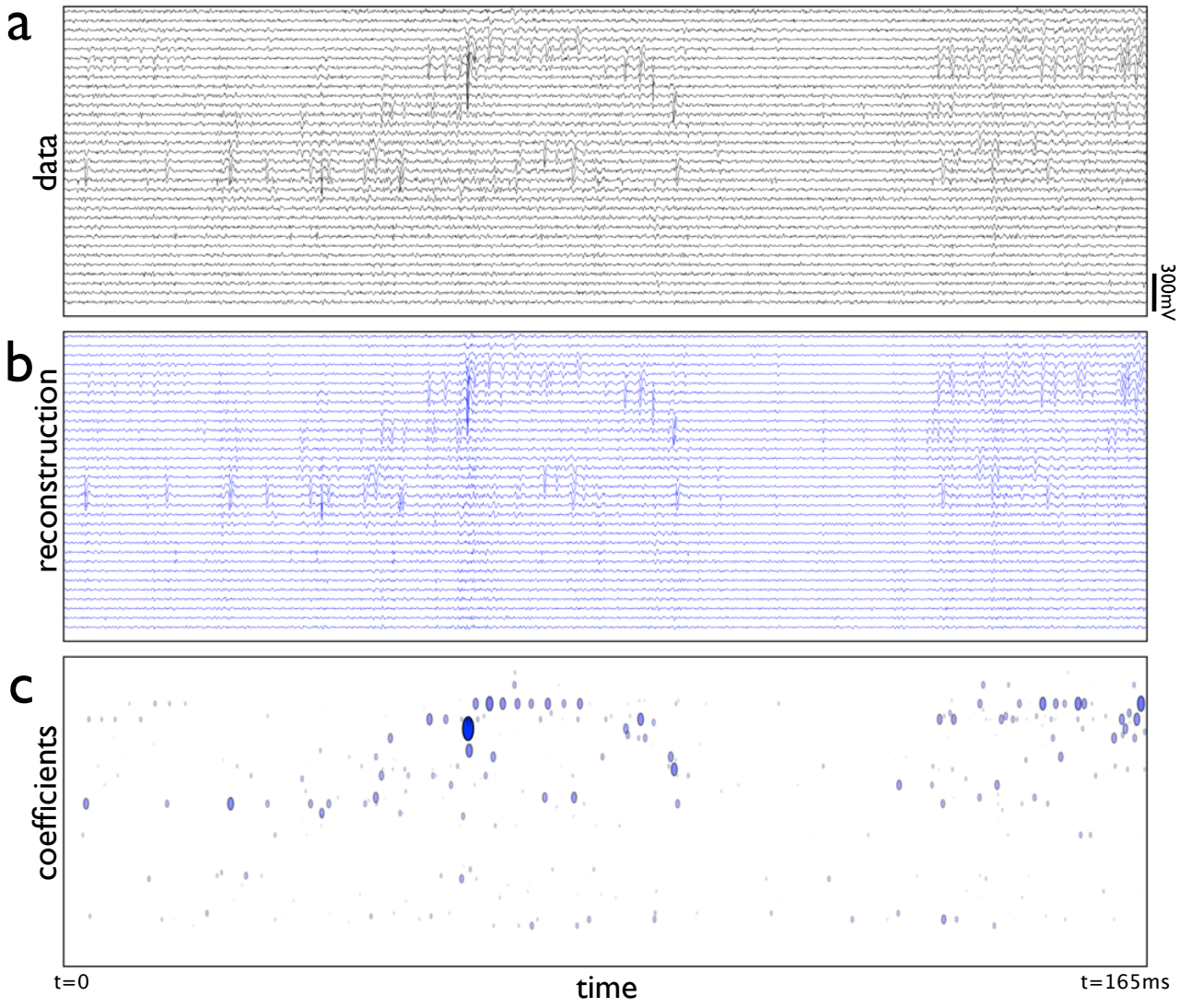

Figure 2.6: **Spike data and reconstruction. (a)** Sample high-pass filtered data, **(b)** reconstruction with basis shown in Fig. 2.5, and **(c)** sparse coefficients with log-magnitudes represented by size of ellipsoid. As the basis functions are sorted by lamina, the position of the coefficient activations correspond well to the spatial arrangement of structure in the real data. Bursting neurons are visible as a quick series of activations of the same coefficient. Neurons in close proximity tend to fire together, generating overlapping waveforms. The sparse coefficients are able to separate out these causes. Despite only .5% of the coefficients being active in this case, much of the structure of the data is represented.

This confound was addressed with modeled LFP data with a high degree of periodic noise in Sec. 2.16. To ensure the learned LFP basis captured true structure in the data, bases were learned from different portions of the data, initialized at random in different ways, and perturbed during the optimization to demonstrate that the learned structure was consistent in all cases.

After learning an LFP basis, inference was performed on the entire dataset. A sample of LFP data and its reconstruction with the learned basis is shown in Fig. 2.8. The LFP is relatively more poorly understood than spiking activity. Consequently, it is difficult to interpret the structure in the basis functions and coefficients based on known physiology. This will be discussed in detail in Chap. 4 where the distribution of spike and LFP coefficients are characterized together in a joint statistical model.

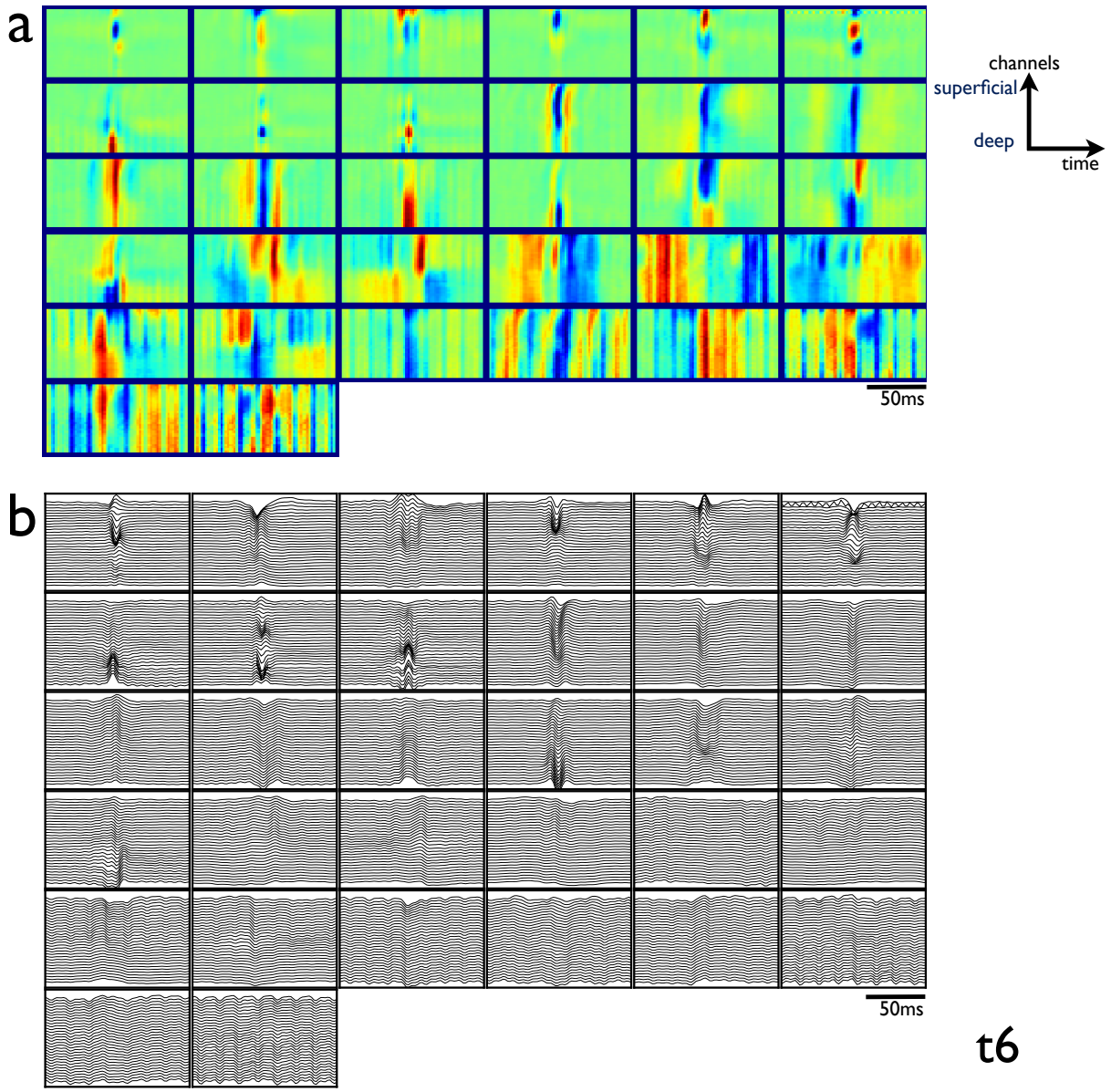

Figure 2.7: **Learned LFP basis. (a)** A convolutional sparse coding basis learned from low-pass filtered data from one recording penetration. Basis functions are displayed as in Fig. 2.5 with 128 ms on the time axis. Basis functions fall into a few general classes, including components that are localized in time with features divided across lamina, localized in time but correlated across lamina, with extend low frequency structure in time, as well as basis functions corresponding to 60 Hz line noise. **(b)** The same basis replotted as a line plot for visualization purposes.

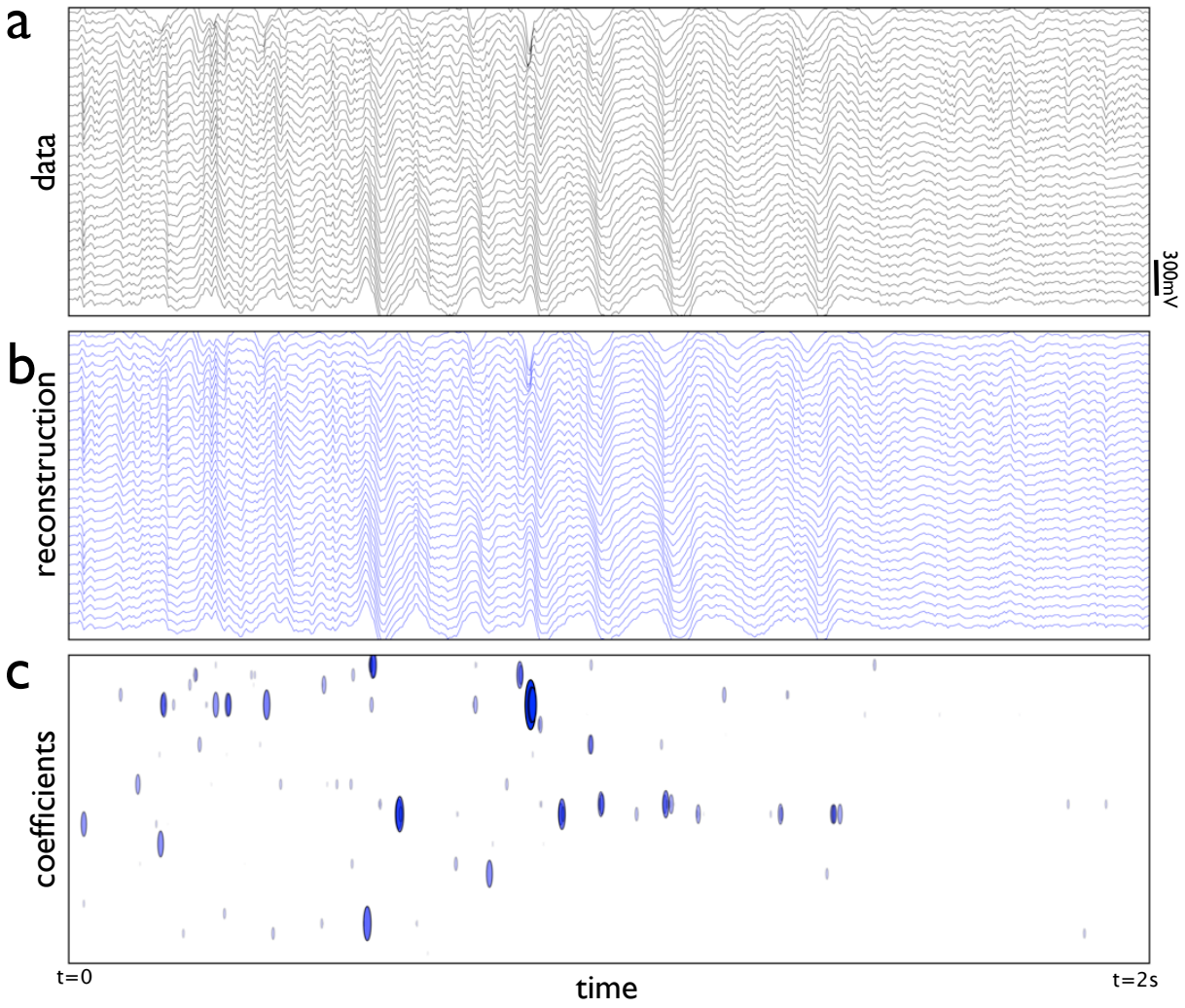

Figure 2.8: **LFP data and reconstruction. (a)** Sample low-pass filtered data, **(b)** reconstruction with basis shown in Fig. 2.7, and **(c)** sparse coefficients with magnitude represented by size of ellipsoid. Only 8% of the coefficients were active in this case.

2.3.2 Applications

2.3.2.1 Spike sorting

To understand the relationship between spike bases learned through sparse coding with actual spiking activity in the recordings, one penetration was spike sorted using a widely used cluster-based algorithm[48], which is referred to henceforth as 'manual' spike sorting. Spike events were detected by checking if any channel in a segment of data had passed a threshold criterion. The waveforms were extracted and a set of low-dimensional features were computed from them. These features included the first few PCA components of all waveforms, the peak-to-peak amplitude and duration of the waveforms as well as other characteristics that reflect the priors an experimentalist would use to categorize waveforms by visual inspection. The features were then clustered using a gaussian mixture model[13], where individual clusters correspond loosely to individual neurons with means and covariances representing the shape and variability of their waveforms in the low-dimensional feature space. As the mixture model performs poorly with increased dimensionality, it was necessary to block the channels on the polytrode into a set of 8 virtual, non-overlapping tetrodes which were sorted separately and merged in the end. The clusters were loaded into a graphical interface and manually modified, including merging clusters that appeared to be the same neuron based on criteria such as auto- and cross-correlograms and inter-spike interval histograms. Some waveforms were assigned to multi-units and others were discarded. The process took several weeks of work with several iterations of refinement and careful inspection by a team of people, including an expert electrophysiologist. Overlapping waveforms and bursting were particular troublesome and took a great deal of manual modification to account for shortcomings of the mixture model. In contrast, learning a spike basis and inferring coefficients for a full recording dataset took on the order of hours when using a moderately sized parallel cluster.

Mean waveforms for the manually sorted waveforms from a 20 minute segment of recorded data are shown in Fig. 2.9 alongside a sparse coding basis learned from the corresponding full recording session. The mean waveforms and basis elements have similar temporal and spatial extents with a characteristic tri-phasic structure. They are distributed across the electrode in a similar manner with a higher density of waveforms in the superficial layers and more localized waveforms in the deep layers. To assess the meaning of the sparse coding coefficients in a similar framework to manual sorting, coefficients for the spike-like bases elements were thresholded at 50% of their peak value and waveforms were extracted from the raw data. Sample waveforms for a subset of the basis functions are shown in Fig. 2.10, demonstrating that this simple strategy could be used as an effective pre-processing step for manual spike sorting. An example is shown in Fig. 2.11, where waveforms were extracted based on thresholded sparse coefficients and loaded into a freely available tool for processing electrophysiology data[49], where standard methods for post-processing can be employed.

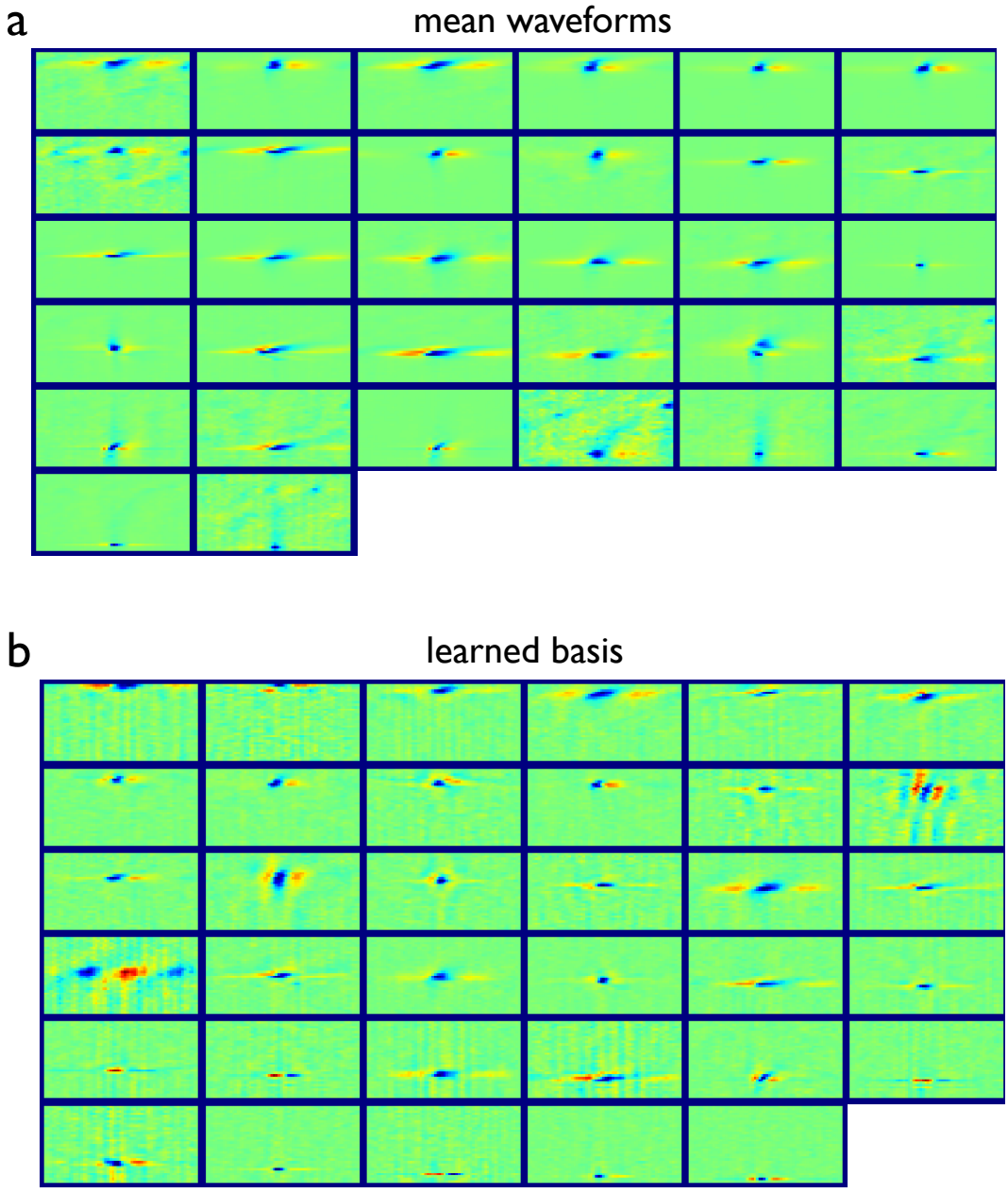

Figure 2.9: **Mean waveforms. (a)** Mean spike waveforms from 20 minutes of collected data using spike times from manually sorted data. **(b)** Spike basis functions learned from the same recording penetration for comparison. Waveforms are distributed across lamina, have similar temporal extent and structure. The spike basis functions include several waveform shapes not found in the manually clustered cells. Additionally, some waveforms in one set do not have a corresponding waveform in the other set.

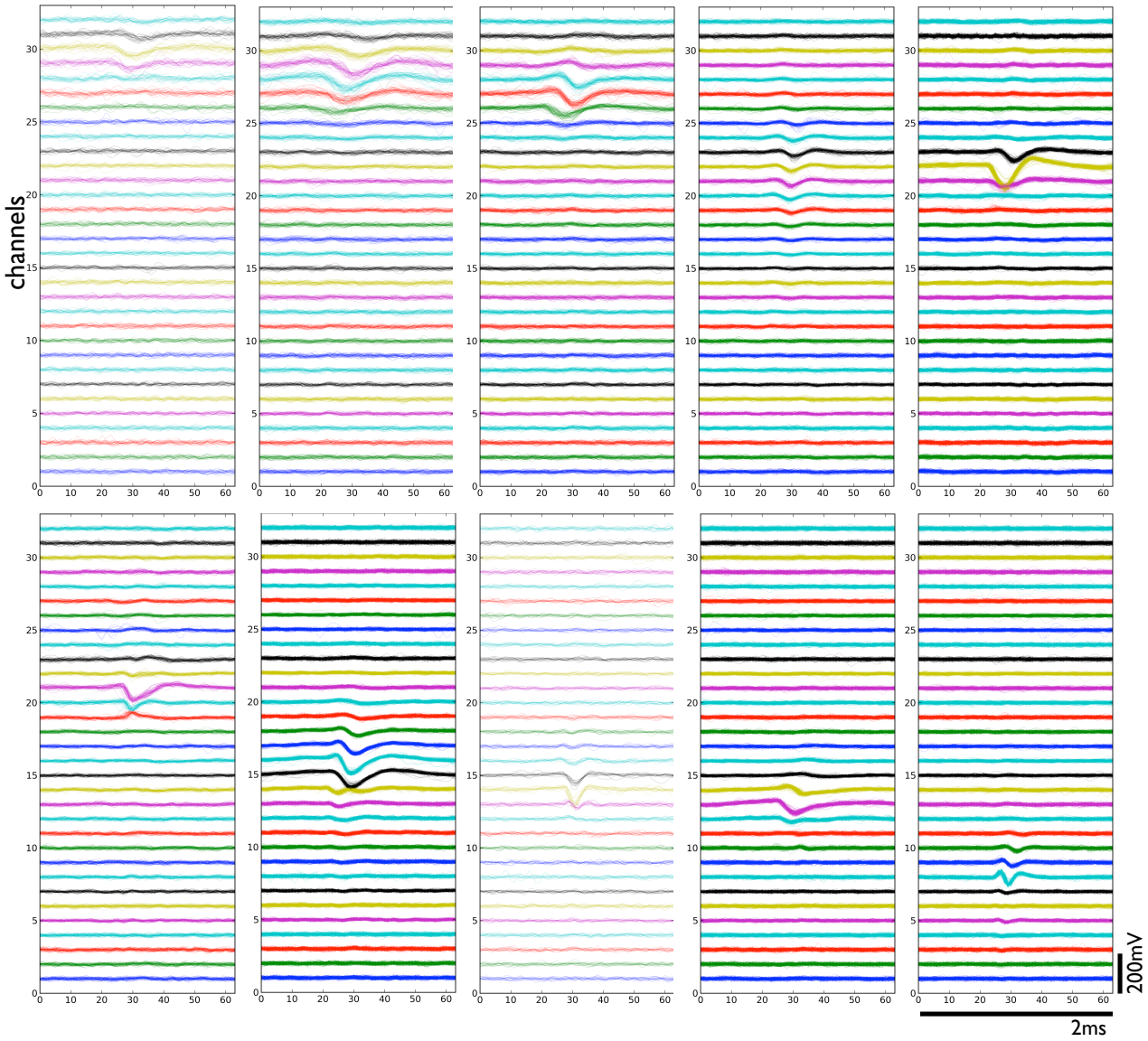

Figure 2.10: **Waveforms extracted using thresholded spike coefficients.** Spike coefficients from 30 s of data was thresholded at 50% of maximum for each coefficient. The waveform for each spike coefficient time was extracted from the raw data and the contribution from all other basis functions was subtracted. A representative sample of several sets of waveforms, each corresponding to a basis function, is plotted. Maximum number of waveforms shown per plot is 200, with some basis functions having fewer than that appearing lighter in color.

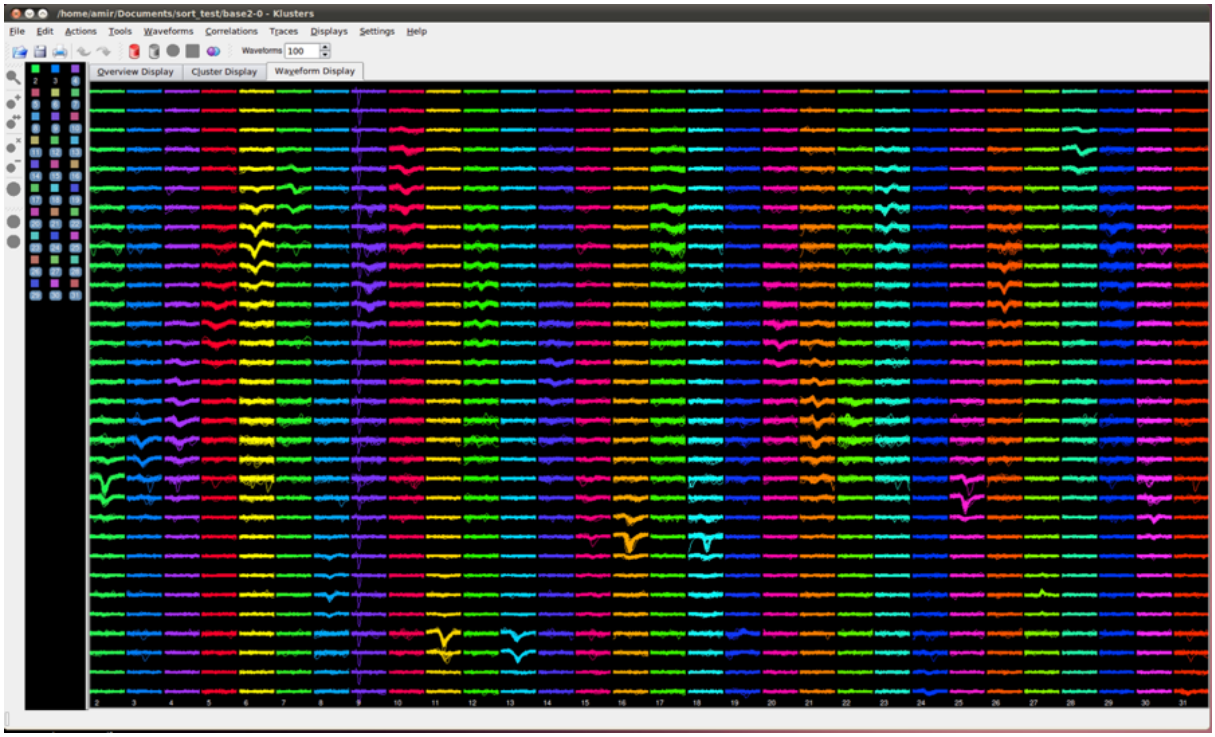

Figure 2.11: **Sparse coding as a preprocessing step for manual clustering.** Klusters[49] is part of a suite of free tools for processing neurophysiological data. Here, extracted waveforms identified from spike coefficients were loaded into Klusters for manual modification such as merging of clusters and removing multi-unit clusters. Displayed are 100 waveforms in columns for 31 spike basis functions.

To assess the correspondence between manually sorted spike times with sparse coefficients, cross-correlograms were computed for all pairs of waveforms and basis functions. Spike and coefficient times were pre-binned to the frame rate as this is the unit of time used in much of the analysis in Chap. 3 and 4. Pairs with normalized cross-correlogram peaks above a threshold with overlapping waveforms were identified. Four such pairs are shown in Fig. 2.12. Waveforms are similar in each example case with spike times and coefficient times agreeing to a high degree. However, many pairs do not show such a close correspondence with the most common confound being that several basis function coefficients were used to give the spike times of one manually sorted neuron. As the ground truth is not know in this case, it is difficult to ascertain which method is better, and this direction will not be pursued further. Instead a framework is developed for generating biophysically plausible model spike data where any methodology can be properly evaluated with known ground truth[33] (Sec. 2.3.3).

2.3.2.2 Applications using subsets of bases

Having learned a sparse coding basis and coefficients for an entire recording dataset, one useful application afforded by this new representation is to remove components of the data that represent noise artifacts or uninteresting structure. Fig. 2.13 shows a reconstruction of high-pass data where only the spike-like basis functions are retained and the remaining basis functions are discarded. The effect is to reduce the noise level in each channel. This illustrates one reason it is highly desirable to learn as much of the structure of data as possible, including noise artifacts, as it is then possible to remove them from the signal. The computational complexity of the reconstruction is trivial and does not require further optimizations. Another example of using a subset of components is given for the LFP in Fig. 2.14. One possible approach to understanding the structure of the LFP is to consider reconstructions with only certain classes of basis functions. In this example, only the basis functions localized in time and having fine structure separated by lamina were used. Subsequent analysis can then be applied to this reconstructed data which has drastically reduced complexity.

2.3.3 Sparse coding on model spike and LFP data

Polytrode data is statistically rich and convolutional sparse coding is an effective means for decomposing its complex structure into a more tractable representation. In order to evaluate sparse coding, particularly in an applied setting such as spike sorting, it is instructive to work with a biophysically faithful model where the mechanism for generating the data is completely known. Presented here is model spike data generated from compartmental models of neurons with know morphology using the LFPy framework[69]. Given a physical description of a recording device and electric currents in all compartments of the model

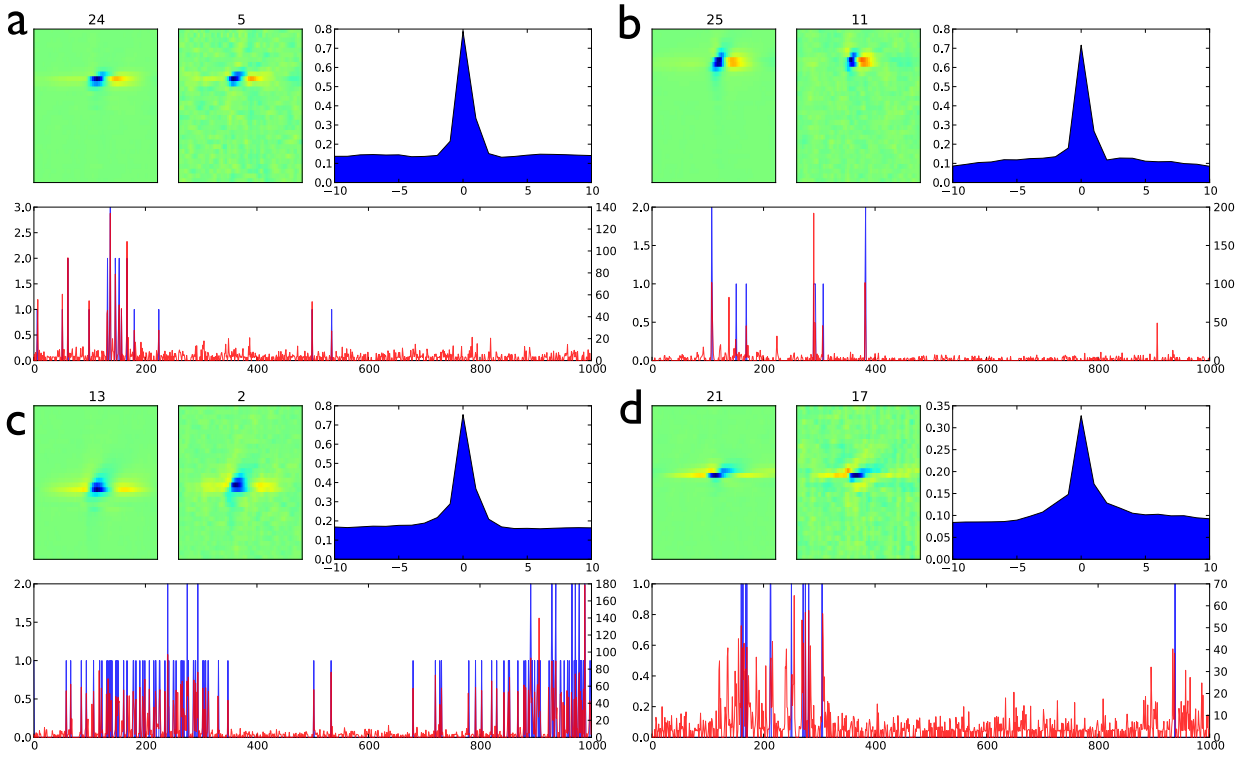

Figure 2.12: **Comparison of spike times and sparse coefficients.** To assess the correspondence between manual spike sorting and sparse coefficients, cross-correlograms were computed between spike times and coefficients between all pairs of manually sorted cells and sparse basis functions. The analysis was done at a temporal resolution of the frame rate of the stimuli (150 Hz) as this binning is used for all analysis in Chap. 3 and 4. **(a-d)** Shown are four sample candidates where normalized cross-correlations were above a heuristic threshold. Shown in each plot, starting clockwise at left, are average waveforms, basis functions, normalized cross-correlograms, and plot of binned spike times in blue and binned coefficient values in red (2nd y-axis). Time units are in the frame rate (6.7 ms). Though the correspondence is good for many pairs, a number of confounds exist that need to be addressed.

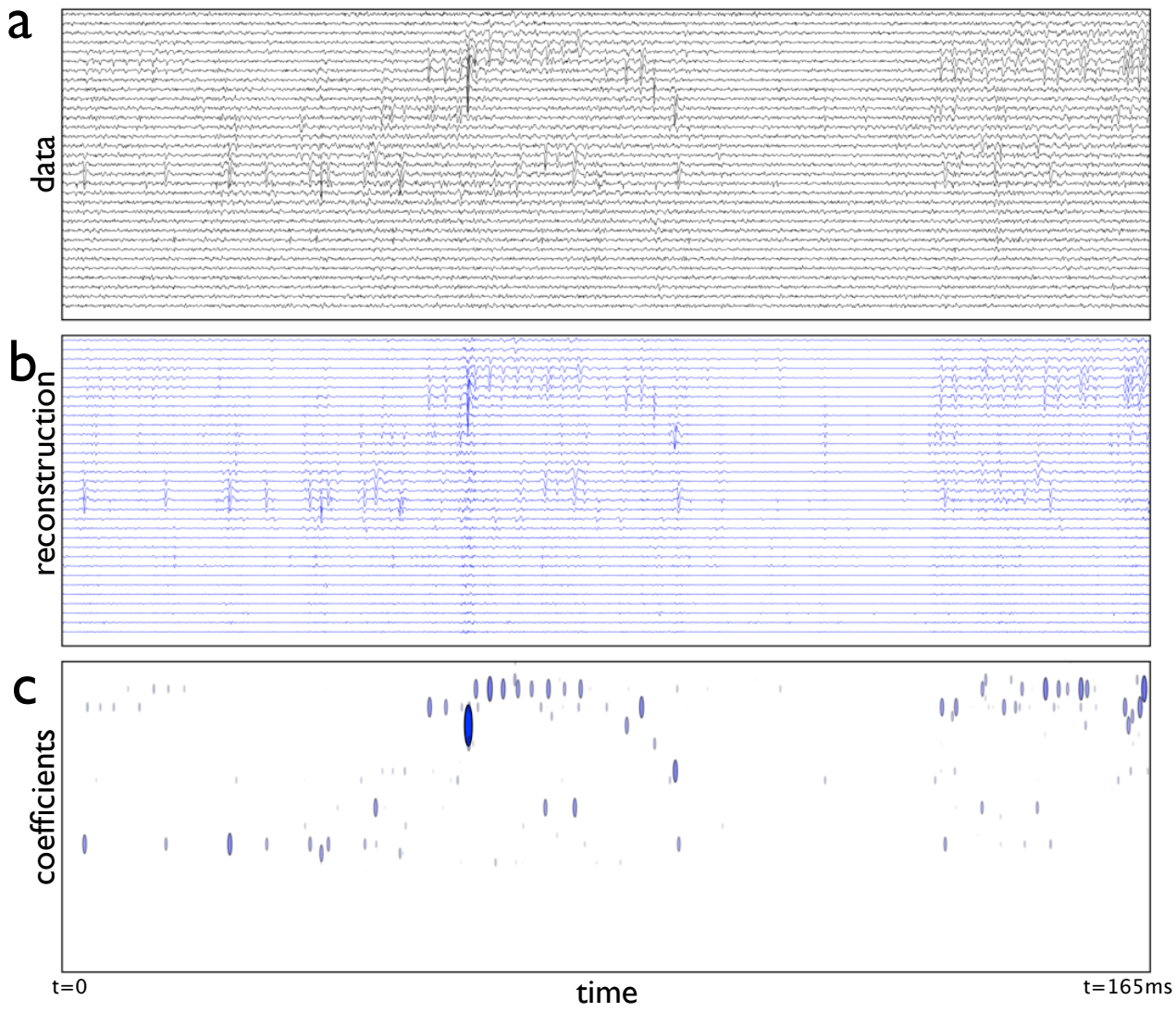

Figure 2.13: **Denoising spike data. (a,b)** Data in Fig. 2.6 reconstructed using a subset of basis functions (Fig. 2.9b) resembling spike waveforms. Such denoising is trivial to implement as it does not require an additional optimization. **(c)** Sparse coefficients for the subset of basis functions.

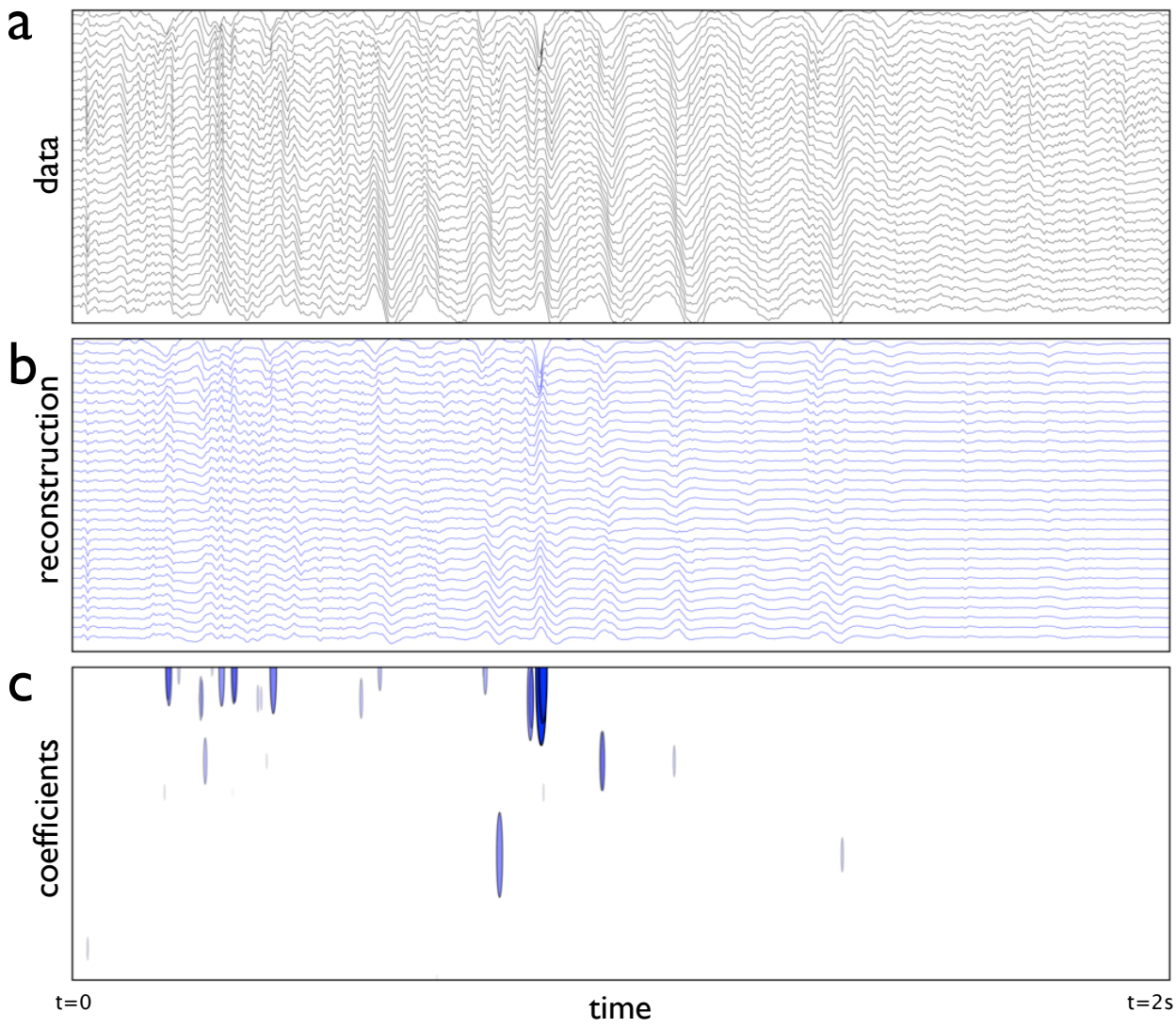

Figure 2.14: **Reduced representation of LFP data. (a,b)** Data in Fig. 2.8 reconstructed using a subset of basis functions with localized kernels and fine laminar structure (Fig. 2.7). Such a reduced representation of complex data can be used for further analyses or applications such as event detection. In this example, a large event in the superficial lamina with a specific structure is readily apparent in the reduced representation but is lost in the complexity of the original data. **(c)** Sparse coefficients for the reduced representation.

neurons, one can calculate the potential at a given time at each recording site using a line source approximation[53]. Twenty neurons were distributed randomly about a model 32-channel polytrode. Currents were injected with random arrival times into a set of 3 channels distributed along the dendritic trees. In this illustrative example, only a single neuron model was used and no noise was added. Results are displayed in Fig. 2.15. Spike waveforms were spatially localized to a greater degree than in real data, which made learning more difficult as knowledge of the geometry of the channels could not be exploited. The model data is useful for characterizing and accounting for several confounds in the representation if the intended application is spike sorting. These include waveforms being represented by more than one basis function (Fig. 2.15a-b) and waveforms with similar shapes but different amplitudes being represented with the same set of basis functions (Fig. 2.15c). Sparse coding parameters in this example could be tuned to eliminate these confounds, but with real data, such tuning can be difficult to perform in a systematic way. One strategy is to formulate more sophisticated sparse coding models to specifically address these shortcomings. Another approach is to account for confounds in processing the coefficients themselves. Model data provides a framework for evaluating both strategies.

Generating biophysically faithful model LFP data is considerably more difficult as the LFP represents the coordinated activity of a large population of neurons. Here, a simplistic model was used to examine the effects of line noise on the learning of LFP bases. The statistics of line noise is quite different from physiological data and is not well represented by the sparse coding model used here. To better understand how a high level of line noise impacts learning, model data was generated using a ground-truth basis of a set of spatio-temporal gabor functions with coefficients sampled from a Bernoulli distribution. Fig. 2.16a shows the ground truth basis. To this generated signal, increasing levels of a rectified sinusoidal line noise was added. Fig. 2.16b shows the results of learning a basis where the line noise is similar in relative magnitude to the line noise in real polytrode data. The basis functions are faithfully recovered. However, if the noise is increased to a degree where it dominates the signal, as with Fig. 2.16c, the basis functions begin to entrain some of the structure of the noise. However, they are still able to faithfully recover to a remarkable degree the structure of the ground truth basis.

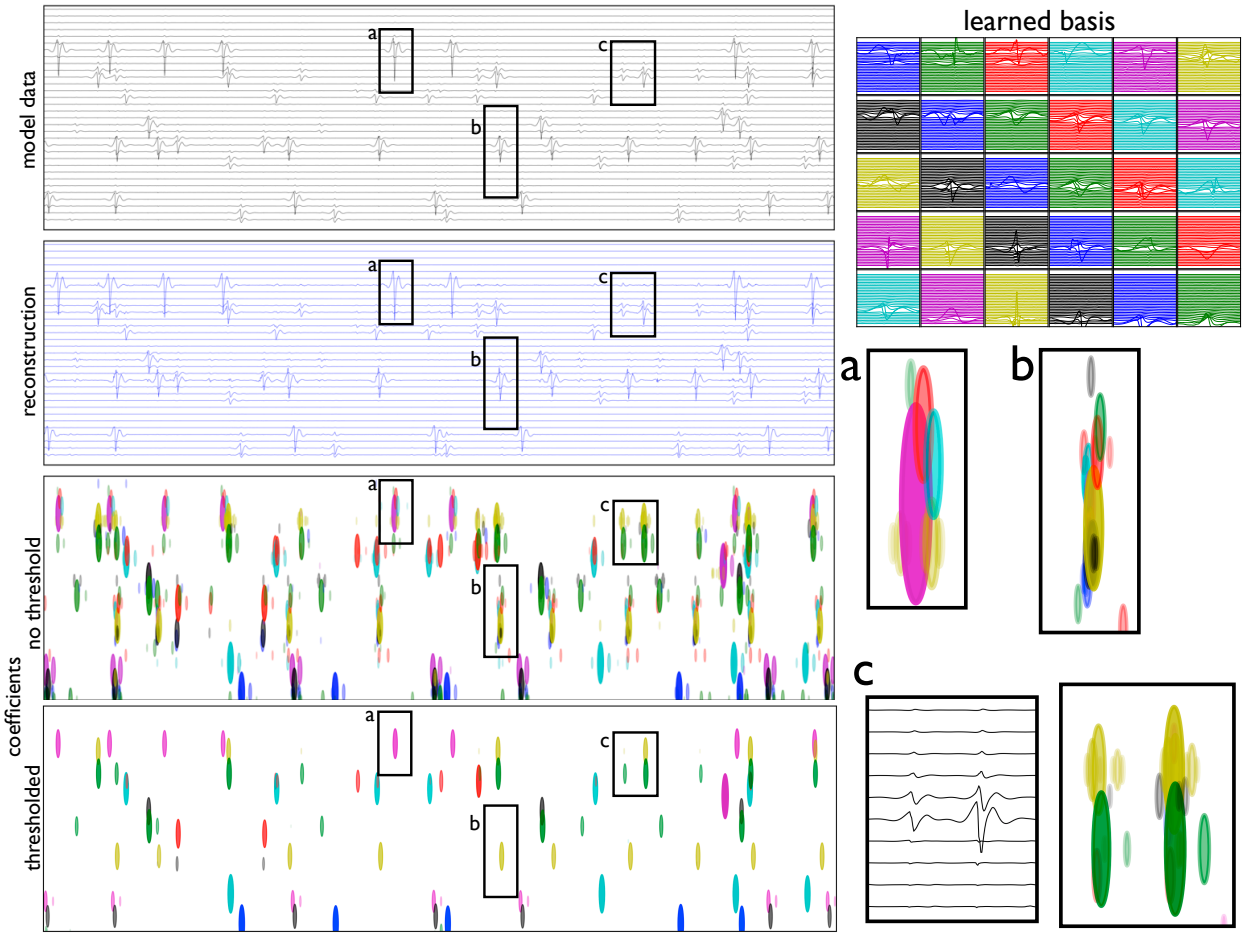

Figure 2.15: **Biophysically faithful model spike data.** To demonstrate several confounds to interpreting spike sparse coefficients as spike times when waveforms are localized to one or a few channels, model spike data was generated from 20 randomly placed pyramidal neurons, high-pass filtered as is done with real data, and learned a 30 component basis. Model data, reconstruction, learned basis, as well as coefficients with no threshold and coefficients thresholded at 50% of their maxima are displayed. Though the learning is robust to added gaussian noise, no noise was added in this example to aid visualization. The inset plots demonstrate 2 common confounds. Colors in the learned basis correspond to coefficient ellipsoid colors. **(a,b)** Several basis functions are learned and used to represent one spike waveform. Through a simple thresholding heuristic this confound is removed for **(a)**, but not for **(b)**. **(c)** The same set of basis functions are used to represent two spike waveforms of similar shape but differing amplitude. This can be addressed in a tool such as Klusters[49] (see Fig. 2.11).

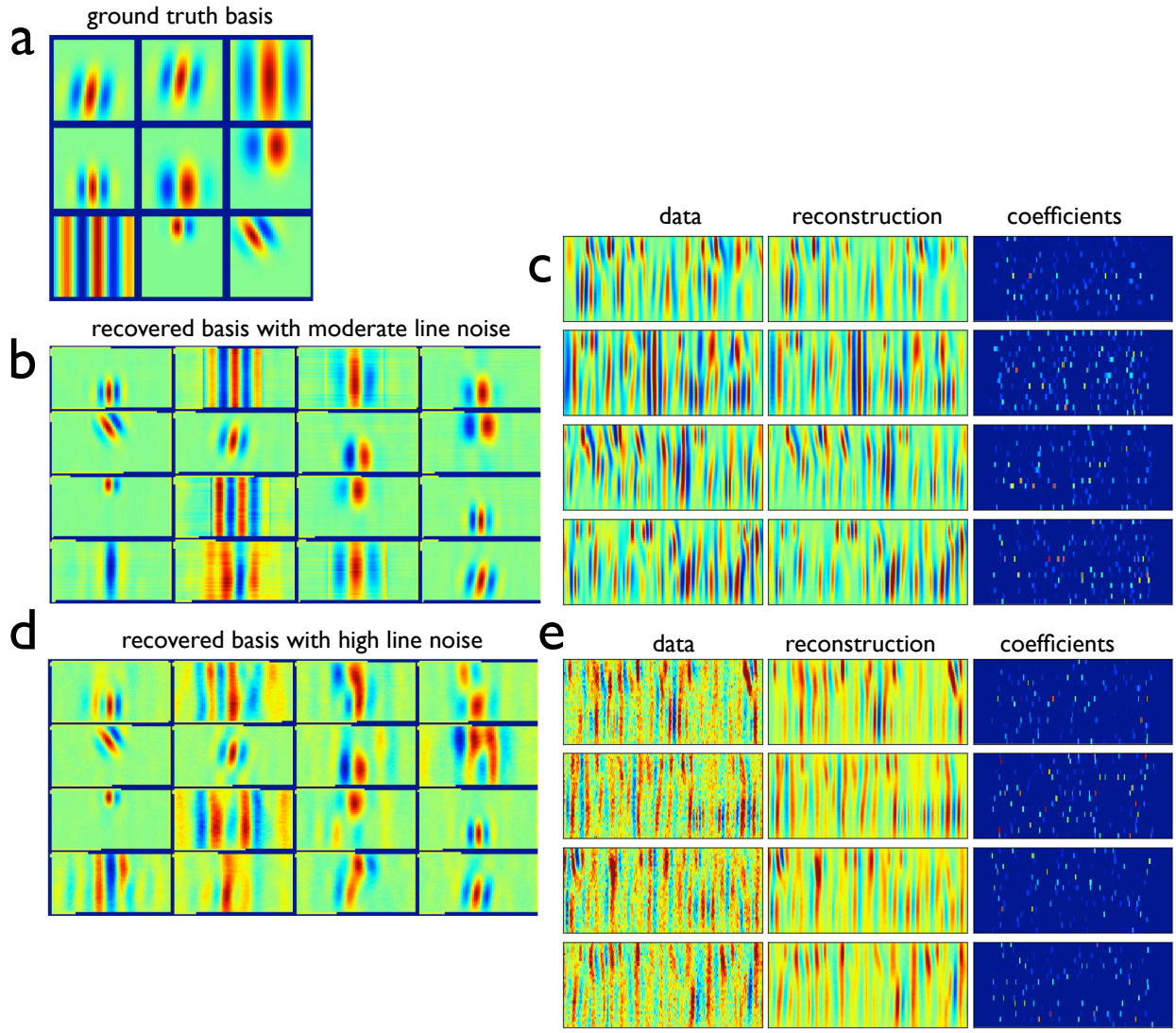

Figure 2.16: **Model LFP data with line noise and recovering ground-truth basis. (a)** A model basis consisting of a set of spatiotemporal gabors. **(b)** A basis learned from model data generated from the ground truth basis with Bernoulli sampled coefficients. A rectified sinusoid was added as model line noise. **(c)** Four model data samples, reconstructions, and inferred coefficients for test cases corresponding to the basis in **(b)**. **(d)** A basis learned from model data which is dominated by line noise as well as gaussian noise. Some basis elements are distorted but faithfully recover much of the basic structure in the ground truth basis. **(e)** Model samples, reconstructions, and coefficients for the case with high noise. Coefficients are plotted as temperature maps with zero values in blue.

2.4 Discussion

A novel application of a convolutional sparse coding algorithm to polytrode data recorded from visual cortex of anesthetized cats was presented. When applied to high-pass filtered polytrode data, the algorithm learns a set of spike waveform components along the length of the polytrode. These components and coefficients have close correspondence to waveforms and spike times extracted using a conventional spike sorting algorithm. When applied to the LFP, the learned representation consists of basis elements that have structure separated across lamina at different temporal scales. Together these methods provide a powerful set of new tools for exploratory analysis of polytrode data. In Chap. 3, this method is applied to all recorded data from several polytrode experiments and its utility is demonstrated in characterizing the laminar distribution of responses to a wide range of stimuli.

Sparse coding as applied to polytrode data provides several immediate applications. Given even the conservative settings of learning and inference of spike basis functions in this chapter, a 100-fold compression can be achieved. Sparse coefficients can be stored with basis functions as metadata in lieu of actual data, reducing hardware requirements and bandwidth needed to transmit data. Compression will become increasingly necessary as the number of polytrode recording contacts grow into the hundreds and thousands in the future. It may even need to be built into the head-stage and acquisition system to be run online. Sparse coding provides a fast way to denoise data for subsequent processing. Denoising involves reconstructing data while omitting a subset of basis functions, which does not require optimizations. Learning basis functions on different portions of the data can be used to identify non-stationarities, such as electrode drift, that can be addressed by performing learning and inference adaptively over a dataset.

The most promising application of sparse coding is either as a pre-processing step for manual spike sorting or as a stand-alone algorithm. Spike sorting is a highly underdetermined, difficult computational problem while also being a necessary processing step for all subsequent analysis of electrophysiology data. Existing methods, such as clustering algorithms based on mixtures of gaussians, extract events from data using heuristic thresholds and use feature spaces that are highly tailored and artificial. These methods have numerous deficiencies such as not being able to model overlapping waveforms, not modeling all the data but just extracted waveforms, and scaling poorly with increased dimensionality of the signal, all areas where sparse coding excels. However, before the methods in this chapter can be incorporated into a full spike sorter, it is essential to handle certain confounds discussed in Sec. 2.3.3, such as single waveforms being represented by multiple basis functions and a single basis function being used with different magnitudes to explain different waveforms of different neurons. Sparse coefficients are analog whereas action potentials are all-or-none events, posing a difficult issue of the interpretability of the magnitude of sparse coefficients, requiring subsequent thresholding that may be difficult to perform in a principled manner. The coefficients could be constrained to be binary, but such inference problems are typically

intractable, though convex relaxations may exist. An additional difficulty in interpreting coefficients is that sparse representations can be brittle. If data is perturbed in a trivial way, coefficients may change substantially. Generating biophysically realistic model data incorporating as much of the complexity of the real data is a promising strategy for evaluating spike sorting performance and will be the subject of future study.

Chapter 3

Population response across lamina

3.1 Introduction

In Chap. 2, an application of a sparse coding algorithm to polytrode recordings was used to discover a representation for the data where the underlying causal structure was made explicit. In the high frequency regime, the learned bases consisted of components that had close correspondence with extracellular spike waveforms. In the low frequency regime, the learned bases displayed structure both in time and across lamina. In the present chapter, this new representation is used to characterize neural population responses across lamina in anesthetized cat visual cortex to a variety of natural stimuli.

The primary motivation for the focus on natural movies is that they are much more ecologically relevant than artificial stimuli such as gratings and noise movies that are typically used in experiments (Sec. 1.2). Given the visual system's high degree of adaptation to the statistics of the environment, it seems implausible that it could be characterized adequately by presenting stimuli that are so foreign to it. Another important factor is the strong non-linearities in the system, which render a reductionist approach useless[85]. However, to take advantage of the enormous body of existing electrophysiology done with single neurons using artificial stimuli, a set of commonly used simple, controlled stimuli such as drifting gratings were included in the stimulus paradigm to allow for a comparison.

A summary of all recorded data used in this study is displayed in Fig. 3.1 in terms of mean spectral power of LFP with time. Particularly striking is the degree of non-stationarity in the time course of the spectra during a full recording session. Changes in the spectral distribution, possibly signifying physiological state changes, occur with transitions between stimuli as well as during a stimulus presentation. It is important to remain mindful of such confounds during exploratory data analysis as they may affect interpretation of variability in responses. Chap. 4 describes a statistical approach that can account for these changes of state.

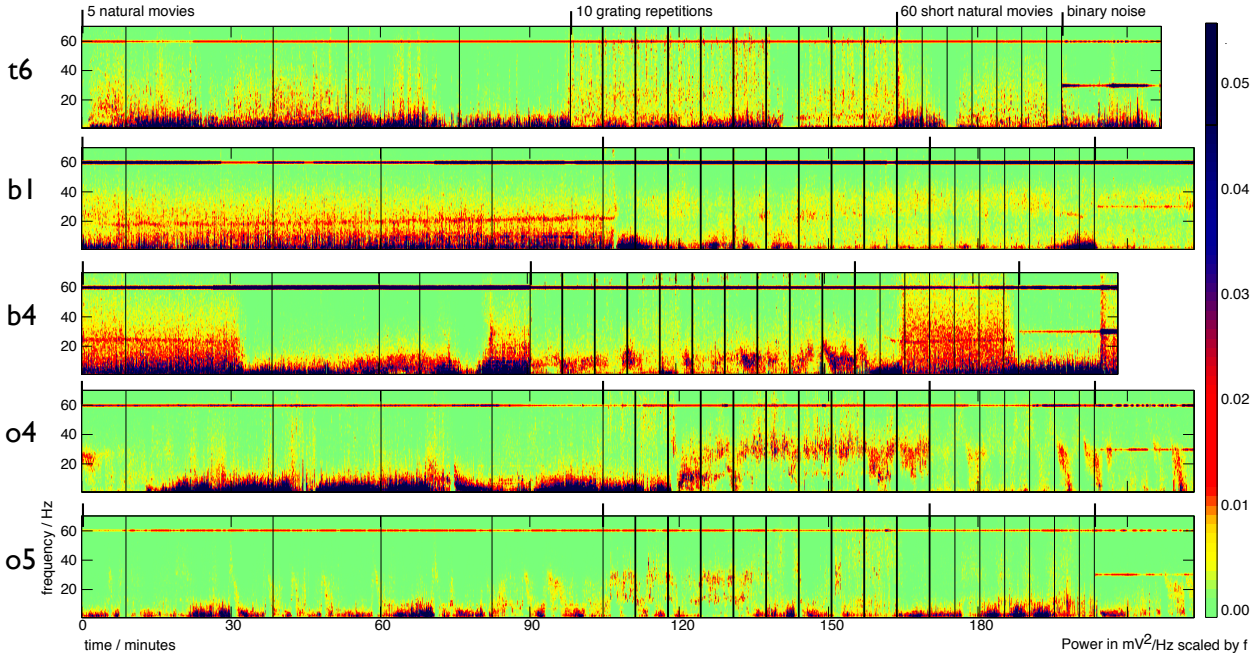

Figure 3.1: **LFP power across several recording sessions.** Power spectra of mean LFP with time across full recording sessions spanning all stimuli for 5 penetrations in 3 animals. Spectra were computed in 1 s time bins and are displayed scaled by frequency to better visualize high frequency features. The stimulus paradigm used consisted of a series of long natural movies, repeated drifting gratings, repeated short natural movies, full-field on-off movies, and binary white noise, played in succession. Some stimulus sessions were truncated for experimental reasons and account for the differing time lengths of the recordings. Many gross features of the recordings are readily visible including state changes between and within stimulus presentations as well as know sources of experimental artifacts such as 60 Hz line noise and 30 Hz locking to frame rate in the binary white noise recordings.

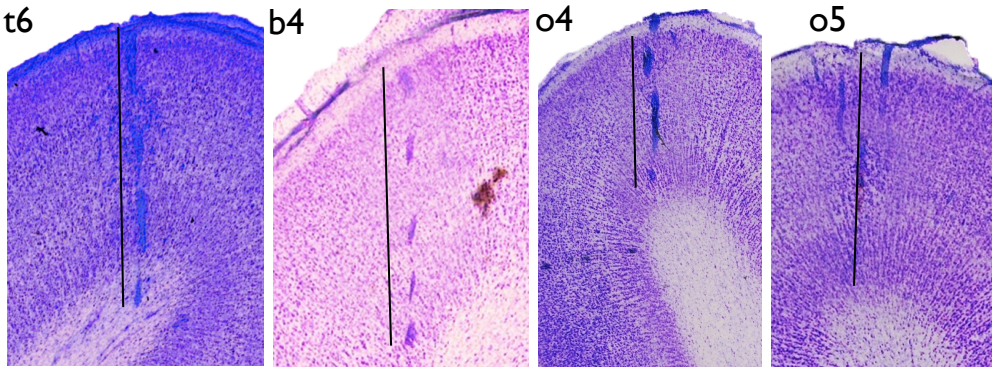

Figure 3.2: **Histology with polytrode track.** Nissl-stained coronal sections of four penetrations in three cats. In each case, several sections are overlaid to show the full extent of the electrode penetration. Labels **t6**, **b4**, **o4**, **o5** correspond to those in Fig. 3.1 and are used in all remaining figures throughout this section. No histological track was visible for session **b1**. In order to study laminar dependence of neural response, it is critical to determine the position of the polytrode with respect to the tissue. Histology was one of several methods used to determine position.

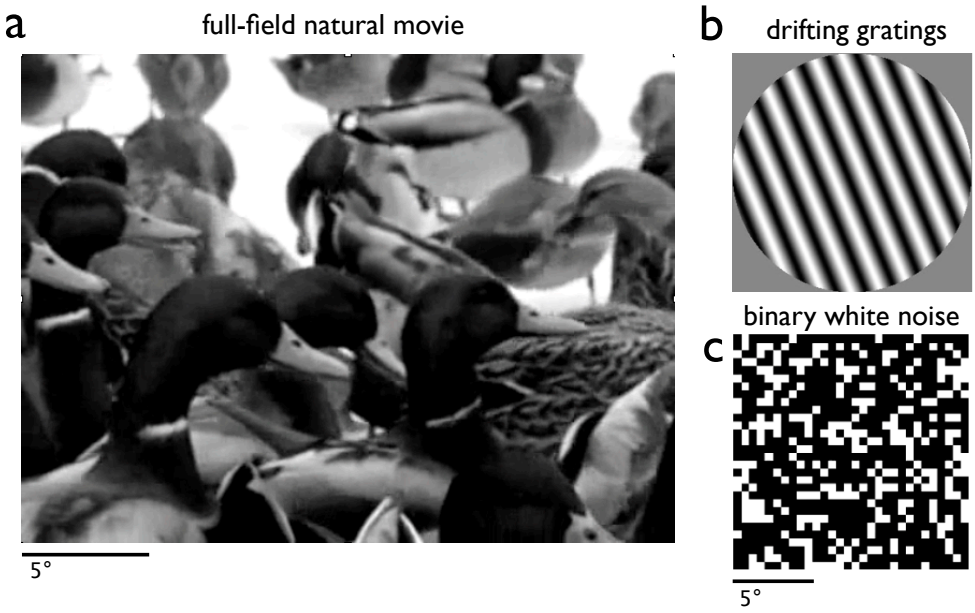

Figure 3.3: **Stimuli used in experiments.** Sample frames from stimuli used in experiments, including **(a)** long natural movies of ducks, **(b)** drifting sinusoidal gratings of different directions, orientations, spatial, and temporal frequencies, and **(c)** binary white noise. All stimuli covered a field of view of greater than 10 degrees.

3.2 Methods

3.2.1 Experimental methods

Polytrode recordings were made from anesthetized cat visual cortex. Surgical methods, recording procedure, and histological methods (Fig. 3.2) are detailed in Sec. 2.2.1.

3.2.2 Stimuli

For each recording session, a series of visual stimuli were played in succession. These included 3 long duration full-field natural movies, 2 long duration movies generated from static images scanned according to a fractal pattern, 10 repetitions of drifting sine wave gratings at various directions, spatial and temporal frequencies, 60 repetitions of a short natural movie, 90 repeats of a full-field on-off step stimulus, and a long duration binary white noise movie (Fig. 3.3). Hereafter, the natural movies, Hilbert movies, and short natural movie repeats will be grouped as 'natural' stimuli while the remaining stimuli will be described as 'artificial'. The combined duration of all movies was approximately four hours.

3.2.2.1 Natural stimuli

Three movies of a controlled natural setting were filmed with a Casio EXILIM Pro EX-F1 digital camera, equipped with a 7.20 mm $\times$ 5.35 mm (1/1.8 inch) CMOS image sensor with 6.8 million pixels. The camera was set for 21 mm focal length (corresponding to 105 mm of 35 mm film equivalent focal length), for a field of view of approximately 15 by 20 degrees and an aperture of $f/3.4$. The field of view was chosen to match the visual angle at which the movies were later presented on the monitor.

Movies were recorded at 512×384 pixel resolution at 300 frames per second and encoded using the H.264 codec at an 8.8mbps data rate. The camera was stationary with constant focus and was not panned or zoomed for the duration of each movie. Because there were no provisions to override automatic gain control on the camera, care was taken to keep the illumination of the scene constant over the duration of movie recording so that no artifacts would be introduced due to gain changes. Natural scenes were selected to contain full-field motion. The high frame rate was chosen to prevent temporal aliasing for playback at various non-standard frame rates such as 100Hz, 150Hz and 200Hz.

The three movies, hereafter referred to as Duck8, Duck30, and Cat20, were of lengths 8.9, 29.6, and 21.7 minutes in duration, respectively. Duck8 depicts an outdoor scene of a collection of ducks eating hand-thrown pieces of bread, Duck30 depicts a scene of ducks swimming in a pond, while Cat20 depicts an indoor scene of several cats feeding and playing.

Several reasons motivated the creation of a new natural movie dataset for this study. Most importantly, recording movies at 300 fps allowed presentation at a frame rate of 150 Hz with one refresh per frame, which is the practical limit on a conventional CRT screen. Even faster

presentation would be possible using this dataset and specialized hardware. This ensures that the movies captured as much of the temporal power spectrum of natural visual stimuli as possible. While it may seem that the visual system should not respond to such high temporal frequencies, it has been shown, for example, that neurons in LGN and striate cortex can phase lock to the screen refresh at a rate of 60 Hz[109,110], an order of magnitude larger than the classical preferred temporal frequencies of visual cortical neurons. Anecdotally, human subjects are able to distinguish cinematic film at 24 fps from video at 60 fps, reporting a different 'feel' for footage at different frame rates without being able to identify the frame rate differences as the reason[2]. It was important to avoid these known effects of the frame rate on the visual system by recording and presenting movies at as high a frame rate as was feasible.

The natural movies used in this study had other advantages over conventional consumer video, which have been used in numerous previous studies. No scene changes interrupt the flow of the movie. These 'cuts' are characterized by a sudden change in luminance and other low order movie statistics, which have a disruptive effect on the LFP, generating evoked potentials and potentially interrupting ongoing oscillatory activity. The new movies contain no camera motion such as panning, which introduces a dominant spatio-temporal frequency in the movie. Due to the high depth of field of the camera used, there was no blur and associated lack of high spatial frequencies. Finally, the high degree of self-motion in the scenes chosen, to some degree, compensated for the lack of saccadic eye movements in the anesthetized preparation and prevented a drop in neural response following adaptation to a stationary stimulus.

For the presentation system consisting of a Nokia 21" CRT monitor running at a resolution of 640×480 and a refresh rate of 150Hz, the camera movies were processed as follows: movies were temporally down-sampled from 300 to 150 Hz, converted to 8-bit gray-scale, smoothed with a spatial gaussian filter to reduce blocking artifacts due to the high compression ratio of the camera, and transcoded from the H.264 codec to DivX to be playable without frame drops through Quicktime^{TM}-based presentation software. An additional filtering step was used to reduce the accumulated quantization artifacts of the conversion process and to smooth out high frequencies. All processing was done with a combination of the open-source transcoders MEncoder[76], FFMpeg[36], and HandBrake[47].

3.2.2.2 Hilbert movies

Two separate Hilbert movies were generated, approximately 22 minutes in duration each, by scanning over regions of a single natural image. Each movie was based on a 512×512 pixel gray-scale version of a single image. The images were the standard computer vision test image Lena and a picture of a log against a background of rocks used in Olshausen and Field[82]. The movies were produced frame by frame by tracing, on each step, a 256×256 pixel aperture along a space-filling path known as a Hilbert curve[51] across the entire image

(Fig. 3.4). This stimulus was inspired by efforts of Otto Creutzfeldt[25] more than 30 years ago. In a series of experiments, he investigated 'transfer' properties of cells by scanning a picture, line by line, over the receptive field of recorded neurons. A single natural image contains a wealth of high-order statistical features, including edges of different spatial frequency and orientation, textures, and occlusive surfaces with complex figure and ground relationships. Scanning over such an image is likely to span a far richer and more ecologically relevant stimulus space than artificial stimuli such as a series of drifting gratings. In contrast to a continuous natural movie, this stimulus allows for a succinct representation of a neuron's preferred stimulus. A neuron's firing rate, for example, can be plotted along the coordinates of the curve to show the relationship of its response with features in the image.

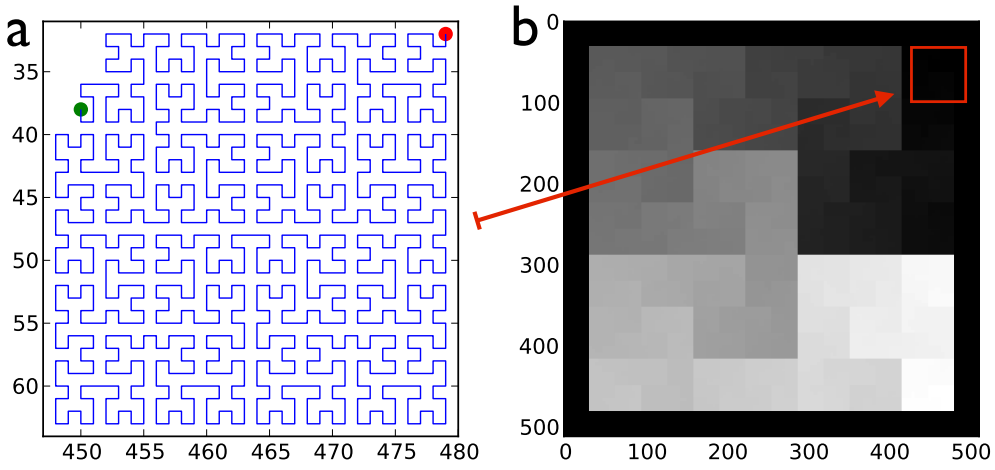

Figure 3.4: **Hilbert path movies.** The Hilbert movies scanned a natural image along a space-filling curve starting at the upper right of the image and ending at the bottom right after visiting each pixel once. **(a)** The first 200 steps of the Hilbert curve with red dot denoting start of path at the upper right of the movie and green denoting the current 200th point. **(b)** A gray scale image of the path over the whole movie with a time ordering of dark to light. The red box corresponds to the region covered in **(a)**.

3.2.2.3 Gratings

Drifting gratings with a circular aperture and a size of 300 × 300 pixels (approximately 15 × 15 degrees) were generated in a pseudorandom order at 100% contrast for 2 seconds each, with 16 orientations, 4 spatial (0.15, 0.3, 6, 1.2 cyc/deg), and 3 temporal (0.5, 1, 2 Hz) frequencies, with a total of 192 gratings in all. This series was saved as an MPEG encoded movie (XviD codec) and was played back 10 times during recording.

3.2.2.4 Short natural movies

A 30 second subset of the Duck8 movie was selected for repeated presentation and saved as a separate movie, with processing identical to the long natural movies.

3.2.2.5 Full-field on-off step movies

To identify laminar depth of the recordings from a current source density (CSD) analysis[75], we showed 90 repetitions of a full-field stimulus switching between black and white at a rate of 1 Hz.

3.2.2.6 Binary white noise

Finally, we included a receptive field mapping stimulus consisting of a 30×30 grid of black and white pixels subtending 15×15 degrees of visual angle. The stimulus was refreshed at 30 Hz (5 refreshes per frame) with each pixel being individually chosen to be black or white with a chance of 50%. The duration of this stimulus was 20 minutes.

3.2.3 Stimulus presentation

Stimuli were presented using the Psychophysics Toolbox[18](PTB) at a frame rate of 150 Hz with one frame per refresh on a Nokia 445XiPlus CRT screen. The maximum brightness of the screen was set to 100 cd/m^2, and the response of the monitor was adjusted to deviate less than 10% from a linear response over the dynamic range of the stimulus. Ambient light in the room was kept below 1 cd/m^2 as measured on the screen. The distance from the screen to the preparations was 57 cm, chosen such that 1 cm on the screen corresponded to 1° of visual angle. Stimuli were verified to be frame accurate by acquiring vertical sync signals using a photodiode as well as using the internal timing diagnostics offered by PTB.

3.2.4 Spatiotemporal whitening

Natural movies have characteristic power-law spectra[30,38]. To compute sparse coefficient-triggered averages for long natural movie sessions, it was necessary to compensate by flattening the spectra such that the power-law statistics did not dominate the analysis. The natural movies were first spatially gaussian filtered and down-sampled either 8 or 16 times. They were then convolved with an acausal, broadband zero-phase whitening filter[3,82], which was constructed in several steps. An average spatio-temporal power spectrum was computed from small blocks of the movie. The spectrum was then radially averaged in space, inverted and multiplied by a fourier-domain lowpass filter to attenuate the highest spatial and temporal frequencies. The filter was then applied to the whole natural movie using an overlap-add method[96] in the fourier domain, implemented using the FFT libraries for the

NVidia GPU for speed. As the filter was acausal, this introduced acausal artifacts into the coefficient-triggered averages of time length of about 6 frames, corresponding to the support of the temporal impulse response of the whitening filter.

3.2.5 Sparse coefficients

For each polytrode penetration, a single convolutional sparse coding basis (2.2.5) was learned for each of the spike and LFP datasets using data from the entire penetration. These bases were then used to infer coefficients on the full spike or LFP dataset using a parallelized block algorithm (2.2.6.1) that scaled approximately linearly in the number of parallel nodes used. Computation time in practice was on the same order as recording time when using a cluster of 8-16 nodes. Polytrode recordings, as with all electrophysiology, are susceptible to non-stationarity that is difficult to account for, such as electrode drift and changes in the health of the tissue. The penetrations used in this study were relatively stable. However, if polytrode drift was suspected, the basis and coefficients could be learned and inferred adaptively in time with gradual changes in the kernels accounting for some of this non-stationarity.

3.2.6 Polytrode position

As it was critically important for characterizing responses in specific cortical layers, we determined the position of the polytrode with respect to the lamina using a combination of strategies. The polytrode channels span 1.55 mm, the approximate extent of a cortical column in cat visual cortex. Despite best efforts to insert the polytrode perpendicularly to the cortical surface, some variation was inevitable due to the curvature of cortex in area 17. Penetrations were made gradually, during which short recording snapshots were made for computing sparse convolutional spike bases. The learned bases helped to assess the position of single units along the probe as well as trace their signal as the probe was moved downwards. When stable units were observed along the length of the probe, the main recordings were then begun. This ensured that the probe was deep enough but not yet in white matter. At the end of recording, lesions were made in the top and bottom polytrode channels. However, these lesions did not appear in the histology.

Polytrode tracks were visible in histological Nissl-stained sections made from the recording area (Fig. 2.1). As the coronal blocks used for sectioning were not perfectly parallel to the polytrode, tracks appeared on multiple sections requiring them to be overlaid to reconstruct a full track. The histology confirmed if the polytrode was inserted perpendicularly, and whether it spanned a single micro-column or crossed several columns. Due to shrinkage of tissue in the histological preparation as well as the angle of the probe relative to the sections, the precise location of the probe could not be determined. This was addressed by using a repeated on-off full-field step stimulus to calculate current source density (CSD)[75,92].

Furthermore, spiking response to this stimulus allowed precise determination of layer 4 and, in some cases, layer 6.

3.3 Results

3.3.1 Responses to on-off step stimuli

To determine the precise location of layer 4 along the polytrode, we plotted rasters of the first 75 ms of spike coefficient response to changes in an on-off full-field step stimulus. Fig. 3.5 shows the onset of a spatially localized spiking response at approximately 30 ms in the upper channels of the polytrode in layer 4[75]. This is followed in some cases by second and third peaks, occurring at different times for the two different step stimuli. More data would be needed to establish if these interesting timing differences occur consistently and can be accounted for by known differences between the ON and OFF pathways, for example. ON bipolar cells in the retina use a second-messenger to invert polarity whereas OFF bipolar cells use ionotropic receptors with faster kinetics. These different mechanisms result in a response timing difference that is passed through to visual cortex[59]. Fig. 3.6 shows spike coefficient rasters for a different penetration in which the polytrode was inserted at a slight angle. Again the response signature is localized to a narrow set of channels, presumably layer 4. A small signature is also seen in the deep layers, which may correspond to layer 6, which is known to have Y- and X-type synaptic inputs from LGN[75]. Fig. 3.7 shows the trial-averaged potentials as well as the CSD computed for both penetrations. The prominent and spatially localized current sink in the first example corresponds to layer 4 and agrees well with the spike coefficient rasters. The CSD tangential penetration is more difficult to interpret, but the current sink agrees well with the position of the spiking response.

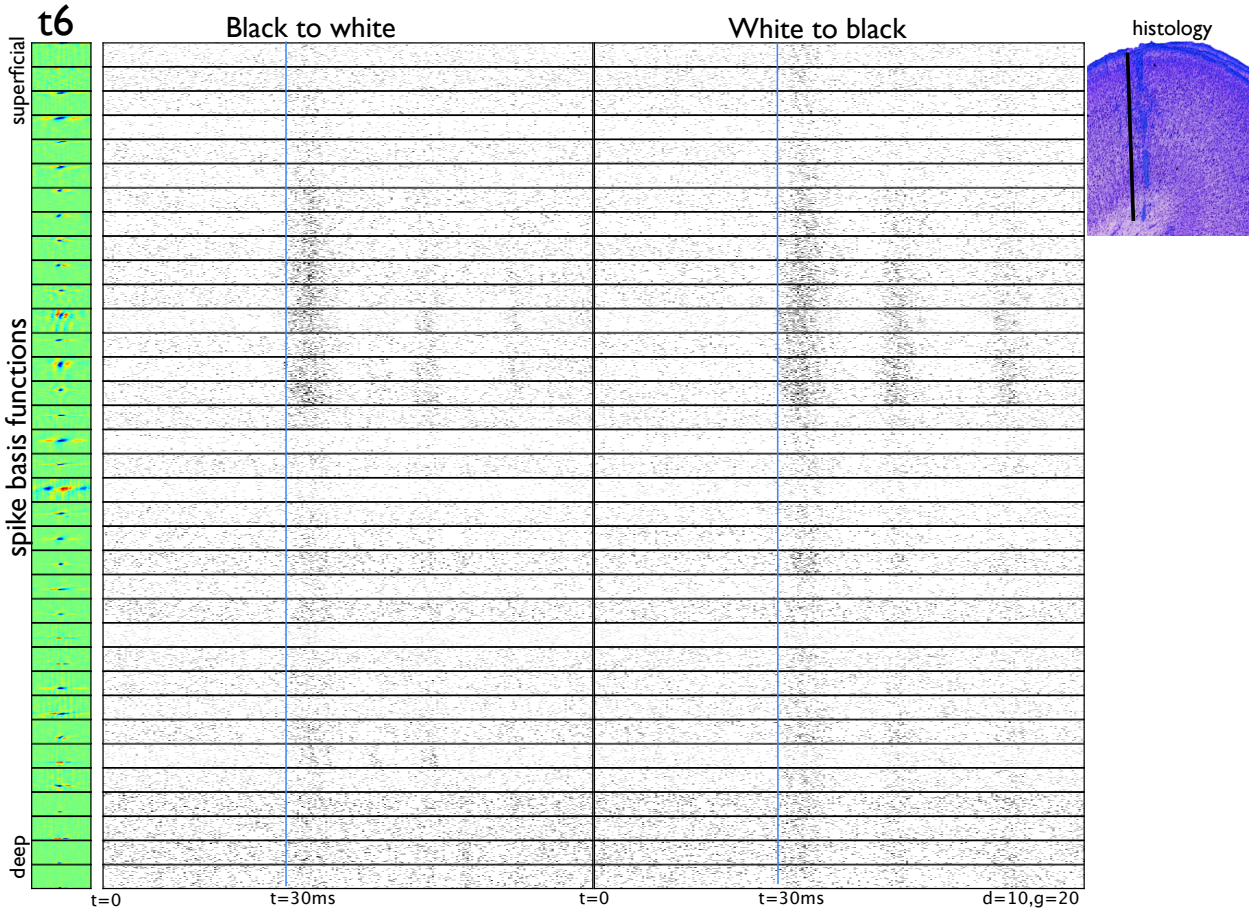

Figure 3.5: **Response to on-off step stimuli.** Spike coefficient rasters for 90 repeats of a 1 second black-to-white, 1 second white-to-black full-field movie. The first 75 ms of each repeat is shown here. The spike basis functions sorted across lamina corresponding to each set of rasters is shown at left in green. Histology for this penetration is shown at top right with position of probe indicated by a black bar to its immediate left. The onset time of response is approximately 30 ms, with the majority of the firing initiated in the middle layers, presumably layer 4. A number of additional peaks are visible after the first response with different lag times for the two stimulus cases.

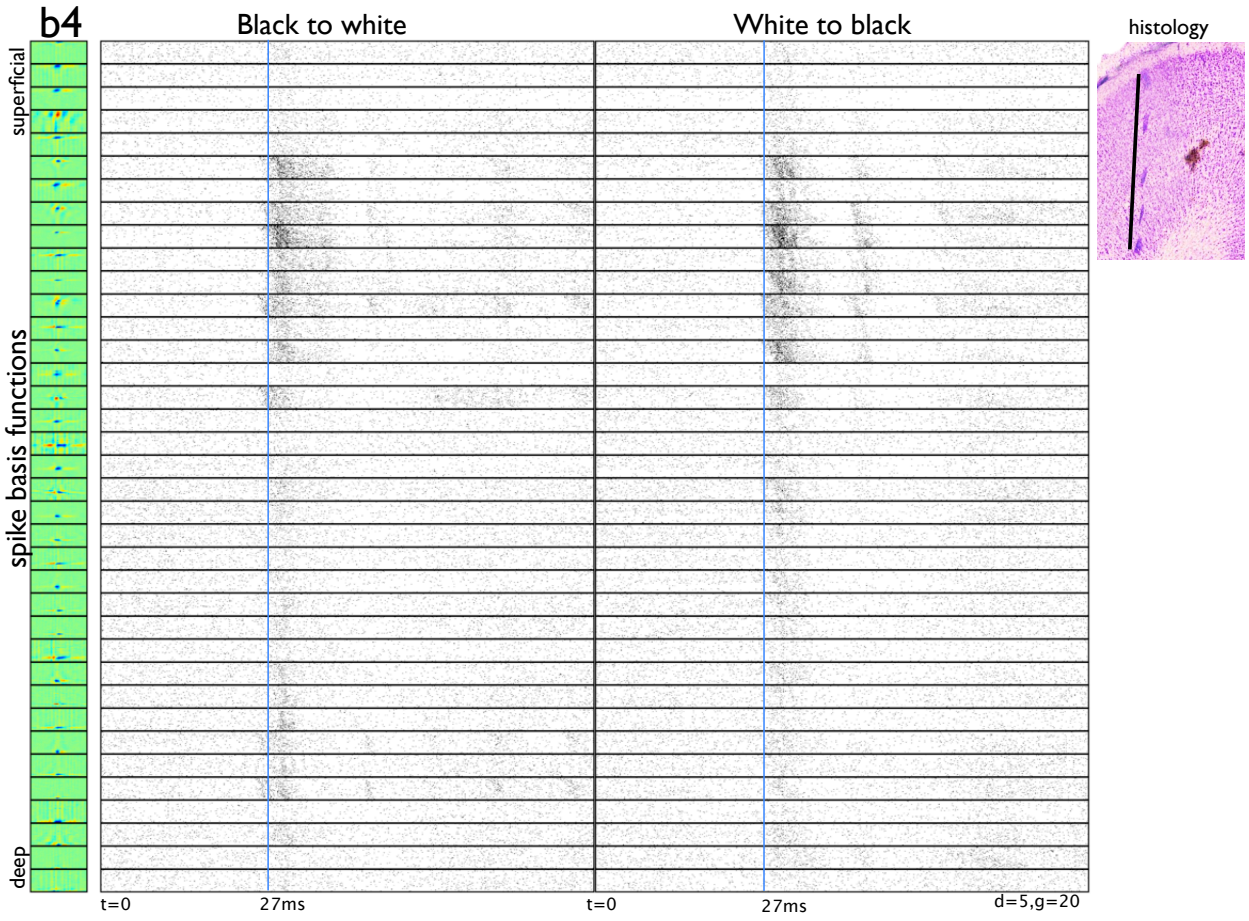

Figure 3.6: **Response to on-off step stimuli.** Spike coefficient rasters (see Fig. 3.5) for a slightly tangential polytrode penetration. A burst of activity in the upper channels likely corresponds to layer 4. Activity in the lowest laminar layers may correspond to layer 6 LGN afferents[75]. Responses to the two stimuli are qualitatively different, with the on-black stimuli showing a more punctate first peak and a shorter lag for the small second peak in activity.

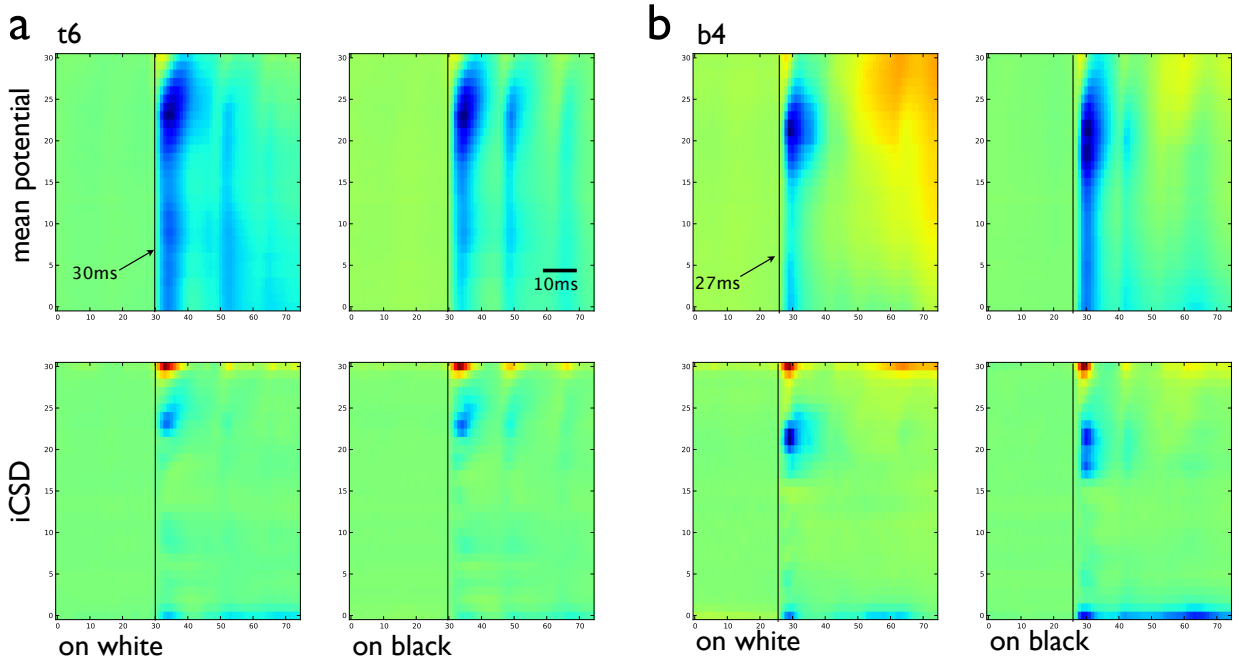

Figure 3.7: **Mean potential and CSD.** Trial-averaged potentials and current source density (CSD) for on-off step movies. On-white and on-black transitions were calculated separately. CSDs were computed using the step iCSD method[92]. Black vertical bars denote onset of response and the full span in time in each plot is 75 ms. **(a)** Corresponds to Fig. 3.5 and **(b)** corresponds to Fig. 3.6. CSD estimates are a good means to confirm position of the polytrode[75] with the prominent current sink agreeing with location of activity in the spike coefficient rasters.

3.3.2 Responses to natural movie repeats

Results of Yen *et al.*[111] from multi-site tetrode probes in anesthetized cat visual cortex showed an unexpected degree of heterogeneity in the response of nearby cells to repeated natural movies. Fig. 1.2 shows spike rasters of a collection of neurons recorded from the same tetrode to natural movie repeats. We wished to extend this result by studying responses to repeated natural movies along an entire cortical column. Fig. 3.8 shows spike coefficient rasters for a slightly tangential penetration. Responses were punctate and repeatable from trial to trial, particularly in the granular and supra-granular layers, but highly heterogenous across lamina. Many cells in the deeper layers showed graded or no responses to the natural movies. Fig. 3.9 shows detail of Fig. 3.8 for a subset of basis functions, highlighting the differences in responses between nearby cells as well as the non-stationarity in trial-to-trial statistics. In Chap. 4, we use a statistical framework to model and account for these distinctive response features. Spike coefficient rasters are also shown in Fig. 3.10 for a vertical penetration. In this case, the responses in the superficial layers are highly consistent, punctate and reliable across trials. Deeper layers again show a more diffuse response.

The structure of LFP responses to repeated natural movies is poorly understood. Here, we show that a subset of LFP basis functions have coefficients that show punctate temporal structure and repeatability from trial-to-trial, but to a smaller degree than the spike coefficients. For the tangential penetration example, basis functions with laminar structure showed a degree of trial-to-trial repeatability whereas basis functions with coherent laminar structure did not. A short segment of LFP rasters to repeated movies for this penetration is shown in 3.11. LFP rasters for a perpendicular penetration is shown in 3.12, showing a wide range of responses for basis functions with different spatial and temporal structure. Fig. 3.13a shows temporal detail for this recording. LFP coefficients are in general more diffuse than spike coefficients with consistent phase relationships between basis functions. This structure will be explored further in Chap. 4. For comparison, Fig. 3.13b shows the mean LFP across trials for the same recording. Although trial structure in the LFP is still visible, it is more difficult to discern. A sparse decomposition affords a new approach to understanding the structure in the LFP by separating it into a more tractable set of independent components, each with possible biophysical correlates that can be studied separately.

3.3.3 Responses to drifting gratings

Sinusoidal gratings are perhaps the most commonly used stimulus in physiological studies of the visual system. These stimuli are simple to generate, have a low dimensional parametrization, and are highly effective at driving cortical activity. However, their ecological relevance is highly questionable. To compare with the large body of work done with this stimuli and to see how those results extend to a large number of neurons recorded simultaneously across a cortical column, we presented repeats of a pseudorandom set of sinusoidal gratings

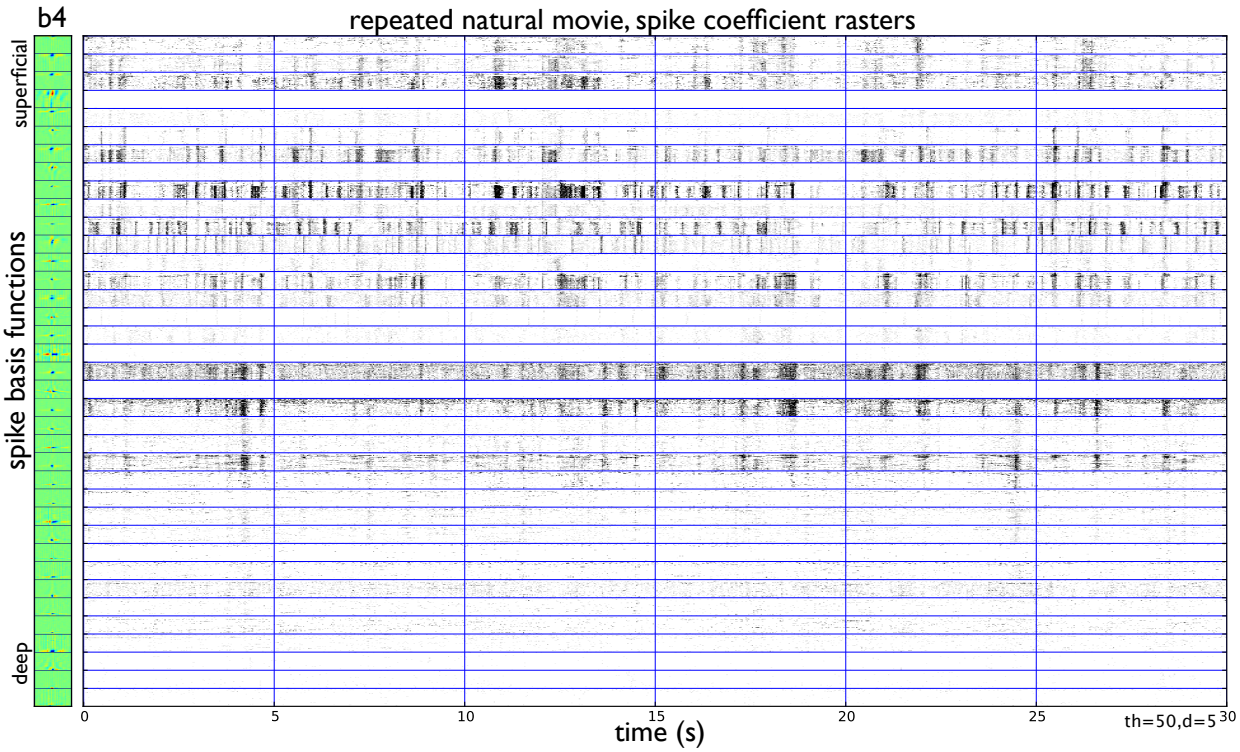

Figure 3.8: **Spike coefficient rasters to repeated natural movies.** Spike coefficient rasters for 60 repeats of a 30 s natural movie. Basis functions in green are ordered by lamina. Horizontal blue lines demarcate division between basis functions. Rasters were thresholded to 50 μV/frame refresh and down-sampled 5 times in time. Vertical blue lines were added to aid visualization. This penetration corresponds to Fig. 3.6 and is slightly tangential.

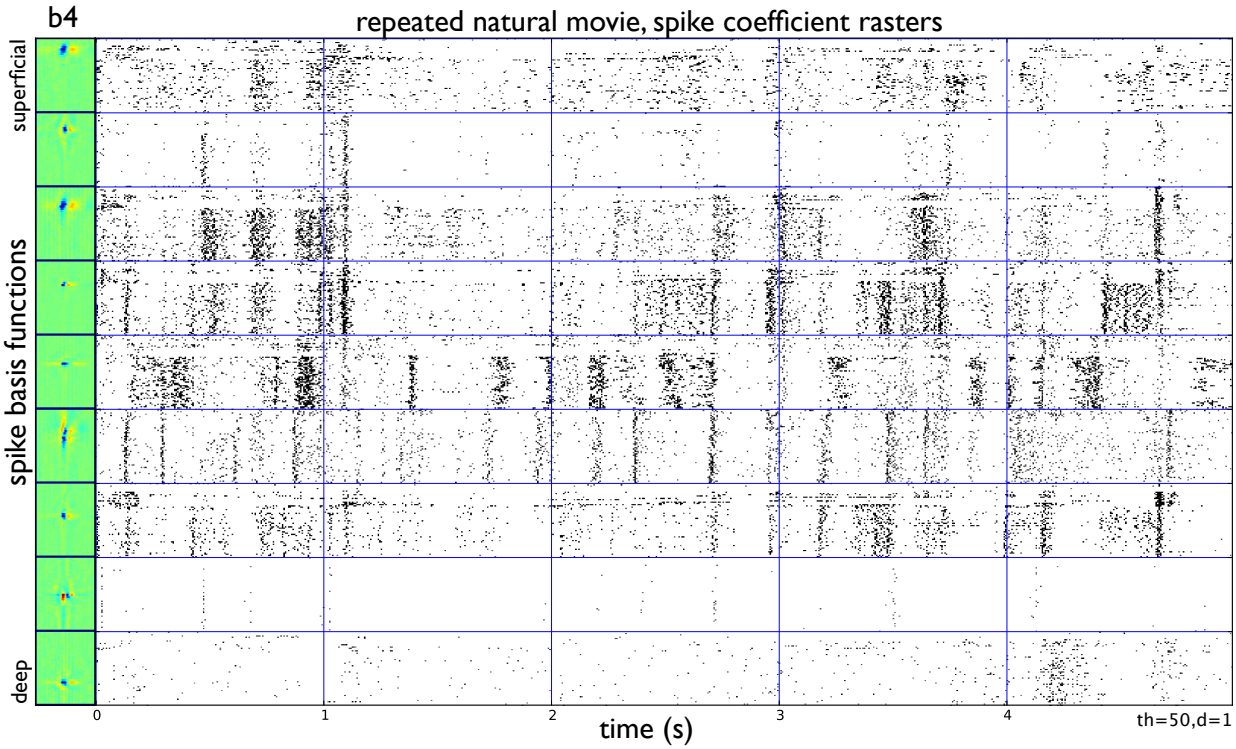

Figure 3.9: **Spike coefficient rasters to repeated natural movies.** Spike coefficient rasters for a subset of the basis functions in Fig. 3.8 for 5 s of the full 30 s repeats showing some of the variety in raster patterns, such as trial-to-trial repeatability, sharp or graded responses to features in the stimulus, and inter-trial non-stationarity.

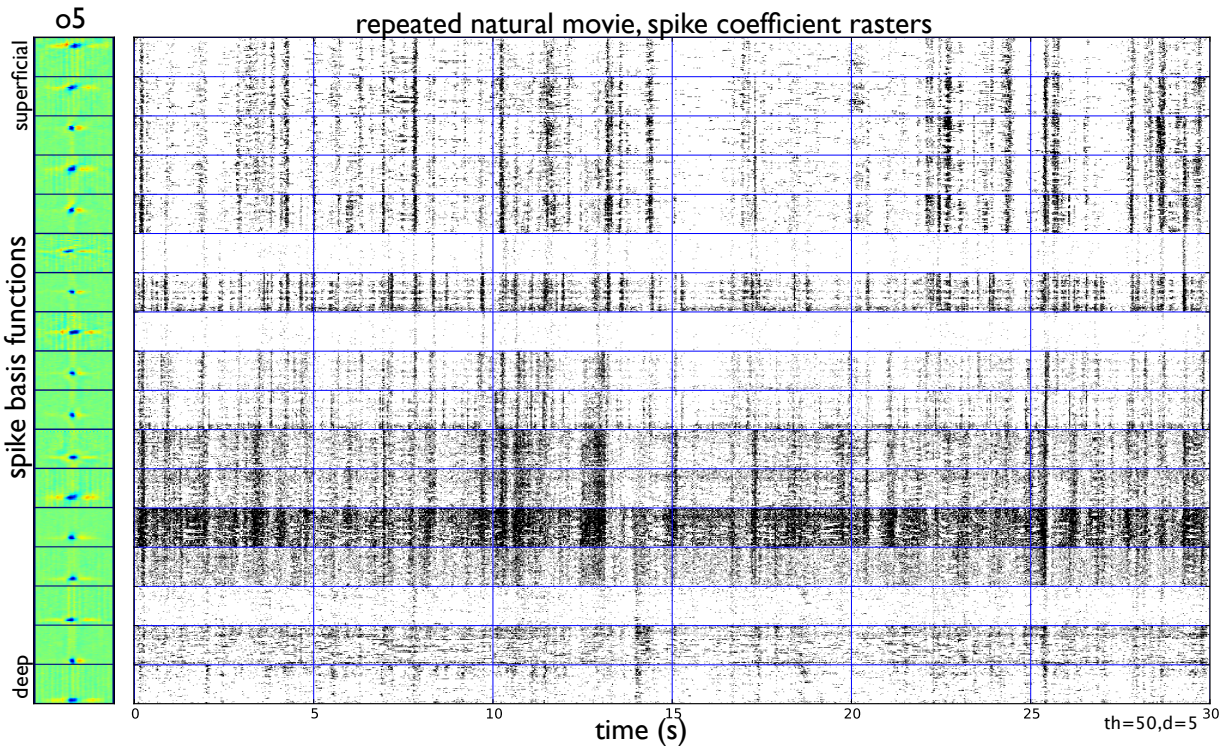

Figure 3.10: **Spike coefficient rasters to repeated natural movies.** Spike coefficient rasters for 60 repeats of a 30 s natural movie (see Fig. 3.8). This penetration was close to perpendicular. Activity in the superficial layers is highly synchronous and punctate whereas activity in the deep layers is more graded, though still responsive to certain frames of the stimulus.

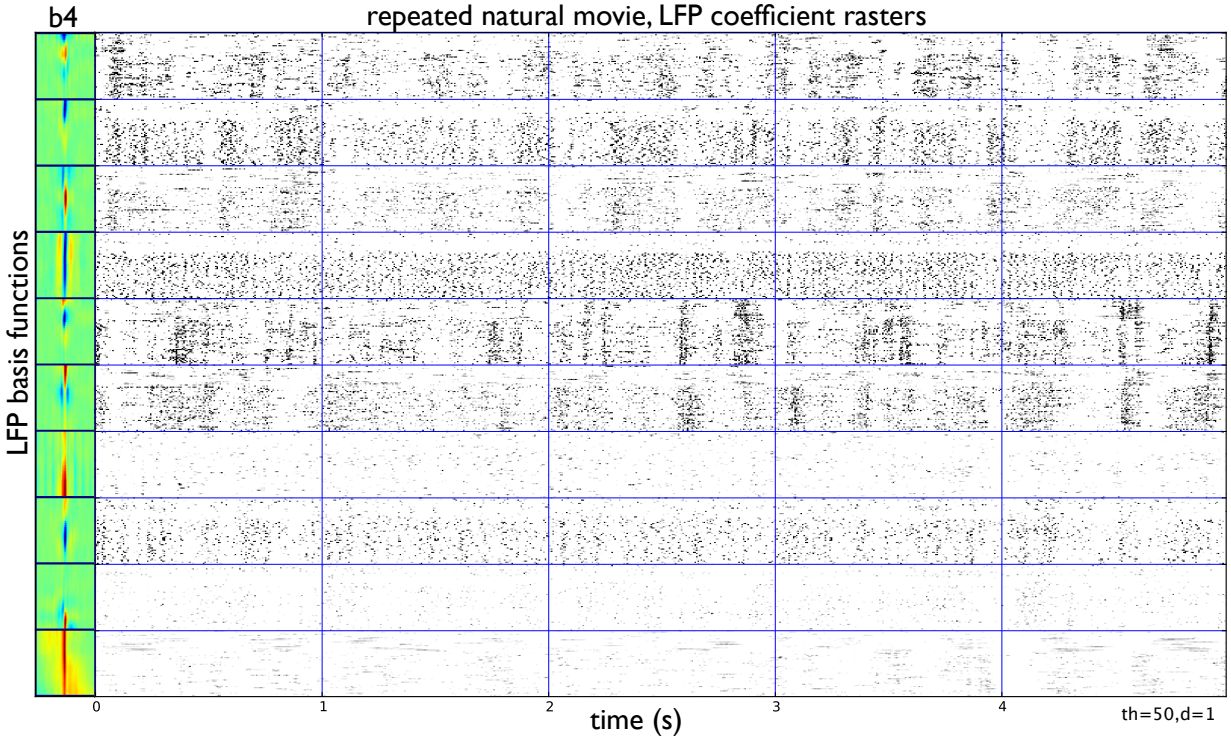

Figure 3.11: **LFP coefficient rasters to repeated natural movies.** A representative subset of sparse coefficient rasters for LFP basis functions. The first 5 s of 60 repeats of a 30 s natural movie are shown here. The timing of the LFP coefficients corresponds the center time of the LFP basis functions. Though the statistics of the coefficients are quite different from spike coefficients, some repeatable structure exists with heterogeneity between basis functions.

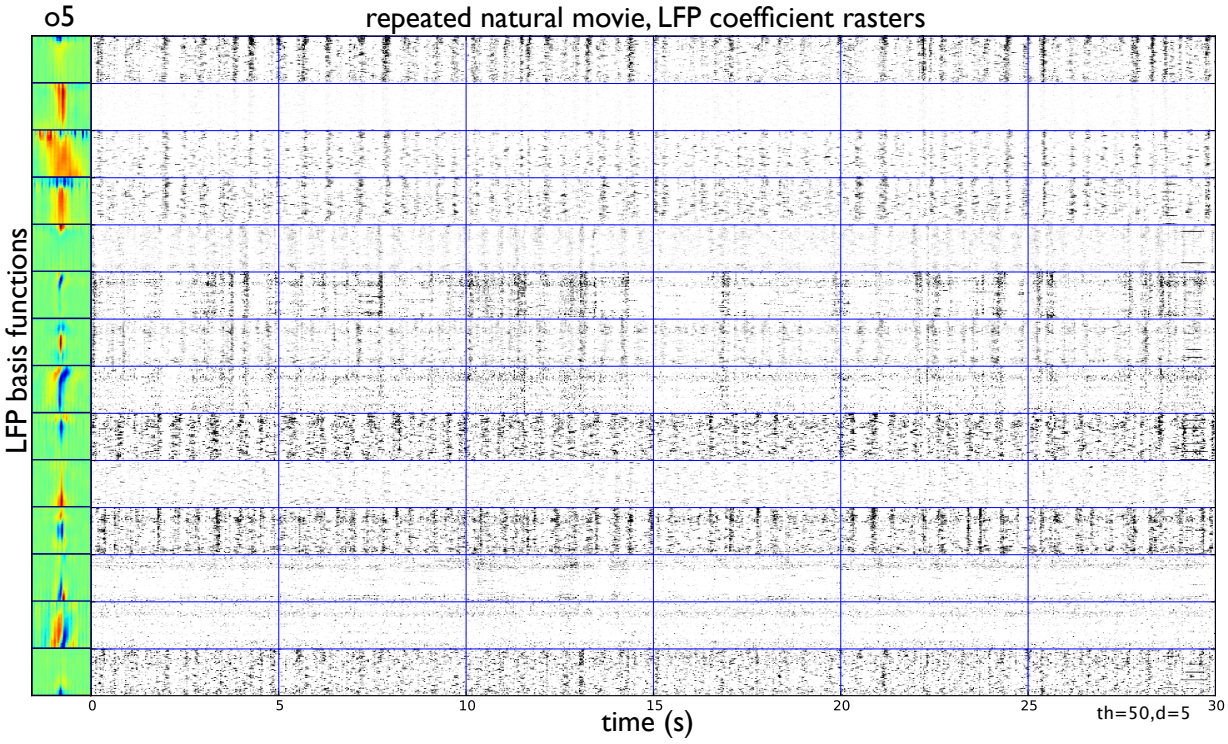

Figure 3.12: **LFP coefficient rasters to repeated natural movies.** LFP coefficient rasters for a vertical penetration to 30 s repeated natural movies. A representative subset of basis functions are shown here. Basis functions with laminar structure have clearly repeated responses to different frames in the stimulus. Horizontal line artifacts near the end of some repeats are due to an acquisition system error.

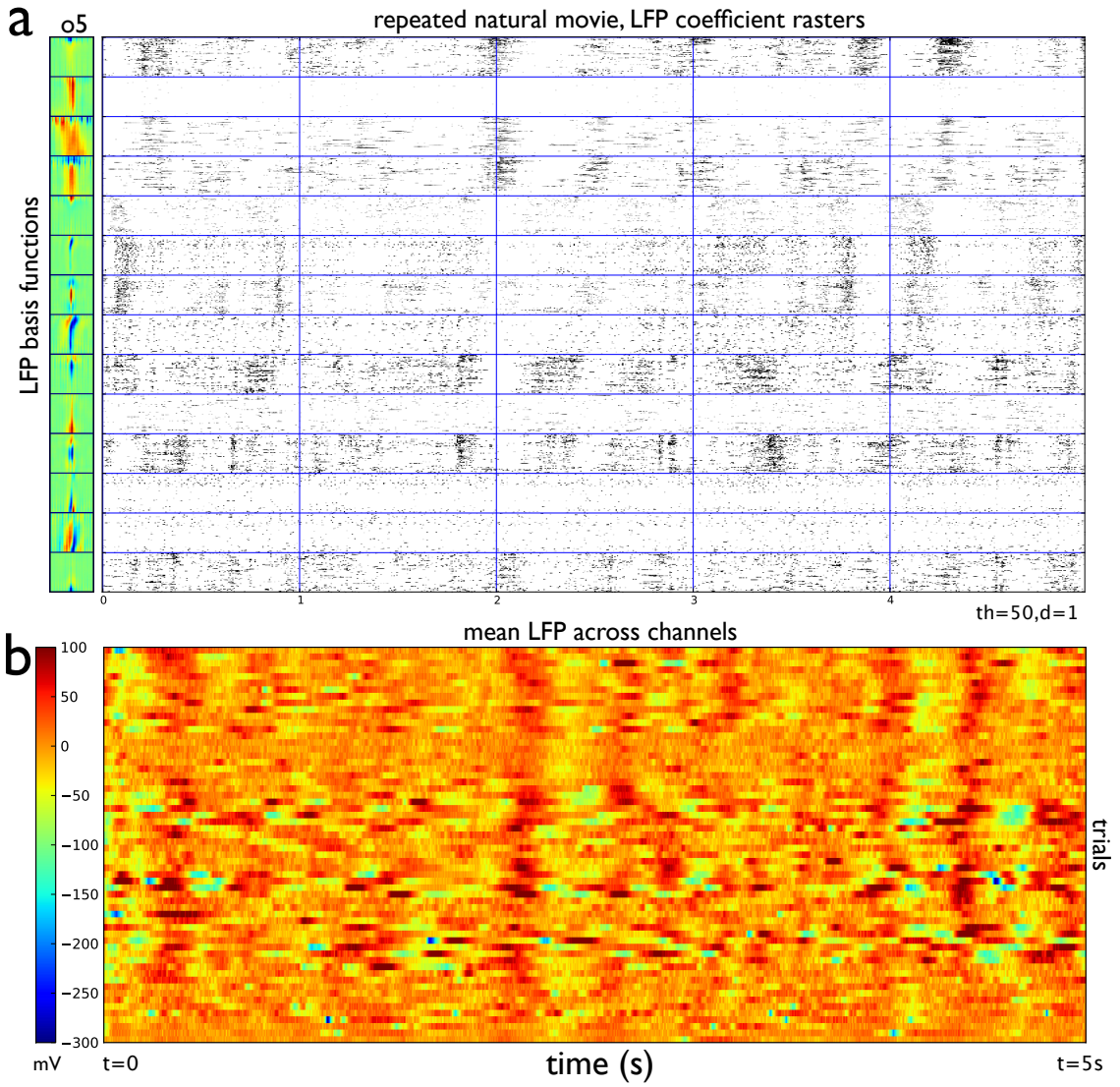

Figure 3.13: **LFP coefficient rasters compared to mean LFP. (a)** First 5 s of LFP coefficient rasters for data in Fig. 3.10 showing additional temporal detail of coefficient statistics. **(b)** First 5 s of mean LFP across all channels for all 60 repeats of the natural movie for comparison with coefficients. Though temporal structure exists in the LFP, it is harder to discern. The sparse coding representation provides a decomposition with clearer separation of causes that can be studied separately, as is done in Chap. 4.

of different orientation, direction, spatial, and temporal frequencies for 10 repeats each. Results for a vertical penetration are shown in Fig. 3.14. Spike basis function responses were highly tuned to orientation for all repeats, particularly in the superficial layers, with a more graded response in the deeper layers. The deepest layers showed suppressed responses to the preferred orientation of the column. For a tangential penetration (Fig. 3.15), orientation selectivity shifted along the column due to the polytrode crossing into neighboring columns. This result is confirmed in the orientation of receptive fields computed from natural movies (Sec. 3.3.4). LFP basis functions with components in the superficial layers also demonstrated a strong orientation preference (Fig. 3.16).

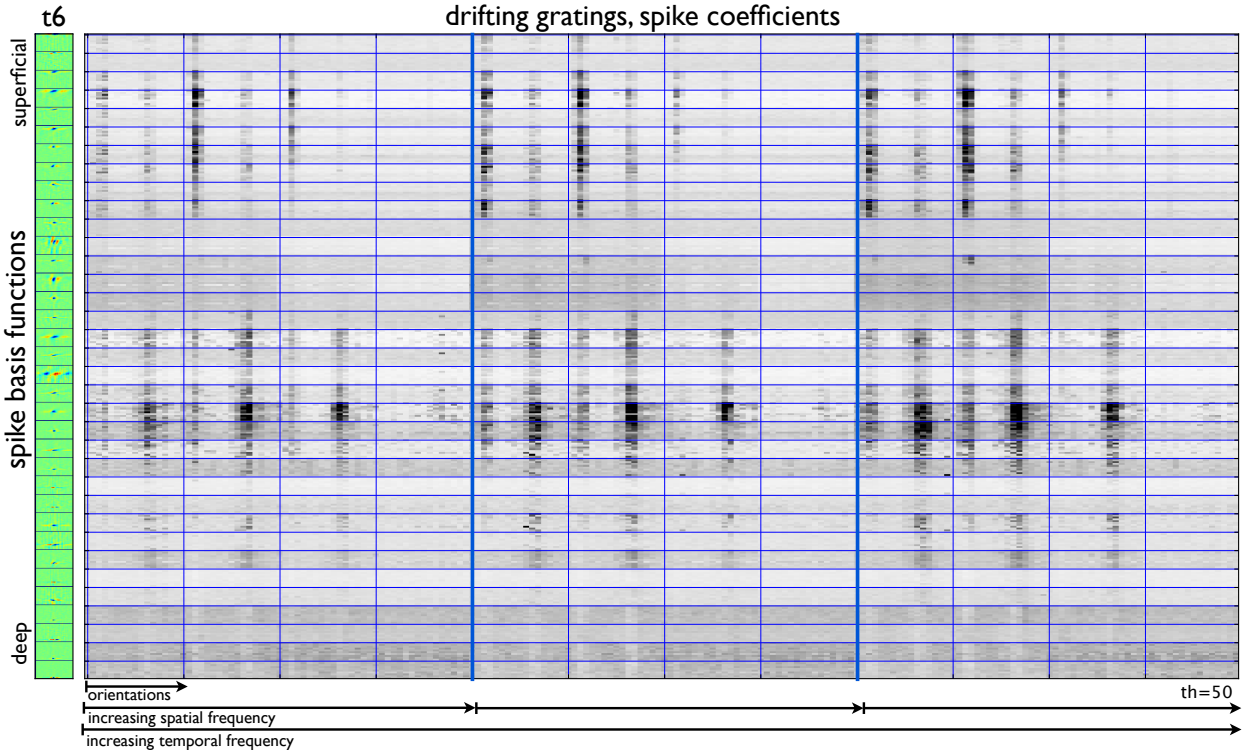

Figure 3.14: **Spike coefficient rasters for drifting gratings.** Spike coefficient rasters for 10 repeats of a sequence of 2 s drifting grating presentations. Each block within blue grid lines consists of 10 repeats (vertical) of 16 orientations and directions (horizontal) consisting of 20 s of data. The remainder of the plot is organized according to 3 large blocks of increasing temporal frequencies. Within each block are 4 subdivisions of increasing spatial frequencies as indicated by the arrows at the bottom of the plot. Activity in the superficial layers is highly tuned to specific orientations with a more graded response in the middle layers. Several basis functions in the deep layers show a suppressed response to the dominant orientation of the column.

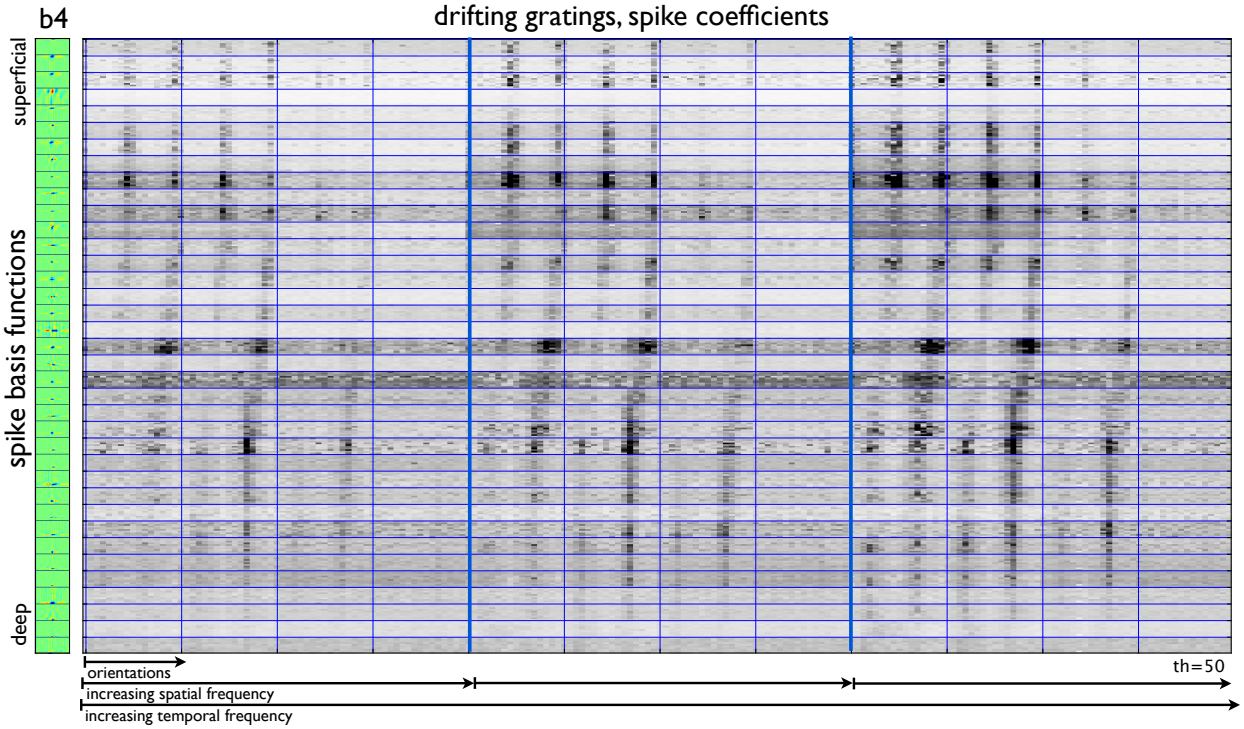

Figure 3.15: **Spike coefficient rasters for drifting gratings.** Spike coefficient rasters for gratings as in Fig. 3.14. This penetration traversed several columns as seen by the shift in orientation preference across lamina. This corresponds well with the change in orientation of receptive fields computed from responses to natural movies.

3.3.4 Responses to long natural movies

One method for characterizing response of visual cortical neurons is to describe the distribution of the spatio-temporal stimulus sequences that precede each spike. The sample mean of this ensemble, the spike-triggered average (STA), can be interpreted as the neuron's preferred stimulus, or receptive field, when the spike-triggered ensemble is peaked and distributed symmetrically around the STA[98]. Here, STAs were computed, weighted by coefficient value, for each basis function to a whitened long natural movie session and compared with STAs computed from binary white noise movies. Fig. 3.17 shows STAs computed from a whitened Duck30 movie played at 150 Hz frame rate for 29.6 minutes. A subset of the basis functions, mostly those corresponding to neurons in the superficial layers, have clearly defined STAs with oriented, bandpass, spatially localized structure. Several basis functions with circular receptive fields with biphasic temporal profile likely correspond to LGN axon terminals[60]. Many basis functions, particular in the deep layers, do not show structure in their kernels, likely due to lack of recorded spiking activity in these layers for this penetra-

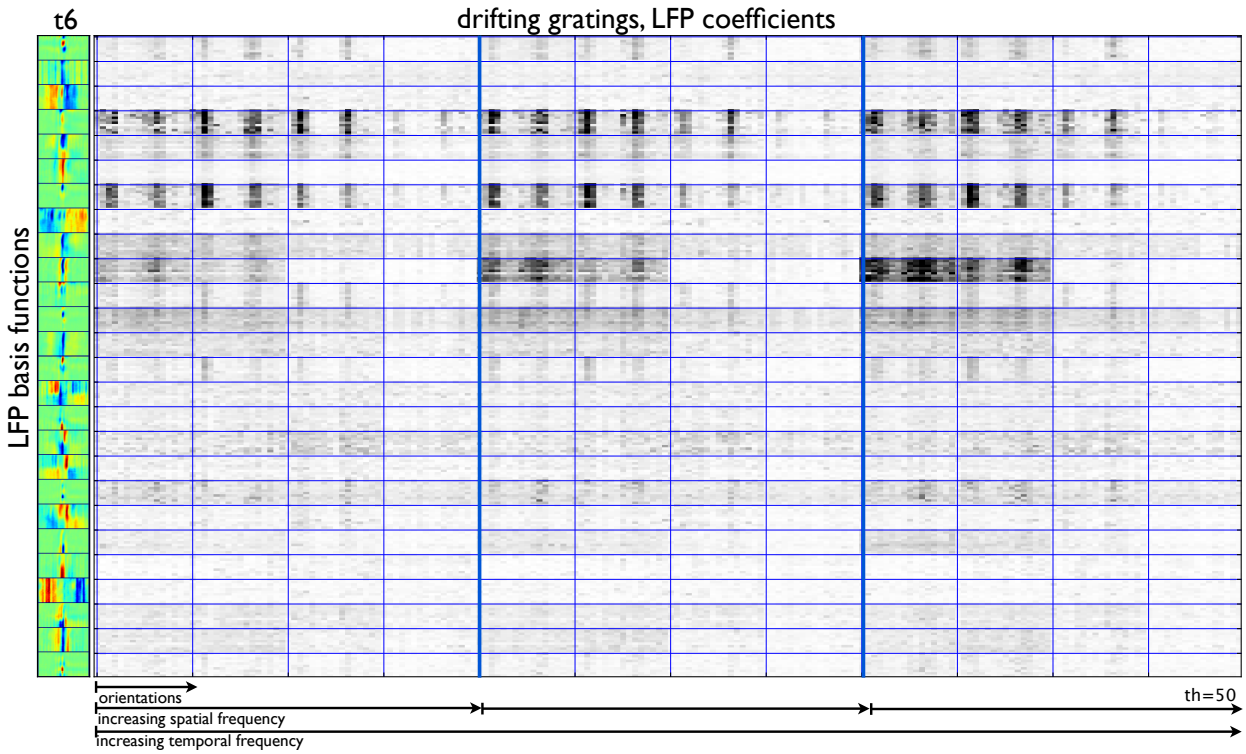

Figure 3.16: **LFP coefficient rasters for drifting gratings.** LFP coefficient rasters for the same session of drifting gratings as in Fig. 3.14 also show orientation tuning. Basis functions with laminar structure in the superficial layers are the most tuned to orientation.

tion. For comparison, STAs were computed using a 20 minute binary white noise movie. Shown in Fig. 3.18, they agree very well with natural movie STAs in detail such as orientation and approximate spatial frequency. Additionally, they reveal certain circular surround receptive field in the deep layers not present in the natural movie case. The orientation of receptive fields gradually change with lamina in agreement with the gratings results in Fig. 3.15. Although a spike-triggered analysis is useful as results can be compared to a large body of existing work, it is not clear to what degree the receptive fields characterize neural responses[85]. Chap. 4 addresses this issue by formulating a predictive statistical model where the degree to which the receptive field can explain the responses of neurons can be more easily assessed.

3.3.5 Responses to Hilbert movies

Natural images have a rich statistical structure to which the visual system is highly adapted[100]. In a series of electrophysiology experiments designed to explore this idea, Otto Creutzfeldt[25] recorded responses of neurons while scanning a set of images, including pictures of natural scenes, line by line, over their receptive fields. He found that different neurons responded to certain features of the stimulus, such as edges at particular orientations or regions with changes of luminance or contrast. Inspired by this strategy, we recorded responses from two natural images scanned according to a fractal path pattern. The intensity of spike coefficients for several representative basis functions for one penetration are shown in Fig. 3.19. The responses across the population of basis functions is highly heterogenous. Some are highly selective to features in the stimulus, particularly in the superficial layers, while others respond to lower order statistics such as changes in contrast or luminance, and many basis functions in the deep layers do not have an easily interpretable response. Fig. 3.20 shows a similar analysis for representative LFP basis functions, which also demonstrate selectivity to features of the stimulus.

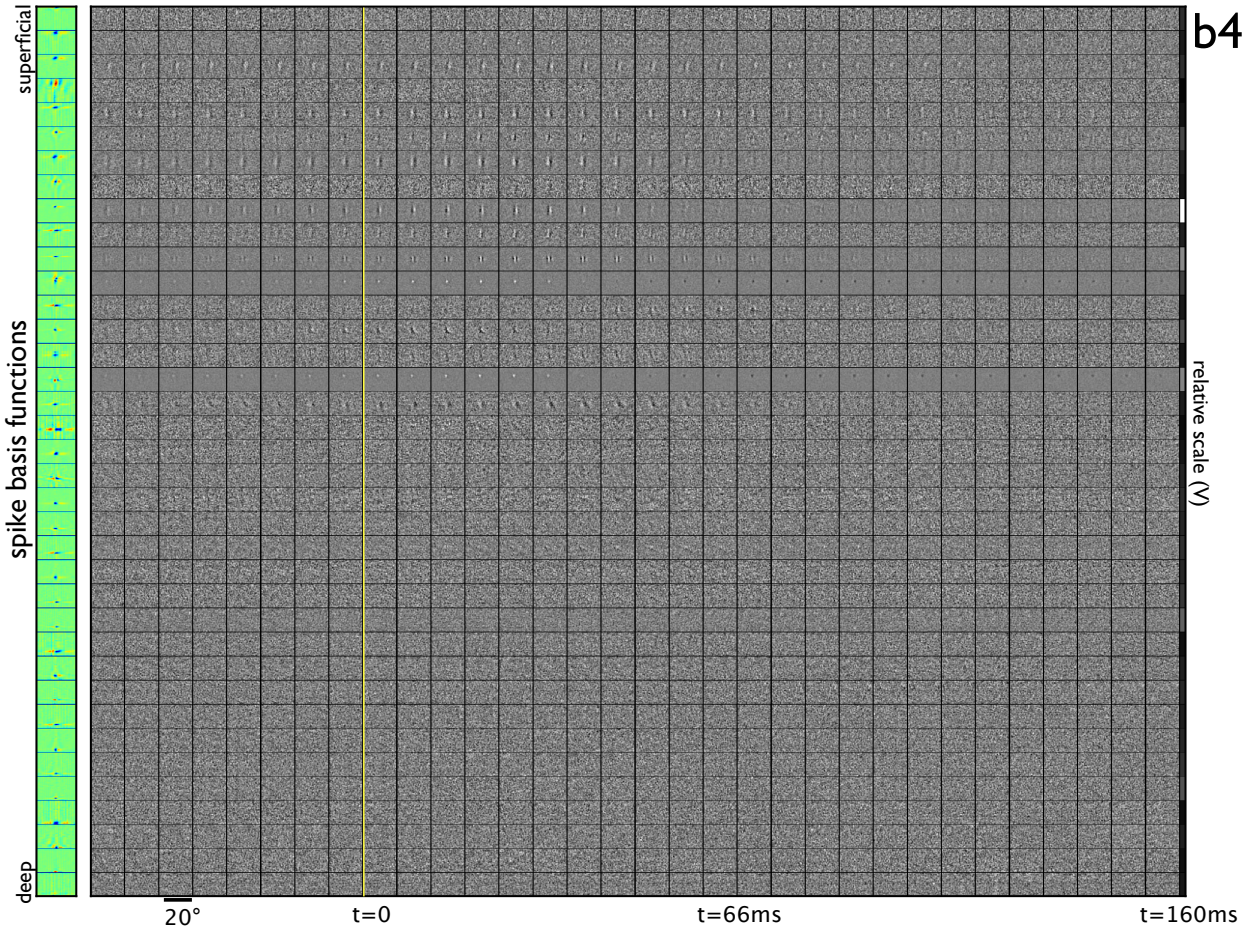

Figure 3.17: **Long natural movie STAs.** Spike coefficient-triggered averages (STAs) to a spatiotemporally whitened Duck30 movie. The STA for each basis function appears to its right and is plotted as a convolution kernel of movie frames. A subset of the basis functions in the superficial layers show oriented, bandpass, localized structure. Each spatial frame of the kernels represents the full 512×384 movie frame downsampled 8 times. Many basis functions, however, do not show any structure. Structure in the averages appear before t=0 due to the acausal natural of the whitening filter and temporal correlations in the movie. Each kernel is normalized. Relative voltages between kernels is indicated by the small band at the right of each kernel, with white indicating larger voltages.

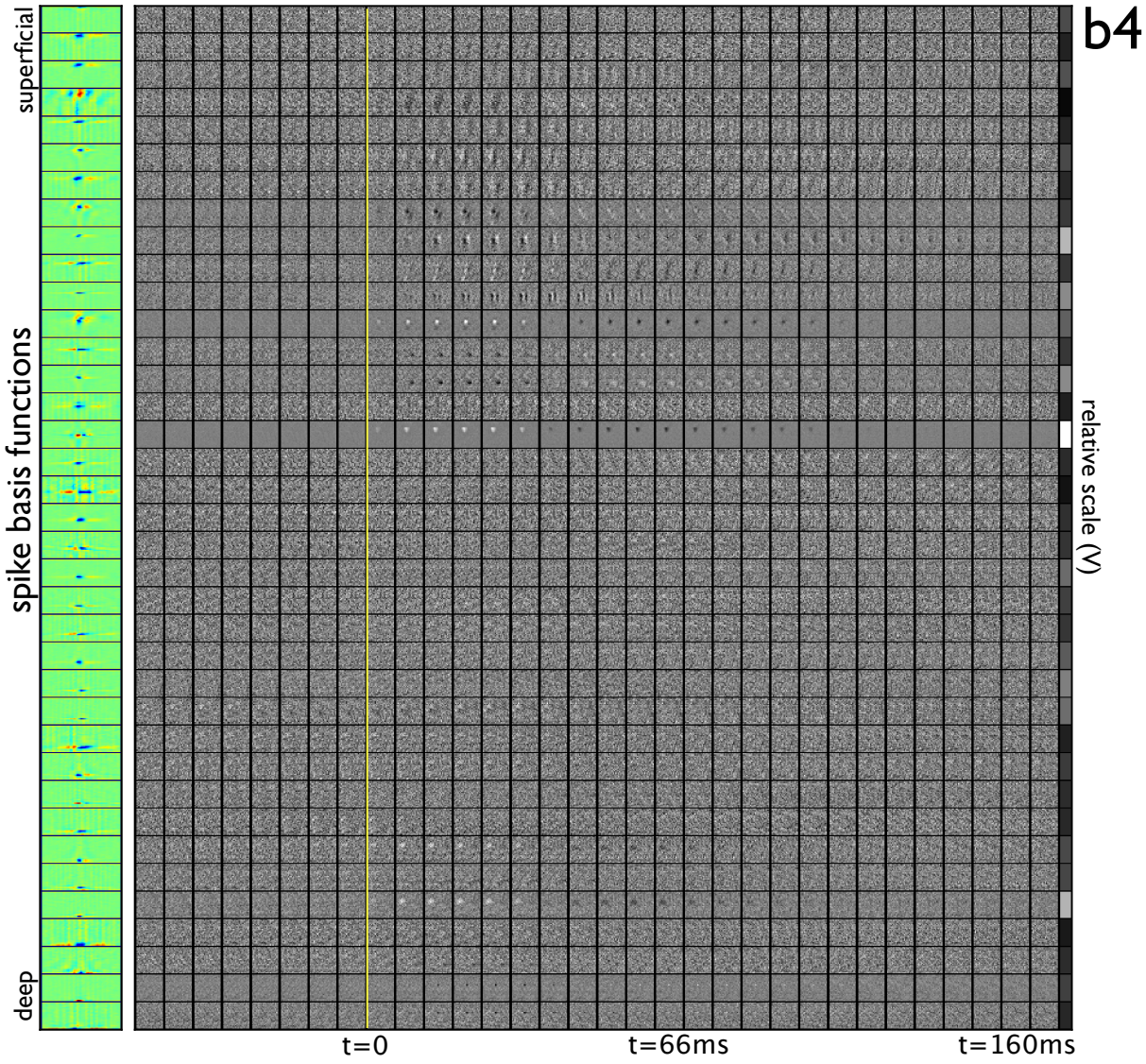

Figure 3.18: **Binary white noise STAs.** Spike coefficient-triggered averages for a binary white noise movie for the same penetration in Fig. 3.17. Receptive fields are shown at the refresh rate for easy comparison to natural movie STAs. Good correspondence exists for many cells between the two stimuli. However, some of the structure visible in the noise is an artifact of averaging over several frames; the frame rate of the white noise movie is 1/5 of the natural movies.

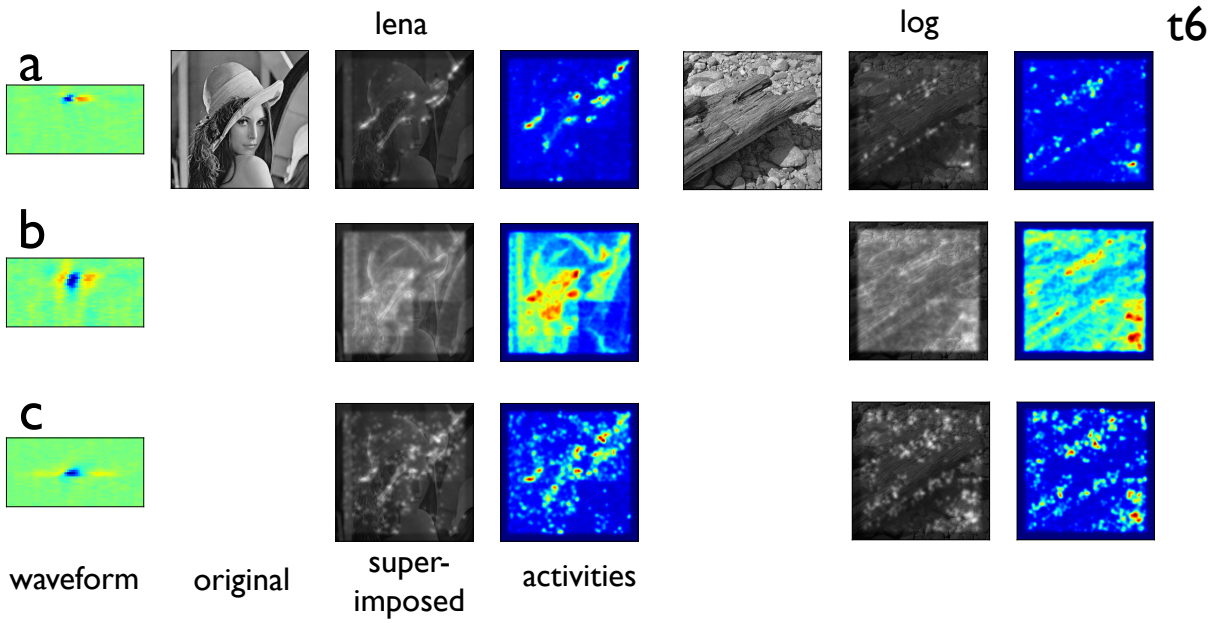

Figure 3.19: **Spike responses to Hilbert movies.** Responses for three spike basis functions are displayed. The columns are the basis function waveform, the original image (lena or log), the original image with coefficient activities superimposed, and a heat map of the coefficients alone. The coefficients are presented at a delay of 45 ms, which corresponds to the peak of the response of a typical striate cell. The activities are blurred with a gaussian to make them easily visible. **(a)** A superficial basis function with remarkable sparseness and selectivity to features of both lena and the log at a certain orientation. **(b)** A likely LGN axon afferent responding to areas with changes in local contrast. Blocking in the response plots is due to the path taken by the Hilbert curve (see Fig. 3.4). **(c)** A deeper basis function with less selective response than in **(a)**. The Hilbert movies provide a succinct description of the response of cells in different lamina by taking advantage of the statistical richness of natural images.

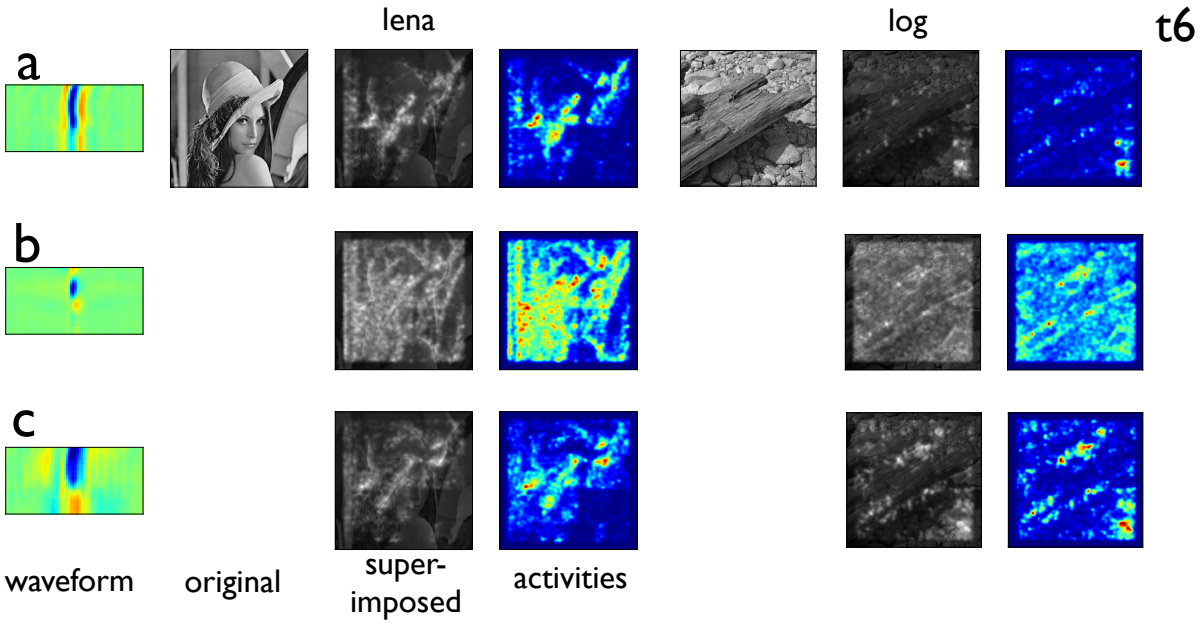

Figure 3.20: **LFP responses to Hilbert movies.** Responses for three LFP basis functions to two Hilbert movies with subplots organized as in Fig. 3.19. Basis functions clearly respond to certain features in the stimulus, such as edges of a particular orientation as well as areas with changes of luminance and contrast. The Hilbert movies provide a novel strategy for exploratory analysis of the relationship of the LFP to visual stimuli, an area that is poorly understood.

3.4 Discussion

We performed silicon polytrode recordings in anesthetized cat visual cortex of several hours in duration while showing a range of stimuli including natural movies, gratings, and white noise. Processing and understanding this data presents an enormous challenge. Here, we applied a sparse coding unsupervised learning algorithm to transform the data into a new representation. We showed the correspondence between activity of spiking and LFP sparse coefficients with a wide range of stimuli. In Chap. 4, we make the step from descriptive exploratory analysis to statistical models of response by using coefficients as regressors in Generalized Linear Models (GLMs).

The importance of using a sparse decomposition of the data for analysis cannot be over-emphasized. Consider the problem of manually or even semi-automatically spike sorting the data from 32 channels. In most labs this could easily consume months of effort. By contrast, the sparse basis function decomposition can be computed in a matter of hours. In fact, we have even used it on the fly during the process of a recording session to determine how best to position the polytrode. For the LFP, it enabled us to break up the complex spatio-temporal waveforms recorded from the polytrode into a set of features that could be directly related to events in the movie sequence. Determining what properties of the movie drive the LFP will require further study.

Chapter 4

A statistical model of population response

4.1 Introduction

In Chap. 3, a sparse coding representation was used to provide a descriptive analysis of the spiking and LFP responses along a cortical column to a variety of visual stimuli. In the present chapter, two statistical models are constructed, one to account for joint spike activity as a function of the stimulus, network spiking interactions, and the LFP, and a second to account for LFP activity as a function of spiking and stimulus. In each case, the degree to which model components contribute to explaining structure are quantified. The resulting models provide a causal description of the relationship between spiking, LFP, and the stimulus in a columnar microcircuit.

The class of models considered in this section have been used recently to study a variety of neural population data. These Generalized Linear Models (GLMs)[88,105] have been applied, for example, to large-scale recordings of a population of retinal ganglion cells and are able to account for a significant fraction of the variance of the population activity[95]. However, few similar studies have been done with cortical populations[63]. Here, a GLM model (Fig. 4.1) is applied to laminar activity recorded using silicon polytrodes to characterize responses to dynamic, natural stimuli.

In previous chapters, neural activity was studied as a function of visual stimuli alone, largely ignoring ongoing cortical dynamics that are not directly related to the stimulus. In primary visual cortex, the majority of axonal inputs are coming from other cortical neurons rather than direct inputs from the thalamus[31]. Within layer 4 alone, it has been estimated that roughly 5% of the excitatory input arises from the LGN, with the majority resulting from intracortical inputs[90,91]. Ignoring the internal state of this system and focusing purely on stimulus driven activity is bound to lead to an incomplete picture. In particular, a growing body of literature suggests that oscillations in the LFP are linked to cortical states and may

mediate communication of global information related to attention and feature binding[41,101]. The GLM framework developed here is an attempt to quantify these relationships in a principled manner. This is illustrated in Fig. 4.2, which shows cross-correlograms of spike coefficients with LFP coefficients. A striking relationship between the timing of spike events and LFP activity in different lamina is readily apparent.

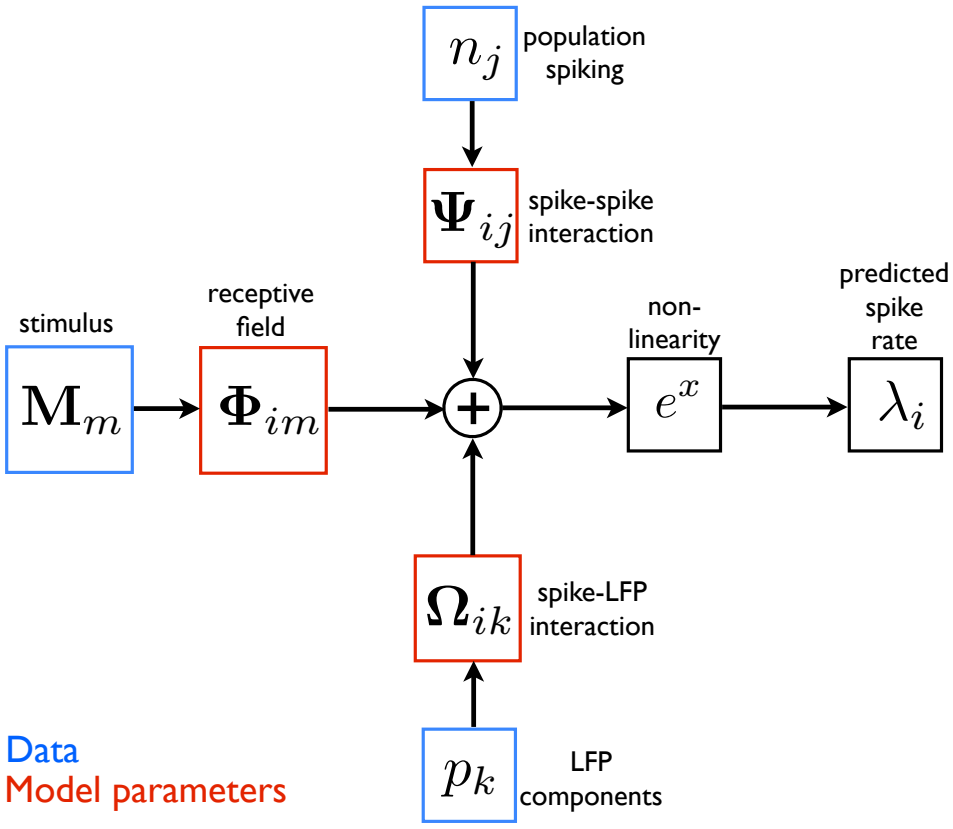

Figure 4.1: **Spike-LFP GLM model.** A schematic of the GLM model used to predict population spike or LFP coefficients. The distribution of coefficients for each basis function was assumed to be exponential conditioned on the data and model parameters. Input data to the model consisted of the stimulus, spike coefficients, and LFP coefficients. The model parameters for the spike GLM model included spatio-temporal receptive fields, spike-spike interactions, and spike-LFP interactions. The LFP GLM model included LFP-LFP interactions instead of spike-spike interactions. The contributions from each data source were summed, mapped by a non-linearity to produce a predicted firing rate $\lambda_i(t)$ for each spike or LFP component. Data is indicated by blue and model parameters by red boxes. This model is described in Sec. 4.2.1.

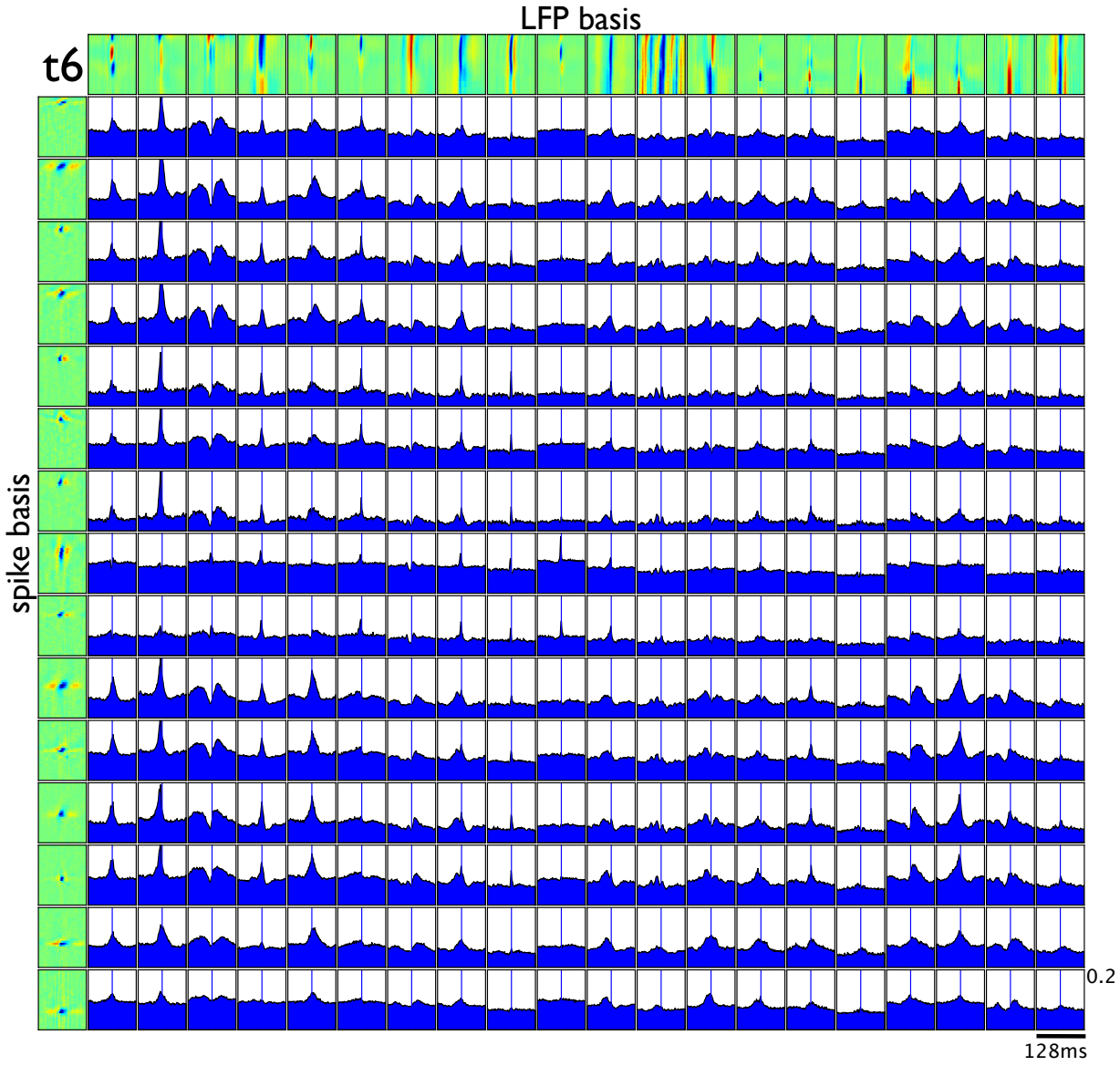

Figure 4.2: **Cross-correlograms of spiking and LFP coefficients.** Spike coefficients have characteristic distributions of activity with respect to LFP coefficients, often firing near the trough of the LFP waveform. On the vertical axis are spike basis functions ordered by lamina. On the horizontal axis are the LFP basis functions. The time scale is 2 ms for spike basis coefficients and 128 ms for LFP coefficients. Cross-correlograms of spike and LFP coefficients are normalized and computed over the course of a long natural movie, with time axis spanning -64 ms to 64 ms, aligned to the corresponding LFP basis function. The vertical scale for cross-correlograms is $[0, .2]$ with 1 indicating perfect correlation. Spike coefficients tend to be active during negative peaks in the potentials in the corresponding lamina of LFP basis functions. Conversely they are suppressed during positive peaks in LFP potential. These statistics demonstrate a high degree of correspondence between spiking in different lamina and the structure of the LFP.

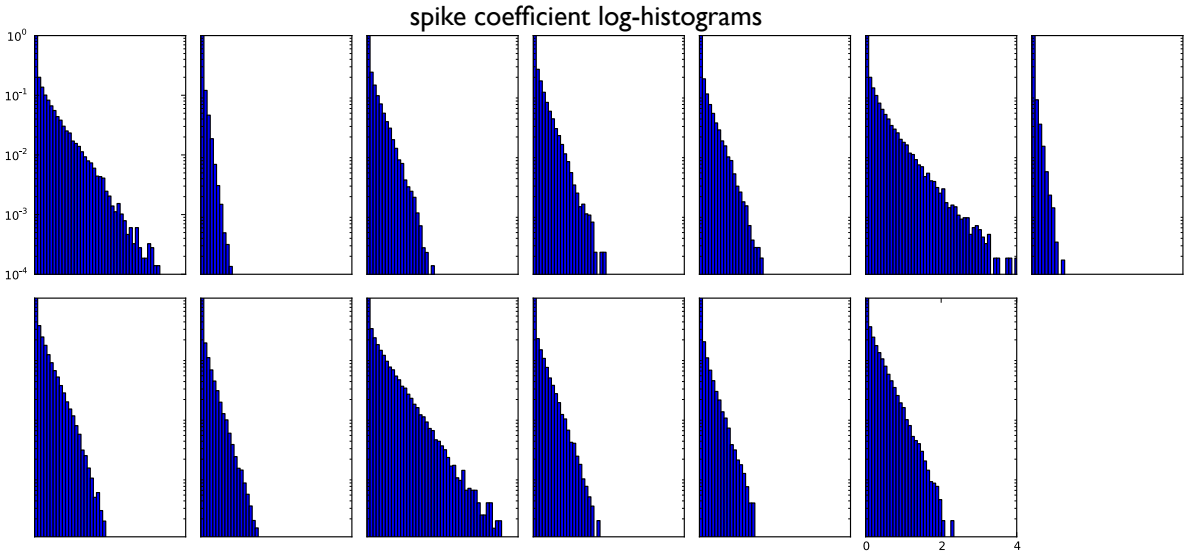

Figure 4.3: **Distribution of coefficients.** Normalized log-histograms of spike coefficients for the 13 basis functions corresponding to Fig. 4.4 in units of μV/frame. Coefficients were downsampled to the frame rate (6.7 ms), the time resolution of the GLM model. The distributions are highly kurtotic, with a large peak at zero. In the GLM model, we approximate these coefficient histograms using an exponential distribution. Most of the log-histograms have linear tails and correspond well to the shape of the model distribution, though some are significantly heavier tailed. LFP coefficient histograms were similarly distributed (not shown).

4.2 Methods

4.2.1 Exponential GLM

The spiking GLM model is described here. The LFP GLM model is highly analogous. The activity $n_i(t)$ of the i-th spike unit in time bin t is modeled as exponentially distributed conditioned on the model parameters $\mathbf{\Theta}$ and data $\mathbf{X}$,

$$P(n_i(t)|\mathbf{\Theta}, \mathbf{X}(t)) = \frac{1}{\lambda_i(t)} e^{-n_i(t)/\lambda_i(t)}$$

where $\mathbf{X}(t)$ indicates all data up to but not including time t. The conditional intensity of coefficient activation is,

$$\lambda_i(t) = f\left(b_i + \sum_m \mathbf{\Phi}_{im}(t) * \mathbf{M}_m(t) + \sum_j \mathbf{\Psi}_{ij}(t) * n_j(t) + \sum_k \mathbf{\Omega}_{ik}(t) * p_k(t)\right) \tag{4.1}$$

where $*$ denotes convolution in time. The model parameters $\mathbf{\Theta} = \{\mathbf{b}, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega}\}$ are,

$$\begin{aligned} b_i &: \text{bias} \\ \mathbf{\Phi}_{im} &: \text{spatio-temporal receptive fields for movie } m \\ \mathbf{\Psi}_{ij} &: \text{spike-spike coupling with spike basis } j \\ \mathbf{\Omega}_{ik} &: \text{spike-LFP coupling with LFP component } k \end{aligned}$$

The data $\mathbf{X} = \{\mathbf{M}, \mathbf{n}, \mathbf{p}\}$ consists of,

$$\begin{aligned} \mathbf{M}_m &: \text{spatio-temporal movie } m \\ n_j &: \text{spike rate for basis } j \\ p_k &: \text{LFP rate for component } k \end{aligned}$$

The point-wise non-linearity was chosen as $f(x) := \exp(x)$, which is convex and log-concave, a requirement for the full model to be convex. This form of the non-linearity outperformed other tested functions in the differentiable, convex, log-concave family. The log-likelihood of the model is,

$$\begin{aligned} \mathcal{L}(\mathbf{\Theta}) &= \sum_t \log p(\mathbf{n}(t)|\mathbf{X}(t), \mathbf{\Theta}) \\ &= -\sum_{i,t} \left(\log \lambda_i(t) + n_i(t)/\lambda_i(t)\right) \end{aligned}$$

which is concave in model parameters $\mathbf{\Theta}$ and could be maximized using standard methods for unconstrained optimization. A model schematic is illustrated in Fig. 4.1. The time scale used for this analysis was the frame rate of the stimulus, or 150 Hz. Spike coefficients $n_i(t)$ and LFP coefficients $p_k(t)$ were binned at the frame rate.

To reduce the number of free parameters in the optimization, we made the simplifying assumption that the spatio-temporal kernels $\mathbf{\Phi}$ are space-time separable, that is,

$$\mathbf{\Phi}_{im}(t) = \Phi^s_{im}(x)\, \Phi^\tau_{im}(t)$$

The log-likelihood in this case is not jointly convex in Φ^s and Φ^τ, requiring alternative convex maximization steps where either set of variables is kept fixed. Though a global optimum is

not guaranteed in this approach, a number of checks were used to test if the optimization had reached a local minimum.

To further reduce the dimensionality of the temporal dimensions of the kernels, we constrained their structure to be represented in a log-cosine bump basis[95]. These basis functions are logarithmically spaced in time to give higher resolution near $t = 0$ and constrain parameters to be smooth in time. For the temporal receptive field Φ^τ over the stimulus, we chose a 5-dimensional basis spanning 240 ms, whereas for the spike-spike and spike-LFP couplings, we chose a 4-dimensional basis spanning 54 ms. In all cases, the kernels were forced to be zero at $t = 0$ to ensure strict causality. Only the temporally localized basis functions for the LFP were included in the model, with their coefficient times shifted by their temporal support to maintain causality.

To aid fitting the spatial receptive fields, Φ^s were represented in a basis learned from natural images using `sparsenet`[82] with a sparse prior imposed on coefficients. In the optimization, this was implemented by projecting the movie onto the `sparsenet` basis and enforcing sparsity on the weights of these regressors by an L_1 regularizer with a single penalization parameter that was chosen in each case using cross-validation. This framework allowed estimating a model over multiple mappings of the same stimulus movie, each projected onto a suitable basis for regularization.

4.2.2 Optimization implementation

The optimization routines were constructed such that any subset of the model parameters could be held fixed while the remaining were optimized. Two initialization optimizations were run prior to the main optimization to pick good starting model parameters. The first determined the bias parameters b_i while the other parameters were kept at zero. This step was necessary to avoid numerical instabilities introduced by the exponential non-linearity, where the bias makes the single largest contribution to the log-likelihood. Then, the spatiotemporal receptive field kernels Φ^s and Φ^τ were initialized by performing a singular value decomposition (SVD) of their STAs (Sec. 3.3.4). The SVD in this case provided the optimal rank-1 least-squares estimate for the STA in terms of an outer product of spatial and temporal components. An optimal scaling of these kernel estimates together with the bias terms were determined with an optimization. Then a series of alternating steps were performed where either Φ^s or Φ^τ was held fixed and all other parameters were optimized. Finally, after the optimization had converged, two additional optimizations were performed to reduced the shrinkage effects of the L_1 regularization on the spatial receptive fields. First, only Φ^s was optimized for a large number of steps to ensure coefficients near zero were pegged exactly to zero. Then a full optimization with all parameters was performed where the support of Φ^s was held fixed, but no sparseness penalty was enforced on the nonzero components. Without such an unbiasing step, the scale of predicted $\lambda_i(t)$ were often too small.

Though the dimensionality reduction in model parameters described above made opti-

mization more tractable, the size of the data posed a challenge. In order to have a sufficiently rich model while using as much of the data as possible in a single optimization, the optimization algorithm needed to be parallelized. A data parallel `cython` implementation of the L_1-regularized limited-memory quasi-Newton method[1] described in Sec. 2.2.5.2 was developed. This allowed the optimization to scale almost linearly with the size of the data and be run on a large cluster of computers. Additionally, significant computational efficiency was achieved by pre-filtering data with their respective log-cosine temporal bases and caching them to disk such that no convolutions needed to be performed online in objective and derivative calculations.

4.2.3 Natural movie data

In order to fit the GLM model, we used a 29.6 minute, 260,000 frame Duck30 movie that was spatially down-sampled 16-times to 32×24 pixels, then whitened (3.2.4). The degree of downsampling used was in line with our knowledge of the spatial frequency preference of the recorded column. We extracted 12×12 or 16×16 central regions of the down-sampled movie which approximately corresponded to the location of the classical receptive field. This down-sampled movie was then projected onto a complete or overcomplete `sparsenet` basis[82]. A second movie was created from the first by squaring of pixels. It was, however, not projected onto any basis. This simple transformation was intended to capture some of the known spatial phase invariances of cortical neurons. The first 200,000 frames of the stimulus were used for fitting and the remainder for validation of the single regularization parameter. For all examples in this section, we chose a subset of spike and LFP basis functions. For the spiking GLM model, 13 of 100 spike basis functions were selected where STA computed from natural movies had discernible structure. STAs for these spike basis functions are shown in Fig. 4.4. 18 of 32 LFP basis functions were selected that had structure localized in time. The model was estimated on data from a single penetration that was verified to span all layers of cortex. After fitting the parameters of the GLM to the long natural movie, the model was used to predict responses to repeated 30 second short natural movies. Predicting repeated movie response assessed the ability of the model to account for trial-to-trial variations in the response.

4.2.4 Modeling LFP coefficients

The GLM framework described above is a description of spike coefficient activity in terms of other data. The same framework was used to study LFP coefficient activity as a function of spiking and the stimulus. The model and optimization steps were similar in all respects. We used 18 of 32 LFP basis functions as in the spiking GLM case and 37 of 100 spike basis

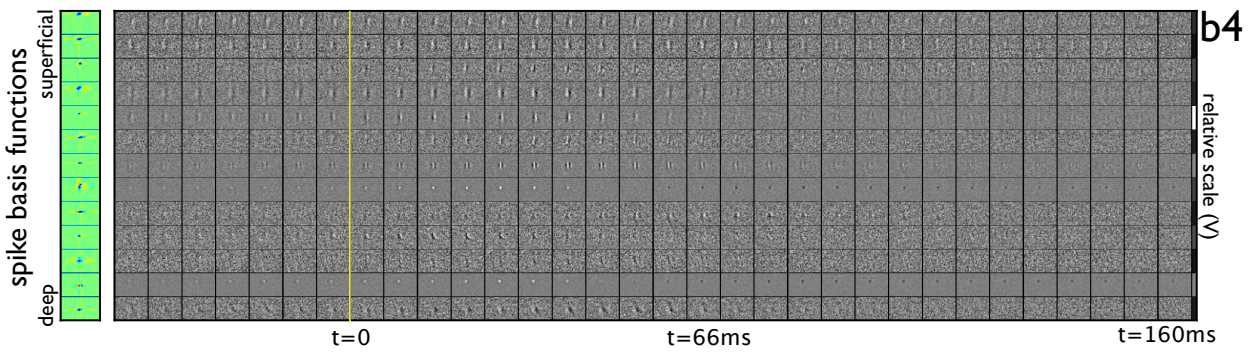

Figure 4.4: **Long natural movie STAs for example population.** A subset of the basis functions from Fig. 3.17. These units were chosen to evaluate the model in this section as their receptive fields had discernible structure that could be easily compared to GLM model fits.

functions. The model parameters $\mathbf{\Theta} = \{\mathbf{b}, \mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Omega}\}$ in this case are,

$$\begin{aligned} b_i &: \text{bias} \\ \mathbf{\Phi}_{im} &: \text{spatio-temporal receptive fields for movie } m \\ \mathbf{\Psi}_{ij} &: \text{LFP-LFP coupling with LFP component } j \\ \mathbf{\Omega}_{ik} &: \text{LFP-spike coupling with spike component } k \end{aligned}$$

The data $\mathbf{X} = \{\mathbf{M}, \mathbf{n}, \mathbf{p}\}$ consists of,

$$\begin{aligned} \mathbf{M}_m &: \text{spatio-temporal movie } m \\ n_j &: \text{LFP rate for basis } j \\ p_k &: \text{spike rate for component } k \end{aligned}$$

and $\lambda_i(t)$ in (4.1) the conditional intensity of activity of LFP component i.

4.3 Results

In previous chapters, a descriptive overview was presented of the laminar structure of the spiking and LFP activity in a column of cat visual cortex, without making attempts at a quantitative description or a mechanistic model of the observations. The GLM framework is a full statistical model of how the stimulus as well as the dynamics of spiking and LFP give rise to the observed data. This statistical framework quantifies dependencies between different parts of the response as well as assessing the degree to which the model is able to predict spiking and LFP.

4.3.1 GLM model of spiking response

Fig. 4.5 shows results of fitting a GLM model to responses of 13 spike basis functions spread across lamina. The stimulus component consisted of two movies, a long whitened natural movie and its pixel-wise square, indicated by x and x^2 in Fig. 4.5a. As a result, each spike unit has two spatiotemporal receptive fields associated with it, corresponding to the two movies. Several of the spatial receptive fields show fine oriented, bandpass structure, similar to the corresponding STA kernels in Fig. 4.4. One unit, an LGN axon terminal, has a circular receptive field with a biphasic temporal response. The squared movie kernels did not show structure in this case. The temporal course is similar for most cells with a full-width half-maximum of approximately 60 ms.

The spike-spike interaction terms are shown in Fig. 4.5b. Inset plots are temporal convolutional kernels, forced to zero at $t = 0$ to ensure causality. These model parameters indicate how likely a cell is to fire given its own past firing as well as the firing of the rest of the population during the previous 54 ms. The dominant structure along the diagonal indicates that many cells have a tendency to fire in bursts. Distinct subpopulations of cells in the granular and deep layers have inhibitory kernels other, whereas other groups between superficial and granular layers had excitatory couplings. These patterns of coupling hint at wiring structure that is in agreement with known neuroanatomy of this circuit[31]. The relative smoothness of kernels across neighboring cells may hint at homogeneity of cells within layers. However, this could also be a confound of cross-talk between neighboring units in the learning algorithm (Sec. 2.3.3).

Fig. 4.5c shows interactions between spiking and the LFP components. A clear pattern emerges where many of the LFP basis functions that have a localized negative peak are strong predictors of spiking in broadly the same layer as the corresponding spike basis function. Contamination of the LFP waveform with spike remnants might be at least partially responsible for this effect. Note, however, that the causality of the model prevents the trough of the LFP basis function to line up exactly with the spike times. Additionally, a significant number of basis functions show strong facilitation across layers.

Direct inspection of the model parameters can give important insights into circuit structure, but to validate if the model is capturing the structure of the data, model predictions were calculated to compare with actual spike rates. Fig. 4.6 shows prediction using GLM model fits across 60 trials of a 30 s natural movie. For two example cells shown, good agreement exists between the observed spike rates and the model predictions. The model is also successful at predicting trial-to-trial variability. It would have been impossible to achieve this kind of fit with a model that is based purely on the stimulus and did not take into account spiking of other neurons and the dynamics of ongoing activity in the LFP. In order to understand how different model components contribute to describing the data, a number of model fits were performed leaving certain components out. Fig. 4.7 shows log-likelihoods for these fits. The population spiking activity contributed far more than spike-LFP interac-

tions. Self-spiking interactions, a dominant feature in the model fits (Fig. 4.5), contributed less than half of the the total spike-spike interactions to the log-likelihood.

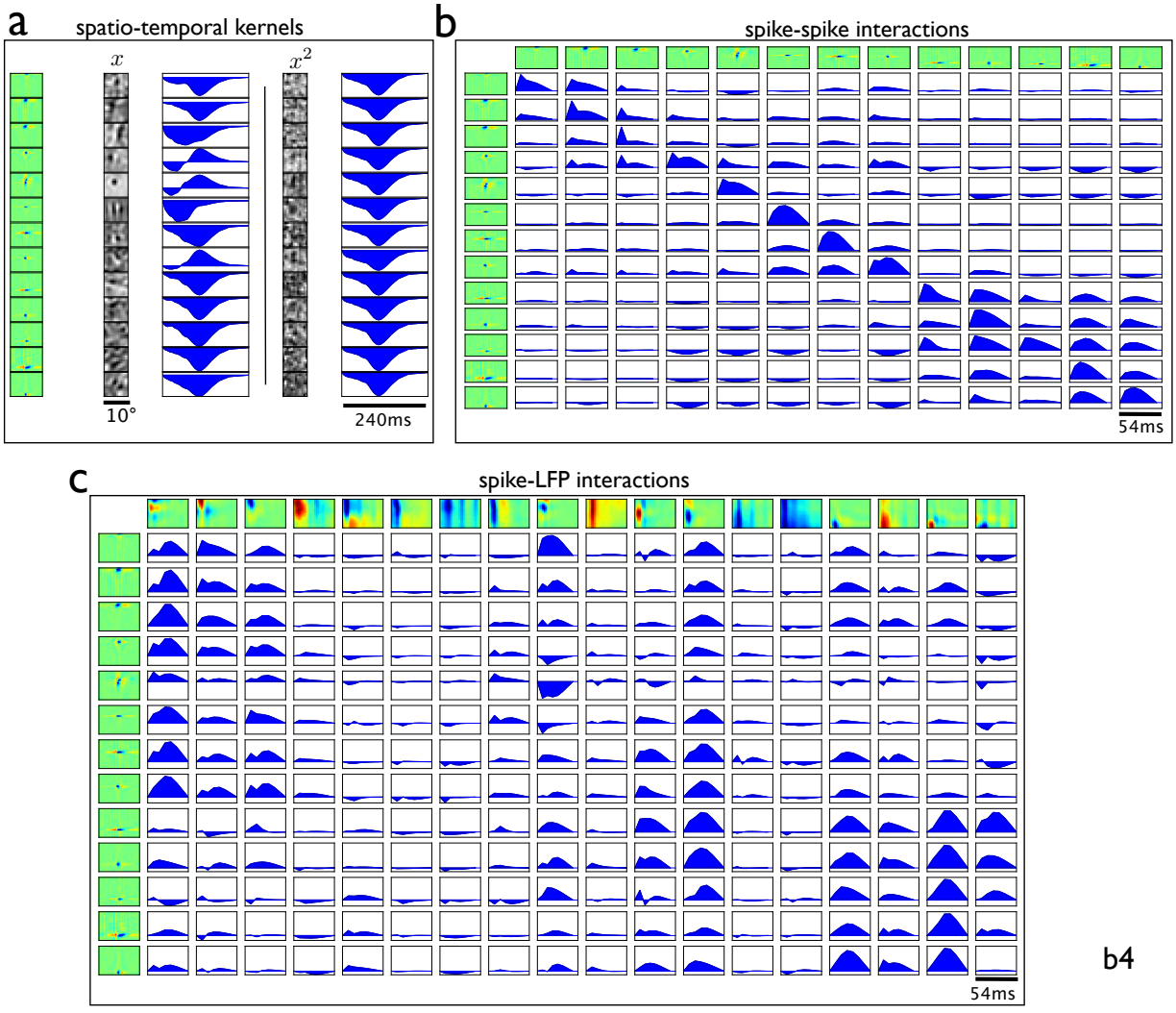

Figure 4.5: **Full spiking GLM model.** A GLM model fit for 13 spike units. **(a)** Two movie regressors were used, a whitened natural movie and its square. This corresponds to the two columns of spatial and temporal kernels. The waveforms of the basis functions are shown in green and are ordered by lamina. The temporal extent of the kernels is 240 ms. **(b)** A matrix of spike-spike interactions with each inset block a temporal convolution kernel of how much the basis functions along the horizontal axis influence activity in basis functions along the vertical axis. Self interactions dominate the spike-spike parameters. The kernels along each row are on the same vertical scale. **(c)** A matrix of spike-LFP interactions. LFP basis functions are plotted as convolution kernels aligned in time with fitted interaction kernels. The temporal span of spike-spike and spike-LFP kernels is 54 ms.

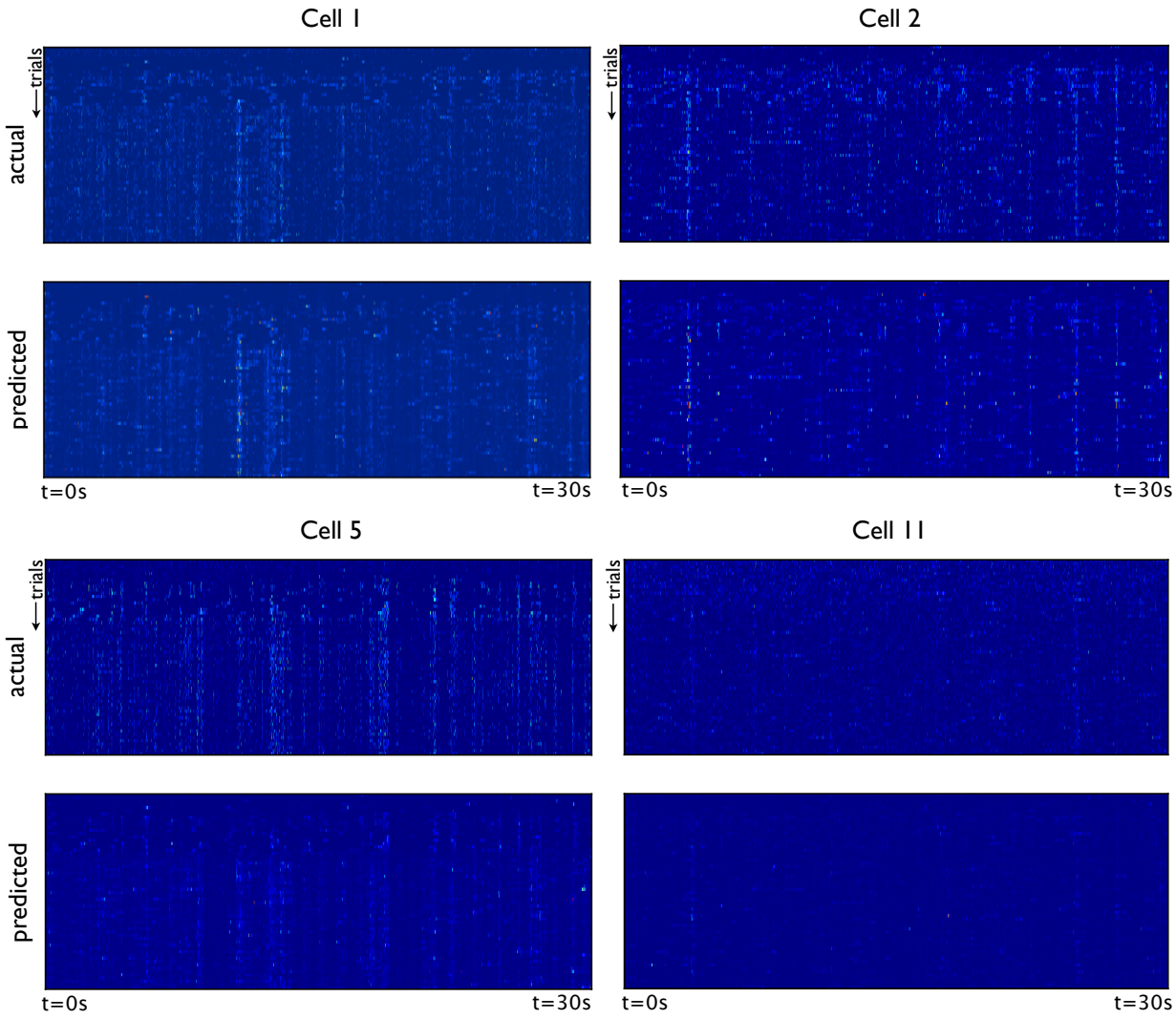

Figure 4.6: **Example model predictions.** A GLM model fit to a long natural movie was used to predict trial-to-trial spike coefficient responses to a different repeated natural movie. Here, the 4 cells from the population in Fig. 4.5 are shown. Each pair of plots show trial-by-trial coefficient activations. Predictions are conditional intensities $\lambda(t)$ computed using the GLM model. The model is able to correctly capture a large degree of the variance of responses, including the inter-trial structure.

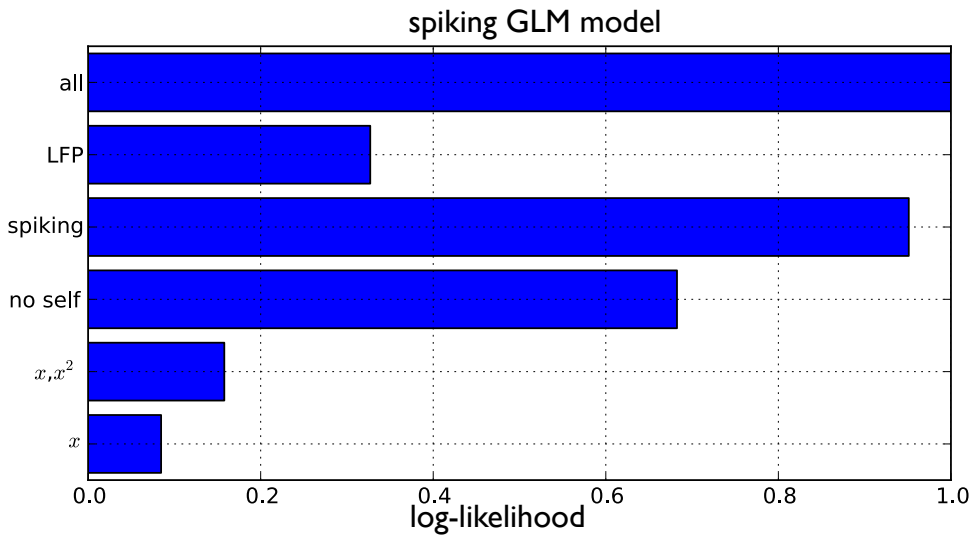

Figure 4.7: **Explanatory power of spike GLM components.** Log-likehoods for GLM fits on 11 minute subsets of the Duck30 recording session for models with different subsets of components, including, $\mathbf{x}$: whitened movie only, $\mathbf{x}, \mathbf{x^2}$: whitened movie and squared whitened movie only, **no self**: receptive field (RF) components $(\mathbf{x}, \mathbf{x^2})$ and spike-spike interactions, but no self-interactions, **spiking**: RF components and spike-spike interactions, **LFP**: RF and spike-LFP interactions, **all**: all model components. The bias portion of the log-likelihood has been subtracted out in all cases and log-likelihoods rescale to a maximum of 1.

4.3.2 GLM model of LFP response

LFP coefficient activity as a function of spiking, stimulus, and other LFP components was modeled using the same GLM framework used above. The LFP is thought to be generated mainly by dendritic processing of synaptic inputs[40] though this is a subject of much debate. Recently, new studies have formulated biophysically realistic forward models of a cortical column to better understand the LFP and its significance[32]. Our strategy is complementary as it provides a statistical approach based on real recordings that can test predictions of these models.

Figs. 4.8-4.9 show GLM model fits for 18 LFP basis functions with local temporal structure. LFP basis functions with extended temporal structure are difficult to interpret in this setting of a causal predictive model and were not considered in the analysis. However, cross-correlograms of these basis functions with spiking (Fig. 4.2) show clear structure and we hope to include them in future work. In contrast to the spiking spatio-temporal kernels, the LFP kernels did not have structure to the whitened movie. However, it did have structure to the squared movie. LFP-LFP coefficient interaction terms show interesting temporal ordering of different LFP components. Basis functions with coherent positive peaks negatively inhibited basis functions with coherent negative peaks. Fig. 4.8c shows a matrix of LFP-spike couplings. LFP componentsare along the vertical axis and spike units ordered by lamina along the horizontal axis. The LFP coefficients with negative peaks in certain lamina tend to be active where there is spiking in the corresponding lamina. The converse is true for LFP coefficients with positive peaks. However, this component of the model does not make a significant contribution to the GLM fits and it is unclear if the fine structure is significant. Fig. 4.10 shows log-likelihoods for model runs with different subsets of model components to assess which components contributed most towards the fit. The LFP coefficients seem to be best modeled by the activity of other LFP basis functions and not by the stimulus or spiking activity across lamina.

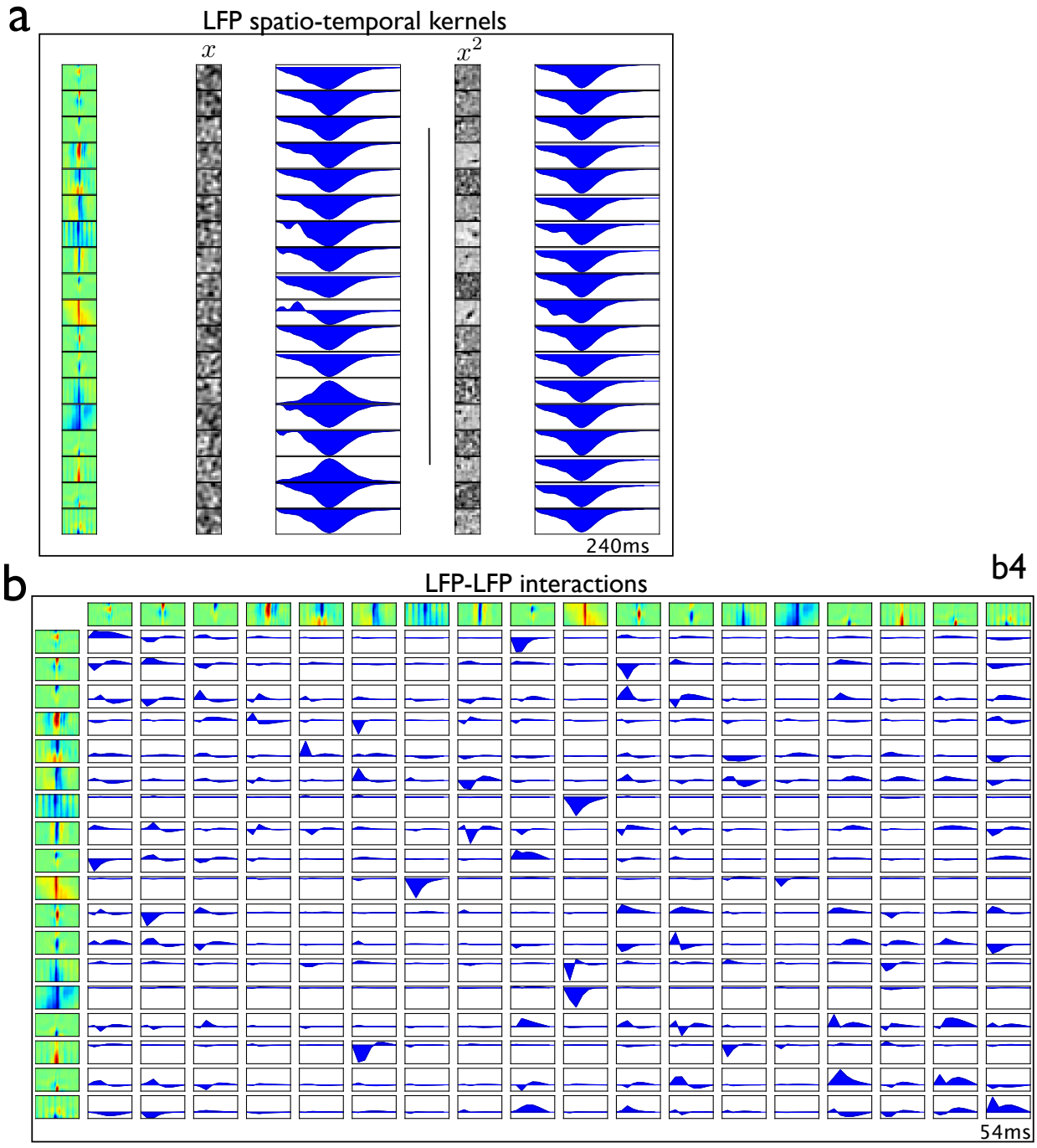

Figure 4.8: **LFP GLM model fit.** A GLM model fit for 18 LFP components. **(a)** LFP spatio-temporal kernels for whitened movie and whitened movie squared. The temporal extent of the kernels is 240 ms. **(b)** LFP-LFP interaction kernels, with time axis spanning 54 ms. LFP activations for basis functions along the horizontal occur at $t = 0$ and are not aligned with the temporal kernels.

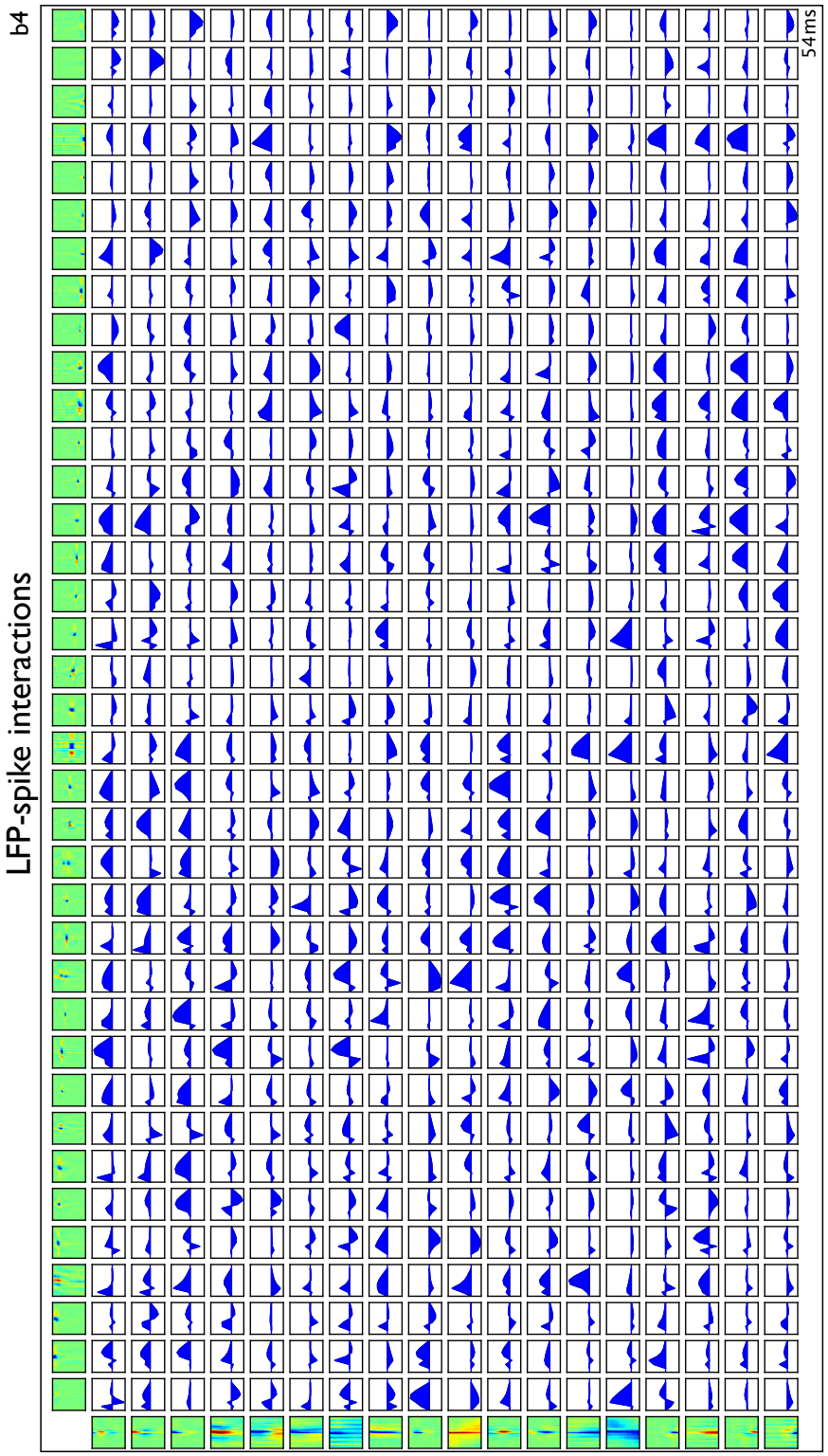

Figure 4.9: **GLM model fit for LFP.** Full LFP-spike interaction matrix for the model show in Fig. 4.8 showing the relationship between LFP components to spiking across all lamina. Spike events occur at $t = 0$ and the spike waveforms *do not* correspond to the temporal kernel as in previous plots and only provide illustration of their laminar locations. The kernels are 54ms in extent.

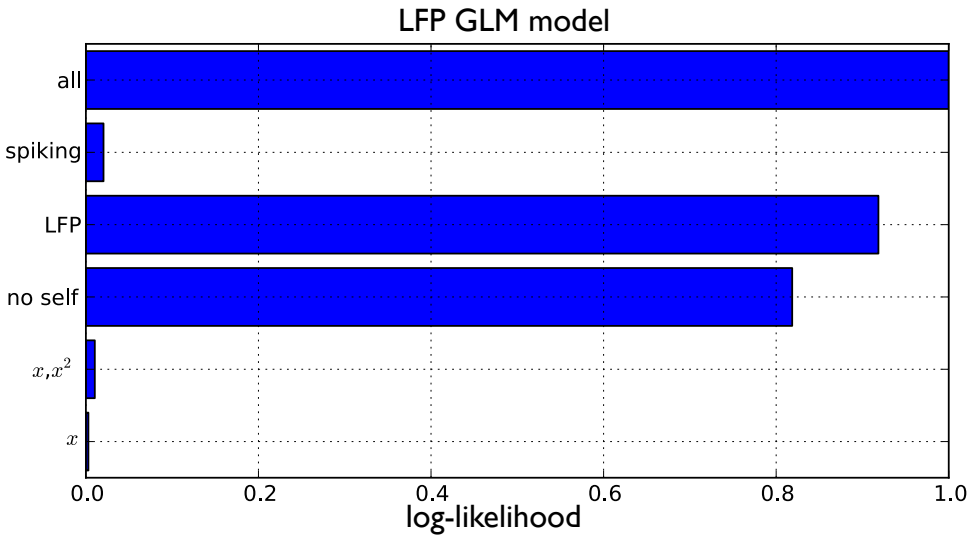

Figure 4.10: **Explanatory power of LFP GLM components.** Log-likehoods for GLM fits on 11 minute subsets of the Duck30 recording session for models with different subsets of components, including, $\mathbf{x}$: whitened movie only $\mathbf{x}, \mathbf{x^2}$: whitened movie and squared whitened movie only, **no self**: RF components and LFP-LFP interactions, but no self-interactions, **LFP**: RF and LFP-LFP interactions, **spiking**: RF and spike-LFP interactions, **all**: all model components. The bias portion of the log-likelihood has been subtracted out in all cases and log-likelihoods rescaled to a maximum of 1.

4.4 Discussion

In this chapter, we developed a statistical model of spiking and LFP activity in two GLM models that incorporate the stimulus as well as interactions between spiking and LFP components. For the spike GLM model, predictions for novel natural movies captured much of the structure including inter-trial differences. Spike-spike interactions contributed the most explanatory power of the model components while the stimulus component contributed surprisingly little. In the LFP GLM, the LFP-spike coupling fits showed interpretable structure, but had minimal explanatory power. This result was not expected given the structure in the spike-LFP cross-correlograms (Fig.4.2). However, if one assumes that the LFP reflects the constructive superposition of massive numbers of synaptic currents, it is unreasonable to expect that one could predict the LFP from the spiking activity of only 37 neurons.

These preliminary results establish the utility of an exponential GLM model of sparse coefficients. However there are a number of open questions and unresolved issues that remain. It would be useful for example to have a richer model of the stimulus that breaks up the image into different spatial and temporal scales and orientations. We also need a way to quantify the significance of different model parameters. In assessing prediction performance, we need to quantify log-likelihood on an absolute scale in terms of nats or bits so that we can compare to other models and datasets.

Chapter 5

Conclusions

5.1 Main contributions

In this work, I contribute several new tools for exploratory analysis of neural population activity which can be used to characterize the structure of laminar responses in visual cortex to dynamic, natural stimuli. Specifically,

1. I show how a novel application of sparse coding to laminar recordings can separate the complex statistical structure of these recordings into tractable and interpretable underlying causes.

2. I use the new representation afforded by the sparse coding algorithm to characterize spiking and LFP response in a cortical column to natural movies.

3. I create a framework for understanding population activity in a statistical model that accounts for network interactions as well as the driving influence of the stimulus.

5.2 Future directions

I plan to expand on the work in this thesis in several directions. The sparse coding model used as the basis for much of our results was formulated deliberately to be as minimal as possible to make interpretation when applied to complex neural data feasible. The only assumptions made were that causes in the data linearly sum and that they are sparsely active. It was intended to demonstrate that even a simple unsupervised learning algorithm that takes into account the actual statistics of the data can have far greater exploratory power than existing methods. However, having established its efficacy in this work, a wide range of augmented formulations are possible that could enhance performance for certain applications such as spike sorting. These include different forms of the objective such as space-time

separable basis functions, autoregressive terms to better model oscillatory phenomenon, and hierarchical models that attempt to directly characterize the statistical dependencies between sparse activations of different components.

The methods in this thesis can be applied to a wide range of large-scale neural recordings. In particular, we have applied them to semi-chronic electrode array recordings in awake monkey visual cortex and recordings from hippocampus using multi-shank polytrodes. All algorithms in this thesis were implemented in parallel, which made it possible to apply to large recordings in a time that scaled with the number of computational nodes used. Therefore, they can be applied to datasets of unprecedented size, removing key constraints in the design of future experiments. For example, it enables use of long natural movies that cannot be loaded into the memory of a single workstation and to perform optimizations and model fits with entire datasets that can be distributed on a large cluster.

As electrophysiological recording technology improves and datasets grow ever larger and more complex, scalable unsupervised learning algorithms will be an essential tool for advancing our understanding of cortical function.

Bibliography

[1] G. Andrew and J. Gao. Scalable training of L_1-regularized log-linear models. In *Proc. Int. Conf. Machine Learning*, pages 33–40. ACM, 2007.

[2] M. Armstrong, D. Flynn, M. Hammond, S. Jolly, and R. Salmon. High frame-rate television. *SMPTE Mot. Imag. J*, (117):54–59, 2009.

[3] J. J. Atick and A. Redlich. What does the retina know about natural scenes? *Neural Comp.*, 4(2):196–210, 1992.

[4] F. Attneave. Some informational aspects of visual perception. *Psychol. Rev.*, 61(3):183–93, 1954.

[5] C. Bachoc and F. Vallentin. New upper bounds for kissing numbers from semidefinite programming. *Arxiv preprint math/0608426*, 2006.

[6] C. Bachoc, G. Nebe, F. Vallentin, et al. Lower bounds for measurable chromatic numbers. *Arxiv preprint arXiv:0801.1059*, 2008.

[7] R. Baddeley, L. F. Abbott, M. C. Booth, F. Sengpiel, T. Freeman, E. A. Wakeman, and E. T. Rolls. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. Biol. Sci.*, 264(1389):1775–83, 1997.

[8] R. Baraniuk, V. Cevher, and M. Wakin. Low-dimensional models for dimensionality reduction and signal recovery: A geometric perspective. *Proc. IEEE*, 98(6):959–971, 2010.

[9] H. Barlow. *Sensory Communication*, chapter Possible principles underlying the transformation of sensory messages, pages 217–234. MIT Press, 1961.

[10] S. Behnel, R. Bradshaw, C. Citro, L. Dalcin, D. Seljebotn, and K. Smith. Cython: The best of both worlds. *Computing in Science & Engineering*, (99):1–1, 2011.

[11] A. Bell and T. Sejnowski. The "independent components" of natural scenes are edge filters. *Vision Res.*, 37(23):3327–3338, 1997.

[12] A. Benucci, D. L. Ringach, and M. Carandini. Coding of stimulus sequences by population responses in visual cortex. *Nat. Neurosci.*, 12(10):1317–24, 2009.

[13] C. Bishop. *Pattern recognition and machine learning*. Springer New York, 2006.

[14] T. J. Blanche, M. A. Spacek, J. F. Hetke, and N. V. Swindale. Polytrodes: high-density silicon electrode arrays for large-scale multiunit recording. *J. Neurophys.*, 93(5):2987–3000, 2005.

[15] D. D. Bock, W.-C. A. Lee, A. M. Kerlin, M. L. Andermann, G. Hood, A. W. Wetzel, S. Yurgenson, E. R. Soucy, H. S. Kim, and R. C. Reid. Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, 471(7337):177–82, 2011.

[16] H. Bokil, P. Andrews, J. E. Kulkarni, S. Mehta, and P. P. Mitra. Chronux: a platform for analyzing neural signals. *J. Neurosci. Methods*, 192(1):146–51, 2010.

[17] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge Univ. Press, 2004.

[18] D. H. Brainard. The psychophysics toolbox. *Spat. Vis.*, 10(4):433–6, 1997.

[19] G. Buzsáki and A. Kandel. Somadendritic backpropagation of action potentials in cortical pyramidal cells of the awake rat. *J. Neurophys.*, 79(3):1587–91, 1998.

[20] C. F. Cadieu and B. A. Olshausen. Learning transformational invariants from natural movies. In *NIPS*, volume 21. MIT Press, 2009.

[21] E. J. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[22] M. Carandini, J. B. Demb, V. Mante, D. J. Tolhurst, Y. Dan, B. A. Olshausen, J. L. Gallant, and N. C. Rust. Do we know what the early visual system does? *J. Neurosci.*, 25(46):10577–97, Nov 2005.

[23] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Jour. Sci. Comp.*, 20(1):33–61, 1999.

[24] A. Collette. Hdf5 for python. http://h5py.alfven.org/, 2008.

[25] O. D. Creutzfeldt and H. C. Nothdurft. Representation of complex visual stimuli in the brain. *Naturwissenschaften*, 65(6):307–18, 1978.

[26] B. J. Culpepper, C. F. Cadieu, J. Sohl-Dickstein, and B. A. Olshausen. Building a better probabilistic model of images by factorization. *ICCV*, 13, 2011.

[27] S. V. David and J. L. Gallant. Predicting neuronal responses during natural vision. *Network*, 16(2-3):239–60, 2005.

[28] G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *Constructive approximation*, 13(1):57–98, 1997.

[29] E. Doi, D. C. Balcan, and M. S. Lewicki. Robust coding over noisy overcomplete channels. *IEEE Trans. Image Process.*, 16(2):442–52, 2007.

[30] D. W. Dong and J. J. Atick. Statistics of natural time-varying images. *Network*, 6(3):345–358, 1995.

[31] R. J. Douglas and K. A. C. Martin. Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.*, 27:419–51, 2004.

[32] G. T. Einevoll, K. H. Pettersen, A. Devor, I. Ulbert, E. Halgren, and A. M. Dale. Laminar Population Analysis: Estimating firing rates and evoked synaptic activity from multielectrode recordings in rat barrel cortex. *J. Neurophys.*, 97(3):2174–2190, 2007.

[33] G. T. Einevoll, F. Franke, E. Hagen, C. Pouzat, and K. D. Harris. Towards reliable spike-train recordings from thousands of neurons with multielectrodes. *Curr. Opin. Neurobio., to appear*, 2011.

[34] G. Felsen and Y. Dan. A natural approach to studying vision. *Nat. Neurosci.*, 8(12):1643–6, 2005.

[35] G. Felsen, J. Touryan, F. Han, and Y. Dan. Cortical sensitivity to visual features in natural scenes. *PLoS Biol.*, 3(10):e342, 2005.

[36] FFmpeg video transcoder. http://ffmpeg.org.

[37] G. D. Field, J. L. Gauthier, A. Sher, M. Greschner, T. A. Machado, L. H. Jepson, J. Shlens, D. E. Gunning, K. Mathieson, W. Dabrowski, L. Paninski, A. M. Litke, and E. J. Chichilnisky. Functional connectivity in the retina at the resolution of photoreceptors. *Nature*, 467(7316):673–7, 2010.

[38] D. J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4(12):2379–2394, 1987.

[39] P. Földiák. Forming sparse representations by local anti-hebbian learning. *Biol. Cybern.*, 64(2):165–70, 1990.

[40] J. A. Freeman and C. Nicholson. Experimental optimization of current source-density technique for anuran cerebellum. *J. Neurophys.*, 38(2):369–82, 1975.

[41] P. Fries. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends in Cog. Sci.*, 9(10):474–480, 2005.

[42] D. George and J. Hawkins. Towards a mathematical theory of cortical micro-circuits. *PLoS Comput. Biol.*, 5(10):e1000532, 2009.

[43] M. X. Goemans and D. P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *J. ACM*, 42:1115–1145, 1995.

[44] C. M. Gray and W. Singer. Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *PNAS*, 86(5):1698–702, 1989.

[45] C. M. Gray, P. König, A. K. Engel, and W. Singer. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 338(6213):334–337, 1989.

[46] C. M. Gray. The temporal correlation hypothesis review of visual feature integration: still alive and well. *Neuron*, 24:31–47, 1999.

[47] Handbrake video transcoder. http://handbrake.fr/.

[48] K. D. Harris, D. A. Henze, J. Csicsvari, H. Hirase, and G. Buzsáki. Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J. Neurophys.*, 84(1):401–14, 2000.

[49] L. Hazan, M. Zugaro, and G. Buzsáki. Klusters, NeuroScope, NDManager: a free software suite for neurophysiological data processing and visualization. *J. Neurosci. Methods*, 155(2):207–16, 2006.

[50] E. B. Hendrickson, J. R. Edgerton, and D. Jaeger. The capabilities and limitations of conductance-based compartmental neuron models with reduced branched or unbranched morphologies and active dendrites. *J. Comput. Neurosci.*, 30(2):301–21, 2011.

[51] D. Hilbert. Über die stetige abbildung einer line auf ein flächenstück. *Mathematische Annalen*, 38:459–460, 1891. 10.1007/BF01199431.

[52] M. Hines and N. Carnevale. The neuron simulation environment. *Neural Comp.*, 9(6):1179–1209, 1997.

[53] G. R. Holt and C. Koch. Electrical interactions via the extracellular potential near cell bodies. *J. Comput. Neurosci.*, 6(2):169–184.

[54] J. Huang and D. Mumford. Statistics of natural images and models. In *IEEE Conf. Comp. Vis. Pattern Rec.*, volume 1. IEEE, 1999.

[55] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, 160:106–54, 1962.

[56] D. H. Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.*, 195(1):215–43, 1968.

[57] D. H. Hubel and T. N. Wiesel. Ferrier lecture. Functional architecture of macaque monkey visual cortex. *Proc. Biol. Sci.*, 198(1130):1–59, 1977.

[58] J. Hunter. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, pages 90–95, 2007.

[59] J. Jin, Y. Wang, R. Lashgari, H. A. Swadlow, and J.-M. Alonso. Faster thalamocortical processing for dark than light visual targets. *J. Neurosci.*, 31(48):17471–9, 2011.

[60] J. Jin, Y. Wang, H. A. Swadlow, and J. M. Alonso. Population receptive fields of ON and OFF thalamic inputs to an orientation column in visual cortex. *Nat. Neurosci.*, 14(2):232–8, 2011.

[61] Y. Karklin and M. Lewicki. Emergence of complex cell properties by learning to generalize in natural scenes. *Nature*, 457(7225):83–86, 2008.

[62] N. Karmarkar. A new polynomial-time algorithm for linear programming. In *Proc. Theory Comp.*, pages 302–311. ACM, 1984.

[63] R. Kelly, M. Smith, R. Kass, and T. Lee. Accounting for network effects in neuronal responses using L1 regularized point process models. *NIPS*, 23:1099–1107, 2010.

[64] A. Klöckner, N. Pinto, Y. Lee, B. Catanzaro, P. Ivanov, and A. Fasih. PyCUDA: GPU run-time code generation for high-performance computing. *Arxiv preprint arXiv:0911.3456*, 2009.

[65] H. Ko, S. B. Hofer, B. Pichler, K. A. Buchanan, P. J. Sjöström, and T. D. Mrsic-Flogel. Functional specificity of local synaptic connections in neocortical networks. *Nature*, 473(7345):87–91, 2011.

[66] M. S. Lewicki and B. A. Olshausen. Probabilistic framework for the adaptation and comparison of image codes. *J. Opt. Soc. Am. A*, 16(7):1587–1601, 1999.

[67] M. S. Lewicki. A review of methods for spike sorting: the detection and classification of neural action potentials. *Network*, 9(4):53–78, 1998.

[68] M. S. Lewicki. Efficient coding of natural sounds. *Nature Neuro.*, 5(4):356–363, 2002.

[69] H. Linden, E. Hagen, S. Leski, E. S. Norheim, K. H. Pettersen, and G. T. Einevoll. LFPy: A tool for simulation of extracellular potentials. In *Neuroinformatics*, 2011.

[70] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Math. Prog.*, 45(3):503–528, 1989.

[71] Z. F. Mainen, J. Joerges, J. R. Huguenard, and T. J. Sejnowski. A model of spike initiation in neocortical pyramidal neurons. *Neuron*, 15(6):1427–39, 1995.

[72] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online learning for matrix factorization and sparse coding. *JMLR*, 11:19–60, 2010.

[73] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Processing*, 41(12):3397–3415, 1993.

[74] M. E. Mazurek and M. N. Shadlen. Limits to the temporal fidelity of cortical spike rate signals. *Nat. Neurosci.*, 5(5):463–71, 2002.

[75] U. Mitzdorf. Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. *Physiol Rev*, 65(1):37–100, 1985.

[76] MPlayer video encoder. http://www.mplayerhq.hu.

[77] I. E. Nesterov and A. S. Nemirovskiĭ. *Interior-point polynomial algorithms in convex programming*, volume 13 of *SIAM Stud. Applied Math.* Soc. Indus. Appl. Math., Philadelphia, 1994.

[78] NeuroNexus Technologies, Inc. http://www.neuronexus.com/.

[79] L. G. Nowak, R. Azouz, M. V. Sanchez-Vives, C. M. Gray, and D. A. McCormick. Electrophysiological classes of cat primary visual cortical neurons in vivo as revealed by quantitative analyses. *J. Neurophys.*, 89(3):1541–66, 2003.

[80] N. Okazaki. libLBFGS: a library of Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS). http://www.chokkan.org/software/liblbfgs/.

[81] T. E. Oliphant. Python for scientific computing. *Computing in Science & Engineering*, 9:90, 2007.

[82] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

[83] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vis. Res.*, 37:3311–3325, 1997.

[84] B. A. Olshausen and D. J. Field. Sparse coding of sensory inputs. *Curr. Opin. Neurobio.*, 14(4):481–7, 2004.

[85] B. A. Olshausen and D. J. Field. How close are we to understanding V1? *Neural Comp.*, 17(8):1665–99, 2005.

[86] B. A. Olshausen, J. Baker, S.-C. Yeng, and C. M. Gray. Receptive field models fail to predict responses of V1 neurons to natural movies. Soc. Neurosc., 2004.

[87] B. A. Olshausen. Learning sparse, overcomplete representations of time-varying natural images. In *Int. Conf. Image Proc.*, volume 1, pages 1–41. IEEE, 2003.

[88] L. Paninski. Maximum likelihood estimation of cascade point-process neural encoding models. *Network*, 15(4):243–62, 2004.

[89] F. Perez and B. E. Granger. IPython: A system for interactive scientific computing. *Computing in Science & Engineering*, 9:90, 2007.

[90] A. Peters and B. R. Payne. Numerical relationships between geniculocortical afferents and pyramidal cell modules in cat primary visual cortex. *Cereb Cortex*, 3(1):69–78, 1993.

[91] A. Peters, B. R. Payne, and J. Budd. A numerical analysis of the geniculocortical input to striate cortex in the monkey. *Cereb Cortex*, 4(3):215–29, 1994.

[92] K. H. Pettersen, A. Devor, I. Ulbert, A. M. Dale, and G. T. Einevoll. Current-source density estimation based on inversion of electrostatic forward solution: effects of finite extent of neuronal activity and conductivity discontinuities. *J. Neurosci. Methods*, 154(1-2):116–33, 2006.

[93] K. H. Pettersen, H. Lindén, A. M. Dale, and G. T. Einevoll. Extracellular spikes & multi-electrode recordings. *Handbook of Neural Activity Measurement, edited by Romain Brette and Alain Destexhe*, pages 1–44, 2008.

[94] J. D. Pettigrew, M. L. Cooper, and G. G. Blasdel. Improved use of tapetal reflection for eye-position monitoring. *Invest. Ophthalmol. Vis. Sci.*, 18(5):490–5, 1979.

[95] J. W. Pillow, J. Shlens, L. Paninski, A. Sher, A. M. Litke, E. J. Chichilnisky, and E. P. Simoncelli. Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–9, 2008.

[96] W. Press, B. Flannery, S. Teukolsky, W. Vetterling, et al. *Numerical Recipes*. Cambridge Univ. Press, 1986.

[97] A. D. Redish. MClust. http://redishlab.neuroscience.umn.edu/MClust/MClust.html.

[98] O. Schwartz, J. W. Pillow, N. C. Rust, and E. P. Simoncelli. Spike-triggered neural characterization. *J. Vis.*, 6(4):484–507, 2006.

[99] D. Seljebotn. Fast numerical computations with cython. In *Proc. 8th Python in Sci. Conf.*, 2009.

[100] E. P. Simoncelli and B. A. Olshausen. Natural image statistics and neural representation. *Annu. Rev. Neurosci.*, 24:1193–216, 2001.

[101] W. Singer and C. M. Gray. Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.*, 18:555–86, 1995.

[102] E. C. Smith and M. S. Lewicki. Efficient auditory coding. *Nature*, 439(7079):978–82, 2006.

[103] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Royal Stat. Soc. B*, pages 267–288, 1996.

[104] J. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Trans. Inform. Theory*, 52(3):1030–1051, 2006.

[105] W. Truccolo, U. T. Eden, M. R. Fellows, J. P. Donoghue, and E. N. Brown. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J. Neurophys.*, 93(2):1074–89, 2005.

[106] W. E. Vinje and J. L. Gallant. Sparse coding and decorrelation in primary visual cortex during natural vision. *Science*, 287(5456):1273–6, 2000.

[107] W. E. Vinje and J. L. Gallant. Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *J. Neurosci.*, 22(7):2904–15, 2002.

[108] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.*, 1:1–305, 2008.

[109] P. E. Williams, F. Mechler, J. Gordon, R. Shapley, and M. J. Hawken. Entrainment to video displays in primary visual cortex of macaque and humans. *J. Neurosci.*, 24(38):8278–88, 2004.

[110] D. E. Wollman and L. A. Palmer. Phase locking of neuronal responses to the vertical refresh of computer display monitors in cat lateral geniculate nucleus and striate cortex. *J. Neurosci. Methods*, 60(1-2):107–113, 1995.

[111] S.-C. Yen, J. Baker, and C. M. Gray. Heterogeneity in the responses of adjacent neurons to natural stimuli in cat striate cortex. *J. Neurophys.*, 97(2):1326–41, 2007.