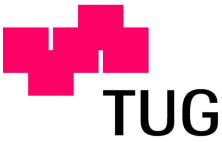
Analysis of Neural Signals: Interdependence, Information Coding, and Relation to Network Models



Malte J. Rasch

Institute for Theoretical Computer Science

Graz University of Technology

Doctoral thesis

Inauguraldissertation

Graz, April 2008

Acknowledgements

Foremost, I would like to thank my supervisors Wolfgang Maass and Nikos Logothetis for inspiring my research, and for their continuous support and guidance.

Many thanks go to Arthur Gretton, Stefano Panzeri, Yusuke Murayama, Karsten Borgwardt, Marcelo Montemurro, Alex Smola, Bernard Schölkopf, Martin Winter, Horst Bischof, and Zoltan Kisvarday for enjoyable and fruitful collaborations.

Special thank to Klaus Schuch for help with Python and PCSIM, Stefan Häusler, for many interesting philosophical discussions, and to all the colleagues at IGI Graz and MPI Tuebingen, for enjoyable working conditions.

I thank Andreas Tolias and Georgios Keliris for generously providing the awake data set in Chapter 2.

Last but not least I would like to thank the Austrian Science Fund FWF, the FACETS project of the European Union and the Graz University of Technology for their generous financial support.

Abstract

Most of the functionality of the brain is thought to emerge from communications between neurons using spikes. Thus, for investigating neural function it is essential to monitor spiking activity of neurons. Experimentally measured signals, however, do often not directly reflect spiking activity of neurons, but instead comprise a mixture of biophysical events from various origins. In this dissertation I investigate the interdependence of two signals: extracellularly measured spiking activity and local field potentials.

First, by means of machine learning techniques I ask to what degree spike trains can be inferred from simultaneously measured local field potentials from primary visual cortex and lateral geniculate nucleus of non-anesthetized and anesthetized macaque monkeys.

Second, using an information theoretic approach I further show that in primary visual cortex of macaque monkeys, spikes are related to LFP oscillations in a stimulus dependent manner. In particular, information about a scene in a movie can be predicted with higher precision from spike trains, when the phase of local field potential oscillations is taken into account.

The structure of experimental spike trains in response to natural scenes is rich, with periods of reliable high activity bursts intermingled with long silent periods. I develop a neural network model based on many anatomical particularities of the primary visual cortex of macaques in order to compare the statistics of the spike trains generated by an artificial neural network under similar stimulus conditions. To achieve a close match to the data free parameters of the model are optimized using a new method for comparing multi-dimensional distributions, called Maximum Mean Discrepancy (MMD).

Zusammenfassung

Es wird im Allgemeinen angenommen, daß die Funktionsweise des Gehirns hauptsächlich auf der Kommunikation mittels Aktionspotentialen zwischen Neuronen beruht. Daher ist es für die Untersuchung neuronaler Funktion essentiell, das Feuern von Neuronen zu beobachten. Experimentell gemessene Signale sind jedoch häufig nicht unmittelbar auf die Aktivität von Neuronen zurückführbar, sondern setzen sich aus verschiedensten biophysikalischen Vorgängen und Quellen zusammen. In dieser Dissertation untersuche ich die Abhängigkeit zwischen zwei solcher Signale: extrazellulär gemessene Aktivität neuronaler Impulse und lokale Feldpotentiale (LFP).

Zunächst untersuche ich mittels Techniken des Maschinellen Lernens inwieweit neuronale Impulse anhand der lokalen Feldpotentiale vorhergesagt werden können. Die neuronalen Daten sind im primären visuellen Areal sowie im seitlichen Kniehöcker von anästhesierten und unanästhesierten Rhesusaffen aufgenommen wurden.

Danach zeige ich mittels eines Ansatzes aus der Informationstheorie, daß im primären visuellen Kortex verschiedene Stimuli eine verläßliche Änderung der Phasenbeziehung neuronaler Impulse zu tiefen Frequenzen der LFPs hervorrufen. Das führt dazu, daß Szenen eines Films weit besser aus der Feuerrate hervorgesagt werden können, wenn zusätzlich die Phasenbeziehungen zu LFPs beachtet werden.

Die Struktur experimentell gemessener Zeitreihen neuronaler Impulse, welche durch die Stimulation mit Filmen natürlicher Szenerien erzeugt werden, ist komplex und variabel: Häufig folgen Phasen verläßlich hoher Aktivität lange Zeiten der relativen Stille. Ich entwickle ein künstliches neuronales Netzwerk, das viele anatomische Details des primären visuellen Areals von Rhesusaffen berücksichtigt, um die Statistik der Impulszeitreihen des künstlichen Netzwerks mit denen der Realität unter ähnlichen Stimulusbedingungen zu vergleichen. Um eine möglichst nahe Übereinkunft zu erreichen, werden freie Parameter des Netzwerks mittels einer neuen Methode zum Vergleich multidimensionaler Wahrscheinlichkeitsverteilungen, kurz Maximale Mittelwert Diskrepanz (MMD), optimiert.

Contents

List of Figures

Chapter 1

Introduction

> "*Daß in der morphologischen Kette der Thiere das Nervensystem and später das Gehirn sich* entwickelt*: giebt einen Anhaltspunkt – es* entwickelt *sich das Fühlen, wie sich später das Bilderschaffen und Denken entwickelt. Ob wir es schon noch nicht begreifen: aber wir sehen, daß es so ist. Wir finden es unwahscheinlich, Lust und Schmerz schon in alles Organische zu versetzen: und es ist immer noch auch beim Menschen der Reiz eine Stufe,* beides *nicht da ist.*"
>
> [Friedrich Nietzsche, 1885]

Due to the small size of neural cells, their electrochemical communications, and the complexity of a neural system, it is impossible for us humans to directly perceive, observe or manipulate neural activity. Thus sophisticated techniques have been developed to indirectly access and visualize brain activity in its functional state. These techniques necessarily rely on measuring a particular aspect of neural systems and therefore not only show an incomplete picture of the underlying processes but may indeed be subject to artifacts not relevant for information processing.

One prominent example is BOLD-fMRI, a popular non-invasive technique that relies on changes in the blood oxygen level (Ogawa *et al.* (1992)). Changes in the blood oxygen is only a very indirect measurement of neural activity and it is not immediately clear how and to what extend it actually reflects brain processes. However, it has been shown by combined electrophysiological and fMRI measurements that the time course of the fMRI signals can be mapped to firing activity in a local area (Logothetis *et al.* (2001)). The correspondence is reflected best in its high correlation to the power in cortical local field potentials (LFP). Similarly, extracellular voltage measurements, such as LFP, not only record voltage fluctuations attributed to a single neuron but instead comprises the activity of many cells around the electrode tip together with various oscillatory activities originating from various sources. LFP is thus only a "crude" measure of neural activity (Henrie & Shapley (2005)). Out of these reasons, there is necessarily a strong interest in establishing the biophysical origins, the information content, and interdependence of different signals.

Commonly, all indirect signals are tried to relate to spiking activity of single neurons, which are often considered the actual carrier of information in the brain. However, in recent years the view is emerging that other signals, such as LFP, may partly originate from elements of the neural system, which are in itself interesting to observe in the sense that they contribute to the computational function of the neural system. For instances one could take advantage that LFP seems to reflect mostly synaptic activity (Logothetis (2003)), which means that also synaptic drives originating from further away or from completely different areas are locally observable.

In this dissertation I investigate some of these questions for two neural signals: extracellularly measured spiking activity and local field potentials from macaque monkeys. In Chapter 2 I analyze the interdependence of LFP and simultaneously recorded spikes by means of support vector machines and linear regression models. I show that

spiking activity is, on a larger time-scale, reliably related to the local field potentials. I use non-linear and multidimensional machine learning techniques to directly infer spiking activity from features of the LFP. By analyzing the features which are most important for this prediction task, I show further that different features of the LFP, i.e. high-γ power and the phase of low frequency oscillations, carry valuable information for spike prediction. Furthermore, both contributions to prediction performance are rather independent, suggesting that both features correspond to partly unrelated aspects of the neural circuit. For instance results suggest that low frequency oscillations are good predictors for clusters of spikes.

Having established and characterized the relationship of LFP to spiking activity in a wide range of data in detail, I ask in Chapter 3 whether this relation is relevant for information processing. Indeed, one finds that the gain in information about naturalistic scenes in a movie stimulus, by hypothesizing a phase-of-firing code rather than by a spike count code, is considerable in the same data of V1 of macaques. This phase-of-firing code relies on the relative position of spikes within a oscillation cycle of very low frequency components of the LFP (< 10 Hz). The extra information available in the phase of firing is crucial to disambiguate between stimuli eliciting high spike rates of similar magnitude. Thus, I challenge the view of LFP as an unimportant epiphenomenon, because the spiking activity in respect to the oscillation cycle of the LFP do indeed code for over 50% more information than can be extracted from the spike count alone.

Machine learning approaches applied to data analysis, as done in the first part of this dissertation, are very important, since they are able to find regularities in great amounts of data with minimal hypothesis bias from the investigator. A complimentary approach to investigate brain function is to understand underlying laws and principles by using mechanistic mathematical models. If a mathematical model captures the computational functions of the real system, it is capable to generalize its behavior and one might be close in truly understand the complex system.

Today, modeling all details of a neural system is still impossible. However, one goal when employing artificial neural networks is that at least the statistics of its activity is realistic. Asserting realistic spiking dynamics in a model network circuit is crucial, if one would like to use the network as a scaffold for investigating realistic information processing. Therefore in Chapter 5, I develop a model of the early visual

pathway (LGN and V1), bearing many anatomical details, and compare this state-of-the art network model with the same experimental dataset used in the previous chapters, which are recorded analogously from V1 and LGN. In both systems, in the model and the experimental data, I characterize statistics of the neural activity and compare trial-to-trial variability.

The V1 model extends over (at least) 5×5 mm cortical surface. It is adapted specifically to the anatomy of macaque monkeys. The vertical connection structure between layers builds on earlier work by Haeusler & Maass (2007), and incorporates realistic functional and structural connection profiles between cortical laminae. Moreover, preferred orientation of neural sites is used for defining the thalamic input projections as well as the superficial long-range connectivity.

Besides developing the network model and characterizing its spiking activity, I optimize parameters of the network model in order to match the firing statistics of the real system, by applying a new method for comparing efficiently multidimensional probability distributions. This new technique, using the Maximum Mean Discrepancy (MMD), is introduced in Chapter 4. Basically, the method finds a function (from a given function space), that, when applied to two random variables, shows greatest difference in the means. Using an (universal) recurrent kernel Hilbert space, I show how to use MMD to perform a statistical test for the two-sample problem, and test its performance against existing methods.

From the modeler's perspective it is interesting to see how far apart current network models are from the reality, as it is represented by experimental data. Since the network model of Chapter 5 and the monkeys in the experimental data are shown the very same movie stimuli, a direct comparison of data and model is possible. However, I find that experimental data is still far from being exactly reproduced by a state-of-the-art network model of the primary visual area, suggesting that many details of this complex system still remain unknown.

1.1 Publications and contributions

The data analyzed in all Chapters were recorded in the N. Logothetis' lab. They were generously made available to me for the research of this dissertation by N. Logothetis (Max-Planck Institute for Biological Cybernetics). Data were mostly recorded by Y. Murayama, but also other employees the lab were involved. For completeness of the presentation, I describe the recording setup in this thesis. Because the subset of data used and the preprocessing methods differ slightly from chapter to chapter, I include a brief experimental methods section describing data acquisition and preprocessing in each chapter. This achieves also a low degree of interdependence between chapters and thus allows to read each chapter rather independently.

The chapters are based on the following publication and contributions:

- Chapter 2 is based on the following publication:

  > RASCH, GRETTON, MURAYAMA, MAASS, & LOGOTHETIS (2008). Inferring spike trains from local field potentials. *Journal of Neurophysiology*, **99**(3), 1461–76

  Y. Murayama and N. Logothetis were responsible for collecting the experimental data. A. Gretton, W. Maass, and N. Logothetis supervised the work.

- Chapter 3 is based on the following publication:

  > MONTEMURRO, RASCH, MURAYAMA, LOGOTHETIS, & PANZERI (2008). Phase-of-Firing Coding of Natural Visual Stimuli in Primary Visual Cortex. *Current Biology*, **18**(5), 375–80.

  The techniques in information bias correction described in the method section of Chapter 3 were developed by the co-authors M. Montemurro and S. Panzeri. They are stated here for completeness. The bulk of the paper was composed in team work by M. Montemurro, M. Rasch and S. Panzeri, under the leadership of the first and the last author, which contributed most. Y. Murayama and N. Logothetis were responsible for collecting the experimental data.

- Chapter 4 is based on the following publications:

BORGWARDT, GRETTON, RASCH, KRIEGEL, SCHÖLKOPF, & SMOLA (2006). Integrating structured biological data by kernel maximum mean discrepancy. *Bioinformatics*, **22**(14), e49–e57.

GRETTON, BORGWARDT, RASCH, SCHÖLKOPF, & SMOLA(2007). A Kernel Method for the Two-sample-problem. *Advances in Neural Information Processing Systems*, eds. Schölkopf, Platt, Hofmann, **19**, Cambridge, MA, The MIT Press.

GRETTON, BORGWARDT, RASCH, SCHÖLKOPF, & SMOLA(2007). A Kernel Approach to Comparing Distributions. *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence (AAAI-07)*, 1637-1641, AAAI Press, Menlo Park, CA, USA.

GRETTON, BORGWARDT, RASCH, SCHÖLKOPF, & SMOLA (2008). A Kernel Method for the Two-sample-problem. *submitted to Journal of Machine Learning Resarch.*

The theoretical development and proofs of the Maximal Mean Discrepancy were done by the co-authors A. Gretton, A. Smola, K. Borgwardt, and B. Schölkopf. They are incorporated in this thesis for completeness. I participated in its refinement, implemented all statistical tests and ran most of the performance experiments.

- Chapter 3 is based on the following manuscript:

  RASCH, SCHUCH, HÄUSLER, LOGOTHETIS & MAASS (2008). Comparison of firing characteristics of a state-of-the-art network model of macaque V1 to experimentally recorded data under semi-natural movie stimulation. *Manuscript.*

  K. Schuch collaborated with the python implementation of the V1 model and S. Häusler contributed the basic generic microcircuit model framework. N. Logothetis was responsible for the experimental data and W. Maass supervised the work.

In the following chapters I will write in the first-person plural pronoun, to emphasize the collaborative character in the presentation.

Chapter 2

Inferring spike trains from local field potentials

We investigated whether it is possible to infer spike trains solely on the basis of the underling local field potentials (LFPs). Employing support vector machines and linear regression models, we found that in the primary visual cortex (V1) of monkeys, spikes can indeed be inferred from LFPs, at least with moderate success. Although there is a considerable degree of variation across electrodes, the low-frequency structure in spike trains (in the 100 ms range) can be inferred with reasonable accuracy, whereas exact spike positions are not reliably predicted. Two kinds of features of the LFP are exploited for prediction: the frequency power of bands in the high γ-range (40-90 Hz), and information contained in low-frequency oscillations (< 10 Hz), where both phase and power modulations are informative. Information analysis revealed that both features code (mainly) independent aspects of the spike-to-LFP relationship, with the low-frequency LFP phase coding for temporally clustered spiking activity. Although both features and prediction quality are similar during semi-natural movie stimuli and spontaneous activity, prediction performance during spontaneous activity degrades much more slowly with increasing electrode distance. The general trend of data obtained with anesthetized animals is qualitatively mirrored in that of a more limited data set recorded in V1 of awake monkeys. In contrast to the cortical field potentials, thalamic LFPs (e.g. LFPs derived from recordings in dLGN) hold no useful information for predicting spiking activity.

2.1 Introduction

In a typical electrophysiology experiment, the signal measured by an electrode placed at a neural site represents the mean extracellular field potential (mEFP) from the weighted sum of all current sinks and sources along multiple cells. If a microelectrode with a small tip is placed close to the soma or axon of a neuron, then the measured mEFP directly reports the spike traffic of that neuron and frequently that of its immediate neighbors as well. If the impedance of the microelectrode is sufficiently low and its exposed tip is a bit farther from a single large pyramidal cell, so that action potentials do not predominate the neural signal, then the electrode can monitor the totality of the potentials in that region. The EFPs recorded under these conditions are related both to integrative processes (dendritic events) and to spikes generated by several hundred neurons.

The two different signal types can be segregated by frequency band separation. A high-pass filter cutoff of approximately 300-500 Hz is used in most recordings to obtain multiple-unit spiking activity (MUA), and a low-pass filter cutoff of ca. 300 Hz to obtain the so-called local field potentials (LFPs). A large number of experiments have presented data indicating that such a band separation does indeed underlie different neural events (for references see for instance Logothetis (2003)).

In summary, depending on the recording site and the electrode properties, the MUA most likely represents a weighted sum of the extracellular action potentials of all neurons within a sphere of approximately 140-300 μm radius, with the electrode at its center (Henze *et al.* (2000)). Spikes produced by the synchronous firings of many cells can, in principle, be enhanced by summation and thus detected over a larger distance (Arezzo *et al.* (1979); Huang & Buchwald (1977)). In general, experiments have shown that large-amplitude signal variations in the MUA range reflect large-amplitude extracellular potentials and that small-amplitude fast activity is correlated with small ones (Buchwald & Grover (1970); Gasser & Grundfest (1939); Grover & Buchwald (1970); Hunt (1951); Nelson (1966)).

The low-frequency range (i.e. the LFPs) of the mEFP signal, on the other hand, represents mostly slow events reflecting cooperative activity in neural populations. Initially these signals were thought to represent exclusively synaptic events (Ajmone-Marsan (1965); Buchwald *et al.* (1965); Fromm & Bond (1964, 1967)). Evidence for their origin was often gathered from current-source density (CSD) analysis and combined field potential and intracellular recordings (Mitzdorf (1985); Nadasdy *et al.* (1998)). Mitzdorf has suggested that LFPs actually reflect a weighted average of synchronized dendro-somatic components of the synaptic signals of a neural population within 0.5-3 millimeters of the electrode tip (Juergens *et al.* (1999); Mitzdorf (1987)). Later studies, however, provided evidence of the existence of other types of slow activity unrelated to synaptic events, including voltage-dependent membrane oscillations (e.g. Kamondi *et al.* (1998)) and spike afterpotentials. Taken together, LFPs represent slow waveforms, including synaptic potentials, afterpotentials of somato-dendritic spikes, and voltage-gated membrane oscillations, that reflect the input of a given cortical areas as well as its local intracortical processing, including the activity of excitatory and inhibitory interneurons.

Given the different natures of LFPs and MUA, we felt that it would be interesting to address the question of whether one can infer spiking of neurons from the locally measured field potentials. Herein we address this question in a straightforward manner. We use methods derived from the field of machine learning in an attempt to infer exact spike timings from the underlying LFPs. We compare the accuracy of spikes trains predicted by supervised learning algorithms on a wide range of recordings from the primary visual cortex (V1) as well from the lateral geniculate nucleus (LGN) of anesthetized and awake macaques and investigate what kind of features of the LFP are important for inferring spikes from LFP.

2.2 Methods

2.2.1 Data acquisition

Electrophysiological data recorded from 9 anesthetized and 2 awake monkeys (*Macaca mulatta*) are included in the present study. All animal experiments were approved by the local authorities (Regierungspraesidium) and are in full compliance with the guidelines of the European Community (EUVD 86/609/EEC) for the care and use of laboratory animals. Surgical procedures are described elsewhere (Logothetis *et al.* (2002)).

To perform the neurophysiological recordings in anesthetized monkeys, the animals were anaesthetized (remifentanil (typical $1\mu\mathrm{g}\cdot\mathrm{kg}^{-1}\cdot\mathrm{min}^{-1}$)), intubated and ventilated. Muscle relaxation was achieved with mivacurium ($5\mathrm{mg}\cdot\mathrm{kg}^{-1}\cdot\mathrm{h}^{-1}$). Body temperature was kept constant, and lactated Ringer's solution was given at a rate of $10\mathrm{ml}\cdot\mathrm{kg}^{-1}\cdot\mathrm{h}^{-1}$. During the entire experiment, the vital signs of the monkey and the depth of anesthesia were continuously monitored. Drops of 1% ophthalmic solution of anticholinergic cyclopentolate hydrochloride were instilled into each eye to achieve cycloplegia and mydriasis. Refractive errors were measured and contact lenses (hard PMMA lenses by Whlk GmbH, Germany) with the appropriate dioptric power were used to bring the animal's eye into focus on the stimulus plane. Simultaneous recording of neural activities were made from the primary visual cortex using 8-16 electrodes configured in 4x4 or 2x8 matrices in a grid of 1-2mm. Electrode tips were typically (but not always) positioned in the upper or middle cortical layers. The impedance

of the electrode varied from 300 to 800 kOhm. In the case of simultaneous LGN recording, an additional set of drives usually comprising 2-4 electrodes was additionally positioned. The electronic interface, including drives, holder and preamplifier, was custom-designed to minimize cross-talk of signals between electrodes (typically about 1ppm). The signals were amplified and filtered into a band of 1-8 kHz (Alpha Omega Engineering, Nazareth, Israel) and then digitized at 21kHz with 16bit resolution (National Instruments Co., TX USA), ensuring enough resolution for both local field and spiking activities. Binocular visual stimulation was provided through a two-fiber optic system (Avotec Inc., FL USA) after fine alignment to each of the animal's foveas by a modified retinoscope coupled with a stimulus projector holder.

In the case of the anesthetized animals we differentiate between two different conditions: spontaneous activity ("spo") and movie-driven data ("stm"). In the former the input screen is blank for about 5 minutes. In the latter a 4-6 minute segment of a commercially available movie is shown. Movie frames were synchronized with the refresh rate of the monitor (60 Hz, two syncs per movie frame) and covered 7-12 degrees of the visual field. Most of the electrodes were confirmed to have a receptive field within the movie presentation area (see Fig. 2.2.1 B for an example). Multiple trials of movie presentations and spontaneous activity are run within one recording session (intermingled with recordings of other stimuli not considered here). From these data we include 1304 recorded time series in the present study, which we call trials throughout this paper. The data set comprises recordings from 9 animals collected in 12 recording sessions. From each session we take 5 repeats of movie presentation and 5 repeats of spontaneous activity trials (with the exception of j97nm1, where only one movie trial is available). To avoid any subjective selection bias all measured electrode channels per session are included. This results in 670 trials for spontaneous activity and 634 for movie stimulus recorded using 134 electrode placements. Movies are identical within a session but may differ between sessions.

In 3 sessions (two anesthetized monkeys) up to 4 electrodes were simultaneously placed in the LGN. Thus stimuli reflected in these data (100 trials) are exactly identical to those in corresponding recordings of the V1 data. Data from awake monkeys are more limited and were only included in the present study to corroborate results for the data described above. They are recorded using 1-4 tetrodes from chronic implants penetrating V1 (see Tolias et al. 2007 for detailed description of surgical methods and

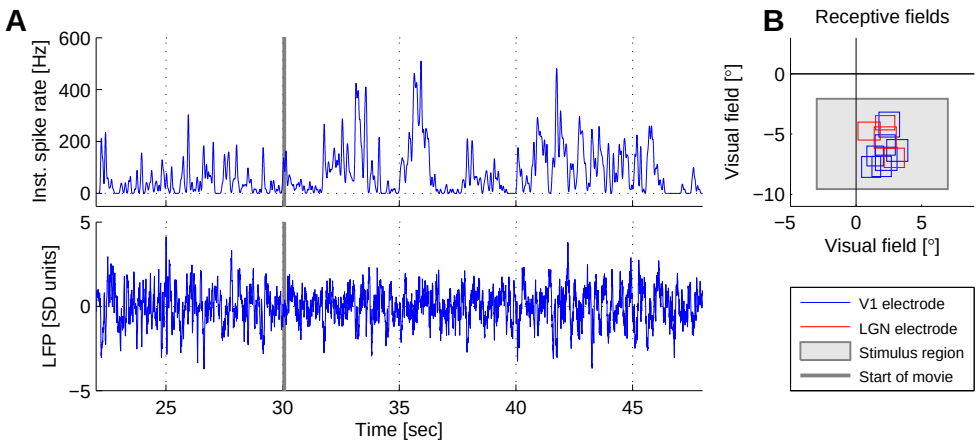


Figure 2.1: **A**: Representative electrode recording from Session a98nm5 of an anesthetized monkey. Upper plot shows the instantaneous firing rate of an experiment during movie presentation for a V1 electrode in a small time region. Movie presentation starts after a blank period. Recording time of 170 sec duration starting 5 seconds after movie onset is used for prediction performance evaluation and is called a trial (see Methods). The lower plot shows the LFP trace corresponding to the V1 electrode above. **B** shows the arrangement of receptive fields relative to the movie presentation area for Session a98nm5, where simultaneous V1 and LGN recordings are available. Other sessions have similar electrode arrangements.

recording setup) in a total of 56 trials. Unlike the data from anesthetized animals the stimulus conditions here are mixed, with spontaneous activity (no task) and a fixation task showing gratings of different orientations. All data are processed in the same way as outlined in the next section.

2.2.1.1 Processing

The data preprocessing steps are as follows. Electrode signals were decimated to 7 kHz. Spiking activity is inferred from high frequencies of the resulting signal (see below). The recording hardware introduces a high-pass filter with a cutoff < 1 Hz; 1 Hz is thus the lowest frequency considered here.

The 7 kHz signal is low-pass filtered with a cutoff frequency of 250 Hz and resampled first to 500 Hz for computational convenience. The resulting signal is low-pass filtered at 90 Hz to derive local field potentials (LFPs). For low-pass filtering we use a custom finite impulse response filter (FIR), namely a Kaiser window FIR filter with 60 db attenuation in the stopband, a 0.01 db passband ripple, and a transition band of 1 Hz. To eliminate possible phase shifts, signals were filtered forwards and backwards (using MatLab filtfilt function). The signal is then resampled to a final sampling rate of 200 Hz.

Good properties of FIR filters are won at the expense of large filter sizes (a few seconds). However, since we discard leading and trailing portions of > 15 seconds of each trial, filter on- and offset artifacts are of no concern here.

2.2.1.2 Spike extraction

Spike times are detected by applying a threshold to the high-pass filtered 7 kHz signal described above (4th order Butterworth, cutoff frequency 500 Hz). Since this MUA signal is usually asymmetric, the detection threshold is automatically applied to that side where spike waveforms exhibit greater deflection. To avoid dependency of the size of spikes the threshold is applied at 3.5 times the standard deviation σ of the "noise component" of the MUA signal. σ is estimated by calculating the standard deviation of the signal neglecting the 4.55% ($> 2\sigma$) absolute highest values divided by the percentage of variance which is kept in general, when setting the probability of values absolute $> 2\sigma$ to zero.

Visual inspection confirms that spikes are detected well. If the assumption of a Gaussian "noise component" is correct, then the rate of wrongly detected spikes is about 1.6 Hz (for $\sigma = 3.5$). Note that the resulting spike trains will most likely include spikes from multiple neurons (see Disccusion 2.4). Because most recordings were done with single tip electrodes we do not employ any kind of spike sorting.

2.2.2 Learning to infer spike trains

The learning algorithm has to learn to map from LFP waveforms (or other LFP features) to spikes. Ideally, the learning algorithm should output all predicted real-valued spike timings at once if the LFP time course is given as input. This task requires too much data, however. Instead, we simplify the task by assuming that spikes are independent and that the spike-to-LFP relationship remains constant over time. With these assumptions one can use a binary classifier, which yields the prediction of a spike (or no spike) at time t. Concatenating the prediction for each t results in a predicted spike train for a given LFP. Note that the independence assumption does not imply that predicted spike trains are necessarily uncorrelated, as temporal correlation can be induced by the underlying LFPs.

In supervised fashion the binary classifier is trained on a set of training examples and tested on a distinct test set. We train a binary classifier on LFP features summarized in the sample vectors $\mathbf{x}_i$, $i = 1, \ldots, L$, to predict the label $y_i \in \{1, -1\}$. i is the ith point of the discrete LFP time series with sampling frequency $1/\Delta$ at recording time $t_i = i\Delta + t_0$. Thus $y_i = 1$ states that there occurs (at least) one spike within time bin t_i, and $y_i = -1$ indicates no spike. In this framework prediction is temporally restricted to the sampling resolution of the LFPs, making it necessary to bin the spike timings. The sampling interval Δ is 5 ms, in accordance with the sampling frequency of the LFP signal (200 Hz).

2.2.2.1 Learning algorithms

A support vector machine (SVM) was used (Vapnik (1999)) as our learning algorithm. For a more detailed introduction to SVMs, see for example (Bishop (2006); Burges (1998); Schölkopf & Smola (2002)). Support vector machines perform binary classification in a supervised fashion.

Briefly, the model can be stated as follows (see Bishop (2006) for details)

$$h(\mathbf{x}) = \mathbf{w}^T \Phi(\mathbf{x}) + b \tag{2.1}$$

where one looks for the decision boundary or weight vector $\mathbf{w}$. b is a bias term and Φ a projection into a space of features. Support vector machines choose the hyperplane which has the widest margin between both classes rather than an arbitrary separating hyperplane. This is achieved by enforcing appropriate constraints in the optimization. For non-separable problems, such as our real-world data, one introduces the concept of soft margins, i.e. in the optimization one now allows for incorrectly classified examples, where an additional parameter C regulates the penalty.

SVMs have the power to do nonlinear separation (seen from perspective of the input space) by choosing an appropriate kernel which implicitly defines the feature map Φ. Herein non-linear radial basis function kernels (RBF kernels) are used.

As a simple alternative to SVMs we used standard linear regression (with a constant bias term) on the label vector and the samples (see e.g. Bishop (2006)). Briefly, using a linear model $h_{reg}(\mathbf{x}_i) = \mathbf{w}^T \mathbf{x}_i + b$ we calculated the optimal weight vector $\mathbf{w}^*$ by minimizing the mean squared error $\langle (h_{reg}(\mathbf{x}_i) - y_i)^2 \rangle$ on the training samples. Class labels on the test set were obtained by thresholding with the sign function, i.e. $y_j = \mathrm{sign}(h^*_{reg}(\mathbf{x}_j))$.

2.2.2.2 Extraction of LFP features

An LFP feature could be any aspect of the LFP which one might deem helpful for inferring whether there is a spike at t_i or not. In our analysis, we used the LFP at different lags with respect to t_i, its power at different frequencies, and the phase of oscillations at particular frequencies (also at different lags). Multiple features are simply concatenated in the sample vector $\mathbf{x}_i$. Note that each dimension of the resulting samples $\{\mathbf{x}_i\}$ is normalized to zero mean and unit standard deviation. Features are extracted prior to dividing the samples into test and training sets.

If $g(t_i)$ represents the (normalized) voltage at sampling bin , a time feature may be defined as $T_k(t_i) := g(t_{i+k})$, where $\tau = k\Delta$ represents the time lag (we neglect boundaries to simplify description). Features $T_k(t_i)$ represent simply the LFP time course relative to sample time t_i.

Another type of feature, which we denote $P_{f,k}(t_i)$, is the estimated power at (center) frequency f of the LFP time course at time t_{i+k}. To obtain an estimate for the power at a given frequency and time, we calculated the spectrogram, employing the multitaper approach introduced by Thomson (Jarvis & Mitra (2001); Percival & Walden (2002); Thomson (1982)). As spikes are single events on the time scale of Δ, we chose a high temporal resolution at the expense of frequency resolution. We set the moving window to 150 ms and the time-bandwidth product to 1.6, which implicitly sets the half bandwidth to $W = 10.67$ Hz. Spectral estimation was averaged over two Slepian tapers. As this window setting does not allow accurate power estimation below 20 Hz we used larger windows for bands below 20 Hz (500 ms) and below 6 Hz (2 s). This reduced the half bandwidth to 3.2 Hz and 0.8 Hz for frequencies below 20 Hz and 6 Hz, respectively. We also tried Morlet wavelets with variable bandwidth per frequency, but this did not alter prediction performance.

To access phase information at particular frequencies of the LFP, we first bandpass filtered the recorded signal with the FIR Kaiser filter (described above) with a bandwidth of 2 Hz (Fig. 2.6) or 4 Hz (Fig. 2.7), and then used the Hilbert transform to extract an instantaneous phase $\varphi_f(t_i)$ at frequency f. From these signals we defined phase features $\phi_{f,k}(t_i) := \cos(\varphi_f(t_{i+k}))$ having a lag of $\tau = k\Delta$. These features have identical phase information as the bandpassed signals but are devoid of any amplitude modulation. Additionally, we used $\hat{\phi}_{f,k}(t_i) := \sin(\varphi_f(t_{i+k}))$ in the feature analysis of Fig. 2.6 to help the classifier linearly extract phase locking at phases where the cosine would be near the zero crossing.

2.2.2.3 Performance measures

The kappa measure was used as a measure of performance (Cohen (1960)). Let $p_{l,r}$ be the fraction of samples having target label $l \in \{-1, 1\}$ and predicted label $r \in \{-1, 1\}$ and let q_l and $\tilde{q_r}$ be the fraction of samples in the test set which have the label l in the target or the label r in the prediction, respectively. Then the chance level for classification is given by $\rho_c = q_{-1}\tilde{q}_{-1} + q_1\tilde{q}_1$. If we define $\rho_0 = p_{-1,-1} + p_{1,1}$ to be the overall fraction of correctly classified samples (both positive and negative), then κ is given by

$$\kappa := \frac{\rho_0 - \rho_c}{1 - \rho_c} \tag{2.2}$$

This measure is a normalized above-chance classification rate. It can be easily seen that κ equals zero if prediction is at chance level, i.e. $\rho_0 = \rho_c$, and equals one if the predicted classification is perfect, i.e. $\rho_0 = 1$.

Another performance measure is the Spearman rank correlation r_σ between smoothed predicted spike trains and target spike trains. This might give a more intuitive picture of the prediction quality. If not stated otherwise, spike trains are smoothed by a Gaussian kernel of width 25 ms.

Yet another measure for prediction quality is the mutual information between class labels. The mutual information (MI) between target spikes S and the prediction outcome of a classifier $C(F)$ using LFP features F and labels $L := \{-1, 1\}$ is:

$$I(S; C(F)) = \sum_{l \in L} \sum_{r \in L} p_{l,r} \log_2 \frac{p_{l,r}}{q_l q_r} \tag{2.3}$$

where we take the probabilities defined above. This estimation of MI is different to non-parametric approaches in that it can only access dependency which is in reach of the classifier; thus, one has to make sure that the classifier captures the main aspects of its dependency. Note that we use the naive estimator for mutual information (without bias correction (Panzeri *et al.* (2007))). Since all MI value calculations involve an identical number of bins – namely two, one for each class – we can nevertheless safely compare results even for classifications with different numbers of features. However, the absolute MI values might be biased.

To access redundancy, synergy, and independence of information (Pola *et al.* (2003); Schneidman *et al.* (2003)) conveyed by two features F_1 and F_2 about the spiking activity S, we estimate mutual information using two classifiers trained on features F_1 and F_2 individually, yielding $I(S; F_1) := I(S; C(F_1))$ and $I(S; F_2) := I(S; C(F_2))$. Then a third classifier is trained on both features jointly, yielding $I(S; F_1, F_2) := I(S; C(F_1, F_2))$. If both features carried independent information from the point of view of the classifier, both features together would convey identical information as individual features, i.e. $I(S; F_1, F_2) = I(S; F_1) + I(S; F_2)$. If both features were related by a one-to-one mapping (completely redundant information about the spikes), then all terms would be equal, $I(S; F_1) = I(S; F_2) = I(S; F_1, F_2)$. If the two features did not carry information individually, i.e. $I(S; F_1) = I(S; F_2) = 0$, but carried information together, $I(S; F_1, F_2) > 0$, they would be termed (completely) synergistic. Thus we

define a normalized degree of synergy of information about the spikes (as measured by the classification algorithm) as (Schneidman *et al.* (2003))

$$\mathrm{syn}(F_1, F_2|S) := \frac{I(S; F_1, F_2) - I(S; F_1) - I(S; F_2)}{I(S; F_1, F_2)} \tag{2.4}$$

This measure ranges from -1 in the case of completely redundant information to 1 for completely synergic information. The measure $\mathrm{syn}(F_1, F_2|S)$ is zero if both features F_1 and F_2 convey independent information about the spikes S.

To analyze prediction accuracy on different time scales, we used spectral coherence (Jarvis & Mitra (2001)). Spectra were again estimated via a multitaper approach designed for point events (Jarvis & Mitra (2001)). Here the time-bandwidth product was set to $TW = 3$ using the average of $K = 5$ tapers, yielding a half bandwidth of $W = 0.001$ Hz for $T = 17$ s.

2.2.2.4 Performance evaluation

We evaluated the prediction performance for each trial separately, using 10-fold cross-validation. We analyze a 170 second region, avoiding the on- and offset of the movie stimulus. Spontaneous activity trials are also restricted to 170 s duration. In the case of tetrode recordings, performance is estimated as the average performance of the four wires of the tetrode.

Hyperparameters for the SVM algorithm were estimated as follows. The RBF kernel width ρ was selected by a heuristic procedure. We took ρ to be 1.77 (or 3.54) times the median distance of all Euclidean distances in the training set. For each trial we chose that C (and ρ) showing the best performance (averaged over 10 cross-validation runs on a logarithmic grid of 25 values from 0.25 to 400). We visually confirmed that this range is appropriate for our data (not shown). We used the libSVM library (http://www.csie.ntu.edu.tw/cjlin/libsvm/) for all SVM calculations.

Since the sample sizes were heavily biased towards the negative (non-spiking) class we randomly picked approximately the same number of samples of both classes from the training region. This effectively changes the loss function from equal loss to higher importance for spikes (about a 5-fold increase, depending on the mean firing rate). We used 1000 and 1200 samples for spiking and non-spiking classes (or the maximum available in the training region with a constant class ratio) and empirically found this to

be a good compromise between prediction quality and computational speed, as more than 1000 samples only marginally improved the results (not shown). Training the classifier on all possible samples was prohibitive due to the enormous sample size. We tried to employ class biasing in the C parameters (Musicant *et al.* (2003)), but this only increased computation time with little gain in prediction quality.

The test set was always a temporally contiguous region to avoid feature correlation of trained and tested samples which might lie nearby in time if a randomized set of samples were used.

2.2.2.5 Feature selection algorithm

We now describe how to determine the usefulness of different features for spike prediction. Whereas features important for the SVM classifier are hard to interpret, in the case of linear regression (with squared loss) one can derive an analytical expression for the prediction error of a set of features involving only the spike triggered averages and correlation among features. Based on this prediction error, we derived an algorithm which forwardly selects a small subset of features out of a much larger pool of features. As explained below, the selected subset will show minimal prediction error compared to other subsets with the same number of features. In that sense the subset of features selected by our algorithm represents the most useful features from a given pool. As this feature selection can be efficiently done for huge feature pools, we restricted feature analyses to linear classification, rather than using the SVM classifier. This is not too restrictive in our case, as linear classification achieves almost 90% of the performance of an SVM classifier (see Results section).

In linear regression we look for the optimal weight vector $\mathbf{w}$ which has minimal error in a mean squared sense,

$$\epsilon(\mathbf{w}) = \left\langle (y_i - \mathbf{w}^T \mathbf{x}_i)^2 \right\rangle \tag{2.5}$$

where the brackets indicate averaging over all samples i. Minimizing the error is straightforward and results in the optimal[1] $\mathbf{w}^*$ of

$$\mathbf{w}^* = \left\langle \mathbf{x}_i \mathbf{x}_i^T \right\rangle^{-1} \left\langle y_i \mathbf{x}_i \right\rangle \tag{2.6}$$

[1]Note that the optimal $\mathbf{w}$ in the above statements has only minimal error for the regression, and that there may be a better weight vector for classification. We neglect this here for the sake of simplicity.

provided that the estimated correlation matrix $A := \left\langle \mathbf{x}_i \mathbf{x}_i^T \right\rangle$ has full rank. We note that we have a binary classification, hence $y_i \in \{-1, 1\}$. Thus

$$\langle y_i \mathbf{x}_i \rangle = -\frac{c_{-1}}{c_1 + c_{-1}} \mathbf{m}_{-1} + \frac{c_1}{c_1 + c_{-1}} \mathbf{m}_1 \tag{2.7}$$

where $\mathbf{m}_{-1}$ and $\mathbf{m}_1$ are the class means for the non-spiking and spiking class, respectively, and c_{-1} and c_1 are the number of samples in each class. With the condition that each feature is normalized to variance one and zero mean, Eq. 2.6 reduces to

$$\mathbf{w}^* = cA^{-1}\mathbf{m}_1 \tag{2.8}$$

with $c := \frac{2c_1}{c_1 + c_{-1}}$. The minimal error is then given by

$$\epsilon(w^*) = 1 - c^2 \mathbf{m}_1^T A^{-1} \mathbf{m}_1. \tag{2.9}$$

In other words, the error of the linear regression is only dependent on the *spike triggered average* $\mathbf{m}_1$ and the correlation among features. We note that if we restrict ourselves to the use of n features (the dimensions of $\mathbf{x}_i$) out of a pool of $N > n$ features, the above equations remain valid if the correlation matrix and the mean are also restricted to these features only (that is, A has size $n \times n$ with rank n and $\mathbf{m}_1$ is n-dimensional).

In order to select a set of n features we employ the following iterative algorithm. We start with the feature which has highest spike triggered average, i.e. $f_1 = \mathrm{argmax}_j |(\mathbf{m}_1)_j|$ (the variance of each feature being normalized to one). Assume now that $n-1$ features are already selected. Then we search through all $N - n + 1$ remaining features and choose one that minimizes the error (Eq. 2.9), where the restriction of the correlation matrix is now enforced on the n features rather than on $n - 1$ (and analogously with $\mathbf{m}_1$). We stop this iteration when the desired number of features $m \ll N$ is selected.

This algorithm is highly efficient in finding a good set of features, since we need only to calculate the correlation matrix between the selected features and all other features (which costs much less effort than calculating it for all pairs).

2.3 Results

This section is organized as follows. After showing the general spike-to-LFP relationships present in our data, we report the population performance for the task of predicting spike trains from LFP, focusing first on data from V1 of anesthetized monkeys

(9 monkeys) collected during the presentation of 5min of commercial movie stimuli and equally long periods of spontaneous activity. We then compare these results with a more limited set of sessions recorded from V1 of awake monkeys (2 monkeys) and with results on data from LGN of anesthetized monkeys (2 monkeys). Finally, we investigate which LFP features are important for the prediction task and which aspect of the spikes they code for.

2.3.1 Average spike-to-phase and LFP-power relation

Figure 2.2.1 shows spiking activity and (normalized) LFPs of a representative electrode. Relationships between spiking activity and underlying LFPs are visualized in Fig. 2.2.

In Fig. 2.2 A the spike-triggered average (STA) of the LFPs of an example electrode during movie stimulus is plotted. Clearly, there is a linear relation between spikes and LFPs. One notes a sharp negativity at spike position at zero time lag and a prominent upswing for positive time lags, i.e. after spiking has occurred. Likewise, in the spike triggered average of the spectrogram (Fig. 2.2 B) power is enhanced in the high-frequency range (40-90 Hz) during spiking activity. Enhancement of power at high LFP frequencies as a response to spikes is common among electrodes, stimulus conditions, and monkeys, as we will see in the next sections.

Figure 2.2 C shows the probability of spiking activity at the oscillation phase of a particular LFP frequency for the same example electrode, but averaged over 30 repeats of the movie presentation (about 120 min recording time). One notes that the phases of all LFP frequencies are at least weakly related to spiking activity (Raleigh test of non-uniform angular distribution). Most strikingly, spikes are relatively tightly locked to phases of low frequencies up to 10 Hz. The generality of this behavior is illustrated in Fig. 2.2 D, where the phase preferred by spikes is plotted as an average across all data from V1 (anesthetized animals). The average preferred phase shifts with frequency from the onset of a positive half wave to the valley of the LFP oscillation (compare to Fig. 2.2 E). This behavior is very consistent, regardless of whether activity is spontaneous or movie-driven. For some electrodes the preferred phase-frequency dependence is slightly different (as in the example of Fig. 2.2 C for high frequencies). For a few electrodes the phase-to-spike relation seems to be mirrored at π (not shown).

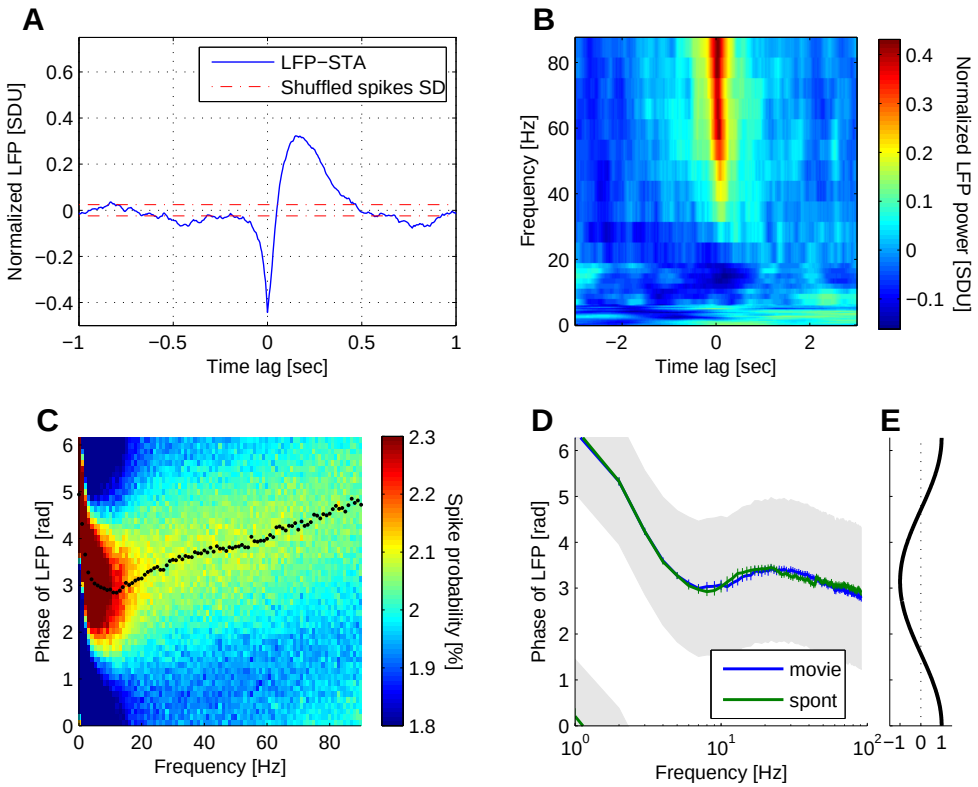


Figure 2.2: Spike-LFP relationships for one electrode in V1 of an anesthetized monkey during movie stimulation (A-C), and across all recordings from V1 in anesthetized monkeys (D-E). **A:** Spike-triggered average LFP. For significance levels inter-spikes intervals are shuffled and the standard deviation of the resulting STA calculated. **B**: Spike-triggered average of the LFP spectrogram (see Methods), with power series normalized to zero mean and standard deviation one. Power at high frequencies is clearly modulated by spiking activity, whereas power at lower frequencies shows only diffuse dependency on spikes. **C:** Probability distribution of LFP phases at a particular spiking position. LFP phases are computed via Hilbert transform (1 Hz bands). Here all spikes over 30 repeats of movie-driven activity are included (same electrode as before). Note that the color map shows only a narrow range of probabilities and that values above or below the limits are truncated. Black dots indicate the preferred (i.e. mean) phase. No phase locking of spikes would result in a uniform distribution at 2% per bin. While locking to low frequencies is strong, locking to high frequencies is only weak (but present). Plot **D** shows the average preferred phases for all electrodes across all anesthetized data individually for movie-driven activity (”movie”) and spontaneous activity (”spont”). Bars indicate standard errors. The phase range containing half of the spikes around the preferred phase is indicated by the shaded area. **E** illustrates the interpretation of phase, showing that spikes are locked at very low frequencies to the onset of a positive half wave and at high frequencies to the valley.

The gray shaded area in Fig. 2.2 D shows the (average) phase range within which 50% of the spikes fall. It would be zero for perfect phase locking and π for no phase-spike relation. One notes that this range is somewhat smaller for low frequencies (0.85π), but approaches 0.98π for frequencies higher than 20 Hz, indicating that the phase locking is far from perfect at all frequencies, and is especially weak for high frequencies.

In summary, we have seen that there is indeed a consistent relationship between LFPs and spiking activity on average. In the next sections we ask to what degree it is possible to exploit these relations (and maybe other information available in the LFP) in a systematic way to infer spikes from the LFP.

2.3.2 Population prediction performance

In Fig. 2.3 a typical example of a predicted spike train is depicted together with the utilized LFP features. Panels A and B show 8 seconds of LFP spectrogram and the time course in the test region, respectively. Small vertical lines in panel B indicate spike times before binning to 5 ms resolution. Several interesting points can be noted. As expected from the LFP-to-spike relations discussed in the last section, spikes preferentially occur in the upswing and valleys of very low and medium LFP oscillations, as seen for instance at times 171 s and 173.4 s in panel B. Additionally, the power of multiple frequencies is enhanced when a burst of spiking activity occurs, as suggested by Fig. 2.2 B, but the frequency response to bursts is diffuse and variable (compare the burst at 173.4 s to that at 174.5 s). The clustering of spiking activity on a time scale of a few hundred milliseconds in this example is actually quite typical in our V1 data (see below). For the single spikes in between the clusters no feature of the LFP is immediately predictive.

Figure 2.3 C shows target and predicted spiking activities. Prediction of spikes is made for individual sample times at a resolution of 5 ms using a set of LFP time course and frequency features (see below). Concatenating the prediction over time yields a predicted spike train which is compared to the target spike train. One notes that the prediction captures at least approximately the overall structure of the spike train. The occurrence of bursts of spiking activity, which are associated with easily seen traces in the LFP time course and spectrogram, is predicted well. Nevertheless, the exact

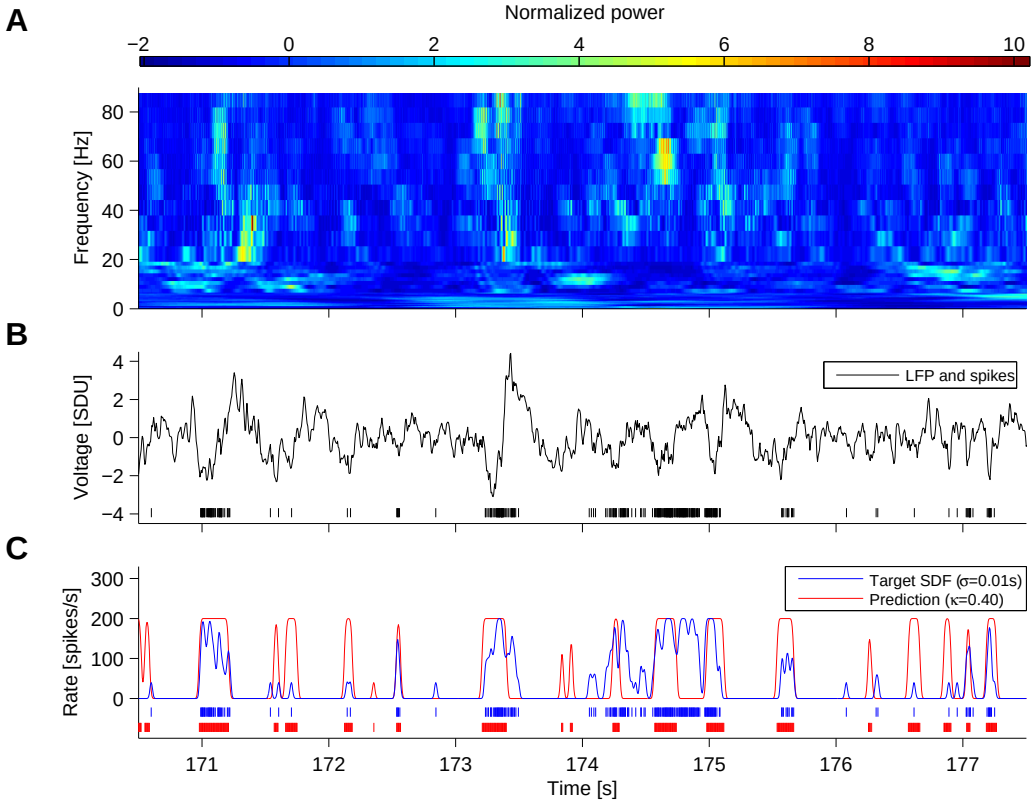


Figure 2.3: Example of spike prediction from LFP (anesthetized monkey, Session a98nm5, spontaneous activity). Panel **A** shows the (normalized) spectrogram of the 8 seconds of LFP activity, the time course shown in panel **B**. Simultaneously recorded spikes are indicated by marks before binning to the LFP resolution (5 ms). Panel **C** shows the binned target spikes and their spike density function (blue) together with the predicted spikes and their spike density function (red). The prediction is relatively good ($\kappa = 0.40$, $r_{25ms} = 0.60$) on this trial, but other trials show even better performance (compare to r_{25ms} values of other trials in Fig. 2.4 B "spo"). One notes that regions of high activity are well predicted, whereas the location of single spikes is less accurate. Classification is done with the SVM-RBF classifier trained on the region 35-160 sec using the same features as for the population analyses (Fig. 2.4).

onsets and offsets of the bursts are somewhat inaccurate in the prediction. Even some smaller bursts and single spikes are closely predicted (172.5 s), although no clear mark in the LFP time course or spectrogram can be seen with the naked eye. However, their length (176.6 s) and exact position (176.3 s) sometimes seem inaccurate. There are also occasions where spikes are simply missed (172.9 s) or fabricated (173.9 s).

Prediction performance is evaluated in different ways. One measure is the κ performance, a measure which is defined on the samples in the test set and is positive for above-chance classification; it equals one for perfect classification (see Methods 2.2 for definitions). In contrast, the correlation measure r_σ (see Methods 2.2) is defined as a local average in the time domain and is therefore less sensitive to small temporal inaccuracies. The performance measure κ of the predicted spike train in Fig. 2.3 C has a value of $\kappa = 0.40$, which is relatively good (but not the best possible, see below). Rank correlation is $r_{25ms} = 0.60$.

In the example of Fig. 2.3, the predicted spike train resembles the original to a certain degree. We ask whether this prediction quality carries over to LFPs recorded from different monkeys, electrodes and stimulus conditions. For that we estimated prediction performance using a large data set (see Methods 2.2). Inferences are made on the basis of a set of LFP features, with which we observed a dependence between spikes and LFPs in the previous section. In the population analysis we include as features the time course around each sample position (in a window of 100 ms before and 300 ms after spike position) and an estimate of the frequency content of LFPs at zero time lag ($P_{f,0}(t_i)$ see Methods 2.2), resulting in a total of 116 features. This feature set generally produced good performance (with a reasonable computational speed) over a wide range of data. For the prediction itself a non-linear support vector machine is employed with radial basis functions such as kernel (SVM-RBF) and linear classification (see Methods 2.2 for details).

In Fig. 2.4 the prediction performance over all trials is evaluated (on 10 cross-validation runs) and averaged. The anesthetized V1 data set is labeled ”spo” for spontaneous activity and ”stm” for movie stimulus driven activity. We shall focus on this data for the moment. The remaining conditions shown in this plot will be discussed below.

Plot A shows the average performance κ for the SVM RBF classifier and for linear classification. From the results we draw the following insights. First, since perfor-

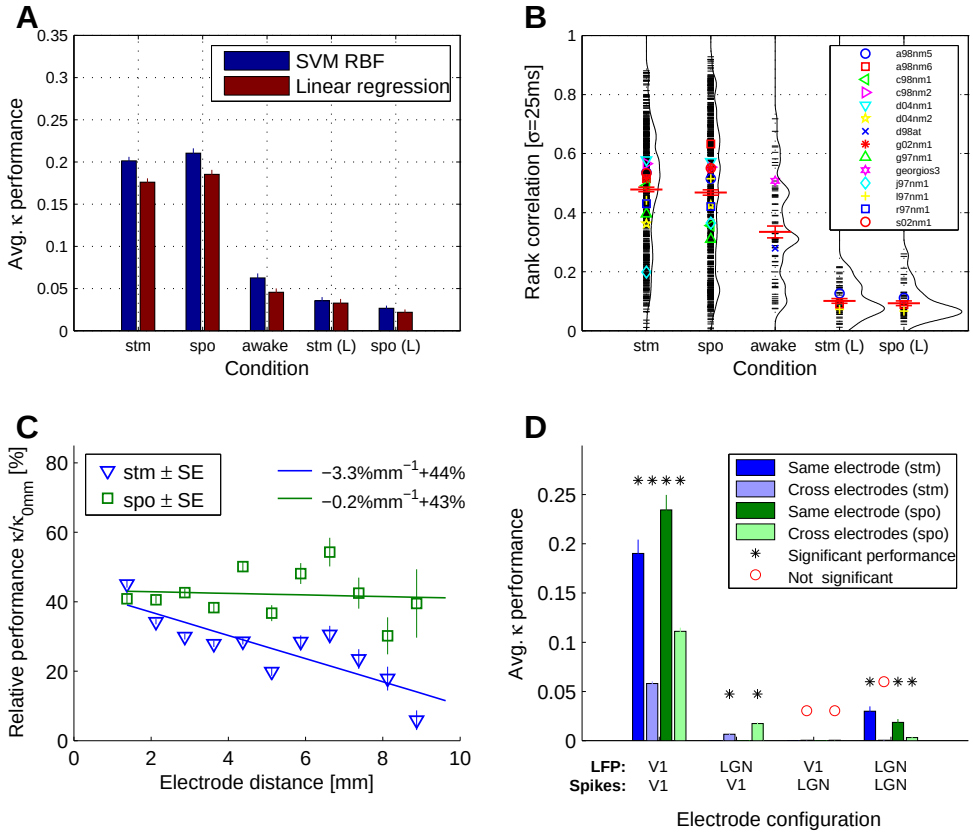


Figure 2.4: Population performance for spike prediction from LFP. **A**: Average prediction performance κ for SVM and the linear regression classifier across conditions (anesthetized monkey V1, movie stimulus (”stm”), anesthetized monkey V1, spontaneous activity (”spo”), awake monkey V1 with mixed stimuli (”awake”), and spontaneous activity or movie-stimulus driven activity in anesthetized monkeys from LGN, ”spo (L)” and ”stm (L)”, respectively). Prediction is above chance level for all conditions (see Results section for significance tests). **B**: Prediction accuracy of the non-linear classifier evaluated by rank correlation between target and predicted spikes train smoothed by a Gaussian kernel of width $\sigma = 25$ ms. Red horizontal lines indicate the average performance within each condition, and its standard error. Small black lines show the quality on individual trials. In some cases prediction yields very accurate results, with correlations as high as 0.8-0.9. Black curves on the sides of each condition are smoothed histograms over trials. Symbols indicate the average performance of individual sessions (i.e. one day of recording). Average session performance clusters near the overall mean, but variance for individual electrodes is high. **C:** Cross-electrode prediction. LFPs are taken from one electrode and spikes from another. Relative prediction decreases much faster with increasing electrode distance in the ”stm” condition than for the ”spo” condition. A linear fit of the data points is also shown. Vertical lines indicate standard errors. **D:** Average prediction performance for simultaneous LGN and V1 recordings (3 sessions, 2 monkeys). Performance is compared for available cross-electrode prediction from different areas and with average performance when using the same electrode for spikes and for LFPs.

mance measure κ is greater than zero for above-chance prediction it can be said that both classifiers can exploit information in the LFP time course to predict spiking activity (all conditions highly significant; T-test $p < 10^{-6}$, Wilcoxon signed rank test for zero median $p < 10^{-6}$). Second, prediction quality for the stimulus condition and for spontaneous activity differs only slightly: indeed, one cannot reject the hypothesis that the underlying distributions have identical means (two-sided unpaired T-test $p = 0.21$, linear $p = 0.18$). However, if one compares pairwise recordings during spontaneous activity and stimulus presentation done with identical electrodes, mean and median prediction performance on spontaneous activity is significantly better than on stimulus-driven activity (one-sided paired T-test $p < 10^{-4}$, linear $p < 10^{-4}$; for the distribution free Wilcoxon matched-pairs signed-ranks test $p < 10^{-4}$, linear $p < 10^{-3}$). Average prediction performance for spontaneous activity is $\kappa = 0.211 \pm 0.006$ (linear $\kappa = 0.185 \pm 0.005$) and $\kappa = 0.201 \pm 0.005$ (linear $\kappa = 0.175 \pm 0.005$) for stimulus driven activity.

Third, non-linear margin classification is consistently better than linear classification (one-sided paired T-test "stm" $p < 10^{-6}$, "spo" $p < 10^{-6}$; Wilcoxon matched-pairs signed-ranks test $p < 10^{-6}$). It amounts to about a 12% increase in performance on average. This suggests that the mapping from LFP features is non-linear. However, since a simple linear regression classifier already achieves almost 90% of the accuracy of the non-linear classifier, one could state that the LFP feature space exploited here seems expressive enough for this task.

We found that for individual trials performance varies widely. For selected trials prediction performance can reach $\kappa = 0.65$. Plot B of Fig. 2.4 shows the rank correlation measure r_{25ms} of the SVM RBF prediction. Each thin short line represents performance for an individual trial. While the correlation for some trials is as high as 0.8-0.9 on this moderately small time scale (25 ms), it is almost zero in others. There are some trials where prediction fails completely in each of the conditions. The failing trials are not all from the same sessions, since the session means (markers) tend to cluster around the overall mean.

There is not much variability in performance over time: the average standard deviation for the κ performance of 5 repeats of 170 ms recordings for the same electrodes is 0.023$\pm$0.002 for stimulus–driven activity, 0.026$\pm$0.002 for spontaneous activity and 0.045$\pm$0.002 for both together. This is in contrast to the variance across electrodes

recorded simultaneously. Here the average standard deviation (in κ) is 0.113±0.004 for stimulus-driven and 0.130±0.005 for spontaneous activity. The roughly 25-fold increase in variance across electrodes compared to within-electrode variance suggests that prediction performance is a matter of which electrode is being observed, rather than stimulus condition or time. Electrode tips might be positioned in a region where the arrangement of current sources and sinks might differ (e.g. in deep or superficial layers), or where active neurons might be less well correlated with the bulk activity of the cortex. Since we cannot distinguish the layers from which electrodes record, we cannot pursue this further.

Up to now we have presented only results for recordings from V1 of anesthetized monkeys. We also have a limited amount of data available from V1 where monkeys were awake and behaving. The stimuli for the awake data set are mixed and include spontaneous activity and fixation tasks. Another pool of data consists of recordings from LGN of anesthetized monkeys (see Methods 2.2).

We see from Fig. 2.4 A that prediction differs quite drastically for the different data types. Spike prediction for the anesthetized monkey data from V1 is more than 5 times better than in the LGN, where performance is hardly above chance: on average $\kappa = 0.035 \pm 0.005$ (linear $\kappa = 0.033 \pm 0.005$) for movie-driven activity and $\kappa = 0.027 \pm 0.003$ (linear $\kappa = 0.022 \pm 0.003$) for spontaneous activity.

As in V1, there is little difference between spontaneous and movie-driven activity in LGN, although there is a reversed tendency for spike prediction to be easier on movie-driven activity than on spontaneous activity. This tendency is barely significant (one-sided paired T-test $p = 0.02$, linear $p = 0.01$; Wilcoxon matched-pairs signed-ranks test $p = 0.05$, linear $p = 0.08$).

We find that average prediction performance on awake data is $\kappa = 0.063 \pm 0.005$ (linear $\kappa = 0.046 \pm 0.005$). This is much worse than on anesthetized data (unpaired T-test $p < 10^{-5}$), but still significantly better than on LGN data (all unpaired one-sided T-tests $p < 0.05$). Figure 2.4 B reveals that individual trials have a correlation of target and prediction similar to that in anesthetized monkeys. There are trials with correlation up to $r_{25ms} = 0.6$, whereas in the case of the LGN, no trial exceeds 0.3 correlation.

2.3.3 Cross electrode predictions

The area of cortex which contributes to the generation of LFPs is different from that producing our spiking signal (see Introduction 2.1). Thus, it might be interesting to see how the relationship between the two signals changes with distance. As recordings were done with multiple electrodes simultaneously (in the data set from anesthetized animal), we tried to infer spikes from LFPs collected with two different electrodes. In Fig. 2.4 C the average performance is plotted against the (3-dimensional) distance of the electrode tips. To facilitate comparison, performance is evaluated relative to the average performance achieved using the spiking signal from the electrode from which the LFPs were taken.

One notes that prediction performance drops to about 40% when electrodes are 1 mm apart (the minimal distance in our recording setup). Interestingly, for stimulus-driven activity performance degrades significantly with distance (rank correlation between distance and relative kappa performance using all measurements: -0.20, $p < 10^{-4}$), whereas for spontaneous activity no significant correlation with distance can be found for distances up to 1 cm (rank correlation 0.015, $p = 0.2$). Note that the number of samples becomes relatively small for distances greater than 6 mm since rectangular electrode grids with 1mm spacing are used for most sessions. However, we can safely compare spontaneous and stimulus-driven activity because the electrode placements do not change with the condition.

As LGN data was collected while other electrodes simultaneously recorded from V1, we can investigate whether the LFPs of V1 can be predicted on the basis of spikes from LGN and vice versa. This is shown in Fig. 2.4 D averaged over data from the 3 sessions recording simultaneous measurements from V1 und LGN (see Methods 2.2). Performance is averaged either across electrode predictions (regardless of distances) or over all predictions using the same electrode for both signals. Although results are difficult to interpret because of the limited size of the data set, one notes that using LFPs from LGN and spikes from V1 results in performance above chance, whereas LFPs from V1 seem to hold no information about spikes in LGN (unpaired Wilcoxon signed rank test for median performance different from zero, significance level 0.05).

2.3.4 Temporal accuracy of predicted spike trains

We found an average κ value of about 0.2, which is well above chance but nevertheless far from perfect prediction at $\kappa = 1$. On the other hand, in example Fig. 2.3 C some features of the target spike trains seem to be captured well by the prediction, especially regions of high and low activity, which alternate on a time scale of about 0.5 s in this example. Thus one might ask at what time scale the predicted spiking activity most closely resembles the target spiking activity, or at what timing accuracy the prediction fails.

To answer this we evaluated the coherence between target and predicted spike train (Fig. 2.5). Coherence is a correlation measure in the frequency domain. Coherence at a particular frequency makes a statement about the exactness of the prediction on a time scale of one over that frequency. We also estimated the temporal accuracy directly in the time domain (by varying the correlation kernel width) where one arrives at similar conclusions (not shown).

In Fig. 2.5 one observes that coherence is low for high frequencies and rises for low frequencies. Thus the general resemblance of a predicted spike train might be adequate, but the exact spike position is often predicted with some jitter. This is also evident in the example of Fig. 2.3.

Coherence drops at about 25 Hz for the anesthetized V1 data. This time scale is comparable to a spike train whose spikes are jittered by Gaussian random noise with a standard deviation of 25 ms. Coherence levels of such surrogate data are indicated by the dashed lines in Fig. 2.5. Since the jitter destroys all information in the high frequencies, the plateau at low coherence for the surrogate data can be taken as a significance level for the coherence estimation. In surrogate data low frequency aspects stay completely intact (thus coherence of 1), but for predicted spike trains this is only partly the case. However, average coherence rises considerably for larger time scales compared to smaller ones, suggesting that at least in a subset of trials, slow structure is well predicted.

Data from the awake monkey is less coherent at low frequencies but much more so than for data from LGN, where almost no significant coherence is observed, even for low frequencies. Note that we have far fewer trials for awake and LGN data, so averaging is less effective in smoothing.

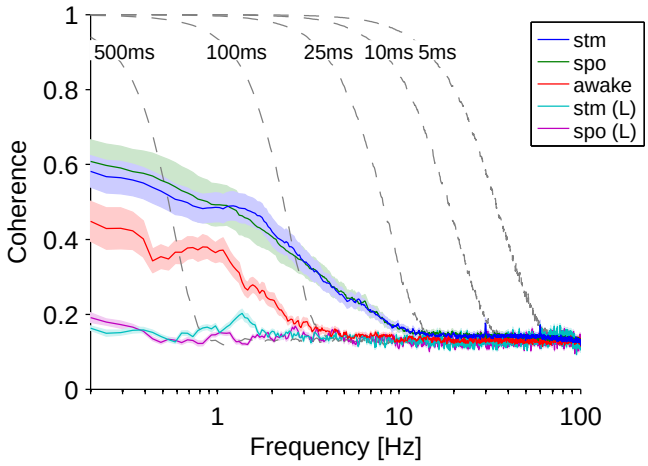


Figure 2.5: Coherence levels of predicted and target spike trains. Coherence is plotted against frequency and is averaged over all trials for each condition. For comparison we included surrogate data, in which spike trains were generated with the same ISI distribution as in the monkey data. Gray dashed lines show coherence between surrogate spike train and its jittered version with Gaussian noise of different standard deviations σ (from 5 ms to 500 ms, as listed in the plot). Coherence drops for higher frequencies, suggesting that on average prediction is only reasonably good for slow structure in the spike trains. Note that the chance coherence level is at about 0.15 here, as shown by the surrogate data. Chance levels do not tend to zero because coherence is estimated on 10 CV-regions (each 17 s duration) and only subsequently averaged over all trials. Colored regions indicate standard errors.

In summary, predicted spike trains are seldom accurate to a spike timing precision of below 25 ms, as suggested by comparison to a jittered version of the original spike train. On the other hand, predicted spike trains capture structure on a larger time scale reasonably well, say for clusters of high spiking activity in the 100 ms range.

2.3.5 LFP features important for inferring spikes

For determining the usefulness of particular LFP features for inferring spiking activity, we iteratively select a small number of features out of a large pool of features. The selected subset shows minimal prediction error for a given number of features, and therefore selected features can be seen as the most important for prediction. As spike prediction in LGN is almost impossible, only V1 data are analyzed in the following.

We consider a feature pool consisting of phase and power features ($P_{f,k}$, $\phi_{f,k}$, and $\hat{\phi}_{f,k}$). Phase is estimated on 45 frequency bands each 2 Hz wide, while the power features $P_{f,k}$ have different frequency resolutions (see Methods 2.2). Setting k appropriately, we include time lags of up to 3 s in both directions (before and after t_i). Out of this pool of features, containing together $N = 138115$ features, only up to $m = 10$ features are selected for each trial individually using the algorithm outlined in the Methods 2.2 section. Figure 2.6 D shows that, on average, selecting only 10 features out of the huge pool is enough for a linear classifier to approach the performance of the linear classifier used previously (Fig. 2.4), which employed 116 general features (dashed line). For the first five selected features the gain in performance is highest.

Figures 2.6 A-C show histograms of $m = 5$ selected, most important features aggregated for all trials. Phase- and power-related features are colored blue and red, respectively. Analogous to previous results, useful features differ only slightly between stimulus-driven activity (Fig. 2.6 A "stm") and spontaneous activity (Fig. 2.6 B "spo"): stimulus induction does not seem to induce a general change in the preference of features for spike-LFP interaction. One notes that in both spontaneous and stimulus-driven activity power fluctuations in the high γ-band (40-90 Hz) are preferred features. Selected frequencies are biased towards high values, with 80-90 Hz being the most likely selection. Indeed, high-frequency power features are selected as the first and most useful feature in about 90% of the trials (and in 82% in awake

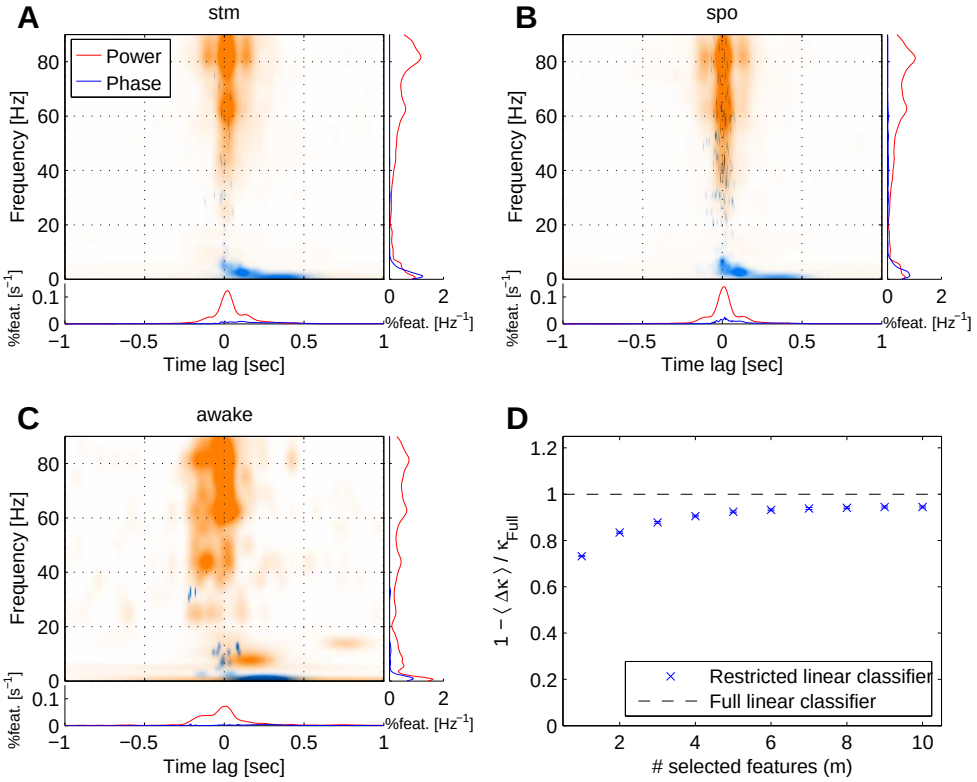


Figure 2.6: Most useful LFP features for spike prediction. Distribution of $m = 5$ selected features per trial, pooled over all trials. Color plots show the density of the selected features in the feature space (frequency and time lag). Each feature is represented by a Gaussian with 95% of its density falling into the width of correlatedness (i.e. the time window and bandwidth of spectral estimation for power features; we have chosen a quarter oscillation period for phase features). Marginal plots show the marginal distribution of the histogram for absolute lags up to 3 seconds, where colors code for feature type (power $P_{f,k}$ (red), phase features $\phi_{f,k}$ and $\hat{\phi}_{f,k}$ (blue)). Panel **A-C** show conditions "stm", "spo" and "awake" (as described in Fig. 2.4). See Results section for discussion. Panel **D**: Performance of linear regression restricted to m selected features relative to the performance of the full linear classifier. To avoid overfitting, performance is tested on a time region of each trial which was not used to estimate the importance of features. $\Delta\kappa$ denotes the difference in prediction performance κ between the restricted linear classifier using up to *m* selected features and the full linear classifier κ_{Full} with 116 features (as in Fig. 2.4). $\langle \Delta\kappa \rangle$ divided by the mean performance of the full classifier is plotted against the number of selected features m. One observes that performance with selected features approaches the full classifier.

animals) (not shown). The time lags of the selected γ power features are almost symmetrically distributed around zero (with a small bias toward positive lags) in a zone spanning approximately 50 ms to either side. There are smaller symmetrical peaks at 150 ms, which may be attributed to the power estimation, where we use a moving window of 150 ms duration. Likewise peaks at 80 Hz and 60 Hz are introduced by spectral estimation as the bandwidth is roughly 20 Hz (see Methods 2.2).

We identified low-frequency information as a second class of useful LFP features, in particular phase information of low-frequency bands below 10 Hz. Time lags of selected phase features are mostly positive, meaning that the time of the feature is most informative *after* the spike. Useful lags vary from -50 ms to 200 ms, depending on the frequency, and they can be as long as 500 ms for the lowest frequency band (up to 2 Hz). Time lags vary according to an oscillation period of the low bands. Power modulations in the low-frequency bands are selected about as often. The time lags of these features are distributed widely, which is caused by the long window setting of 2 seconds needed to estimate power at low frequencies (see Methods 2.2).

Bands from 10 to 40 Hz, especially 15 Hz to 30 Hz, seem to be much less important for inferring spikes. Despite a small number of scattered features in the ”spo” condition, phase information for γ-bands (e.g. > 40 Hz) does not play a role, either.

For the awake animals results are hard to interpret, given the limited amount of data (see Fig. 2.6 C). However, it seems that the overall structure is similar to the V1 data of anesthetized monkeys in having high-frequency power features as well as very low-frequency phase features for positive lags. However, there seems to be an increase of selected power features for intermediate frequencies.

Both feature types, meaning high-frequency power features around zero lag and low-frequency information, either low-frequency power or low-frequency phase features with positive lags, are often jointly selected among the 5 optimized features. This shows that individual trials have similar features. We found that in 75% (”stm”), 59% (”spo”), and 72% (”awake”) of the trials both types of features are jointly selected, more specifically a high-frequency power feature (above 40 Hz) with absolute lags of less than 250 ms and a low-frequency phase feature (below 10 Hz) with positive lags or a low-frequency power feature below 10 Hz. In absolute terms high-frequency powers are preferred over low-frequency features (”stm” 61% vs. 24%, ”spo” 62% vs. 19%, awake 47% vs. 24%). In ”stm” and ”spo” conditions low-frequency phase features

with positive lags are selected slightly more often in combination with high gamma powers than low-frequency power features (”stm” 49% vs. 44%, ”spo” 38% vs. 33%), whereas low-frequency powers are preferred in the awake condition (28% vs. 60%). Neither of the low-frequency features is present in about 25% of the trials when the first 5 selected features are considered. However, this value drops to about 5% when the first 10 selected features are examined.

In summary, our analysis shows that two feature types are most useful for predicting spikes: power in the higher bands (> 40 Hz) and (to a lesser degree) low-frequency information (< 10 Hz), which can be power modulation or phase information with lags around and after the spike.

2.3.5.1 Population statistics of low frequency phase features

Low-frequency phase features indicate spike positions relative to the low-frequency oscillations of the LFP. This feature thus carries the information of the phase locking to lower bands (Figures 2.2 C and D). However, from the point of view of phase locking it is surprising that the informative lags are asymmetrically distributed around the spiking position. This indicates that it is not merely the locking to a phase that is important, but that instead the LFPs at low frequencies display a consistent slow oscillation *following* spiking activity, and spikes are locked to the onset of that oscillation. In contrast LFPs before spikes are less well determined on average. This asymmetry of the phase locking to lower bands can be seen in the spike-triggered average LFP (STA), as pointed out earlier (Fig. 2.2 A).

The form of the STA is stereotypical for the majority of electrodes in V1. As we showed in the Methods section, a high value in the STA is a good feature for classification (given some covariance constraints). Thus the typical form of the STA explains why the phase features for positive lags are consistently selected among the best features. To show the generality of the form of the STA in our data, we select two particular phase features, ϕ_A and ϕ_B. Feature ϕ_A is determined by the position of the first maximum (peak) for positive lags of the 1-4 Hz bandpass filtered LFP (cross in Fig. 2.7 A). To reject any amplitude modulation in that band we take the cosine of the time course of the phase instead of that of the LFP (see Methods). The second phase

features ϕ_B is the valley nearest to zero lag in the 4-8 Hz band of the LFP (star in Fig. 2.7 A).

In Fig. 2.7 B the distribution of these two features is shown in a scatter plot across all trials. Note that the feature ϕ_A (black crosses) lies very consistently at mean lag around 112 ± 1 ms, although the height differs somewhat (see marginal distributions in attached plots). The feature ϕ_B (green stars) is likewise consistent across trials. However, the height distribution of this second feature (right margin plot) is more skewed to lesser values than for feature ϕ_A, suggesting that it might be less useful for prediction. There is also a minority of trials in which neither feature was well expressed, indicated by the scattered outliers.

2.3.5.2 Information conveyed by low-frequency bands and high-frequency power features

To compare information conveyed by different features about spikes we use the mutual information between target spikes train and predicted spikes train (see Methods 2.2). Mutual information between the class labels is a lower bound for the mutual information contained between the signals under consideration (Natschläger & Maass (2005)). It is only a lower bound, since a classification method might fail to use all the information contained in the signal. However, it is unlikely that a non-linear SVM classifier would miss much of the dependency in our data, since the relationship between features and spikes seems to be mostly simple proportionality. Recall that for our data a linear classifier already achieves about 90% of the performance of a non-linear classifier.

First we tested prediction performance for single-frequency power and low-phase features individually using the SVM classifier. In Fig. 2.7 C we show the average information about the spikes at different frequencies for anesthetized monkeys V1 (blue bar plot). The information change with frequency closely resembles the number of selected power features per frequency from our selection algorithm in the previous section. We note that on average frequencies around 80 Hz convey the most information about the spikes. If one uses all the features tested here simultaneously, average performance reaches 0.037 ± 0.002 bits, which is 35% higher than when the best individual feature is used.

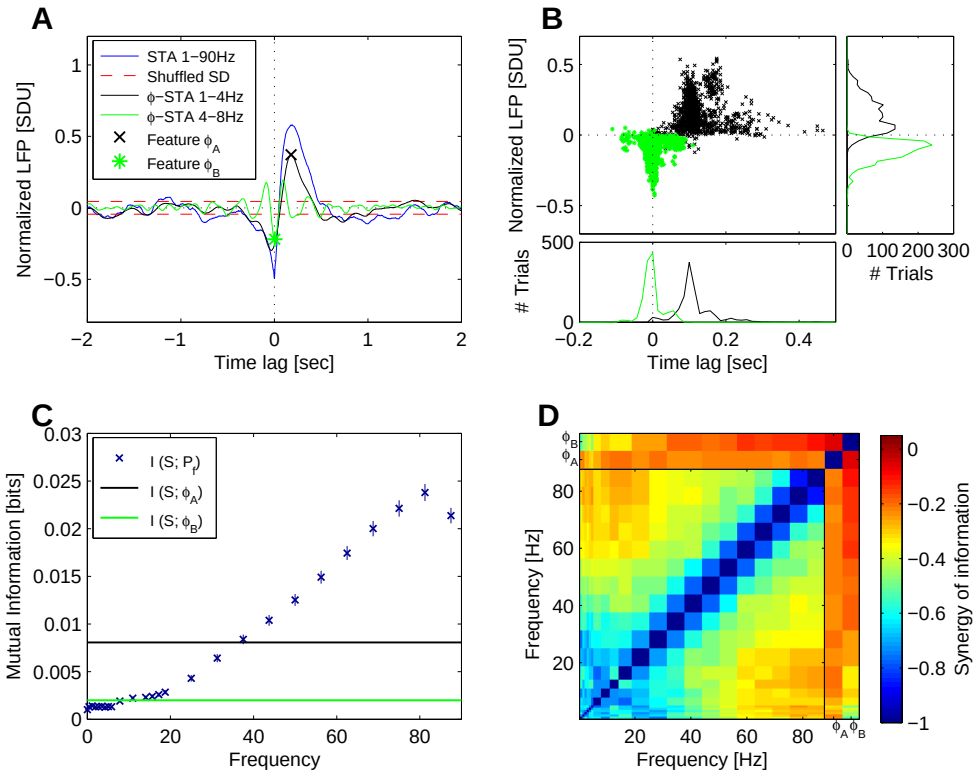


Figure 2.7: Information about spiking activity conveyed by phase features and frequency power features. **A:** Spike-triggered averages (STA). LFP STA (blue line) as well as the STA for the 1-4 Hz (black line) and 4-8 Hz (green line) band-pass filtered LFP. In the latter cases all power modulation is discarded (see Methods), yielding purely phase-related signals. The definitions of phase features ϕ_A (black cross) and ϕ_B (green star) are indicated. **B**: Distribution of lags of features ϕ_A and ϕ_B for all trials of the anesthetized V1 data. **C:** Information about spikes conveyed by single features. **D**: Redundancy of information about spikes conveyed by any combination of LFP features. Color coding indicates the amount of information synergy (see Methods 2.2, Eq. 2.4). See Results 2.3 for discussion.

Information contained in the power decreases monotonically with frequency. This decrease can also be seen on the level of individual trials (not shown). If one uses either one of the low-frequency phase features ϕ_A and ϕ_B on its own, information drops to about half for ϕ_A and much lower for ϕ_B in comparison to the best power feature (black and green lines in Fig. 2.7 C, respectively). Despite the usefulness of low-frequency power modulation in combination with high-frequency powers (as shown in the feature selection), low-frequency power exhibits poorer performance as a single feature than the phase features (in particular in comparison to ϕ_A, see Fig. 2.7 C). Note that the timing resolution of the phase features is much higher as phase is defined at any moment in time, whereas power has to be estimated within a window of sufficient length. The induced temporal correlation of nearby time points for the low frequency power seems to be too high to predict spike times on its own.

If two LFP features F_1 and F_2 conveyed independent information about spiking activity S, the normalized measure for synergy of information $\mathrm{syn}(F_1, F_2|S)$ (Eq. 2.4) would be zero. In general, this measure ranges from minus one for completely redundant information to one for completely synergistic information (see Methods 2.2 for details).

Figure 2.7 D shows the average normalized synergy of information about the spikes for all combinations of features. Here synergy of information is calculated on the basis of single trials, where trials having joint information not significantly above zero information are excluded (Wilcoxon signed-rank test, p-value above 0.1). Generally, information conveyed by high-frequency bands is mainly independent from information contained in low-frequency bands. The information in individual high-power features is more redundant (e.g. for 87 Hz and 50 Hz, $\mathrm{syn}(P_{50Hz}, P_{87Hz}|S) = -0.40 \pm 0.02$) than between high-power features and phase features, where information is nearly independent (synergy values with high-frequency power features around -0.2; for instance $\mathrm{syn}(\phi_A, P_{81Hz}|S) = -0.21 \pm 0.01$ and $\mathrm{syn}(\phi_B, P_{81Hz}|S) = 0.14 \pm 0.02$). Phase feature ϕ_B, becomes more redundant with power for decreasing frequency (e.g. $\mathrm{syn}(\phi_B, P_{2Hz}|S) = 0.45 \pm 0.03$), whereas ϕ_A redundancy is relatively low even with low-frequency powers (e.g. $\mathrm{syn}(\phi_A, P_{2Hz}|S) = 0.23 \pm 0.03$). However, redundancy between any two low-power features is much higher. Both phase features convey almost independent information about spikes ($\mathrm{syn}(\phi_A, \phi_B|S) = -0.05 \pm 0.02$).

Note that the high redundancy of information in two high-frequency powers which are less than 20 Hz apart is a result of spectral estimation, which is done on the bandwidth of 21 Hz (see Methods 2.2).

In summary, both feature types, high-frequency power and low-frequency information, seem to code for mostly independent information, while two high-power features convey more redundant information.

2.3.5.3 Prediction performance is related to clusters of spikes

We noticed that generally prediction performance is superior on data where spikes tend to cluster to bursts of activity with relatively long silent periods in between. This can be seen if one correlates the inter-spike interval coefficient of variation (ISI-CV) with prediction performance. If spikes are temporally clustered, many short intervals are interspersed with few very large intervals, causing a large value for ISI-CV, the ratio of standard deviation to mean of the inter-spike interval distribution. Thus ISI-CV can be seen as an approximate measure for the degree of temporal clustering of spike trains. One notices a strong correlation between the ISI-CV and prediction performance (rank correlation 0.86), whereas prediction performance is only poorly correlated with the firing rate (0.47). This behavior can be seen also for individual features. High-frequency power features have the highest correlation with ISI-CV (0.92), whereas correlation with single phase features is lower (e.g. 0.64 for ϕ_A). Correlation with rate for both features is much lower (high gamma frequency 0.34, phase feature ϕ_A 0.44). We found that if one defines larger clusters of spike or burst events directly and discards single spikes in between (see Fig. 2.8), low phase features are locked to the timings of such bursts and the performance of phase feature ϕ_A is highly correlated with the burst rate (rank correlation 0.82). In our burst definition (see caption of Fig. 2.8) the average burst length is 122 ± 1 ms, which suggests that low frequency phase information preferentially codes for (rather sustained) bursts of activity.

2.4 Discussion

Local field potentials are the best indicators of integrative activity in an area. They reflect the area's input activity in terms of population excitatory and inhibitory post-

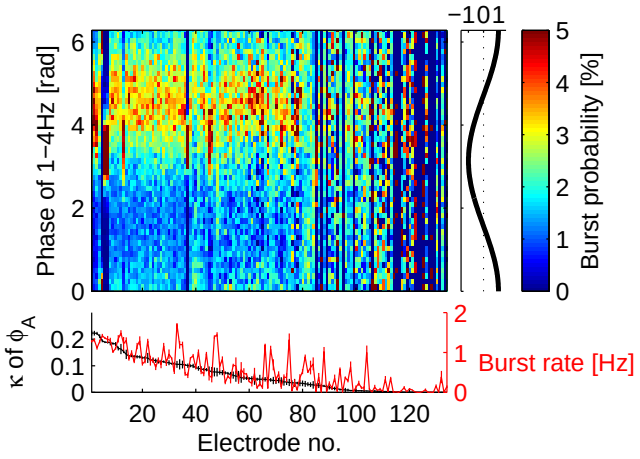


Figure 2.8: Locking of high spiking activity events to phases of slow LFP oscillations. Here a high spiking activity event (burst) is defined as having at least 10 spikes. Spikes constituting a burst have to occur within a maximal mean inter-spike interval of 5 ms, and spikes not contained in a burst are deleted. Bursts have to occur at least 25 ms apart from each other, otherwise they are regarded as one continuous event. The middle position is taken as the timing of an event. The probability of event times occurring in a particular phase of the 1-4 Hz oscillation of the LFP is plotted for individual electrodes (averaged over ca. 10 trials recorded at the same electrode site). Electrodes are sorted according to the performance κ of the low-frequency phase feature ϕ_A. For about 60% of all electrodes burst times are locked consistently to the upswing of a wave in the low LFP band. For most of the rest of the electrodes too few or no bursts are found when the above definition of a burst event is applied. Phase is illustrated in the right margin plot. Burst rates correspond well to the performance of ϕ_A (rank correlation 0.82, lower plot).

synaptic potentials, but also the area's regional processing, because they are directly affected by dendritic spikes, voltage-gated oscillations and various after-potentials – all markers of diverse neural computations (see Buzsáki (2006) for an overview). Not surprisingly, an increasing number of studies report their specificity and usefulness in the search for neural correlates of behavior (Kreiman *et al.* (2006); Lee *et al.* (2005); Liu & Newsome (2006); Osipova *et al.* (2006); Rubino *et al.* (2006); Scherberger *et al.* (2005)). Although these studies show that LFPs convey information that is to some degree independent of spiking activity, it has been suggested and demonstrated by many researchers that spikes synchronize to - or that synchronization gives rise to - specific oscillation frequency of the LFPs, in particular γ-bands (> 40 Hz) in visual cortex and θ-band (4-8 Hz) in the hippocampus (see Buzsáki (2006) and references therein).

Herein we investigated the relation of spiking activity to LFP on a more fundamental level by asking which aspects of the LFP can generally be exploited to predict spike times. Unlike other approaches in which simple linear interaction between both signals is tested (e.g. with means of coherence or correlation), the classification approach used in our study can exploit multiple different features of the LFP simultaneously in a non-linear way to infer spiking activity from LFPs. We found that when the best single feature is used, performance drops to about 70%, showing that multiple features are essential. The non-linear component in the code is rather small, but there is still an average increase of 12% when SVM classifiers are employed.

In contrast to V1, the prediction of spikes from LFPs is (almost) impossible in the LGN. The reason for this could be that spiking activity might be less correlated and thus have less effect on the LFP, or alternatively that the geometrical arrangements of current sources and sinks in the thalamus generates field potentials of lower spatial specificity than those observed in cortex. Furthermore the fact that 80% of the LGN input is cortical and modulatory is an additional potential reason for the LFP-spiking decoupling in this nucleus. Neuromodulation might explain our observation in simultaneous thalamus-cortex recordings that V1 LFP cannot predict LGN spiking. LGN-LFP, on the other hand, is a reasonable predictor of V1 spiking, likely because the former is a better indicator of local LGN activity that is correlated with the V1 LFPs, which in turn can predict the spiking of this cortical area.

From the point of view that LFP is mostly generated by the totality of synaptic input and local processing in a region one might ask whether it is possible to predict spikes

solely on the basis of information preceding the actual time of the spike. To evaluate this possibility we recomputed prediction performance for the same feature types but with shifted lags, so that only causal information of the LFP is included. When we did this, prediction performance dropped to 68% of the average non-causal classifier used for the population analyses, which is still well above chance level. However, conclusions about the relation between synaptic input and spiking output are difficult to draw from this number. For instance it is likely that due to temporal correlation in the spike train (and various LFP bands) neighboring times in the time series are good predictors for each other.

2.4.1 LFP features related to spiking activity

Besides having generated and visualized concrete spike trains inferred solely from information contained in the LFPs, our analysis revealed those features which are the most important carriers of information about spiking activity in the LFP and estimated their relative importance and redundancy properties.

The first and most useful feature for inferring spike trains from LFPs is the power modulation of high-frequency components in the upper γ-band from 40 Hz to (at least) 90 Hz. This is in good agreement with other studies, which have established a link between gamma frequency bands and spiking activity (Csicsvari *et al.* (2003)). The biophysical origin of gamma frequencies in the LFP remains a topic for current research. The rather fixed relationship between spikes and LFPs over a wide range of data and conditions could reflect physical constraints (layered organization of the cortex, distribution of sinks and sources) or inherent properties of the neural network topology and function. For instance, gamma activity might be the effect of fast inhibitory circuits on the LFPs (see Bartos *et al.* (2007) for a recent review).

In contrast to power modulations, phase in the γ-range is much less important for prediction than one might have expected from the well-documented fact that spikes are synchronized in this range (Gray *et al.* (1989); Kreiter & Singer (1992)). Our stimuli consist of cinema movies of several minutes duration and containing a mixture of objects, faces, actions, colors, and edges. Thus, specific object encoding might be only weak and cluttered with other aspects of the code, as the stimulus comprises multiple objects at any given time. Because we assume a stationary spike-LFP relation for the

duration of a trail (170 s), more subtle aspects of the relation such as transient rapid neuron-to-neuron synchronizations would be averaged out and therefore not detected as a useful feature. Another possibility for the relative unimportance of gamma phase features might be that different subgroups of neurons present in the MUA signal lock to different phases in the gamma cycle, as suggested for inhibitory and excitatory neurons (Hasenstaub *et al.* (2005)) In general, if two subgroups of units present in the MUA signal locked to different feature of the LFP, our classification method should not degrade in performance, as it can exploit many features simultaneously. However, the relative usefulness of both features would indeed decrease in the average features analyses.

Locking to phases of higher frequencies is difficult to detect because even a small amount of jitter in spike time precision abolishes locking. Such jitter might be introduced during binning of the spike timings to LFP sampling resolution (5 ms). This is in contrast to the effect of a small amount of jitter at lower frequencies. Accordingly, the phases of low frequency components are indeed useful for predicting spiking activity. Moreover, our analysis suggests that the oscillation phases of low frequencies code for larger bursts of temporally correlated spiking activity. In fact, the high probability of spikes occurring in clusters in our V1 data helps to infer spikes from LFPs. Prediction quality is highly correlated with ISI-CV, because one can predict spiking activity on a time scale of > 25 ms, rather than reaching single spike precision. In the LGN data, where the spikes are less clustered (low ISI-CV, average of 1.1 ± 0.05 compared to 2.4 ± 0.05 in anesthetized V1 and 1.8 ± 0.05 in awake V1), it is almost impossible to predict spikes from LFPs. Since we observe stronger clustering for anesthetized data, we conclude that this might be partly an effect of anesthesia (Steriade *et al.* (1993)).

The low-frequency power modulations which are selected as useful features in a part of trials probably have origins similar to those of the slow phase features, and may code for the relative amount and size of clusters of activity. Thus low-frequency power modulations might provide a slow changing state variable which is useful in combination with the fast-changing high-frequency powers, as indicated by the low redundancy values.

About half a century ago, several studies were conducted which attempted to relate electrical encephalographic signals (EEG) to spiking activity. Then it was found that spikes occur preferably at the negativity of 0.2-2 Hz waves (Fromm & Bond (1964)).

Since EEG has a similar basis of origin as LFPs (Nunez & Shrinivasan (2006)), this confirms our findings, where spikes occur at the minimum to rising phase of the slow oscillation. It is not just the negativity that is important, however, but also the peak that is seen following the spikes. This is similar to slow-wave sleep of cats as seen in the spike-triggered average (Destexhe *et al.* (1999)) and is taken to the extreme in spike-wave complexes, which can be observed in cases of epilepsy (e.g. Destexhe *et al.* (2001a)). In the latter case, it was suggested that the slow positive wave form after a burst of activity is accompanied by neuronal silence and can be explained by slow inhibitory effects mediated by $GABA_B$ (Destexhe *et al.* (2001a)). Although, we observe neither clear up and down states nor any pathological periodic activity in the LFPs of our data, the similarity of the STA suggests that such processes may play a role during more physiological states, albeit in a much weaker form.

2.4.2 Effects of spike detection method

The spiking signal used in this paper is generated by a simple threshold-based procedure for detecting spiking events from the recordings. This detecting procedure is prone to false-positive detection, as well as to a smaller fraction of missed spikes originating at larger distances or from neighboring interneurons. With our method we expect a false-positive rate of 2 Hz, if one assumes that non-spike values (i.e. noise) in MUA amplitudes are subject to Gaussian distribution (see Methods 2.2). Thus the spiking signal contains not only the activity of multiple neurons and possibly of different cell types, but also noise spikes. Since noise spikes should be independently distributed with respect to time, false positive labels should actually reduce performance compared to the ”true” spiking signal. We tested higher spike detection thresholds, where the contribution of noise becomes negligible (e.g. for 5 times SD only 0.05 Hz), and found instead reduced prediction performance, although performance still remained well above chance (not shown). Since it was estimated that in principle spikes arising from up to 1000 neurons could detected by a single electrode (Henze *et al.* (2000)), a spiking signal generated with a higher threshold will include fewer smaller spikes from neurons that are further away, and incorporate only those neurons which happen to be in the immediate neighborhood of the electrode tip. Thus the spiking signal becomes more local and its relationship to the relatively global LFP signal

is naturally weaker. Therefore noise spikes are not likely to artificially enhance our performance results, but influencing the number of neurons whose spikes are detected via threshold will have an effect.

Similarly, spike sorting yielding multiple single unit activities rather than multi-unit activity would naturally decrease prediction performance, since spiking signals become more local and, additionally, the prediction task would change from binary classification to a more difficult multi-label classification task.

2.4.3 Encoding of the stimulus

We found that the relationship between LFPs and spikes is almost unchanged during stimulus-driven activity compared to spontaneous activity. This finding corresponds well to the results of others (Fiser *et al.* (2004); Kenet *et al.* (2003); Vincent *et al.* (2007)), who found that the structure of spontaneous activity is rich and sometimes even resembles stimulus-driven activity. While there is no general change in the structure of features, there are of course transient aspects of the stimuli encoded by spikes (and LFPs) that have an effect on the spike rate, for example, which increases during stimulus presentation (not shown). Because of the encoding of the changing movie stimulus over time, the temporally contiguous training and test sets might differ in their LFP-spikes relation, i.e. in their sample distributions. Thus movie encoding might explain why prediction performance on spontaneous activity is slightly better than on stimulus-driven activity.

Effects of stimulus encoding might also explain the observation that the LFP is ”more global” during spontaneous activity in that prediction of spikes degrades much more slowly with increasing cortical distance than during stimulus presentation. Thus the stimulus actually decorrelates neural activity spatially in comparison to spontaneous activity.

Note that we do not analyze features which actually encode information about the *stimuli*. Information contained about stimuli will be analyzed in a forthcoming paper (Belitski *et al.* (2007)). Belitski et al. analyze the information content of frequency power about the stimulus (movie) in the very same data set that we investigate here. Although we show here that the structure of spikes to LFP does not change considerably during spontaneous activity, Belitski et al. nevertheless show that very low and

very high oscillation powers of the LFP are highly informative about the stimuli. They find that high LFP power series (50-120 Hz) contain information about the stimulus which is partially redundant with that in power series derived from the spike trains. In this respect, their results complement ours.

2.4.4 Conclusion

We conclude that to a certain degree spikes can be inferred from LFPs; a fact that reflects the interaction of these signals. However, we find that millisecond precision, which has been shown to be used for temporal coding (Mainen & Sejnowski (1995)), cannot be inferred from LFP. The temporal aspects of neural spiking used for information coding, rate coding, or coding on spike timing, remain a topic of current research (Rieke *et al.* (1999)). We might conservatively say that whether or not they are important for coding, time-varying rates on the scale of about a hundred milliseconds can be moderately well inferred from the LFPs, but that exact timings cannot. Thus, given our results, it should in principle be possible to develop an appropriate methodology that permits the extraction of certain spiking features from signals measured by methods relying on LFP-like signals, such as fMRI (Logothetis *et al.* (2001)) or optical recordings (Grinvald (1985); Grinvald & Hildesheim (2004)). Nevertheless, the strong dependence of spike predictability on electrode position suggests that the reliability of such predictions may depend on the brain site. Finally, the fact that in the thalamus is practically impossible to predict spikes from the LFP suggests that computations based on input, local processing and output - as instantiated in the different frequency bands of the mEFP - can only be helpful for structures with the appropriate element geometry (e.g. fascicles of pyramidal cells vs. potentially close-field arrangement of thalamic neurons) and the proportion of driver to modulator afferents.

Chapter 3

Phase-of-firing coding of natural visual stimuli in primary visual cortex

"Die Wahrnehmung der Sinne geschieht uns unbewußt: alles, was uns bewußt wird, sind schon bearbeitete Wahrnehmungen."

[Friedrich Nietzsche, 1885]

3.1 Abstract

We investigated the hypothesis that neurons encode rich naturalistic stimuli in terms of their spike times relative to the phase of ongoing network fluctuations, rather than only in terms of their spike count. We recorded Local Field Potentials (LFPs) and multi-unit spikes from the primary visual cortex of anaesthetized macaques while binocularly presenting a color movie. We found that both the spike counts and the low-frequency LFP phase were reliably modulated by the movie and thus conveyed information about it. Moreover, movie periods eliciting higher firing rates also elicited a higher reliability of LFP phase across trials. To establish whether the LFP phase at which spikes were emitted conveyed visual information that could not be extracted by spike rates alone, we compared the Shannon information about the movie carried by spike counts to that carried by the phase of firing. We found that, at low LFP frequencies, the phase of firing conveyed 54% additional information beyond that conveyed by spike counts. The extra information available in the phase of firing was crucial to disambiguate between stimuli eliciting high spike rates of similar magnitude. Thus, phase coding may allow primary cortical neurons to represent several effective stimuli in an easily decodable format.

3.2 Methods

3.2.1 Experimental Methods

Four adult rhesus monkeys (Macaca mulatta) participated in these experiments. All procedures were approved by the local authorities (Regierungsprsidium) and were in full compliance with the guidelines of the European Community (EUVD 86/609/EEC) for the care and use of laboratory animals. Prior to the experiments, form-fitted head posts and recording chambers were implanted during an aseptic and sterile surgical procedure (see e.g Logothetis *et al.* (2002)). To perform the neurophysiological recordings, the animals were anaesthetized (remifentanil (typical 1microgram/kg/min)), intubated and ventilated. Muscle relaxation was achieved with mivacurium (5mg/kg/h). Body temperature was kept constant, and lactated Ringer's solution was given at a rate

of 10ml/kg/h. During the entire experiment, the vital signs of the monkey and the depth of anesthesia were continuously monitored. For the protocol used in these experiments, we had previously examined the concentration of all stress hormones (catecholamines) (Logothetis *et al.* (1999)) and found them to be within the normal limits. Drops of 1% ophthalmic solution of anticholinergic cyclopentolate hydrochloride were instilled into each eye to achieve cycloplegia and mydriasis. Refractive errors were measured and contact lenses (hard PMMA lenses by Whlk GmbH, Germany) with the appropriate dioptric power were used to bring the animal's eye into focus on the stimulus plane. The electrophysiological recordings were performed with electrodes that were arranged in a 4 x 4 square matrix (interelectrode spacing varied from 1 mm to 2.5 mm) and introduced each experimental session into the cortex through the overlying dura mater by a microdrive array system (Thomas Recording, Inc., Giessen, Germany). Electrode tips were typically (but not always) positioned in the upper or middle cortical layers. The impedance of the electrode if the last spike occurred more than one millisecond earlier. All results in the main text were obtained from the spikes detected with this simple threshold method, which is good to detect spike times but not to isolate single units. Thus the spikes used for the analysis represented the spiking activity of a small population of cells rather than well separated spikes from a single neuron. Unless otherwise stated, all results reported were obtained using these unsorted spikes.

3.2.1.1 Spike sorting for control analysis

To check that the amount of information in the phase-of-firing did not change when considering only spikes from well-isolated single neurons (see Information Analysis Section below), we sorted the spikes from a selected number of channels which appeared to be suitable for the task. For spike sorting we used the method described by Quiroga *et al.* (2004). The spike waveforms were extracted around the detection times as described above (in a region of 0.75ms before to 1.75ms after the detected spike). These spike forms were interpolated and 10 wavelets features (with 4 scales) were extracted (Quiroga *et al.* (2004)). From these feature pool the 10 most salient features (KS-test) were used as input for the clustering algorithm. We then sorted the spikes using the paramagnetic algorithm of Quiroga *et al.* (2004). For each electrode a few

reasonable clusters were selected by visual inspection of the spike waveforms ensuring a reasonable distinguishable average waveform among clusters. After this initial selection, spikes which initially were not classified in a particular cluster (or belonging to not selected clusters) were forced to belong to the nearest selected cluster (Mahalanobis distance (Quiroga *et al.* (2004))). A cluster which maintained very similar waveforms after this step was deemed to be a well-isolated cluster and was considered for further analysis. Otherwise the cluster was not considered further for spike sorting. In this way, we obtained 71 well isolated clusters from 37 channels.

3.2.1.2 The power spectrum of LFPs

To document the spectral content of the LFPs, and assess how the LFP power changed over different segments of the movies, we divided each movie into non-overlapping time windows of length $T = 2.048$ s comprising $L = 1024$ datapoints sampled at 500 Hz. The power of the neural oscillation in each time window was quantified by computing the power spectrum at each frequency, independently for each trial. The power spectrum in each window was obtained using the multitaper technique (Percival & Walden (2002)), because it provides a very efficient way to simultaneously control the bias and variance of spectral estimation by using multiple Slepian data tapers. The Slepian functions are defined in terms of their length L in time and their bandwidth W in frequency. We varied LW between 1 and 8 obtaining essentially identical spectra. In the following, we will present results obtained using $LW = 2$. To reduce the spectral bias, the average over tapers was computed using the adaptive procedure described in Percival & Walden (2002).

Figure 3.1 A shows the average spectrogram across all movie trials and windows. The highest LFP power was at low frequencies ($= 8$ Hz), and the power decreased steeply at increasing frequencies. We compared the averaged LFP spectrogram evoked during the movie to the LFP spectrogram of the same electrode during spontaneous activity (measured in the absence of visual stimulation). The evoked and spontaneous LFP spectrograms were similar at frequencies lower than 25 Hz, but there was an increase of power during movie stimulation at very low frequency ($= 4$ Hz). Consistent with previous studies (Frien *et al.* (2000); Henrie & Shapley (2005)), we found a very substantial power increase over spontaneous activity of the movie-evoked responses

in the gamma frequency region 40-120 Hz. This power increase was proportionally higher at 70-100 Hz.

To compare the LFP spectrum with the spectrum of the stimulus dynamics, we also computed the power spectrum of the contrast variations in each 2.048 s window of the movie. Contrast in each frame was computed by taking the standard deviation of luminance of all pixels in a frame and divided it by the mean of luminance in that frame. The resulting spectrogram was then averaged over time windows. The resulting average spectrum of contrast variations in the movie dropped with increasing frequency and at low frequencies it could roughly be approximated as $1/f^2$ (where f is the frequency).

It is interesting to consider at which LFP frequencies the power is more modulated by the movie, and at which frequencies the power is more reliable on a trial-by-trial basis. To address this question, we computed, independently for each frequency and electrode channel, the coefficient of variation (CV) across the movie time (i.e. across stimulus) of the mean power. This (denoted as "Signal CV") quantifies the degree to which the signal changes with the stimulus, and thus can potentially encode the stimulus. We then computed, again for each frequency and channel, the "Noise CV": this is the CV across trials of the fluctuations about the mean for each stimulus. In other words, it quantifies the average unreliability of the power, across trials at fixed stimulus. The magnitudes of the noise and signal CV cannot be directly compared, because they are computed over a different number of data points. However, noise CV or signal CV at different frequencies can be directly compared: frequencies of higher signal CV have a better capacity of following the stimulus features in the movie, and thus to transmit information. Frequencies with lower Noise CV are more reliable across trials and thus can represent the stimulus more reliably. Figure 3.1 B shows that (on average across all channels) the noise CV is relatively stable across frequencies (apart from an increase in the range 16-24 Hz). The signal CV, on the other hand, is maximal for low frequencies (< 12 Hz) and high gamma frequencies (60-100 Hz).

3.2.1.3 Spike probability as function of phase across all LFP frequencies

Figure 3.2 reports the population average of the probability of a spike being emitted at a given phase, for each frequency band. We found that the spike probability was modu-

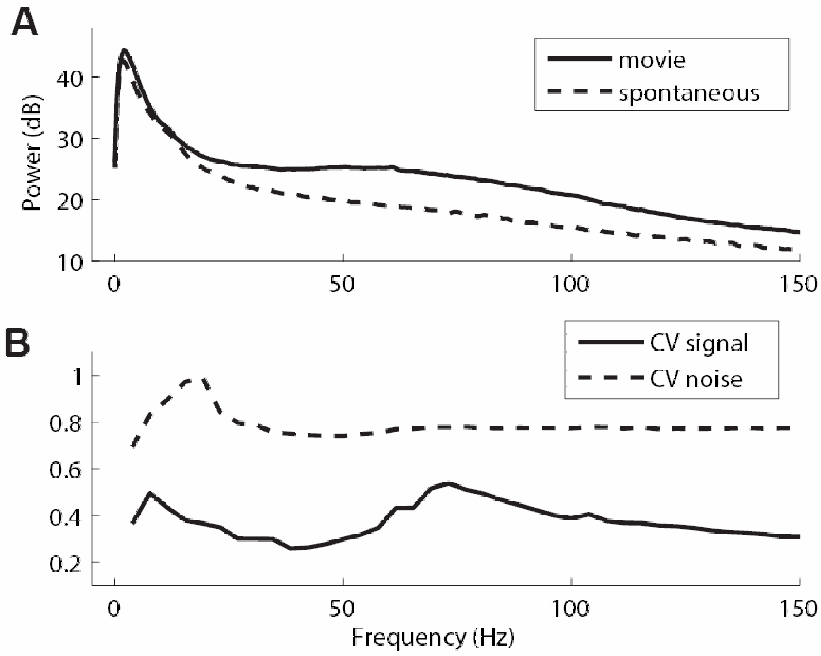


Figure 3.1: Spectral properties of LFPs. **A**: The spectrogram averaged over the movie time is plotted (solid black line). The dashed line is the average spectrogram during spontaneous activity, showing an increase of power at very low frequencies (< 4 Hz) and in the 20-150 Hz bands during visual stimulation. **B**: The average signal and noise CV of LFPs as function of frequency.

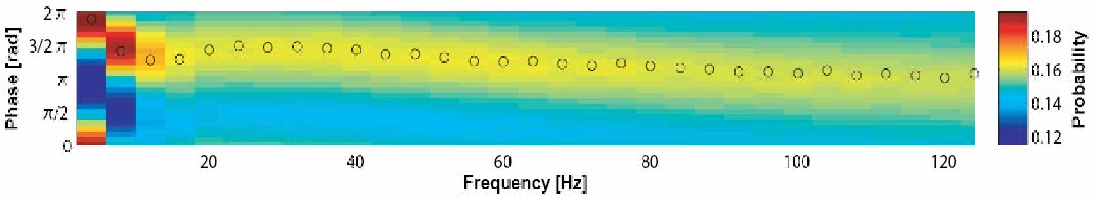


Figure 3.2: Phase locking properties of V1 neurons under natural visual stimulation. The color plot shows the probability (per unit angle) of finding a spike locked at a certain phase as function of the LFP frequency, averaged across the whole dataset. The circles represent the preferred phase in each frequency class (angular mean).

lated by LFP oscillations at all frequencies. The preferred phase for spiking was dependent on frequency. The preferred phase in the 1-4 Hz frequency range was (on average across the population) 1.9π (middle of the raising phase). The population-averaged preferred phase value decreased as the LFP frequency increased, and saturated around 1.2π (just before the trough of the oscillation) for high frequencies.

3.2.1.4 Circular statistics

The angular mean across trials of the LFP at a given time window during the movie was computed as (Fisher (1993))

$$\bar{\phi} = \arg \langle \exp(i\phi_j) \rangle, \tag{3.1}$$

where ϕ_j is the series of phase values at each trial j, $\arg$ denotes that we are taking the argument of the complex number, and $\langle \ldots \rangle$ is the average across trials at fixed time window.

The circular variance (Fisher (1993)) across trials (CVAR) of the phase was computed as

$$\mathrm{CVAR} = 1 - |\langle \exp(i\phi_j) \rangle| \,. \tag{3.2}$$

CVAR values range from 0 (perfectly repeatable phases) to 1 (random phases over the circle).

3.2.2 Information theoretic analysis

3.2.2.1 Definition and meaning of Information

The mutual information (abbreviated to "information" in this paper) between a set of stimuli S and a set of neural responses R is defined as (Cover & Thomas (1991); Shannon (1948)):

$$I(S;R) = \sum_{r,s} P(s)P(r|s)\log_2 \frac{P(r|s)}{P(r)} \tag{3.3}$$

where $P(s)$ is the probability of presenting stimulus s, $P(r|s)$ is the probability of observing response r given presentation of stimulus s, $P(r)$ is the probability of observing response r across all trials to any stimulus. $I(S;R)$ quantifies how much of the information capacity provided by stimulus-evoked differences in neural activity is robust to noise. An alternative but equivalent interpretation of $I(S;R)$ is that it quantifies the reduction of uncertainty about the stimulus that can be gained from observation of a single trial of the neural response. When base-2 logarithms are used, as in Eq. 3.3, information is measured in *bits*. One extra bit of information reduces the uncertainty about the stimulus after one observation of the neural response by a factor of 2. $I(S;R)$ is zero only when the stimulus and the response are statistically independent quantities, indicating that no knowledge about the stimulus can be gained by observing the response. When stimulus and response present statistically significant covariations, then the value of information is positive. The more the responses and the stimuli are mutually predictable, the larger $I(S;R)$. Unlike other simpler correlation measures, information has the unique advantage of capturing all nonlinear dependences of any statistical order that may be present in the data.

3.2.2.2 Quantifying the information carried about a movie stimulus

To quantify how neural responses encode movie stimuli, we computed the information that the neural responses convey about *which part of the movie* was being presented. This was done (following de Ruyter van Steveninck *et al.* (1997); Strong *et al.* (1998)) by dividing the movie presentation time into a number of non-overlapping windows of size Δt, and considering each different movie window as a different stimulus s (because the associated neural response is evoked by a combination of visual features

happing in its immediate past). Since our definition of "stimulus" does not need to spell out which visual features (e.g. contrast, orientation, etc) are encoded, the information we compute is about all the possible visual attributes that may activate the neuronal signals, and does not rely on any ad-hoc assumptions about stimulus selectivity (de Ruyter van Steveninck *et al.* (1997)). This makes this formalism particularly suited to the analysis of neural responses to complex, naturalistic stimuli.

We note that, since the response collected in individual windows is considered separately, this quantification of information ignores the information potentially present in the correlation between the times of the spikes emitted in different stimulus windows.

3.2.2.3 Representation of the spike count and phase-of-firing codes

Although the evaluation of information does not require any *ad-hoc* assumption about what are the relevant stimulus features, it does require an explicit representation of the neural response r. By specifying the neural response information can be used to probe different neural codes and compare their efficiency in capturing information about the stimulus. In particular, here we estimated the amount of information about the stimulus conveyed by two candidate neural codes: the *spike count code* and the *phase-of-firing code*.

We proceeded as follows. Each movie had a duration T of 3.5-6 mins, and was repeated N_s times (N_s varying from 12 to 44 depending on the experimental session; 57 out of 78 recordings were obtained with $N_s = 30$). Each of the spike trains corresponding to those stimulus presentations was divided into non-overlapping time bins of duration Δt, which were labeled by an ascending number s. Then, the actual response r at a particular time bin s, was determined according to which of the two codes was analyzed.

When considering the *spike count code* the neural response r was equal to the number of spikes n fired in that time window. In this case, Eq. 3.3 becomes:

$$I_{\text{count}}(S;R) = \sum_{n,s} P(s)P(n|s)\log_2 \frac{P(n|s)}{P(n)} \tag{3.4}$$

where $P(n|s)$ is the probability of observing n spikes in window s. As we used short stimulus windows ($\Delta t = 4$ ms unless otherwise stated; see below) most of the

time windows contained either 1 or zero spikes. We verified the information loss resulting from assuming a binary spike train, rather than considering the full spike count ($n = 0, 1, 2, \ldots$) in each window, was negligible and statistical insignificant ($p > 0.9$; bootstrap test). For simplicity, we therefore present results obtained with a binarized 0-1 code.

In the case of the *phase code* the neural response r was coded as a spike train where at the position of each spike we added a "label" ϕ indicating the phase ϕ of the LFP oscillation at that particular time. In practice, this was achieved as follows. Since the phase is a continuous quantity, to sample the required probabilities from the finite dataset we discretized the possible values of the phase by diving the cycle 0-2 π into a number Φ of uniformly spaced bins. (We note that, for each given recordings site, frequency range and number of equispaced phase bins to be used, we set the exact boundaries of the phase bins so that to extract maximal information from the phase code under these conditions. We found that, in general, the maximal values of information were obtained when setting the starting point of one of the bins to be equal to the preferred phase of firing. This corresponded to a division of phase in quadrants as reported in Fig. 3.8 H.) The phase label ϕ ($\phi = 1, \ldots, \Phi$) thus indicated the number of the bin to which the continuous phase value belonged. If there was one or more spike at the time bin corresponding to time t, then the response was set equal to the phase bin at the time of the occurrence of the first spike in the stimulus window. Later spikes were rare as there was typically only up to one spike per stimulus window (see above). However, neglecting later spikes can only reduce information due to the "data processing inequality"(Cover & Thomas (1991)). Therefore this procedure ensures that the information value estimated for phase code is a lower bound to its real value. In summary, symbols "$1 - \Phi$" were used to code the angle bin at which a spike was emitted, and symbol "0" was used to denote the absence of a spike; thus the neural response in a given time bin can take one of $\Phi + 1$ different symbols, and the information in Eq. 3.3 becomes as follows:

$$I_{\text{phase}}(S; R_\Phi) = \sum_s P(s) \sum_{\phi=0}^{\Phi} P(\phi|s) \log_2 \frac{P(\phi|s)}{P(\phi)} \tag{3.5}$$

In Eq. 3.5, $P(\phi|s)$ is the probability of observing the symbol ϕ in response to stimulus s. In this chapter (unless otherwise stated), the phase-of-firing information

values were obtained by dividing the phase into quadrants (i.e. using $\Phi = 4$ bins), which amounts to registering the phase with a precision of $\pi/2$ rad. However, the number of phase bins is a somehow arbitrary quantity. To estimate the phase precision needed to read out the phase information, we varied the precision with which the phase was quantized for the information calculation and we assessed the effect of the choice of the number of phase bins. Results are reported in Fig. 3.3. When dividing the period in two halves (i.e phase precision $= \pi$; $\Phi = 2$), the information in the 1-4 Hz LFP phases was reduced by a factor of 22% with respect to the phase coding information obtained with a $\pi/2$ precision (quadrants). When registering the phase with precision $\pi/4$ rad (period divided into $\Phi = 8$ bins), no information was gained with respect to the $\pi/2$ rad precision case. This indicates that a precision of π was not enough to read the whole information in phase, whereas a precision of $\pi/4$ did not add anything useful to what already obtained with $\pi/2$. Thus, $\pi/2$ appears to be the precision needed to read out the phase code. This is compatible with the phase reliability values reported in Fig. 3.9. These results extended to higher LFP frequencies as well: results for the 4-8 Hz LFP were reported in Fig. 3.3.

3.2.2.4 Effects of variation of the size of the stimulus window

The size Δt of the "stimulus" time window (in which the response is computed with the aim of reconstructing by which movie part it was provoked) is a free parameter of the analysis. In all the results and figures in this chapter we used $\Delta t = 4$ ms unless otherwise stated. We tested the effect of varying Δt on the 1-4 Hz LFP phase of firing. We found that the extra amount of information (beyond that carried by spike counts) obtained by labelling the spikes with the phase of the1-4Hz LFP increased from 55% with $\Delta t = 4$ ms, to 62% with $\Delta t = 8$ ms and to 73% with $\Delta t = 16$ ms. Average information results at other LFP filtering frequencies are reported in Fig. 3.4 for the $\Delta t = 4$ ms and $\Delta t = 8$ ms cases. The increase of phase-of-firing information with the size of Δt is compatible with previous analytical results (Panzeri & Schultz (2001)) showing that (for relatively small stimulus windows) the extra information in a spike time code typically increases monotonically with the size of the time window considered (either linearly or as the third power of the time window depending on the relation between spike timing precision and the temporal scales of stimulus-induced responses

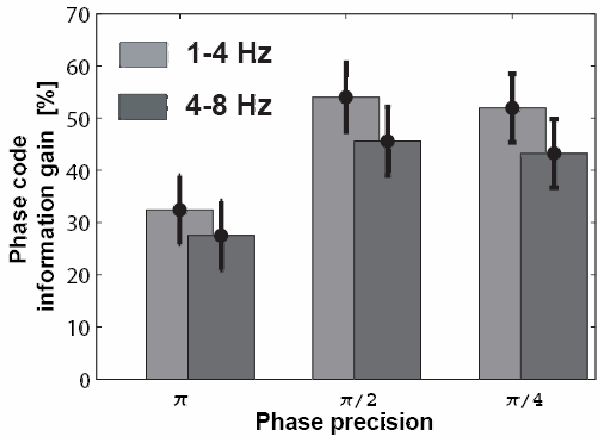


Figure 3.3: Effect of phase precision of phase-of-firing information. The percentage of extra information (above and beyond spike counts) of the phase code as function of the precision used to sample the phase. Results are reported as mean ± SEM over the dataset. Grey and black histograms report the phase code information gain obtained using 1-4 Hz and 4-8 Hz LFPs, respectively.

(Panzeri & Schultz (2001); Panzeri *et al.* (2001))). We decided to focus our presentation mainly on the results obtained using a stimulus window of $\Delta t = 4$ ms because (i) this fine resolution permitted a computation of phase coding information even for high frequency LFPs up to the gamma range (ii) we aimed at providing conservative estimates of the extra information that can be extracted from the phase of firing.

3.2.2.5 Estimation of information from limited samples

In order to estimate information it is necessary to estimate the conditional probability $P(r|s)$ from the experimental dataset. Given that the amount of available data is limited, the estimation of the probabilities suffers from random statistical errors. These errors translate into a bias in the estimation of the information (Panzeri & Treves (1996); Panzeri *et al.* (2007)), which for limited data sets can represent a significant fraction of the total information. If information is computed directly from Eq. 3.3, the bias is positive and is bigger for the phase-of-firing code than for spike count code. If this positive bias remains uncorrected, it may lead to misleading conclusions of important

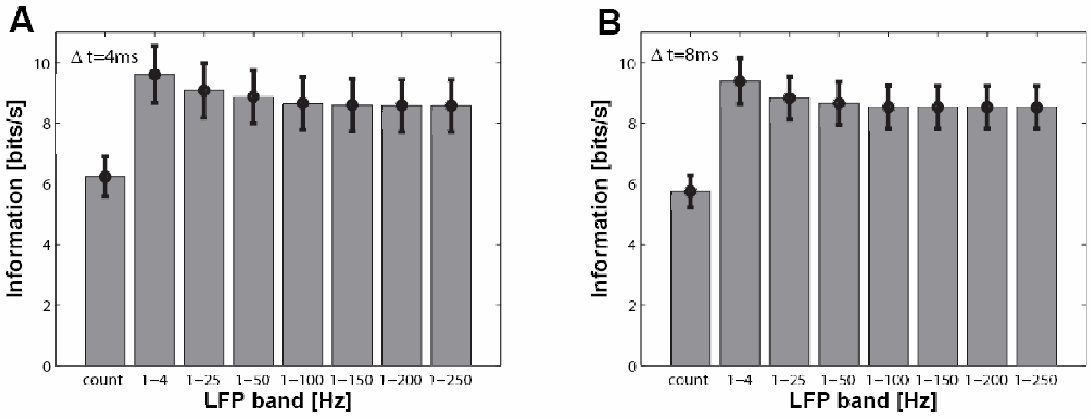


Figure 3.4: Broad band vs narrow band phase-of-firing information. Histograms plot the information rates (mean $\pm$ SEM over the dataset) carried by the spike count (denoted as "count" in the x-axis label) and the LFP phase at which spikes were emitted as function of the considered LFP frequency band (indicated in the x-axis label). **A:** Results obtained using a stimulus window $\Delta t = 4$ ms. **B:** Results obtained using a stimulus window $\Delta t = 8$ ms.

phase-of-firing effects just because of a bias artifact. To correct for this problem, we have implemented a two step procedure (developed following closely ideas presented in Montemurro *et al.* (2007a,b)) that has two advantages: (i) it gives a much tighter estimation of the information than that obtained using Eq. 3.3 (ii) in conditions in which there are not enough data to fully remove the bias, the small residual bias is negative. Consequently, our information estimations typically represent tight lower bounds that are tighter for spike count than for spike timing codes. This makes our conclusion about the extra information in phase code conservative: the actual amount of extra information in the phase code is likely to be slightly higher than the one reported.

The first step in the bias correction is given by a quadratic extrapolation (Strong *et al.* (1998)) of the information values. This method assumes that the estimated information, can be approximated by a second order polynomial in $1/n_s$ (where n_s is the number of repetitions of the movie used to estimate the information), that is $I = I_{\rm raw} + a/n_s + b/n_s^2$, where $I_{\rm raw}$ is the result of the direct application of Eq. 3.3 to the data, and a and b are two parameters that depend on the stimulus-response prob-

abilities. For a particular experiment consisting of N_s repetitions of the movie, we obtained different raw estimates of the information. The first one was the result of applying Eq. 3.3 to the full original dataset, where $n_s = N_s$; a second set of estimates corresponded to random selections of $n_s = N_s/2$ of the trials; and the third corresponded to the estimation of the information from random selections from $n_s = N_s/4$ of the total available trials. Then, these estimates were plotted against $1/n_s$, and a quadratic polynomial was fitted by least squares. The asymptotic value of the information was then taken as the intersection of this polynomial with the vertical axis (which is equivalent to taking the limit $1/n_s \to 0$). This correction was applied both to the spike count code and the phase code information, giving the corresponding values I_{count} and I_{phase}.

The second step consisted in improving this estimate by means of a bootstrap procedure. For the estimation of the information in the spike count code the bootstrap procedure consisted in obtaining an estimate of the information, $I_{\mathrm{count-boots}}$, from a dataset created from the original one by randomly permuting all the spike counts across all time windows for all trials independently. When the number of trials is very large, the value of the information obtained from the transformed dataset must be zero, since all the information that the neural response carried about the stimulus was completely destroyed by shuffling. However, due to finite sampling effects, the quantity $I_{\mathrm{count-boots}}$ will be nonzero. The final corrected estimate of the spike count information $I_{\mathrm{count-corrected}}$ was obtained subtracting out this small residual bias revealed by bootstrap: $\mathrm{I}_{\mathrm{count-corrected}} = I_{\mathrm{count}} - I_{\mathrm{count-boots}}$, where $I_{\mathrm{count-boots}}$ was also corrected with the quadratic extrapolation method.

The bootstrap procedure was also applied to the phase-of-firing information, as follows. We created a shuffled dataset in which, in each individual trial, we left all the positions of the spikes unaltered but paired them with randomly shuffled values of the phases observed in correspondence to a spike. We called $I_{\mathrm{phase-boots}}$ the information estimated from these shuffled data. The information contained in the spike counts was unaffected by this procedure, but all the extra information by the phase of firing was completely erased. However, due to a residual bias $I_{\mathrm{phase-boots}}$ will still be slightly larger than $I_{\mathrm{count-corrected}}$ for a dataset with a limited number of trials resulting from the use of several phase symbols that determine the response outcome in each stimulus time window. Therefore, we computed an estimation for the bias of the phase

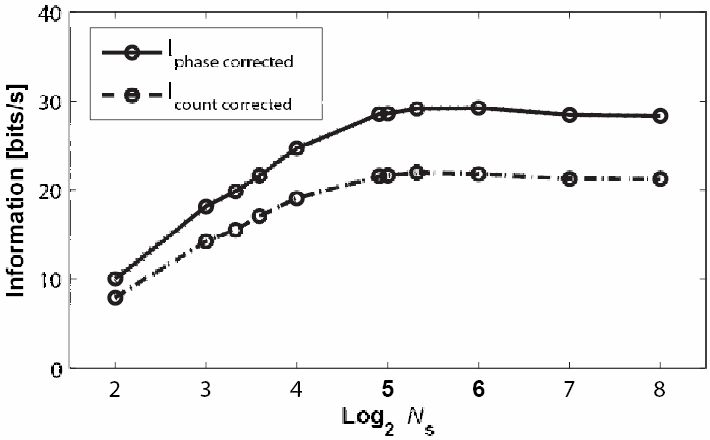


Figure 3.5: Performance of the information bias correction method on simulated data. To assess the bias correction methods we simulated spike trains and phase labeled spikes that preserved the same statistical structure of real experimental datasets. We then applied the information estimation methods described in the text to compute the values of the information (normalized in bits/s) available in the spike count code (lower curve) and the phase code (upper curve) as a function of the logarithm of the number of trials. The overall remaining bias is negative, affecting more the information in the *phase code* than that in the *spike count code*.

code as $\mathrm{bias}_{\mathrm{phase}} = I_{\mathrm{phase-boots}} - I_{\mathrm{count-corrected}}$, where again we previously corrected $I_{\mathrm{phase-boots}}$ using the quadratic extrapolation method. Finally, the corrected value of the information for the phase code was computed as $I_{\mathrm{phase-corrected}} = I_{\mathrm{phase}} - \mathrm{bias}_{\mathrm{phase}}$.

We verified the performance of the bias correction method by means of the following numerical simulation. For a given experiment we measured the time varying firing rate from the available spike trains. From the set of phase-labeled spikes, we also measured the probability of each phase in correspondence to a spike for every time window. Then, a simulated set of spike trials was generated using an inhomogeneous Poisson process with the same spike rate as in the real experiment. In order to create the phase-labeled set of spike trains we labeled each spike according to the real probability of each label for every time window. In this way we could create any number of trials of a simulated set of spikes and labeled spikes that conserved the statistics of a real experimental data set. In generating these data, we binned the phase into 4 bins, exactly as in the phase-of-firing information calculations reported in Fig. 3.10.

Figure 3.5 reports the behaviour of the bias correction methods on simulated data with the statistics of one of the available experimental sessions (animal A98). For a number of trials less than 30 the overall remaining bias is negative, affecting more the estimation of the phase code information than the spike count code information. This means that when the number of trials is less than 30, our estimation of the advantage of the phase code over the spike count code is very conservative. (The mathematical reasons why the bias is mostly negative are investigated in detail in Montemurro *et al.* (2007b); Panzeri *et al.* (2007) and originate from the fact that bootstrap corrections tend to overestimate the bias). When the number of trials is larger or equal than 30, any relative residual error in the corrected information estimates is negligible and thus the estimates are in practice unbiased. This suggests that in the data-size range collected experimentally our information estimates are precise, and only potentially affected by a small residual bias which is slightly higher in the phase code information. This leads to conservative estimates of the amount of extra information in the phase code and to conservative conclusions about the importance of the phase of firing in information processing.

3.2.2.6 Information at fixed spike rate and independence of spike count and phase coding information

An interesting and still highly debated Harris (2005); Harris *et al.* (2002); Mehta *et al.* (2002) question is whether phase and spike rate are just being produced by the same mechanism and thus reflect largely redundant information, or if instead the phase-of-firing information contributes information about the stimuli which is genuinely novel with respect to that carried by spike counts. The finding (reported in Fig. 3.10 A and Fig. 3.4) that the information in phase-of-firing is higher than the information in spike counts is not enough to address this question. In fact, suppose we are labeling the spikes with an informative phase signal ϕ, but the value of the phase ϕ in response to the stimulus is determined deterministically by the value of the mean spike count $\bar{n}$ to that stimulus, i.e. $\phi = \phi(\bar{n})$. In this case, spike count and phase-of-firing cannot convey information about independent or different aspects of the stimulus. However, even in this situation the information in phase of firing may still be bigger than the information in spike counts if the phase reports the stimulus more reliably than the

spike rate itself. Thus, it is important to develop measures that determine whether the information conveyed by phase-of-firing and counts is to some extent novel and unrelated.

One way to demonstrate that phase of firing encodes features of the stimulus not encoded by spike counts is to show that the phase of firing enables to discern between stimuli which cannot be possibly distinguished on the basis of spike counts alone. Stimuli that cannot be distinguished on the basis of firing rate alone are stimuli with the same spike count distribution in response to the stimuli. In the short-stimulus windows regime considered here, this amounts to the stimuli having the same mean spike count. Therefore, for each recording channel, we took only movie parts eliciting exactly a certain spike rate $\bar{n}$, and we computed what we called "the phase-of-firing information at fixed spike rate $\bar{n}$", i.e. the information between the phase of firing and the stimuli at fixed spike rate $\bar{n}$, as follows:

$$I_{\bar{n}-\text{phase}}(S;R_\Phi) = \sum_{s'} \frac{1}{S_{\bar{n}}} \sum_{\phi=0}^{\Phi} P(\phi|s') \log_2 \frac{P(\phi|s')}{P(\phi)} \tag{3.6}$$

where in Eq. 3.6 the sum over s' is restricted to stimuli with a spike count exactly equal to $\bar{n}$ (whose total number is $S_{\bar{n}}$). The spike count code conveys no information about these fixed-rate movie parts. If the phase of firing is totally redundant to the spike rate, then it will too convey zero information about fixed-rate movie parts. If instead the phase of firing conveys significant information about the fixed-rate movie parts, then the phase code must contains novel and independent information to that provided by spike counts, because stimuli undistinguishable from spike count alone would become distinguishable from the phase of firing. Thus, a useful test that some of the information convey by the phase of firing is genuinely novel with respect to that carried by spike counts is to test that $I_{\bar{n}-\text{phase}}(S;R_\Phi) > 0$ for some range of spike rate $\bar{n}$. Results for the 1-4 Hz LFP phase were reported in Fig. 3.10 B of the main text. As reported in the main text, we found that $I_{\bar{n}-\text{phase}}(S;R_\Phi)$ was significantly positive for all firing rates > 1 Hz, and it was as high at 30 bits/s at spike rates > 80 spk/s. This demonstrates that the phase of firing conveys some novel information with respect to spike rates, and suggests that this information is particularly useful to disambiguate stimuli eliciting sustained spike responses with equal strength.

It is intuitive that more information is to be gained by labeling the spikes with a phase which is not redundant to the spike rate than when labeling spikes with a phase totally redundant to the spike rate. To quantify how much additional phase-of-firing information about all stimuli we gained because the phase labeling is not a redundant function of the rate, we introduced a further measure that we called the "phase-of-firing information in the redundant count-phase (RCP) case" $I_{\mathrm{RCP-phase}}(S;R_\Phi)$, which is the information that would be obtained if we kept the same overall distribution of observed phase values but we destroyed any possible way for the phase to carry some information about the stimulus which is independent from that of rates. This phase distribution redundant with the spike rate can be constructed by forcing the phase distribution to become a function of just the spike rate and not of the individual stimuli at fixed rate. In practice, the redundant phase distribution can be built by randomly associating, within each trial, the observed phase responses to a randomly selected stimulus with the same mean spike rate rather than to the stimulus that elicited the considered phase response. In this way, the phase distribution to any stimulus becomes only a function of the mean spike rate to that stimulus, i.e. $P(\phi|s) = P\left(\phi|\bar{n}(s)\right)$, and $I_{\mathrm{RCP-phase}}(S;R_\Phi)$ can be expressed as:

$$I_{\mathrm{RCP-phase}}(S;R_\Phi) = \sum_{\bar{n}} \frac{1}{S_{\bar{n}}} \sum_{\phi=0}^{\Phi} P(\phi|\bar{n}) \log_2 \frac{P(\phi|\bar{n})}{P(\phi)} \tag{3.7}$$

where the sum over stimuli in the information definition was replaced by a sum over all possible spike rate values $\bar{n}$. $P(\phi|\bar{n})$ is the probability of observing phase ϕ in any window eliciting a mean spike rate $\bar{n}$, and is constructed by pooling together all phase values obtained with the different stimuli at fixed rate. If $I_{\mathrm{RCP-phase}}$ (Eq. 3.7) is considerably smaller than I_{phase} (Eq. 3.5), it means that a substantial part of phase-of-firing information about all stimuli is due to the fact that the phase labeling is not a redundant copy of the rate.

We computed $I_{\mathrm{RCP-phase}}$ for our dataset considering the 1-4 Hz LFP phase and we found that $I_{\mathrm{RCP-phase}}$ was 40% smaller than I_{phase}, suggesting that a substantial part of the extra amount of information gained by the knowledge of the phase of firing reflects novel information about the stimulus with respect to that provided by spike counts.

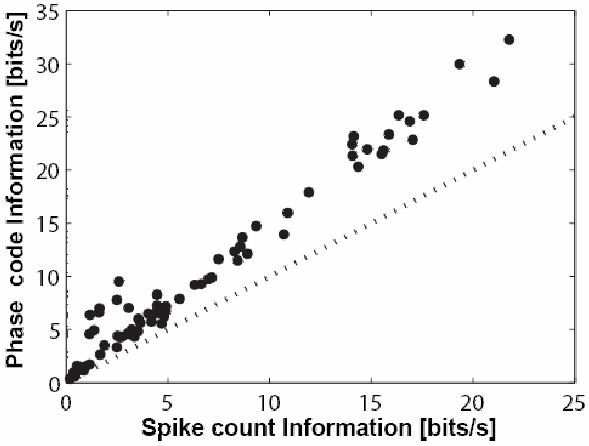


Figure 3.6: Distribution of phase-of-firing information. Population scatterplot of the phase code information of the 1-4 Hz LFP vs the spike count information. The dashed line plots the line in which the two informations would be equal.

3.2.2.7 Population average vs population distribution of the information in the phase of firing

Figure 3.10 A of the main text shows that, on average over the population, the information in the 1-4 Hz LFP phase of firing was 55% higher than the information in spike counts. How typical is phase coding across the whole population of sampled neurons? Figure 3.6 is a plot for all neurons of the information in spike counts versus the information in the 1-4Hz LFP phase of firing. All channels but one had significantly higher information (bootstrap test, $p < 0.01$) in the phase code than in the spike count code, and most channels presented a large amount of extra information in the phase of firing. This suggests that phase coding is common across the entire sampled population, and not restricted to a few highly informative neurons.

3.2.2.8 Information obtained when labeling spikes with LFP energy-of-firing as well as phase

In this study we decided to consider whether the LFP phase at which spikes were emitted is informative. The choice of considering the spike timing relative to the LFP phase and not other characteristics of the LFP was somehow arbitrary and motivated by the

theoretical and experimental literature that previously concentrated on phase coding and its advantages. However, since the LFP is related to some degree to the synaptic input to a cortical area and thus reflects the "context" in which spikes were generated, it is plausible that spikes are informative also with respect to others features of the LFP. To investigate this issue, we computed, from the band-passed LFPs, also the instantaneous amplitude ϵ of the LFP fluctuation by taking the absolute value of the Hilbert transform in each time point. We then computed, as before, the information that is obtained by labeling the spikes with the instantaneous amplitude ϵ. We found, although there was more information in the LFP amplitude at the time of spiking than in spike counts, that the information in the LFP amplitude at the time of firing was always less than the corresponding information in the phase of firing. Results for the 1-4 Hz LFP band are reported in Fig. 3.7. An interesting question is whether the information in the amplitude at the time of of firing is independent or redundant to that carried by the phase of firing. To address this issue, we computed the information that is obtained by labeling the spikes simultaneously with both the LFP phase and the amplitude at the time of firing. Figure 3.7 shows that the information in the "amplitude and phase" at the time of firing was equal to the information in the phase of firing. This means that the LFP amplitude at the time of firing, though informative in itself, conveys information which is redundant to the information carried by the phase of firing. Consistently, we found that LFP amplitude and phase tended to be correlated in the presence of a spike (Fig. 3.7 B). Interestingly, amplitude and phase were only correlated during firing, and they were not correlated in the absence of a spike (Fig. 3.7 C). Thus, it appears that it is natural and sensible to refer spike times to LFP phases rather than to other LFP variables.

3.2.2.9 Information in the phase of firing of isolated neurons

As reported above, the spikes used for the analysis represented the spiking activity of a small population of cells rather than well separated spikes from a single neuron. For a fraction of channels, it was possible to separate some single units. As reported above, we could isolated 71 units from 37 channels. We checked whether the phase-of-firing conveyed information also when considering the isolated units. The results were as follows. For $\Delta t = 4$ ms, the 1-4Hz LFP phase of firing information about the movie

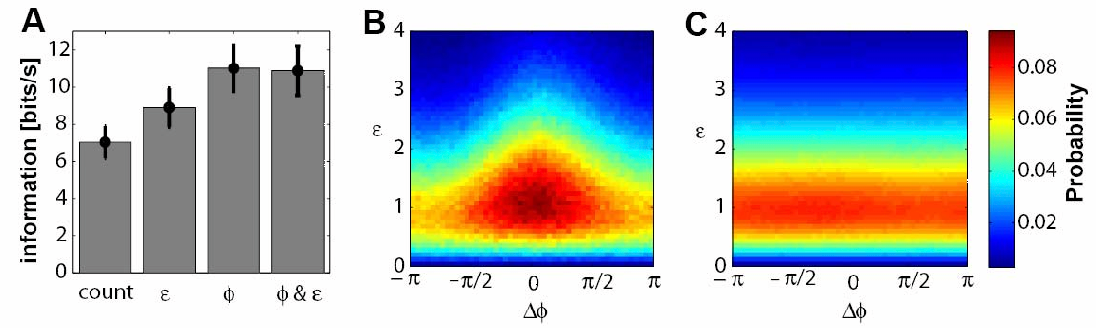


Figure 3.7: Labeling spikes with different LFP features. **A:** Histograms plot the information (mean $\pm$ SEM over the dataset of channels for which ≥ 30 movie repetitions could be collected) carried by the spike count (denoted as "count" in the x-axis label); by the LFP amplitude at which spikes were emitted (denoted by "ϵ" in the x-axis label); by the LFP phase at which spikes were emitted ("ϕ" in the x-axis label); and by the simultaneous phase and amplitude at which spikes were emitted ("ϕ & ϵ" in the x-axis label). **B:** Joint probability density (per unit angle and per unit ϵ) of phase and amplitude in correspondence of a spike. Results are expressed in terms of the angular distance $\Delta\Phi$ from the neuron's preferred phase in the corresponding window and of the amplitude ϵ. Results are averaged over all channels and time windows containing a spike. **C:** Joint probability density (per unit angle and per unit epsilon) of phase and amplitude in absence of a spike. Results averaged over all channels and time windows where no spike was observed. In plotting Panels B and C, the amplitude was normalized to its average standard deviation in a trial.

was 53% bigger than spike counts for the 37 multiunit channels and 46% bigger than spike counts for the 71 well sorted units. For $\Delta t = 8$ ms, the 1-4Hz LFP phase of firing information about the movie was 62% bigger than spike counts for the 37 multiunit channels and 54% bigger than spike counts for the 71 well sorted units. For $\Delta t = 16$ ms, the 1-4Hz LFP phase of firing information about the movie was 75% bigger than spike counts for the 37 multiunit channels and 62% bigger than spike counts for the 71 well sorted units. The small decrease of information from multi-unit to single unit is that the residual error in correcting for the information sampling bias (which is negative, see Fig. 3.5) has approximately the same absolute size for the multiunit spikes and the well sorted spikes (Panzeri *et al.* (2007)), but the well sorted spikes have overall less information per channel because information is approximately proportional to the mean firing rate for short stimulus windows (Panzeri & Schultz (2001)). Overall, the conclusion is that phase-of-firing is informative also when considering well isolated spikes.

3.2.2.10 Robustness of the information in the phase of firing with respect to approximate knowledge of LFP fluctuations

Can the phase-of-firing information still be accessed by decoders affected by limitations similar to that likely encountered by downstream neural networks? A neural decoder may not be able to finely filter the broad-band circuit oscillations. We verified (Fig. 3.4) that a substantial amount of phase-of-firing information was preserved even if the neural signal was filtered within very broad frequency bands containing the informative low-frequency components.

3.3 Results

The most established hypothesis on how sensory information is represented in the brain is the "spike count coding" hypothesis (Adrian (1928)), which suggests that neurons represent information by the number of spikes discharged over some relevant time window. However, the timing of spikes may add important information to that already carried by spike counts (de Ruyter van Steveninck *et al.* (1997); Gollisch & Herz (2005); Hopfield (1995); MacKay & McCulloch (1952); Optican & Richmond

(1987); Schnupp *et al.* (2006); Victor & Purpura (1996)). In particular, information may be encoded in the spike times relative to the phase of some concurrent network oscillation (Bullock (1993); Buzsaki & Draguhn (304); Fries *et al.* (2007); Hopfield (1995); Lisman (2005)): this is called "phase coding". In this Chapter, we investigated which of these coding strategies is used by the primary visual cortex (V1) during the presentation of rich natural-like visual stimuli.

3.3.1 Spike and LFP responses to natural movies

We recorded neural signals with an array of extracellular electrodes from primary visual cortex of 4 anaesthetized macaques in response to a binocularly presented 3.5-6 min long natural color movie. Each recording site corresponded to a well defined V1 visual receptive field within the field of movie projection. From each electrode, we measured both spiking activity and Local Field Potentials (LFPs). Spikes were detected by threshold-crossing of the 500-3500Hz band-passed neural signal, and represented the spiking activity of a small population of cells rather than well separated spikes from a single neuron. LFPs, which were recorded as the 1-250 Hz band-passed neural signal, reflect the fluctuations in the input and the intracortical processing of the local cortical network, including the overall effect of population synaptic potentials (Juergens *et al.* (1999); Mitzdorf (1987)) and other types of slow activity such as spike afterpotentials and voltage-dependent membrane oscillations (Buzsaki (2002); Harada & Takahashi (1983); Kamondi *et al.* (1998); Logothetis (2003)). Each movie was repeated 12-40 times in order to sample the probability distribution over the neural responses to each part of the film. We obtained a dataset of 78 V1 electrode channels recorded during stimulation with a particular movie. Unless otherwise stated, we focused on the relationship between LFPs and spikes recorded from the same electrode.

Figure 3.8 reports responses recorded from one example recording site (electrode 2 in monkey A98) over repeated presentations of the same movie fragment. The presentation of the movie elicited patterns of spikes that were modulated over time in a reliable manner across trials (Fig. 3.8 D). As a consequence, and as summarized by the trial-averaged instantaneous spike rate (Fig. 3.8 F), some movie scenes elicited a high and reliable firing rate, and some other scenes elicited instead a low (but still reliable)

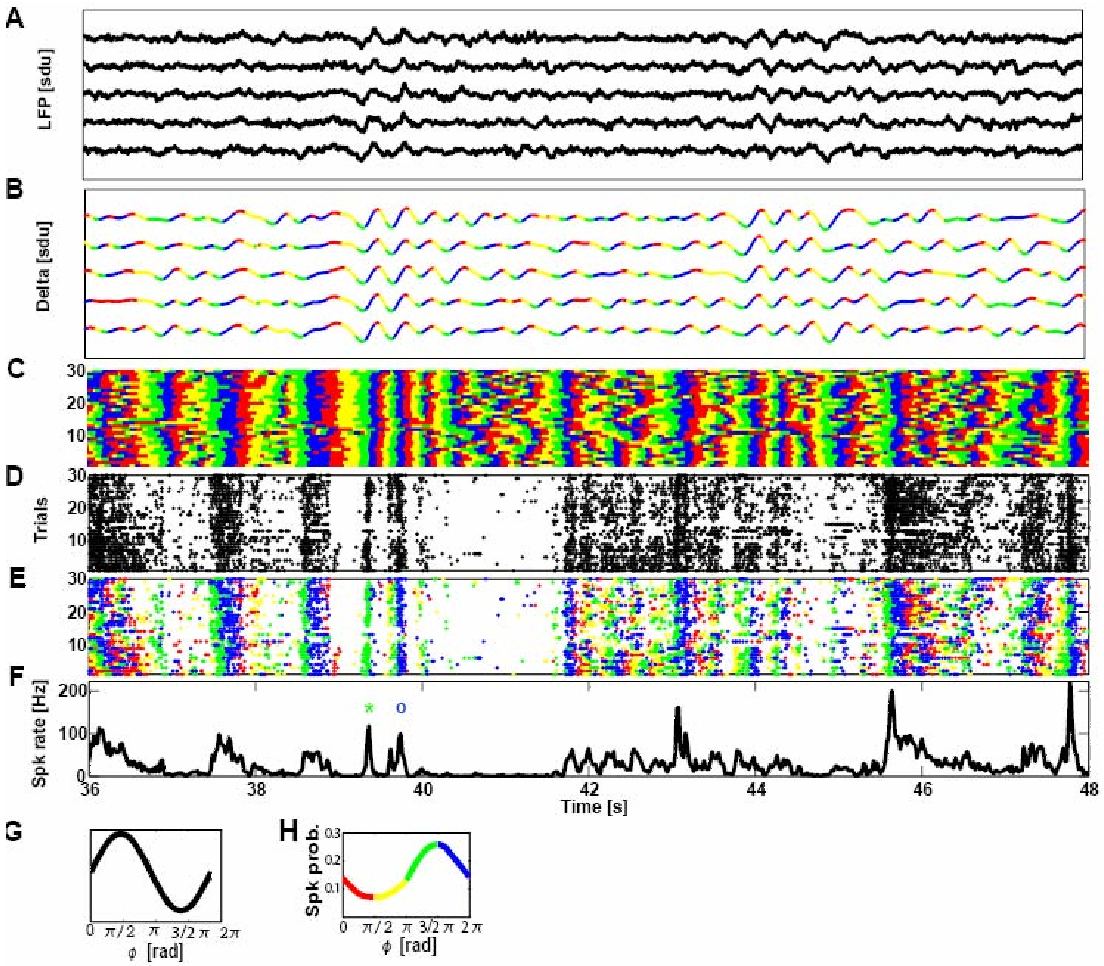


Figure 3.8: Illustration of the time course of the LFP and the spike phase, and of the difference between the spike count and phase code. These data were recorded from electrode 2 in monkey A98 in response to a movie. **A:** LFP traces from five presentations of a 12s long movie extract. Traces were displaced on the vertical axis to make them distinguishable. **B:** Time courses of the 1-4 Hz band-passed LFP to the same five presentations of a 12s long movie extract as in Panel A. Traces were displaced on the vertical axis to make them distinguishable. Different colors correspond to the phase of the 1-4 Hz LFP being in one of four quadrants set as shown in Panel H (on the base of the phase locking properties of spikes). **C:** Time course of the phases 1-4Hz LFP signal phases over 30 repetitions of the movie extract. Phase were color-coded into quadrants as indicated above and illustrated in Panel H. The top 5 trials in Panels C-E correspond to the 5 trials in Panels A-B.

firing rate. Thus counting the number of spikes emitted within some time interval reduces the observers uncertainty about which movie scene was being presented. This illustrates that spike counts encode movie scenes.

As for the spiking activity, the waveform of LFPs in single trials (Fig. 3.8 A) showed fluctuations that were reliable across trials and modulated by the movie. During movie presentation, the power of the LFP spectrum was highest at low frequencies (< 4 Hz) and then dropped with increasing frequency (Fig. 3.1). We thus started by considering the behavior of the phase of LFPs fluctuations in the highest-power band, namely the 1-4 Hz frequency range (delta band). The single-trial 1-4 Hz band-passed LFP traces during movie presentation (Fig. 3.8 B) show that 1-4 Hz LFPs too were reliably modulated by the movie. To extract the instantaneous value of the phase of the LFP fluctuations in each trial and at each time during the movie, we first computed the Hilbert transform of the band-passed traces in each trial, and we then computed the phase as the argument of the Hilbert transform (the resulting phase convention is plotted in Fig. 3.8 G).

To visualize how LFP phases were modulated by the movie, we divided the phase range into four equi-spaced quadrants and labeled each with a different color (Fig. 3.8 G). It was apparent that the 1-4 Hz LFP also encoded the movie, because the phase values were modulated by the movie time and this modulation was extremely reliable across

Figure 3.8 (continued): D: Raster plot of spike times (indicated by dots) resulting from repeated presentation of the selected 12 s movie extract. **E:** Raster plot of the same spike times as in panel E, but with the dots representing the spikes color-coded according to the 1-4 Hz LFP phase quadrant at which they were emitted. These colored spike times illustrate the "phase code" whereas the colorless spike times in Panel D illustrate the spike count code.**F:** Spike rate, averaged over all 30 trials and computed in 4 ms long sliding time bins, during the 12-s movie extract. The green star and the blue circle indicate movie points that elicit similar spike rate responses but different and reliable phase values. These two movie points can be much better discriminated from each other by considering the phase at which spikes were emitted rather than just counting spikes. **G:** The sinusoidal convention used for phase. **H:** The probability (per unit angle) of observing a spike at a given phase value. The curve is plotted with the color-code used to label phase quadrants.

trials at several times during the movie (Fig. 3.8 C).

Moreover, there was a correspondence between spike rate (Fig. 3.8 F) and across-trials reliability of phase (Fig. 3.8 C): scenes eliciting a high spike rate also elicited a highly repeatable phase. The distribution of 1-4 Hz LFP phases at spike times from this example channel during the presentation of the movie (Fig. 3.8 F) was non-uniform (Rayleigh test of non-uniform angular distribution (Fisher (1993)), $p < 0.01$): the probability of spiking at the preferred phase ($3\pi/2$ rad) was twice the probability of spiking at the anti-preferred phase ($\pi/2$ rad). The phase values around the preferred one for firing (coded as green and blue in Fig. 3.8) appeared to be the ones encoding the movie more reliably during periods of firing (Fig. 3.8 C).

Since both spikes and LFP phases carry information about the movie, and since phases are particular reliable in the presence of spikes, it is possible that, in addition to encoding information by spike count, neurons may encode extra information about the movie by the phase at which they fire. This hypothesis means that if we label the spikes with a "color" (as in Fig. 3.8 E) reporting the phase quadrant at which they were emitted ("the phase-of-firing code"), we can predict better which visual feature elicited the firing than if we just count the "colorless" spikes expressing the spike count code (Fig. 3.8 D). Figure 3.8 F illustrates how phase may play a role in representing information about the visual scenes. Two scenes of the movie eliciting comparable firing rates (e.g. those occurring at times marked respectively by a green star and blue circle in Fig. 3.8 F) could not be discriminated by their "colorless" spike count (Fig. 3.8 D), but could be discriminated when taking into account their phase label (green vs. blue colored spikes in Fig. 3.8 E). The exact extent to which knowledge of the phase of firing helps stimulus discrimination will be determined below using information theory.

3.3.2 LFP phase reliability and spike-phase relationships

Having illustrated the phase-of-firing coding with an example recording channel and a selected LFP frequency range, we next characterized the behavior of the entire datasets over a wide range of LFP frequencies. To do so, we divided the LFP frequency range into small frequency intervals (1-4 Hz; 4-8 Hz; and up to 124 Hz in 4-Hz-wide non-overlapping intervals). We then computed band-passed LFPs in each such frequency

interval, from which we extracted the instantaneous band-passed phase during the course of the movie.

A first fundamental condition for phase coding is that LFP phases are reliable and repeatable across different trials at fixed sensory input (i.e. at fixed time during the movie). We investigated phase reliability by computing the circular variance across trials of the phase at each time during the movie. Circular variance is a measure of angular dispersion (Eq. 3.2), and its values range from 0 (perfect reproducibility across trials) to 1 (total unreliability). To be useful for phase-of-firing stimulus coding, the reliable phase values must be observed during periods of firing: if the reliable phases occur during silence, their information cannot be used to tag the spikes. Thus, in Fig. 3.9 A we examined how the circular variance of the LFP phase depends upon the spike rate observed in the same window. Fig. 3.9 A shows that, in the high-frequency LFP regions (> 50 Hz), the circular variance is very high (population average: approximately 0.85) across all spike rate levels. Thus, the phase in the high-LFP frequency range is not reliable enough to support stimulus discrimination. In contrast, and for all spike rate values, phase was far more reliable in the low LFP frequency bands. In the LFP frequency regions below 12 Hz, movie segments eliciting high firing rates also elicited substantially more reliable LFP phases: on average across the dataset, the circular variance across trials of the 0-4 Hz LFP phase was approximately 0.5 in low spike rate windows (< 25 spk/s) and approximately 0.3 in high (> 150 spk/s) spike rate windows. A circular variance of 0.3 would be that achieved by a phase distributed uniformly over a 0.65π-wide interval. This suggests that, in order to extract information from the phase of firing, we need to measure phases with a precision of approximately $\pi/2$ (i.e. a quarter of a phase cycle). A very similar dependence of phase reliability on rate was obtained with 4-8 Hz and 8-12 Hz phases (data not shown). The increase of phase reliability with firing strongly suggests that the role of phase-related spike times may be to discriminate between stimuli each eliciting similarly high spike rates.

A second crucial requirement for phase coding is that the different movie time windows elicit diverse reliable phase responses. For example, in Fig. 3.8 we observed that several movie scenes elicited a reliable phase in the green-coded range, and other different movie scenes elicited a reliable phase in the blue-coded range; it is the color difference in the reliably colored phase ranges that makes it possible to use phase to tag successfully different movie parts. We found that the results obtained for the example

channel plotted in Fig. 3.8 extended to the whole dataset. In particular, for the vast majority of all channels, the distribution of phases at spike times was significantly non-uniform (Fig. 3.2 A), and we could define a preferred phase of spike times for each channel. The preferred phase of the 1-4 Hz LFP varied from channel to channel, and was located in most cases in the $3\pi/2$ to 2π range (between the trough and the middle of the rising phase of the LFP oscillation). For the 1-4 Hz LFP, the average over the dataset of the preferred phase was 1.9π (Fig. 3.2 C). We qualitatively observed that for the vast majority of channels phase values close to the preferred phase were more reliably during firing, much as for the example channel reported in Fig. 3.8. Thus, to quantify which phase range is more reliable during firing at the population level, we realigned the 1-4 Hz LFP phase values for each channel so that the preferred phase was set to a "zero" reference value, and we expressed each phase value as $\Delta\Phi$, the difference between the phase value and the channel's preferred phase. We then examined how the circular variance of the phase depended upon the mean across trials of the phase (computed as "angular mean"; see Eq. 3.1) and upon the trial-averaged spike rate observed in the same window. Fig. 3.9 B reports the population average of the phase circular variance as a function of the value of the trial-averaged phase (expressed as $\Delta\Phi$) and the spike rate. We found that (Fig. 3.9 B) at high spike rates (> 100 spk/s) all phase values within angular distance $\Delta\Phi$ of $\pi/2$ from each channel's preferred phase were reliable (with circular variation in the range 0.3-0.4). This makes it possible to disambiguate different stimuli eliciting an equally strong firing rate by tagging them using some of the different reliable phase values occurring within angular distance $\Delta\Phi$ of $\pi/2$ from the preferred phase. In this way, stimuli indistinguishable from firing rates alone may become distinguishable after being labeled by their phase of firing, and this permits the phase of firing to convey information about the movie which is genuinely novel with respect to the one provided by spike rates.

3.3.3 The sensory information conveyed by spikes times relative to LFP phase

Finally, we used Shannon's information (Shannon (1948)) to address directly the issue of how much additional information, beyond that available in spike counts, is conveyed by phase of firing. Shannon's information (see Eq. 3.3) between a set of stimuli and

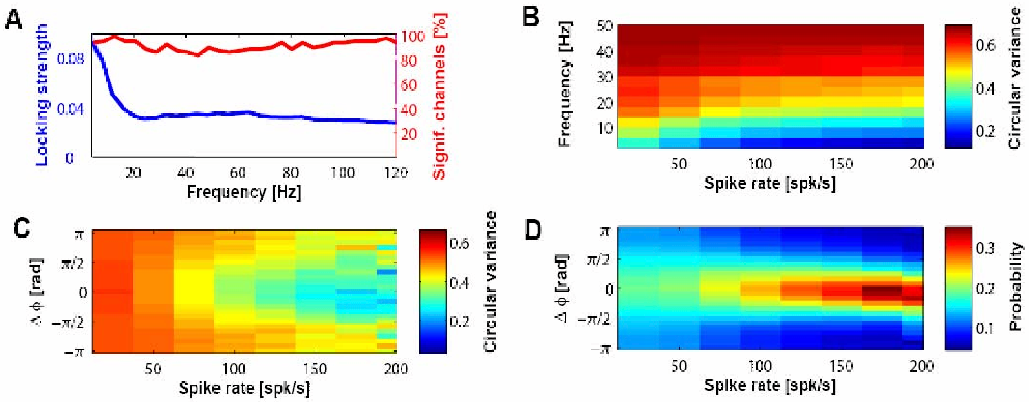


Figure 3.9: Phase locking properties of V1 neurons under natural visual stimulation. **A:** The blue curve plots the locking strength as a function of the frequency, averaged over the dataset (solid blue line: mean; light-gray area: $\pm$ 1 SEM confidence region). The locking strength was computed for each channel as the difference between maximal and minimal probability of spike locking across all channels. The red curve plots the number of channels showing significant spike locking (Rayleigh test, $p < 0.01$) at fixed frequency. **B:** The value of the circular variance across trials at fixed time windows of the phase of LFPs band-passed in a given frequency range is plotted as function of spike rate in the corresponding window. Results averaged over all channels and time windows. **C:** The value of the circular variance across trials of the phase of 1-4Hz LFPs at fixed time windows is plotted as function of the spike rate and of the angular distance $\Delta\phi$ from the neuron's preferred phase in the corresponding window. Results averaged over all channels and time windows. **D:** Probability of observing a preferred phase of 1-4Hz LFPs at a given angular distance $\Delta\phi$ from the neuron's preferred phase as function of the spike rate of the neuron. Results were averaged over the whole dataset.

the neural responses is a principled measure of single-trial discriminability. It quantifies (in units of bits) the reduction of uncertainty about the stimulus which is gained by a single-trial observation of a neural response: one bit corresponds to a reduction of uncertainty of a factor of two. In the following we computed the information that the neural response conveys about which section of the movie was being presented. This characterization of information about the movie does not need any assumption about which features in the movie made the neuron respond (de Ruyter van Steveninck *et al.* (1997)): thus we are computing information about all possible visual attributes occurring in the movie. Information values were expressed in bits/s by dividing the information value by the time window length (typically 4 ms wide; see Methods Section 3.2) in which the neural response is computed.

The amount of information transmitted by a neural response depends on the way the response is quantified, which in turn reflects our assumption on what is the "neural code". We considered and compared the information carried by the two candidate neural codes: the spike count code (the neural response to a part of the movie was quantified as the total number of spikes emitted in the selected response time window), and the phase-of-firing code (the neural response is quantified as the LFP phase at which each spike was emitted). The latter was equivalent to computing information from the "colored" spikes as in Fig. 3.8 E; the former to computing information counting "colorless" spikes as in Fig. 3.8 D. We registered the phase of firing with a $\pi/2$ precision (i.e. phase divided into quadrants) because (consistently with the above phase-reliability analysis) using a precision finer than $\pi/2$ did not increase the information further (Fig. 3.3).

We first investigated how well spike counts encode the movie. We found that, across the entire dataset, spike counts conveyed 6.23 ± 0.66 bits/s of information about the movie (mean $\pm$ SEM). We then considered the information about the movie that is carried via the phase of firing. We found (Fig. 3.10 A) that there was considerably more information in the spike times relative to phase than in the spike counts, and that the amount of phase-of-firing information strongly depended on the considered LFP frequency range. The 1-4Hz LFP phase of firing carried 9.6 ± 0.94 bits/s of information about the movie: thus the phase code in 1-4Hz LFP band conveys 54% extra information that it is not possible to obtain in any way from spike counts. As

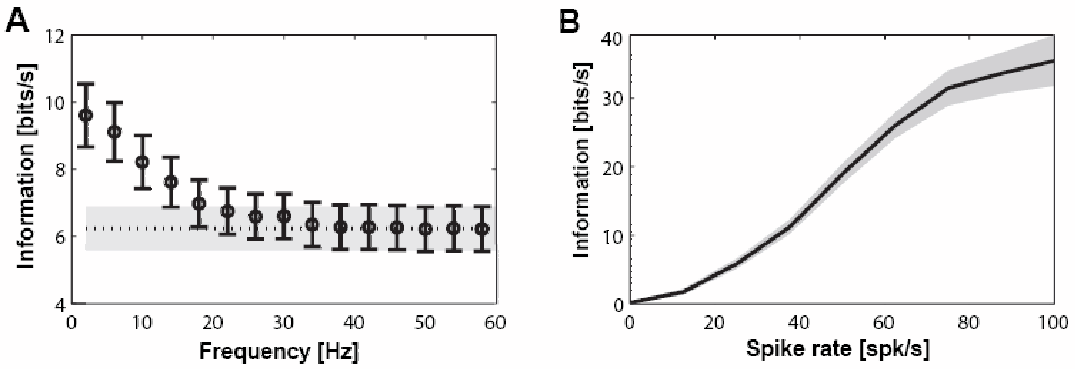


Figure 3.10: The information about the movie conveyed by spike counts and by phase of firing. **A:** Black dots plot the information carried by the LFP phase at which spikes were emitted as function of the considered LFP frequency (mean $\pm$ SEM over the dataset). The black dashed line plots the mean over the dataset of the spike count information (SEM over dataset indicated as grey area) **B:** The phase code (1-4 Hz LFP) information in movie points eliciting exactly the same spike rate is plotted as function of the firing rate (mean $\pm$ SEM across population reported as full line and shaded area respectively). For the purpose of plotting the line in Panel B, the information values computed between stimuli at exactly the same rate were averaged over 10 spk/s wide spike rate bins.

reported in Methods (Section 3.2.2.9), the extra amount of information in the phase-of-firing did not change much when considering only spikes from well-isolated single neurons. The phases of firing in the 4-8Hz and 8-12 Hz LFPs were less informative than the 1-4 Hz LFP phase, but still carried much more information than spike counts (46% and 32% respectively). The amount of information in the phase of firing then rapidly decreased at higher LFP frequencies, and it became equal to the spike count information for LFP frequencies > 24 Hz. Further, additionally labeling the spikes with LFP amplitude instead than just with LFP phase did not lead to any increase of information (Fig. 3.7). Taken together, these results suggest that spike times are only informative with respect to the phase of low-frequency LFPs.

Can the phase-of-firing information still be accessed by decoders affected by limitations similar to that likely encountered by downstream neural networks? On the one hand, a neural decoder may not be able to finely filter the broad-band circuit fluctua-

tions.

To address this concern, we computed the phase-of-firing information obtained when the phase is extracted from LFPs filtered between 1 Hz and a low-pass frequency cutoff that was progressively varied between 4 and 250 Hz. We found that the amount of extra information in LFPs decreased from 54% for 1-4 Hz LFPs to 46%, 39% and 38% for 1-25 Hz, 1-100 Hz and 1-250 Hz LFPs, respectively (Fig. 3.4). Thus, a substantial amount of phase-of-firing information was preserved even if the neural signal was filtered within very broad frequency bands containing the informative low-frequency components. On the other hand, although cortical low-frequency oscillations have high spatial coherence, a further-away decoder may not receive a perfect copy of the phase signal. To verify if this leads to a loss of decoded information, we paired the spikes with 1-4 Hz LFPs recorded simultaneously from electrodes up to 4 mm away, rather than pairing them with the same-electrode 1-4Hz LFP. We found that there was less than 1% loss of information (not significant $p > 0.2$; bootstrap test) when pairing spikes with LFP phases at other electrodes, at all inter-electrode distance considered ($\leq$ 2 mm; $\leq$ 3 mm; $\leq$ 4 mm). Thus, the phase-of-firing information was robust to limitations of the downstream decoder.

Is some of phase-of-firing information genuinely novel with respect to that carried by spike counts? In the following we demonstrate that this is the case, by showing that the phase of firing enables to discern between stimuli which cannot be possibly distinguished on the basis of spike counts alone. For each recording channel, we took only movie parts eliciting exactly a certain spike rate. The spike count code conveys no information about these fixed-rate movie parts. If the phase of firing is totally redundant to the spike rate, it will also convey zero information about fixed-rate movie parts. If instead the phase of firing conveys significant information about the fixed-rate movie parts, then the phase code must contain novel and independent information to that provided by spike counts, because stimuli undistinguishable from spike count alone become now distinguishable from the phase of firing. Results for the 1-4 Hz LFP (reported in Fig. 3.10 B as population average) show that the phase code information about movie parts at fixed rate was small at low spike rates and steeply increased at higher spike rates: it reached a value of 30 bits/s at rates > 100 Hz. The phase-of-firing information about movie parts at fixed rate was significantly positive ($p < 0.0001$; bootstrap test) for all rates > 10 Hz. Similar results (although with an overall scaling

down of the information value) were obtained for 4-8 Hz and 8-16 Hz LFPs. This proves our hypothesis: the phase of firing conveys information about the movie which is not redundant to that of spike rates, as it disambiguates stimuli which elicit equally high spike rates and thus cannot possibly be distinguished from spike counts alone. (See Methods Section 3.2 for further studies on the independence between spike count and phase-of-firing information)

3.4 Discussion

The hypothesis that neurons may encode information by the phase at which they fire has received renewed attention in recent years. Evidence has been reported that spatial-navigation- and memory-related structures encode some information by phase-of-firing (Huxter *et al.* (2003); Jensen & Lisman (2000); Lee *et al.* (2005)). However, the extent to which firing rate and phase encode genuinely different information, rather than just being produced by the same mechanism and thus reflecting largely redundant information, has remained debated (Harris (2005); Harris *et al.* (2002); Mehta *et al.* (2002)). Further, it has been unclear whether phase coding represents a fundamental currency for cortical information exchange right from the primary sensory representation, and if it is a robust enough coding mechanism to represent complex stimuli.

Here we have addressed some of these open questions about the nature of phase coding by demonstrating for the first time that, in primary visual cortex of anaesthetized monkeys, a substantial amount of information about natural stimuli is carried by the phase of firing, and that some of this phase-of-firing information is genuinely different from the information carried by spike counts. In fact, phase-of-firing permits the discrimination of stimulus features that elicit an equally high spike rate and thus cannot be distinguished when considering firing rate alone. This coding mechanism provides neurons in sensory cortex with a mean to represent more than one effective stimulus by "tagging" several effective stimuli with similar firing rates with different values of the phase of the network fluctuations.

We found that only low-frequency (< 12 Hz) LFP phases were reliable enough during periods of firing to be useful for coding, whereas > 40 Hz gamma-range phases (Fries *et al.* (2007)) were not. Gamma oscillations were present in our data and are

stimulus-driven (the highest increase of LFP power from spontaneous activity to movie stimulation was in the gamma range; Fig. 3.1), but they did not generate a reliable and informative spike-phase relationship. An interesting question is how the information-rich phase of low-frequency LFPs is generated. On the one hand, it could reflect the intrinsic dynamics of the sensory or neuromodulatory processing pathways commonly associated to the generation of these low-frequency oscillations. On the other hand it could be driven by the dynamics of natural movies, which contains the highest power in the low frequency range (see Section 3.2). To clarify the origin of the informative phase signal, an important direction for future research is to change the stimulus dynamics by using faster stimuli than natural movies and study how this affects the phase-informative LFP frequency range, as well the accompanying spiking precision (Butts *et al.* (2007)).

Previous reports have documented that the timing of individual spikes with respect to the stimulus onset (such as "latency codes" (Panzeri *et al.* (2001))) is very informative. One objection to such individual-spike-timing codes is that their information may not be relevant because the brain does not have a separate representation of stimulus onset which could allow the interpretation of this spike timing code. Our results demonstrate that the timing of individual spikes is not only informative relative to stimulus onset, but is also informative relative to slow fluctuations in the input and the intracortical processing of the local cortical network. Since the latter signal is presumably available to a downstream area decoding the stimulus attributes, our demonstration of phase-of-firing coding suggest that the visual cortex can access and use the information available in the timing of individual spikes. In this respect, one particular advantage of using low frequency oscillations for phase coding is that such low frequency oscillations are those with greater spatial coherence, and can thus be made more widely available to decoding networks.

The type of biophysical mechanisms needed to decode the phase-of-firing information depends on the origin of the informative phase signal. A simple scenario is that low-frequency LFP fluctuations reflect coherent membrane potential oscillations of populations of neurons (such as transition between up and down states). In this case, different phase-of-firing values may be decoded on the basis of their different post-synaptic responses. A more complex scenario is that the reliable LFP phase is the reflection of a very precise interaction of large cell assembly (Harris (2005)). In such

case, a precisely wired circuit may be needed to detect assembly activation, but the information advantage offered by relative time of firing can be even more quantitatively prominent.

Chapter 4

A kernel approach to comparing distributions

"Den rechten Handschuh könnte man an die linke Hand ziehen, wenn man ihn im vierdimensionalen Raum umdrehen koennte."

[Ludwig Wittgenstein, 1921]

We describe a technique for comparing distributions without the need for density estimation as an intermediate step. Our approach relies on mapping the distributions into a Reproducing Kernel Hilbert Space. We apply this technique to construct a two-sample test, which is used for determining whether two sets of observations arise from the same distribution. We use this test in attribute matching for databases using the Hungarian marriage method, where it performs strongly. We also demonstrate excellent performance when comparing distributions over graphs, for which no alternative tests currently exist.

4.1 Introduction

We address the problem of comparing samples from two probability distributions, by proposing a statistical test of the hypothesis that these distributions are different (this is called the two-sample or homogeneity problem).

Here, we propose to test whether distributions p and q are different on the basis of samples drawn from each of them, by finding a smooth function which is large on the points drawn from p, and small (as negative as possible) on the points from q. We use as our test statistic the difference between the mean function values on the two samples; when this is large, the samples are likely from different distributions. We call this statistic the Maximum Mean Discrepancy (MMD).

Clearly the quality of MMD as a statistic depends heavily on the class $\mathcal{F}$ of smooth functions that define it. On one hand, $\mathcal{F}$ must be "rich enough" so that the population MMD vanishes if and only if $p = q$. On the other hand, for the test to be consistent, $\mathcal{F}$ needs to be "restrictive" enough for the empirical estimate of MMD to converge quickly to its expectation as the sample size increases. We shall use the unit balls in universal reproducing kernel Hilbert spaces Steinwart (2002) as our function class, since these will be shown to satisfy both of the foregoing properties. On a more practical note, MMD is cheap to compute: given m points sampled from p and n from q, the cost is $O(m + n)^2$ time.

We develop a non-parametric statistical test for the two-sample problem, based on the asymptotic distribution of an unbiased empirical estimate of the MMD.

We demonstrate the good performance of our test on problems from bioinformatics and attribute matching. In addition, we are able to successfully apply our test to graph data, for which no alternative tests exist.

4.2 The Two-Sample-Problem

Let p and q be distributions defined on a domain $\mathcal{X}$. Given observations $X := \{x_1, \ldots, x_m\}$ and $Y := \{y_1, \ldots, y_n\}$, drawn independently and identically distributed (i.i.d.) from p and q respectively, we wish to test whether $p \neq q$.

To start with, we must determine a criterion that, in the population setting, takes on a unique and distinctive value only when $p = q$. It will be defined based on (Dudley, 2002, Lemma 9.3.2).

Lemma 4.1 *Let $(\mathcal{X}, d)$ be a separable metric space, and let p, q be two Borel probability measures defined on $\mathcal{X}$. Then $p = q$ if and only if $\mathbf{E}_p(f(x)) = \mathbf{E}_q(f(x))$ for all $f \in C(\mathcal{X})$, where $C(\mathcal{X})$ is the space of continuous bounded functions on $\mathcal{X}$.*

Although $C(\mathcal{X})$ in principle allows us to identify $p = q$ uniquely, it is not practical to work with such a rich function class in the finite sample setting. We thus define a more general class of statistic, for as yet unspecified function classes $\mathcal{F}$, to measure the discrepancy between p and q, as proposed in Fortet & Mourier (1953).

Definition 4.2 *Let $\mathcal{F}$ be a class of functions $f : \mathcal{X} \to \mathbb{R}$ and let p, q be defined as above. Then we define the maximum mean discrepancy (MMD) as*

$$\mathrm{MMD}\left[\mathcal{F}, p, q\right] := \sup_{f \in \mathcal{F}} \left(\mathbf{E}_{x \sim p}[f(x)] - \mathbf{E}_{y \sim q}[f(y)]\right). \tag{4.1}$$

We must now identify a function class that is rich enough to uniquely establish whether $p = q$, yet restrictive enough to provide useful finite sample estimates (the latter property will be established in subsequent sections). To this end, we select $\mathcal{F}$ to be the unit ball in a universal RKHS $\mathcal{H}$ Steinwart (2002); we will henceforth use $\mathcal{F}$ only to denote this function class. With the additional restriction that $\mathcal{X}$ be compact, a universal RKHS is dense in $C(\mathcal{X})$ with respect to the L_∞ norm. It is shown in Steinwart (2002) that Gaussian and Laplace kernels are universal.

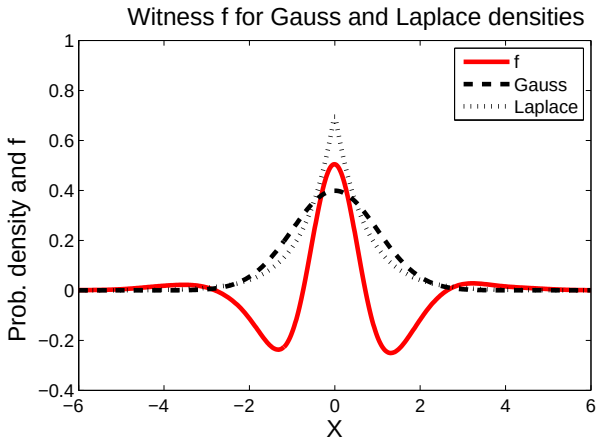


Figure 4.1: Illustration of the function maximizing the mean discrepancy in the case where a Gaussian is being compared with a Laplace distribution. Both distributions have zero mean and unit variance. The function f that witnesses the MMD has been scaled for plotting purposes, and was computed empirically on the basis of 2×10^4 samples, using a Gaussian kernel with $\sigma = 0.5$.

Theorem 4.3 *Let $\mathcal{F}$ be a unit ball in a universal RKHS $\mathcal{H}$, defined on the compact metric space $\mathcal{X}$, with associated kernel $k(\cdot, \cdot)$. Then $\mathrm{MMD}\left[\mathcal{F}, p, q\right] = 0$ if and only if $p = q$.*

See Gretton *et al.* (2007) for more detail. We plot the witness function f from Definition 4.2 in Figure 4.1, when p is Gaussian and q is Laplace, for a Gaussian RKHS kernel.

We next express the MMD in a more easily computable form.

Lemma 4.4 *Given x and x' independent random variables with distribution p, and y and y' independent random variables with distribution q, the population MMD^2 is*

$$\begin{aligned}\mathrm{MMD}^2\left[\mathcal{F}, p, q\right] = \mathbf{E}_{x,x' \sim p}\left[k(x, x')\right] \\ - 2\mathbf{E}_{x \sim p, y \sim q}\left[k(x, y)\right] + \mathbf{E}_{y,y' \sim q}\left[k(y, y')\right].\end{aligned} \tag{4.2}$$

Let $Z := (z_1, \ldots, z_m)$ be m i.i.d. random variables, where $z_i := (x_i, y_i)$ (i.e. we assume $m = n$). An unbiased empirical estimate of MMD^2 is

$$\mathrm{MMD}_u^2\left[\mathcal{F}, X, Y\right] = \frac{1}{(m)(m-1)} \sum_{i \neq j}^{m} h(z_i, z_j), \tag{4.3}$$

which is a one-sample U-statistic with $h(z_i, z_j) := k(x_i, x_j) + k(y_i, y_j) - k(x_i, y_j) - k(x_j, y_i)$.

Proof [Eq. (4.2) in Lemma 4.4] In an RKHS, function evaluations can be written $f(x) = \langle \phi(x), f \rangle$, where $\phi(x) = k(x, .)$. Denote by $\mu_p := \mathbf{E}_{x \sim p(x)} [\phi(x)]$ the expectation of $\phi(x)$ (assuming that it exists),[1] and note that $\mathbf{E}_p[f(x)] = \langle \mu_p, f \rangle$. Then

$$
\begin{aligned}
&\mathrm{MMD}^2[\mathcal{F}, p, q] \\
&= \left(\sup_{\|f\|_{\mathcal{H}} \leq 1} \mathbf{E}_p [f(x)] - \mathbf{E}_q [f(y)] \right)^2 \\
&= \left(\sup_{\|f\|_{\mathcal{H}} \leq 1} \mathbf{E}_p [\langle \phi(x), f \rangle_{\mathcal{H}}] - \mathbf{E}_q [\langle \phi(y), f \rangle_{\mathcal{H}}] \right)^2 \\
&= \left(\sup_{\|f\|_{\mathcal{H}} \leq 1} \langle \mu_p - \mu_q, f \rangle_{\mathcal{H}} \right)^2 \\
&= \|\mu_p - \mu_q\|_{\mathcal{H}}^2 \\
&= \langle \mu_p, \mu_p \rangle_{\mathcal{H}} + \langle \mu_q, \mu_q \rangle_{\mathcal{H}} - 2 \langle \mu_p, \mu_q \rangle_{\mathcal{H}} \\
&= \mathbf{E}_p \langle \phi(x), \phi(x') \rangle_{\mathcal{H}} + \mathbf{E}_q \langle \phi(y), \phi(y') \rangle_{\mathcal{H}} \\
&\quad - 2\mathbf{E}_{p,q} \langle \phi(x), \phi(y) \rangle_{\mathcal{H}},
\end{aligned}
$$

where x' is a random variable independent of x with distribution p, and y' is a random variable independent of y with distribution q. The proof is completed by applying $\langle \phi(x), \phi(x') \rangle_{\mathcal{H}} = k(x, x')$. ■

The empirical statistic is an unbiased estimate of MMD^2, although it does not have minimum variance (the minimum variance estimate is almost identical: see (Serfling, 1980, Section 5.1.4)). Intuitively we expect $\mathrm{MMD}_u^2[\mathcal{F}, X, Y]$ to be small if $p = q$, and the quantity to be large if the distributions are far apart. Note that it costs $O((m+n)^2)$ time to compute the statistic.

Having defined our test statistic, we briefly describe the framework of statistical hypothesis testing as it applies in the present context, following (Casella & Berger, 2002, Chapter 8). Given i.i.d. samples $X \sim p$ of size m and $Y \sim q$ of size n, the

[1] A sufficient condition for this is $\|\mu_p\|_{\mathcal{H}}^2 < \infty$, which is rearranged as $\mathbf{E}_p[k(x, x')] < \infty$, where x and x' are independent random variables drawn according to p.

statistical test, $\mathcal{T}(X,Y) : \mathcal{X}^m \times \mathcal{X}^n \mapsto \{0,1\}$ is used to distinguish between the null hypothesis $\mathcal{H}_0 : p = q$ and the alternative hypothesis $\mathcal{H}_1 : p \neq q$. This is achieved by comparing the test statistic $\mathrm{MMD}[\mathcal{F}, X, Y]$ with a particular threshold: if the threshold is exceeded, then the test rejects the null hypothesis (bearing in mind that a zero population MMD indicates $p = q$). The acceptance region of the test is thus defined as any real number below the threshold. Since the test is based on finite samples, it is possible that an incorrect answer will be returned: we define the Type I error as the probability of rejecting $p = q$ based on the observed sample, despite the null hypothesis being true. Conversely, the Type II error is the probability of accepting $p = q$ despite the underlying distributions being different. The level α of a test is an upper bound on the Type I error: this is a design parameter of the test, and is used to set the threshold to which we compare the test statistic. A consistent test achieves a level α, and a Type II error of zero, in the large sample limit. We will see that the test proposed in this paper is consistent.

4.3 An unbiased test based on the asymptotic distribution of the U-Statistic

We now propose a statistical test of whether $p \neq q$, which is based on the asymptotic distribution of MMD_u^2. This distribution under $\mathcal{H}_1$ is given by (Serfling, 1980, Section 5.5.1), and the distribution under $\mathcal{H}_0$ is computed based on (Serfling, 1980, Section 5.5.2) and (Anderson *et al.*, 1994, Appendix); see Gretton *et al.* (2007) for details.

Theorem 4.5 *We assume* $\mathbf{E}\left(h^2\right) < \infty$. *Under* $\mathcal{H}_1$, MMD_u^2 *converges in distribution to a Gaussian according to*

$$m^{\frac{1}{2}}\left(\mathrm{MMD}_u^2 - \mathrm{MMD}^2\left[\mathcal{F}, p, q\right]\right) \xrightarrow{D} \mathcal{N}\left(0, \sigma_u^2\right),$$

where $\sigma_u^2 = 4\left(\mathbf{E}_z\left[\left(\mathbf{E}_{z'} h(z,z')\right)^2\right] - \left[\mathbf{E}_{z,z'}\left(h(z,z')\right)\right]^2\right)$, *uniformly at rate* $1/\sqrt{m}$ *(Serfling, 1980, Theorem B, p. 193). Under* $\mathcal{H}_0$, *the U-statistic is degenerate, meaning* $\mathbf{E}_{z'} h(z,z') = 0$. *In this case,* MMD_u^2 *converges in distribution according to*

$$m\mathrm{MMD}_u^2 \xrightarrow{D} \sum_{l=1}^{\infty} \lambda_l \left[z_l^2 - 2\right], \tag{4.4}$$

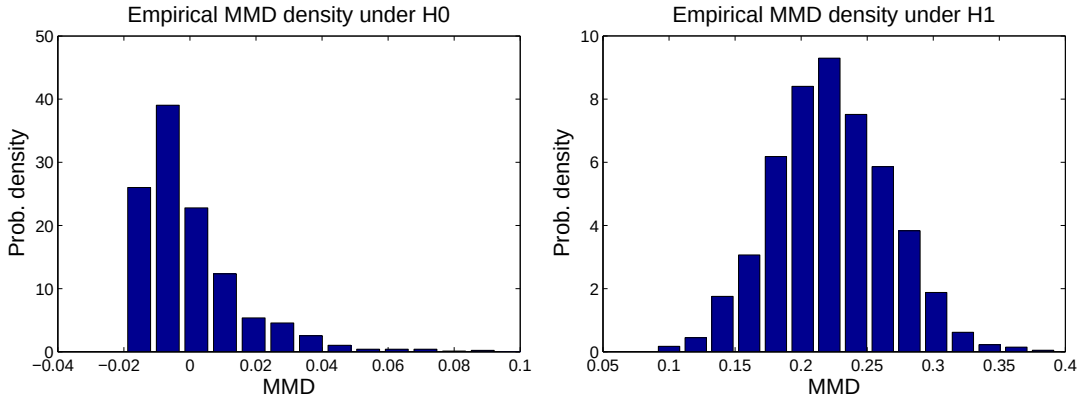


Figure 4.2: Left: Empirical distribution of the MMD under $\mathcal{H}_0$, with p and q both Gaussians with unit standard deviation, using 50 samples from each. **Right:** Empirical distribution of the MMD under $\mathcal{H}_1$, with p a Laplace distribution with unit standard deviation, and q a Laplace distribution with standard deviation $3\sqrt{2}$, using 100 samples from each. In both cases, the histograms were obtained by computing 2000 independent instances of the MMD.

where $z_l \sim \mathcal{N}(0, 2)$ i.i.d., λ_i are the solutions to the eigenvalue equation

$$\int_{\mathcal{X}} \tilde{k}(x, x')\psi_i(x)dp(x) = \lambda_i \psi_i(x'),$$

and $\tilde{k}(x_i, x_j) := k(x_i, x_j) - \mathbf{E}_x k(x_i, x) - \mathbf{E}_x k(x, x_j) + \mathbf{E}_{x,x'} k(x, x')$ is the centred RKHS kernel.

We illustrate the MMD density under both the null and alternative hypotheses by approximating it empirically for both $p = q$ and $p \neq q$. Results are plotted in Figure 4.2.

Our goal is to determine whether the empirical test statistic MMD_u^2 is so large as to be outside the $1 - \alpha$ quantile of the null distribution in (4.4) (consistency of the resulting test is guaranteed by the form of the distribution under $\mathcal{H}_1$). One way to estimate this quantile is using the bootstrap Arcones & Giné (1992) on the aggregated data. Alternatively, we may approximate the null distribution by fitting Pearson curves to its first four moments (Johnson *et al.*, 1994, Section 18.8).

4.4 Experiments

We conducted distribution comparisons using our MMD-based tests on datasets from bioinformatics and database applications. We applied tests based on both moment matching to Pearson curves (MMD_u^2 M) and the bootstrap (MMD_u^2 B). For our kernel, we used a Gaussian with σ set to the median distance between points in the aggregate sample, besides on the graph data, where we used the graph kernel for proteins from Borgwardt *et al.* (2005). We also compared against several alternatives from the literature (see below): the multivariate t-test, the Friedman-Rafsky Kolmogorov-Smirnov generalisation (*Smir*), the Friedman-Rafsky Wald-Wolfowitz generalisation (*Wolf*), the Biau-Györfi test (*Biau*), and the Hall-Tajvidi test (*Hall*). Note that the Biau-Györfi test does not apply to very high-dimensional problems (since it requires partitioning of the space into a grid), and that MMD is the only method applicable to structured data such as graphs.

Overview of previous approaches to statistical hypothesis testing We give a brief overview of previous approaches to the two sample problem for multivariate data. A generalisation of the Wald-Wolfowitz runs test to the multivariate domain was proposed and analysed in Friedman & Rafsky (1979); Henze & Penrose (1999) (*Wolf*), which involves counting the number of edges in the minimum spanning tree over the aggregated data that connect points in X to points in Y. The resulting test relies on the asymptotic normality of the test statistic, and this quantity is not distribution-free under the null hypothesis for finite samples (it depends on p and q). The computational cost of this method using Kruskal's algorithm is $O((m+n)^2 \log(m+n))$, although more modern methods improve on the $\log(m+n)$ term (see Chazelle (2000); note also that Friedman and Rafsky, the authors of Friedman & Rafsky (1979), state that calculating the matrix of distances, which costs $O((m+n)^2)$, dominates their computing time; this may not be the case for large sample sizes, however). Two possible generalisations of the Kolmogorov-Smirnov test to the multivariate case were studied in Bickel (1969); Friedman & Rafsky (1979). The approach of Friedman and Rafsky (*Smir*) in this case again requires a minimal spanning tree, and has a similar cost to their multivariate runs test.

A more recent multivariate test was proposed in Rosenbaum (2005). This entails computing the minimum distance non-bipartite matching over the aggregate data, and using the number of pairs containing a sample from both X and Y as a test statistic. The resulting statistic is distribution-free under the null hypothesis at finite sample sizes, in which respect it is superior to the Friedman-Rafsky test; on the other hand, it costs $O((m+n)^3)$ to compute. Another distribution-free test was proposed by Hall & Tajvidi (2002) (*Hall*): this requires us to aggregate the data as $Z = \{X, Y\}$, find the j points in Z closest to each point in X for all $j \in \{1, \ldots, m\}$, count how many of these are from Y, and compare this with the number of points expected under the null hypothesis (the procedure is repeated for each point in Y w.r.t. points in X). As we shall see in our experimental comparisons, the test statistic is costly to compute; Hall & Tajvidi (2002) consider only tens of points in their experiments.

Yet another approach is to use some distance (e.g. L_1 or L_2) between Parzen window estimates of the densities as a test statistic Anderson *et al.* (1994); Biau & Gyorfi (2005), based on the asymptotic distribution of this distance given $p = q$. One problem with the L_1 approach of Biau & Gyorfi (2005) (*Biau*), however, is that it requires the space to be partitioned into a grid of bins, which becomes difficult or impossible for high dimensional problems. Hence we do not use this test for high-dimensional problems in our experiments.

Toy Example: Two Gaussians In our first experiment, we investigated the scaling performance of the various tests as a function of the dimensionality d of the space $\mathcal{X} \subset \mathbb{R}^d$, when both p and q were Gaussian. We considered values of d up to 2500. The levels for all tests were set at $\alpha = 0.05$, $m = 250$ samples were used, and results were averaged over 100 repetitions. In the first case, the distributions had different means and unit variance. The percentage of times the null hypothesis was correctly rejected over a set of Euclidean distances between the distribution means (20 values logarithmically spaced from 0.05 to 50), was computed as a function of the dimensionality of the normal distributions. In case of the t-test, a ridge was added to the covariance estimate, to avoid singularity (the ratio of largest to smallest eigenvalue was ensured to be at most 2). In the second case, samples were drawn from distributions $\mathcal{N}(0, \mathbf{I})$ and $\mathcal{N}(0, \sigma^2\mathbf{I})$ with different variance. The percentage of null rejections was averaged over

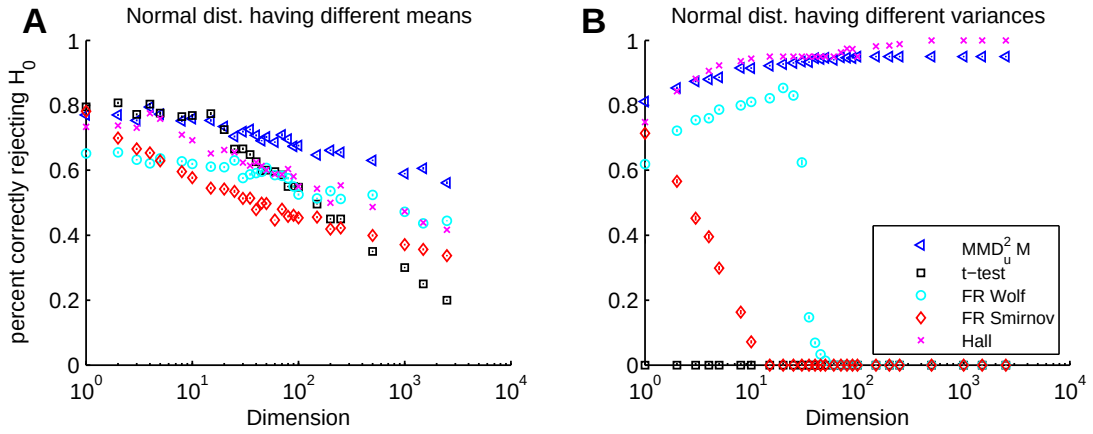


Figure 4.3: Type II performance of the various tests when separating two Gaussians, with test level $\alpha = 0.05$. **A** Gaussians have same variance and different means. **B** Gaussians have same mean and different variances.

20 σ values logarithmically spaced from $10^{0.01}$ to 10. The t-test was not compared in this case, since its output would have been irrelevant. Results are plotted in Figure 4.3.

In the case of Gaussians with differing means, we observe the t-test performs best in low dimensions, however its performance is severely weakened when the number of samples exceeds the number of dimensions. The performance of MMD_u^2 M is comparable to the t-test for low sample sizes, and outperforms all other methods for larger sample sizes.

In the case of Gaussians of differing variance, the *Hall* test performs best, followed closely by MMD_u^2. *FR Wolf* and (to a much greater extent) *FR Smirnov* both have difficulties in high dimensions, failing completely once the dimensionality becomes too great.

Data integration As a first application of MMD, we performed distribution testing for data integration: the objective is to aggregate two datasets into a single sample, with the understanding that both original samples are generated from the same distribution. Clearly, it is important to check this last condition before proceeding, or an analysis could detect patterns in the new dataset that are caused by combining the two different source distributions, and not by real-world phenomena.

We chose several real-world settings to perform this task: we compared microarray data from microarray data from normal and tumor tissues (Health status), microarray data from different subtypes of cancer (Subtype), and neural data recordings with and without spike events (Neural Data I and II). In all cases, the two data sets have different statistical properties, but the detection of these differences is made difficult by the high data dimensionality.

We applied our tests to these datasets in the following fashion. Given two datasets A and B, we either chose one sample from A and the other from B (*Attributes = different*); or both samples from either A or B (*Attributes = same*). We then repeated this process up to 1200 times. Results are reported in Table 4.1. We see that MMD_u^2 is consistently the best test over all these data, always detecting differences where they occurred while still getting a lower Type I error probability than any other test besides the t-test and MMD. However the latter two tests were much too conservative, and failed to detect the vast majority of differences (besides for the data set with the largest sample size, i.e. Neural Data I).

Dataset	Attr.	MMD_u^2 B	MMD_u^2 M	t-test	Wolf	Smir	Hall
Neural Data I	Same	96.5	96.5*	100.0	97.0	95.0	96.0
	Different	**0.0**	**0.0***	42.0	**0.0**	10.0	49.0
Neural Data II	Same	94.6	95.2*	100.0	95.0	94.5	96.0
	Different	3.3	3.4*	100.0	**0.8**	31.8	5.9
Health status	Same	95.5	94.4	100.0	94.7	96.1	95.6
	Different	1.0	**0.8**	100.0	2.8	44.0	35.7
Subtype	Same	99.1	96.4	100.0	94.6	97.3	96.5
	Different	**0.0**	**0.0**	100.0	**0.0**	28.4	0.2

Table 4.1: Distribution testing for data integration on multivariate data. Numbers indicate the percentage of repetitions for which the null hypothesis (p=q) was accepted, given $\alpha = 0.05$. Sample size (dimension; repetitions of experiment): Neural I 4000 (63; 100) ; Neural II 1000 (100; 1200); Health Status 25 (12,600; 1000); Subtype 25 (2,118; 1000). * approximation to 4th moment used.

Dataset	Attr.	MMD_u^2 B	MMD_u^2 M	t-test	Wolf	Smir	Hall	Biau
BIO	Same	93.8	94.8	95.2	90.3	95.8	95.3	99.3
	Different	**17.2**	17.6	36.2	**17.2**	18.6	17.9	42.1
FOREST	Same	96.4	96.0	97.4	94.6	99.8	95.5	100.0
	Different	**0.0**	**0.0**	0.2	3.8	**0.0**	50.1	**0.0**
CNUM	Same	94.5	93.8	94.0	98.4	97.5	91.2	98.5
	Different	2.7	**2.5**	19.17	22.5	11.6	79.1	50.5
FOR10D	Same	94.0	94.0	100.0	93.5	96.5	97.0	100.0
	Different	**0.0**	**0.0**	**0.0**	**0.0**	1.0	72.0	100.0

Table 4.2: Attribute matching on univariate (BIO, FOREST, CNUM) and multivariate data (FOR10D). Numbers indicate the percentage of accepted null hypothesis (p=q) pooled over attributes. $\alpha = 0.05$. Sample size (dimension; attributes; repetitions of experiment): BIO 377 (1; 6; 100); FOREST 538 (1; 10; 100); CNUM 386 (1; 13; 100); FOR10D 1000 (10; 2; 100).

Attribute matching Our experiments address automatic attribute matching. Given two databases, we want to detect corresponding attributes in the schemas of these databases, based on their data-content (as a simple example, two databases might have respective fields Wage and Salary, which are assumed to be observed via a subsampling of a particular population, and we wish to automatically determine that both Wage and Salary denote to the same underlying attribute). We use a two-sample test on pairs of attributes from two databases to find corresponding pairs.[1] This procedure is also called *table matching* for tables from different databases. We performed attribute matching as follows: first, the dataset D was split into two halves A and B. Each of the n attributes in A (and B, resp.) was then represented by its instances in A (resp. B). We then tested all pairs of attributes from A and B against each other, to find the optimal assignment of attributes $A_1, \ldots, A_n$ from A to attributes $B_1, \ldots, B_n$ from B. We assumed that A and B contained the same number of attributes.

As a naive approach, one could assume that any possible pair of attributes might correspond, and thus that every attribute of A needs to be tested against all the attributes

[1] Note that corresponding attributes may have different distributions in real-world databases. Hence, schema matching cannot solely rely on distribution testing. Advanced approaches to schema matching using MMD as one key statistical test are a topic of current research.

of B to find the optimal match. We report results for this naive approach, aggregated over all pairs of possible attribute matches, in Table 4.2. We used three datasets: the census income dataset from the UCI KDD archive (CNUM), the protein homology dataset from the 2004 KDD Cup (BIO) Caruana & Joachims (2004), and the forest dataset from the UCI ML archive Blake & Merz (1998). For the final dataset, we performed univariate matching of attributes (FOREST) and multivariate matching of tables (FOR10D) from two different databases, where each table represents one type of forest. Both our asymptotic MMD_u^2-based tests perform as well as or better than the alternatives, notably for CNUM, where the advantage of MMD_u^2 is large. The next best alternatives are not consistently the same across all data: e.g. in BIO they are *Wolf* or *Hall*, whereas in FOREST they are *Smir*, *Biau*, or the t-test. Thus, MMD_u^2 appears to perform more consistently across the multiple datasets. The Friedman-Rafsky tests do not always return a Type I error close to the design parameter: for instance, *Wolf* has a Type I error of 9.7% on the BIO dataset (on these data, MMD_u^2 has the joint best Type II error without compromising the designed Type I performance).

4.5 Summary and discussion

We have established a simple statistical test for comparing two distributions p and q. The test statistic is based on the maximum deviation of the expectation of a function evaluated on each of the random variables, taken over a sufficiently rich function class. We do not require density estimates as an intermediate step. Our method either outperforms competing methods, or is close to the best performing alternative. Finally, our test was successfully used to compare distributions on graphs, for which it is currently the only option.

Chapter 5

Relation of electrophysiological data to state-of-the art V1 model

We present a cortical circuit model for a patch of primary visual cortex (V1), that is based on detailed anatomical data and adapted, wherever possible, to the physiology of macaque monkeys. The model is complemented by a standard model for the retina and the LGN. Parameters of the model are fitted to an inter-spike interval distribution (including higher order statistics) of the experimental data using MMD *as a divergence of distributions. Stimulating both, the monkey and the circuit model, with the same semi-natural movie stimulus the statistics of firing activity of the circuit model is compared with multi-electrode data simultaneously recorded from areas V1 and LGN. We find that the model reasonable resembles electrophysiological data in its firing statistics. However, a direct comparison of model spike trains to recorded electrodes shows only weak correlation, suggesting that current state-of-the-art network models do not capture all computational aspects of V1 under natural movie stimulation.*

5.1 Introduction

Processing of visual information in the primary visual cortex (V1) has been a subject of extensive research. Nevertheless, many aspects of its computational role and the mechanisms underlying it remain poorly understood (Olshausen & Field (2005)). Numerical simulations of detailed biophysical models provide powerful tools for investigating the computational function of cortical microcircuits. In general, approaches along this line attempt to incorporate the known anatomy and physiology of the primary visual cortex to replicate experimental data on emergent functional properties as for instance the structure of preferred orientation maps (Adorjan *et al.* (1999); Bartsch & van Hemmen (2001); Blumenfeld *et al.* (2006)), direction selectivity maps (Ernst *et al.* (2001); Wenisch *et al.* (2005)) and simple and complex cells (Chance *et al.* (1999); Tao *et al.* (2004); Wielaard *et al.* (2001)).

When trying to understand aspects of a complex system, such as the primary visual system, it is necessary and inevitable to simplify and specialize the model to aspects under consideration. Here we take a different approach, in that we want to analyze in what sense a state-of-the-art model of the early visual pathway, is compatible with electrophysiological recordings from the monkey.

Network models incorporating many anatomical details of cortical systems still contain approximations and abstractions in respect to the real system. One aspect not

captured in common connectionists models is the interaction of LFP and spikes (see Chapter 2), which nevertheless seem to be important for stimulus encoding (see Chapter 3). Thus it is necessary but usually difficult to benchmark the "realisticness" of such network models. Here we develop a tool for optimizing some aspect of the *statistics* of spiking in a few chosen parameters in comparison to real data obtained from electrophysiological recordings. This method also enables the systematic evaluation of components of the model which are necessary to achieve a close match to realistic dynamics. We exemplary fit a detailed network model in a few meta-parameters and characterize the spike statistics of the resulting optimized model in detail. We then try to pin down statistical aspects of the real data which are not modeled by our generic model.

For this purpose we develop a model of a patch of V1 with an area of 25 mm^2 cortical surface comprising several hypercolumns. The model is based on the cortical microcircuit model described in Haeusler & Maass (2007) that implements experimental data from Thomson & Bannister (2003) on lamina-specific connection probabilities and connection strengths between excitatory and inhibitory neurons of three cortical layers and data from Markram *et al.* (1998) and Gupta *et al.* (2000) regarding stereotypical dynamic properties (such as paired pulse depression and paired pulse facilitation) of synaptic connections between excitatory and inhibitory cortical neurons. We here extend this model laterally and incorporate the anatomical particularities of V1 of macaques. Due to its lateral extend it will be possible to include lateral (long-range) connections, which are thought of to be essential for important computational functions of V1 such as spatial integration of extra-classical receptive field context (Gilbert *et al.* (1996)).

Furthermore, in contrast to Haeusler & Maass (2007), we here explicitly modeled the output of the lateral geniculate nucleus (LGN) in response to photoreceptor activity on the retina evoked by visual stimuli. For the retina and the LGN we used the model described in Dong & Atick (1995) which is based on the assumption that in the visual system natural inputs are decorrelated spatially at the level of the retina and temporally in the LGN so that signals that arrive in the visual cortex are encoded in an efficient form. This model accounts for lagged and non-lagged cells, which have been observed experimentally (Humphrey & Weller (1988a,b); Mastronarde (1987)). We also incorporated a tonic LGN response and a tendency to produce a structured spiking output for

high firing rates (as opposed to a pure Poisson spikes train), based on a more recently suggested phenomenological model of the (cat) LGN (Gazeres *et al.* (1998)).

Using the thalamic input model together with the V1 model, we are in principle able to present the same semi-natural movie stimulus to the model, which were used during experiment in our available dataset from V1 in macaque monkeys. Thus, conclusively, we try to directly compare model output with the measured spikes in the experiment.

5.2 Methods

5.2.1 Electrophysiological recordings

For details on experimental setup we refer to Chapter 2 (Section 2.2.1) and Chapter 3 (Section 3.2.1), where it is described in detail. In this Chapter we used V1 and LGN data from 6 sessions (4 monkeys) of Chapter 2 ("a98nm5", "a98nm6", "d04nm1", "d04nm2", "c98nm1", "l97nm1"). Briefly, in all sessions an identical semi-natural movie with about 5 minutes duration was shown repetitively. Data was recorded using multi-electrodes arrays penetrating V1 (and simultaneously LGN in 3 sessions) of anesthetized macaque.

5.2.1.1 Spike extraction

As in Chapter 2, spike times are detected by applying a threshold to the high-pass filtered 7 kHz signal described above (4^{th} order Butterworth, cutoff frequency 500 Hz). However, in this chapter the threshold was set to a high value of 7 times the standard deviation σ of the "noise component" of the high-pass filtered signal. Therefore only very prominent spikes, or spikes originating from neurons very close to the electrode tip, are likely to be present in the spike trains. Because most recordings were done with single tip electrodes we do not employ any kind of spike sorting.

5.2.2 Model

In this section we describe a biological realistic V1 model developed for comparison of electrophysiological recordings from macaque. It consists of an input model

(equivalent to retina and thalamus) and a V1 model for a patch of cortex, receiving and processing the input. In the following we will first describe the V1 model and subsequently the input model in detail.

5.2.2.1 V1 Model

We developed a biological realistic model of a small patch of V1 based on various experimental data. The core of this model is similar to a model of a generic cortical microcircuit suggested by Haeusler & Maass (2007). The model of Haeusler & Maass (2007) consisted of three layers with neurons assigned to layer 2/3, layer 4 and layer 5. Each layer contained a population of excitatory neurons and a population of inhibitory neurons. Inter-layer connectivity (probabilities and strength) were chosen according to experimental data assembled in Thomson & Bannister (2003).

In this Chapter this generic model was extended to match the anatomical particularities of the V1 cortex of macaques. However, since a detailed description of the layer specific connectivity structure is not available for monkeys to our knowledge, we decided to use the same data derived from cat cortex slices (Thomson & Bannister (2003)). Although there exist differences, the primary connectivity structure in macaque is similar to that of the cat (Callaway (1998)). In particular, if one identifies layer 2/3 and 4 in cat with 2-4B and 4C in macaque, respectively, the major geniculate input in both species first reaches layer 4C, is projected to layer 2-4B, which in turn projects it further to layer 5 (and layer 6 via layer 5), where feedback connections are made to layers 2-4 (see Callaway (1998) for a review). Thus using the connectivity data from cat seems to be a first but reasonable approximation.

Three layer population of neurons (excitatory and inhibitory in a ratio of 4:1, Beaulieu *et al.* (1992); Markram *et al.* (2004)) were modeled separately (and spaced on a cuboid grid). As in Haeusler & Maass (2007), we used conductance based single compartment neuron models. However, due to a considerable gain in computational speed we employed a simpler neuron model suggested by Izhikevich (2003), which can be fitted to a wide range of firing dynamics (Izhikevich (2006)). We randomly draw the parameters for each neuron in the network from within the bounds provided by Izhikevich *et al.* (2004). In this parameter range the excitatory pool consists of a mixture of regular spiking, intrinsically bursting, and chattering cells, with a bias to regular

spiking cells. The inhibitory pool of neurons comprises fast spiking and low-threshold spiking cells. Thus individual neurons and hence firing behaviors within our model will have richer and more realistic diversity than in Haeusler & Maass (2007), where all neuron had fixed parameter setting. The scale of synaptic input conductances will be optimized to match firing statistics of electrophysiological data (see Section 5.2.3).

In addition to synaptic input from fellow neurons in the network, each neuron receives synaptic background noise reflecting the bombardment of synaptic inputs from a large number of more distal neurons which causes a depolarization of the membrane potential and a lower input resistance commonly referred to as "high conductance state" (Destexhe *et al.* (2001b)). Short term synaptic dynamics was implemented according to Markram *et al.* (1998), with synaptic parameters chosen as in Maass *et al.* (2002) to fit data from microcircuits in rat somatosensory cortex (based on Gupta *et al.* (2000) and Markram *et al.* (1998)). For details we refer to Haeusler & Maass (2007).

In contrast to Haeusler & Maass (2007), we set the relative amount of neurons per layer to 33%. These numbers corresponds to experimental data from macaques (Beaulieu *et al.* (1992); O'Kusky & Colonnier (1982); Tyler *et al.* (1998)), although we slightly adjusted the relative amount of neurons compared to the experimental values (where layer 4 has about a third more neurons), because our model neglects the magnocellular and koniocellular pathways in favor of the parvocellular pathway (Callaway (1998)). The three layers of the model can be identified with layers 2-4B, 4Cβ and 5-6. To avoid confusion in analogy to Haeusler & Maass (2007) we will nevertheless call them 2/3, 4 and 5 in the following text.

In macaques each of our three layers would contain approximately 50000 neurons under a surface area of 1 mm^2 (Beaulieu *et al.* (1992)). In our model we neglected that neuron density varies with layer about 1.5 fold (Beaulieu *et al.* (1992)) and instead assumed that positions are uniformly distributed throughout the cortex. Thus, for simplicity, we positioned all neurons on a cuboid grid with fixed grid spacings. Using the true neuron density, e.g. for layer 2/3, the grid spacing would amount to 20μm in all directions. Because simulating such a dense network would take too much time, we diluted the neuron density by increasing the lateral grid spacing to 80μm and the vertical spacing to about 200μm (see Results Section 5.3).

Note that sublaminar organization, such as the 2/3 blobs and inter-blob regions (Callaway (1998)) are neglected for simplicity and for the lack of precise data. How-

ever, as described in the following sections, the model contained a realistic thalamic input, smooth orientation map and patchy long-range connections in the superficial layer.

Lateral connectivity structure Whereas the generic microcircuit model of Haeusler & Maass (2007) was restricted to a few 100 μm with uniform connectivity per layer and neuron type, we here extended the model laterally to the order of several millimeters. Thus connection probabilities in our model depend on lateral distance. For intra-cortical connections we generally used a bell-shaped (Gaussian) probability distribution for determining the lateral extend. The standard deviation of the Gaussian was set to $200\,\mu$m for excitatory neurons (Blasdel *et al.* (1985); Buzas *et al.* (2006); Lund *et al.* (2003)) and somewhat smaller, namely $150\,\mu$m, for inhibitory neurons to incorporate the occurrence of extremely narrow inhibitory dendritic and axonal spreads observed ($70\,\mu$m, Lund *et al.* (2003)). Arborization of excitatory neurons in layer 5 seems to be wider and more diffuse and has a spread of more than 500 μm laterally from the soma (Blasdel *et al.* (1985)). Thus for these connections we set the standard deviation to 300μm. Note that the value for the standard deviation is about half the expected maximal extend of 95% of the arborizations.

To ensure consistency with the inter-layer connectivity data of Thomson & Bannister (2003), which were obtained from cortical slices of around 100μm thickness, we scaled the Gaussian profile such that the peak probability corresponds to the experimentally measured connection probabilities. Therefore the connectivity data is locally preserved. All lateral connections are subject to toroidal boundary conditions yielding positional independence in synaptic drive throughout the lateral extend of the circuit.

Patchy lateral long-range connections In both cat and macaque, many pyramidal cells in layer 2/3 of the striate cortex (and elsewhere in cortex (Lund *et al.* (2003))) send characteristic long-range projections targeting laterally 80% excitatory and 20% inhibitory cells (McGuire *et al.* (1991)) which are up to 6 mm and more away (Buzas *et al.* (2006); Gilbert *et al.* (1996); Lund *et al.* (2003)). Moreover, targeted neurons tend to have similar feature preference as its origin, resulting in patchy connections linking similar preferred orientations (Buzas *et al.* (2006); Gilbert *et al.* (1996)). Combining anatomical reconstruction of neurons and optical imaging of orientation maps, Buzas

et al. (2006) proposed a formula for the button density ρ of a typical layer 2/3 pyramidal cell:

$$\rho(r, \Delta\phi) = Z \left(e^{-\frac{r^2}{2\sigma_1^2}} + m\, e^{-\frac{r^2}{2\sigma_2^2}}\, e^{\kappa \cos 2(\Delta\phi - \mu)} \right) \tag{5.1}$$

Here r is the distance and $\Delta\phi$ the difference of preferred orientation of the pre- and postsynaptic neurons. Parameter m is the scaling of importance between a local orientation independent and a long-range orientation dependent term. The orientation dependent term is a product between a Gaussian and a von Mises distribution, and accounts for the higher likelihood of connecting neurons with preferred orientation difference near μ degree. Standard deviations σ_1 and σ_2 regulate the spatial width of the non-oriented and oriented term, respectively. Parameter κ signifies the "peakiness" of the density on the orientation axis. Z is a normalization constant.

If one assumes that in first approximation dendrites (or equivalently axons) have no spatial extend (i.e. they collapse to a point at the position of the soma), the button density defined by Eq. 5.1 can be seen as an estimate for the connection probability of a neuron to neurons at distance r having preferred orientation difference $\Delta\phi$. Since we defined preferred orientation in a hard-wired manner via "oriented" input connections (see Section 5.2.2.3) one can readily apply Eq. 5.1 for the lateral connections in layer 2/3. Thus, we applied Eq. 5.1 to projections from excitatory cells targeting excitatory and inhibitory cells (McGuire *et al.* (1991)).

Parameters were set as follows. For the local non-oriented term we took $\sigma_1 = 200\mu$m, as for the connections between other layers (see above). In Buzas *et al.* (2006) values for μ jitter around zero degree for individual cells. We therefore set $\mu = 0^\circ$ implicating that the connection probability is highest for iso-oriented cells. We set other parameters to values in between the two populations described by Buzas *et al.* (2006), but in the vicinity of their population 1, where the experimentally mapped cortical surface region is bigger. Thus we set $\sigma_2 = 1000\mu$m, $\kappa = 1$, and $m = 10$. As before, the connection probability was scaled according to the Thomson & Bannister (2003) data (by setting Z to appropriate values). Thus, locally, i.e. for a neuron of the same lateral position (and orientation preference), such as a neuron located in the same layer beneath or above the pre-synaptic neuron, the connection probabilities are preserved. However, the weight distribution of the long-range connection is not constraint by

Thomson & Bannister (2003). Therefore we obtained the mean of the synaptic weight distribution by fitting the firing statistics to experimental data (see Section 5.2.3).

Distance dependent synaptic delay Synaptic delays differ for inhibitory and excitatory neurons and were set according to measurements by Gupta *et al.* (2000) (as in Haeusler & Maass (2007)). These delays stem from molecular processes of synaptic transmission. In addition, a second delay originating from finite spike propagation velocity of the fibers was included. This delay is dependent on the (Euclidean) distance of pre- and post-synaptic neurons. Girard *et al.* (2001) measured spike propagation orthodromically as well as antidromically and found a median conduction velocity of 0.3 m/s for the upper half and 1 m/s for the lower half of V1 in macaque. Thus for each excitatory synapse in layer 2/3 we sampled velocity from a Gaussian distribution with mean 0.3 m/s and standard deviation 0.5 m/s (with enforced lower and upper bounds 0.1 and 5 m/s, respectively). For the other layers conduction velocities were drawn from a Gaussian with mean 1 m/s and standard deviation 0.9 m/s (with bounds as before). Due to myelination, conduction velocities of inhibitory fibers are generally higher than for excitatory cells (Thomson & Bannister (2003)). Thus for all inhibitory cells we sampled velocities from a distribution with a mean and standard deviation twice as high as for excitatory neurons in the deep layers.

5.2.2.2 Input model

In the available data set, recordings were done while monkeys watched an semi-natural movie stimulus of several minutes duration. Although our modeling effort was concentrated on the V1 model, to successfully compare these data with the model, one also needs a sufficiently realistic transformation of movie stimulus to (V1 input) spike trains. Therefore retina and lateral geniculate nucleus (LGN) were modeled according to Dong & Atick (1995) as a spatio-temporal filter bank with nonlinearities, which seems to be a good compromise between simplicity and realism (Gazeres *et al.* (1998)). The filter bank converted time varying input signals on the retina, such as movies, into firing rates of LGN neurons. From these firing rates V1 input spike trains were generated. We neglected for simplicity that the ganglion cells typically react to color opponency rather than to pure luminance differences (Perry *et al.* (1984)). Thus all

color information from the movie was converted to gray scale prior to application to the input model.

Retina model The two-dimensional retinal inputs (movie frames) were filtered by a "Mexican hat" (difference of Gaussians) spatial filter (Dong & Atick (1995); Enroth-Cugell & Robson (1966); Rodieck (1965)). Filter sizes (describing the receptive fields of ganglion cells) were adapted to the geometry of parvocellular cells of macaque, where the standard deviations of the Gaussian for center and surround was estimated to $\sigma_{\text{center}} = (0.0177° + 0.00196\epsilon)$ and $\sigma_{\text{surround}} \approx 6.67\sigma_{\text{center}}$ at eccentricity ϵ, respectively (in visual degrees; estimated from Figure 4 a and b in Croner & Kaplan (1995)). After convolution of the stimulus luminance portrait with these kernels (yielding S_{center} and S_{surround}), the response of a retinal on-cell at visual field position r can be described by

$$R_{\text{ON}}(r) = C(r)\,[S_{\text{center}}(r) - \omega S_{\text{surround}}(r)]_+ \tag{5.2}$$

Following Croner & Kaplan (1995) we set the ratio of center to surround $\omega = 0.642$. The positive parts of the center and surround interaction (indicated by the brackets $[\ldots]_+$) were assigned to the response of on-cells and analogously the absolute value of the negative part to the response off-cells (Dong & Atick (1995)). For simplicity and in accordance with the established Difference-of-Gaussian model, we assumed that the origins of the center and surround summation fields are identical, although a recent study suggests that there might be an offset between them (Conway & Livingstone (2006)).

Applying the Difference-of-Gaussians model to the luminance of a stimulus results in a quantity called "contrast gain" (Croner & Kaplan (1995); Enroth-Cugell & Robson (1966); Rodieck (1965)). To calculate the firing rate of ganglion cells one has to multiply the "contrast gain" with the *local* contrast $C(r)$ (as done in Eq. 5.2), if one neglects non-linear saturation in the high contrast regime. Locality is important because the concept of a global contrast, easily defined for full-field grating stimuli commonly used in experiments, is not applicable for real world images and movies (Tadmor & Tolhurst (2000)). Following Tadmor & Tolhurst (2000) we estimated the local contrast using the same kernels as

$$C(r) = \frac{|S_{\text{center}}(r) - S_{\text{surround}}(r)|}{S_{\text{center}}(r) + S_{\text{surround}}(r)} \tag{5.3}$$

where we additionally set the contrast to be zero in the case of darkness. Note that applying Eq. 5.3 results in a response $R_{\mathrm{ON}}(r)$, which is sparser than for a constant global contrast, since the response is now quadratic in the center and surround luminance difference (see Eq. 5.2).

LGN model The retinal output is filtered by the LGN model using a temporal kernel. Our temporal kernel combines a phasic (taken from Dong & Atick (1995)) and a tonic component (as in Gazeres *et al.* (1998)), i.e. $k_{\mathrm{LGN}} = k_{\mathrm{phasic}} + k_{\mathrm{tonic}}$. It is for non-negative times

$$k_{\mathrm{phasic}}(t) = t\,(1 - \pi w_c t)\,\exp(-2\pi w_c t) \tag{5.4}$$

and

$$k_{\mathrm{tonic}}(t) = A\,\exp(-t/\tau)/\tau. \tag{5.5}$$

Parameter $A = 0.3$ is the fraction of tonic activation (in respect to the peak firing rate) for a given stimulus integrated over a time scale of $\tau = 15$ ms. Parameter $w_c = 5.5\ \mathrm{sec}^{-1}$ defines the shape of the phasic kernel (Dong & Atick (1995)).

Analogously to the retina, the positive part and the absolute value of the negative part of the temporal convolution were assigned to lagged cells and non-lagged cells, respectively. Altogether, there are four different time-varying rate outputs, i.e. that of any combination of non-lagged or lagged cells in the LGN with either on- or off-cells from the retina (Dong & Atick (1995)). Following Gazeres *et al.* (1998), a so-called "switching Gamma renewal process" used these time-varying rates to generate spike trains. This process, which was suggested to fit experimental data from cat LGN X-cells (Gazeres *et al.* (1998)), adopts a higher spike time regularity for high input rates (≥ 30 Hz; regularity parameter $r = 5$) and switches to a Poisson process for low rates (< 30 Hz). Spontaneous background activity of each LGN input neuron was set to a low value of 0.15 Hz. The peak LGN spike rate $f_{\max}$ was adjusted to achieve a mean firing rate of about 7 Hz under movie stimulation, when the four input channels are combined. The 7 Hz mean rate was estimated from our electrophysiological data from macaque monkey. Applying a typical 50 seconds movie section, we found that a mean rate of 7 Hz was achieved for $f_{\max} = 250$ Hz. The peak response would be evoked by a dot of highest contrast filling the center region of a ganglion cell with optimal

duration. This value is in good agreement with Gazeres *et al.* (1998), who report peak rates in the range of 50 to 400 Hz.

Input connectivity to V1 The visual field is retinotopically arranged on the cortical surface. However, while there exists one retinal ganglion cell per LGN cell corresponding to the same visual field position at all eccentricities in macaque, there is a considerable magnification in density of cortical neurons in V1 per degree visual field (Schein & de Monasterio (1987); Tootell *et al.* (1982)). Comparing several earlier studies, Schein & de Monasterio (1987) estimated the cortical magnification factor (CMF) at eccentricity ϵ to be (in mm cortex per degree visual field)

$$\mathrm{CMF} = \frac{12.2}{\epsilon + 0.94}. \tag{5.6}$$

Note that the definition of the cortical magnification factor (Eq. 5.6) is very convenient: for a fixed eccentricity and distance between adjacent neurons (grid spacing), one immediately gets the lateral extend of the network needed to cover a given visual field size.

The main input from LGN parvocellular pathway is projected into layer 4Cβ. There is still an ongoing debate to which extent oriented input shape the orientation selectivity exhibited by neurons in primary visual cortex or to what extent local cortical processing is involved (see Teich & Qian (2006) for a review). It seems that in macaques, orientation selectivity is thought to arise from the interaction of cells with gradually shifted input characteristics across the sublaminas of the layer 4C (Callaway (1998); Lund *et al.* (2003)), whereas the inputs to a single cell might not be oriented in macaques as suggested for the cat (Hubel & Wiesel (1977)). However, since we did not model sublaminas, we simplified the circuitry by nevertheless assuming that input connections to each neurons generate orientation tunings. This not only includes orientation selectivity in our model but allows to define orientation maps in a straightforward "hard-wired" manner (see Section 5.2.2.3).

Therefore thalamic input connection probability to a cell in the circuit was modeled as an oriented Gabor function, i.e. a Gaussian multiplied by a cosine function. The height of the Gabor function corresponds to the connection probability of LGN neurons with a cortical cell positioned at the cortical equivalent position of the origin of the

Gabor patch in the visual field. Negative and positive regions correspond to the connection probabilities of LGN on- and off-response cells, respectively. Lagged and non-lagged cells connected equally likely to cortical cells. Following Troyer *et al.* (1998), we expressed the Gabor function in parameters defining the number of subregions n_s, the aspect ratio of width and height of the Gaussian envelope a, the orientation ϕ, the offset of the cosine ψ, and the frequency of the cosine f. From these parameters one calculates the standard deviation of the Gaussian envelope as (see Troyer *et al.* (1998))

$$\Sigma^{\frac{1}{2}} = \frac{1}{4 \cdot 2.448\, f} \begin{pmatrix} n_s & 0 \\ 0 & a \end{pmatrix} \tag{5.7}$$

while using coordinates rotated by ϕ. The advantage of using these parameters is that the frequency implicitly defines the size of the Gabor patch (as the number of subregions is kept constant). Therefore the much smaller receptive fields of macaque V1 as compared to the cat, can be easily included in this framework. We used data from Bredfeldt & Ringach (2002) and chose the frequency f from a Gaussian distribution with mean and standard deviation $3.7\mathrm{deg}^{-1}$ and $2.1\mathrm{deg}^{-1}$, respectively (and enforcing a minimum and maximum of $0.7\mathrm{deg}^{-1}$ and $8\mathrm{deg}^{-1}$, respectively). Phase shift ψ and the number of subregions n_s were drawn from uniform distributions in the ranges $(1.85, 2.65)$, and $(0, 2\pi)$ (from cat as in Troyer *et al.* (1998)).

To incorporate the smooth maps for the preferred orientation ϕ and orientation preference q depending on cortical position u (see Section 5.2.2.3) we set $\phi = \phi(u)$ and the aspect ratio to $a(u) = (a_{\max} - a_{\min})\, q(u) + a_{\min}$, where we used values reported by Troyer *et al.* (1998) for the bounds $a_{\min} = 3.8$ and $a_{\max} = 4.54$ for excitatory neurons, and for the generally less well tuned inhibitory neurons $a_{\min} = 1.4$ and $a_{\max} = 2$.

Lastly, the overall connection probability, defined by the Gabor functions, was scaled to achieve an average number of input synapse of 24 for both excitatory and inhibitory neurons, which is the estimated number of parvocellular afferent connection per cortical neuron in layer 4C of macaques (Peters *et al.* (1994)). There is evidence that layer 6 receives occasional collaterals of the LGN input to layer 4 (Callaway (1998)). Thus we set the connection probability to excitatory neurons of layer 5 of our model (comprising layer 5 of 6 in the macaque nomenclature) to 20% of that of the input to layer 4. These values are in good agreement with the data from Binzegger *et al.* (2004) estimated from cat. In macaques, layer 2/3 receives only koniocellular

input (Callaway (1998)). As we have omitted the koniocellular pathway, layer 2/3 of our model does not receive any thalamic input.

Due to finite conduction velocities of the fibers, signals from the retina reach V1 with a characteristic delay of about 30 ms (Maunsell *et al.* (1999)). As the retina and LGN model did not include any delay so far, we sampled the delay of the LGN input synapses according from a Gaussian distribution with mean 31 ms and standard deviation 5 ms (and additionally enforced delays below 24 ms and above 50 ms to a value uniformly in the latter range). These values were taken from Figure 3 of Maunsell *et al.* (1999).

Top-down connections Besides thalamic input V1 receives multiple feedback connections from extra striate cortices (Felleman & Essen (1991)), especially from V2, where the feedback connections are almost as numerous as the feedforward connections (see Sincich & Horton (2005) for a review). Feedback projections predominantly project to targets in the upper layers but also to layer 5 (Rockland & Virga (1989); Sincich & Horton (2005)), although altogether there is little known about the exact (sublaminar) targets of these feedback projections (Sincich & Horton (2005)). As our model is restricted to V1 we model these projections as an additional input to layers 2/3 and 5. This second input stream has (arbitrarily) the same number of cells as the first (LGN) input stream and is aligned on a virtual grid of the same extent (in mm) as the V1 model. Connections from the input cells are established to their targets by a Gaussian probability profile in their lateral distance with 200μm standard deviation (Lund *et al.* (2003)). The Gaussian connectivity was scaled to yield a peak connection probability of 10% for layer 2/3 and 5% for layer 5. Only excitatory neurons receive the second input stream. In the simulations each input cell emits a Poisson spikes train of step-wise constant rates estimated from our data (see Section 5.2.3.1). Synaptic delays are sampled from a Gaussian with mean 1.25 ms and standard deviation 1.53 ms (with bounds 0.5 ms and 5 ms). These values correspond to delays measured antidromically for feedback connections from V2 (Girard *et al.* (2001)).

5.2.2.3 Orientation map

It is well established that orientation preference and other features (such as visual field position, ocular dominance, or direction preference) form intertwined maps, where neighboring neurons tend to respond to similar features (Hubel & Wiesel (1977); Obermayer & Blasdel (1993)).

We employed Kohonen's Self-Organizing Map algorithm (Kohonen (1982)) for the creation of realistic orientation maps across cortical surface. An orientation attribute for each neuron is necessary for defining thalamic inputs, as well as for preferred orientation dependent patchy lateral long-range connections (see above). The algorithm has been used to generate feature maps which very well resemble cortical measured feature maps in overall appearance as well as e.g. in the structure of occurrence of pinwheels (Brockmann *et al.* (1997); Erwin *et al.* (1995); Obermayer & Blasdel (1993); Obermayer *et al.* (1990, 1992)). Basically, the algorithm tries to map a low dimensional manifold (a horizontal sheet of neurons) to a high dimensional feature space while ensuring that neighboring points on the manifold exhibit similar feature preference. Let $\mathbf{z} = (x, y, q\cos(2\phi), q\sin(2\phi))^T$ define a feature vector, where $0 \leq x, y < k$ are the positions in visual space, $0 \leq q < 1$ the orientation preference (or tuning strength) and $0 \leq \phi < \pi$ the preferred orientation (we do not model ocular dominance because our V1 model is essentially one-eyed). If one uses the low-dimensional variant of the learning rule (Erwin *et al.* (1995); Obermayer & Blasdel (1993)), one attributes to each point on the manifold, i.e. each neurons having cortical 2D surface coordinates $u = (u_1, u_2)^T$, its current "optimal" feature vector $\mathbf{w}(u)$. Relations between neurons u and v are enforced by the neighborhood function $h(u, v) = \exp\left(-\frac{|u-v|^2}{2\delta^2}\right)$. With that the weight update for a neuron v can be written as

$$\Delta\mathbf{w}(v) = \alpha h(u_*, v)\left(\mathbf{z} - \mathbf{w}(u_*)\right). \tag{5.8}$$

One notes that in each learning step the neuron u_* showing maximal response to the current input $\mathbf{z}$ is updated in the direction of the input, weighted by a learning rate α. Depending on the *cortical* distance to the maximally activated neuron, the preferred features of the remainder of the neurons will be updated to a lesser extend in the same direction (mediated by the neighborhood function). In this rule we took the maximally activated neuron to be the nearest in feature space to the current input,

$u_* = \operatorname{argmin} |\mathbf{w}(u) - \mathbf{z}|$. We sampled the input features from uniform distribution (within the above bounds). k regulates the hierarchy between different features (Obermayer *et al.* (1992)) and was set to $k = 5$. If one starts from a retinotopic initial condition, a high value for k ensures that cortical position corresponds to visual space in an approximate one-to-one map. The characteristic length scale δ was set to match the experimental observed correlation length in cortical orientation maps (or pinwheel center distance) of $d_{\text{pin}} = 660\mu$m (Obermayer & Blasdel (1993)). We used the approximate formula $\delta = \sqrt{k}\, d_{\text{pin}}/D/8$, where D denotes the lateral extent of our V1 model. See Figure Fig. 5.1 C for a typical orientation map generated by this algorithm.

5.2.3 Fitting parameters of the V1 model

Many parameters such as most synaptic weight scalings, i.e. the mean peak conductance of synaptic weight distributions of the intra-laminar projections, are constraint by the literature (see Section 5.2.2). However, other parameters are not constraint and therefore have to be adjusted. For example, the overall synaptic weight depends on the number of synapses received by each neuron, and has to balance the effect that there are less neurons per volume in the V1 model than in reality. In the following we attempted to estimate free parameters by fitting the statistics of extracellularly recorded spike data.

Comparing model circuit dynamics to the electrophysiological recordings is difficult for various reasons. For one, although detailed, our model circuit is still lacking many aspects of the cortex, which nevertheless contribute to the recorded signal, e.g. the cortical spatial arrangement in gyri and sulci, biophysical properties of the tissue, 3-D structure of the axonal and dendritic tree, or a high temporal and spatial dynamic of current sources and sinks (Koch (2004); Nunez & Shrinivasan (2006)). Conversely, spike trains do not constrain many aspects of the model, as for instance the exact layer position of the electrodes or the neuron type recorded from (inhibitory or excitatory) as well as its synaptic targets and weights etc are unknown.

Even if one accepts the simplistic few of an abstract circuit model, it is not clear which spikes of the circuit should be compared with the recorded signal. Most importantly, the recordings are made with extracellularly applied single tip electrodes, and thus necessarily record from multiple neurons simultaneously (see also Chapter 2 for

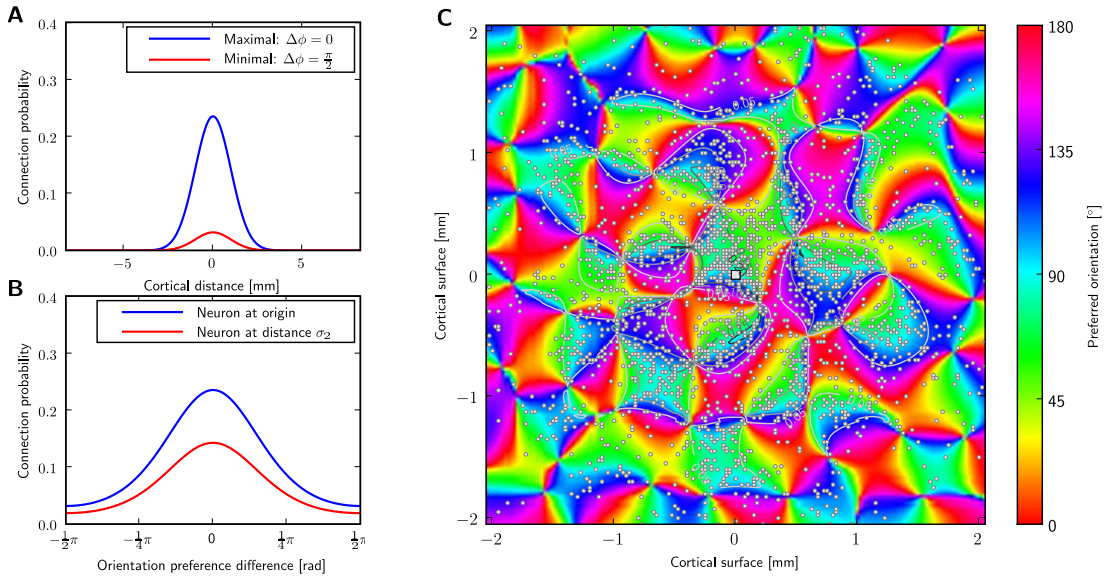


Figure 5.1: Long-range patchy connectivity of an example neuron implemented in a model circuit having $165 \times 165 \times 3$ neurons in layer 2/3 positioned on a cuboid grid with a spacing of 25 μm. (Note that these dimensions are different from that used in the simulations of the Results Section 5.3) **A** and **B**: Conditional probability that the neuron (marked with a white square in the center of plot C) is connected to a neuron having lateral distance r or orientation selectivity ϕ, respectively. Connection probability to a post-synaptic neuron at zero lateral distance and same orientation preference was scaled to experimental data ($C \approx 0.24$; Thomson & Bannister (2003)). **A**: Blue and red curves show the connection probabilities for neurons which have aligned or orthogonal preferred orientation to the pre-synaptic neuron, respectively. **C**: Connections established according to the probability distributions for a presynaptic neuron in the origin of the circuit (white square). Small white dots represent lateral positions of post-synaptic neurons. Colors code for orientation tuning of each neuron (generated by a Kohonen's Self-organizing map). Conditional connection probability is indicated by contour lines. One notes that the connection probability rises for regions with similar orientation as the pre-synaptic neuron (about $90°$) thereby generating a patchy appearance. Only the orientated (long-range) part of Eq. 5.1 (second term) is used for establishing connection in this example plot. However, because of the high weighting factor $m = 10$ (see Eq. 5.1) only very few local connection will be added when considering both terms in the simulations. The orientation map additionally determines the orientation of thalamic input connections (see Methods Section 5.2.2.2).

a detailed discussion). It has been theoretically estimated that the number of possible neurons contributing to the measured spike train is up to 1000 and higher (Henze *et al.* (2000)). However, in practice when using tetrodes, where reliable unit sorting is possible, one is typically able to clearly distinguish a few, typically up to five or six, neuron clusters (see also Chapter 3). To restrict our analysis to as few neurons as possible, we set the threshold for spikes detections to a high value, namely 7 SD of the noise component (see Methods Section 5.2.1.1). Thus the spike trains contained only prominent spikes, which most likely originated from neurons in the immediate vicinity of the electrode tip. Nevertheless, if we would compare single neurons of the model directly to the extracellular recorded spike trains, firing rates would be overestimated. Therefore we add the number of neurons comprised in a single electrode channel to the list of parameters fitted (see below). By doing this we obtain much better agreement of the inter-spike interval distributions between model and recordings.

5.2.3.1 Fitting the relative strength of the thalamic and top-down inputs

In recorded spike trains, the mean firing rate of multiple trials (5 min duration) across monkeys and V1 electrode channels is 11.7 ± 1.7 Hz (mean $\pm$ standard error of the mean) during movie stimulation and 9.9 ± 1.6 Hz during spontaneous activity (blank screen). Thus one could state that due to the thalamic input the mean firing rate of the circuit increases by about 2 Hz or enhances it by about 20 %. From simultaneous extracellular recordings in LGN (see Methods), we analogously find a mean firing rate of 7.1 ± 2.9 Hz during visual stimulation and 4.4 ± 2.1 Hz during absence of visual stimulation. Hence in LGN the movie stimulus increases the mean firing rate by about 60 % as compared to spontaneous activity.

We used these values for determining the synaptic input weight scaling and the overall weight scaling of inter-laminar connections, i.e. the overall scaling of the peak conductances, in the following manner. In the absence of *all* inter-cortical connections, the weight scaling of the two input streams were set to values achieving closest match to a given target mean firing rate r_{target} in each neuron population (minimal Euclidean distance). Assuming that the main input drive to V1 (during visual stimulation) is from the thalamus we set the target mean rate for layer 4 to 4 Hz and layer 2/3 and layer 5 to 1 Hz. Since each layer consisted of same amounts of neurons both inputs then drive the

whole model to a mean firing rate of 2 Hz which corresponds to the activity increase seen during visual stimulation in our experimental data. For this fit the thalamic input resembled the input during movie stimulation. More specifically, we applied a typical 10 sec segment of the movie stimulus to our LGN model, which generated spike trains as described above (Section 5.2.2.2).

Since no data from V2 or higher areas was simultaneously recorded, the top-down input stream was approximated by the firing statistics of V1 neurons as follows. Each top-down input neuron was independently set to emit a Poisson spike train with a mean rate drawn from a mean rate distribution estimated from the electrophysiological data of V1 under spontaneous activity condition. To allow for rate changes, the duration of the constant rate periods for each individual input neuron is drawn from a second Poisson distribution with mean of 0.5 Hz corresponding to the 2 s interval used for the rate estimation on electrophysiological data. Note that in this approximation the top-down inputs were therefore not modulated by the stimulus and merely act as a background noise source to V1 neurons.

5.2.3.2 Fitting the average weight of all intra-cortical projections and the relative weight of lateral long-range synapses

Having established the synapse strength of the inputs, the overall weight scale of all recurrent (inter and intra-laminar) synapses are to be determined. Additionally, we have to establish the mean of the synaptic weight distribution for the superficial long-range connections, because the measurements of Thomson & Bannister (2003) are restricted to thin cortical slices and therefore not applicable.

Optimization criterion via MMD To find reasonable values for both parameters, we fitted the firing statistics to the real data. In particular, we tried to match inter-spike interval distributions between electrophysiological data and simulated spike trains. As error function to be minimized we used the Maximum Mean Discrepancy (MMD) developed in Chapter 4. It is shown in Chapter 4 that the MMD of two distributions p and q is zero iff both distributions are identical. In practice, if uses a quadratic estimate of MMD, positive values (which are above the test threshold) indicate a mismatch of the two distributions under consideration, and intuitively, MMD will be the bigger the

more different the distributions are. Note that the MMD does not necessarily define a metric on the space of distributions. In that sense it is similar to the Kulback-Leibler divergence. However, since no density estimation is needed for computing the MMD, it is superior in various respects to Kulback-Leibler divergence, for instance in the ease of use and computational speed. Moreover, using the two-sample tests of Chapter 4 we readily obtain information whether the two distributions are *significantly* different based on a *finite* sample or not.

To incorporate possible higher order spike correlations we compared the k-ISI distributions of spike trains. The k-ISI distributions $p(\tau_1, \ldots, \tau_k)$ defines the probability of occurrence of k sequential spike intervals of lengths $\tau_1, \ldots, \tau_k$. Thus our goal finding the optimal parameter vector $\mathbf{a}^*$ can formally be written as

$$\mathbf{a}^* = \underset{\mathbf{a}}{\operatorname{argmin}}\ \mathrm{MMD}[\mathcal{F}, p\left(\tau_1, \ldots, \tau_k | \, \mathbf{a}\right), q\left(\tau_1, \ldots, \tau_k\right)] \tag{5.9}$$

where p and q denote the k-ISI probability distribution of the model and from the target data, respectively, and $\mathcal{F}$ denoted the function space (see Chapter 4 for details). For fitting we employed the unbiased estimate of MMD (Eq. 4.3), averaged over 10 trials with samples size 2000, and using RBF kernels of fixed size ($\sigma = 0.1$). Different kernel sizes or Laplace kernels produced similar results. We set $k = 2$ and compared the log-ISIs distribution (i.e. $\tau_i := \log t_i$, where t_i is the i-th interval in the sequence of k intervals).

We used only one model random seed for the fitting process, to reduce computational costs. Thus the neuron-to-neuron connections were identical, only the synaptic weights are changed. Since our networks contains a great amount of synapses (about $4 \cdot 10^7$; see below), it is likely that random effects are averaged already within a model. Indeed, even for a smaller network ($20 \times 20 \times 9$ grid size) the fitted parameters were very similar for different construction seeds (not shown).

Generating multi-unit spike trains from the model circuit For generating multi-unit spike trains from the model circuit we simply combined spike trains of neurons situated on nearby grid positions within the circuit. In detail, at a random "electrode" position with the circuit (uniformly drawn from all neuron positions), spike trains from n neurons were combined to yield an pseudo-electrode signal. Exactly n unique model neurons were drawn from a Gaussian sphere around the "electrode" position. The

diameter of the sphere is dynamically set to $\sigma = \frac{(4n)^{\frac{2}{3}}}{2\pi\beta_z^{\frac{1}{3}}}$, meaning that for higher n units might be combined from further apart and for low n, only neighboring neurons are considered candidate units. β_z is the fraction of the lateral distance to the vertical distance of neurons within the cuboid grid and was set to $\beta_z = 0.4$. This approximate value resulted because our model had higher neuron density in lateral area than in depth (see Section 5.2.2.1). The number of neurons n is drawn from a Poisson distribution of mean λ_n. The parameter λ_n was a free parameter to be fitted (see Section 5.3.1).

5.3 Results

5.3.1 Fitting the model response to experimentally measured data

We developed a V1 model based on many anatomical details from macaque monkey. As described in the Methods Section 5.2, it consists of inhibitory and excitatory pools of point neurons arranged in 3 different layers with probabilistic rules for connectivity taken from the literature. Here we compared the spike statistics from the V1 model to electrophysiological data recorded from from V1 and LGN of anesthetized macaque monkeys. In experiments, 5 minutes lasting commercial color movies were repetitively presented and responses were recorded with extracellular multi-electrode arrays. Movie frames were centered at the fovea and covered an area of $10^\circ \times 7^\circ$ degree visual field.

In simulations, it was taken care that the stimulus presented to the V1 model resembled that presented to the monkeys as much as possible. However, modeling the whole $10^\circ \times 7^\circ$ degree visual field was not feasible because of computational speed. Therefore we trimmed the movie frames to a smaller size, covering $3 \times 3^\circ$ visual degree. The center of the extracted region was aligned at the center of a receptive field of one of the electrode (channel 7) of a particular session ("d04nm1"). Since the approximate diameter of the receptive field of that electrode was experimentally determined to be 1.2° degree, the reduced stimulus should at least contain all direct input information available for neurons recorded by that electrode. On the retina this receptive field was centered at $(0.69^\circ, -2.39^\circ)$ eccentricities relative to the fovea. In the model we set the eccentricity nevertheless to 5°, since otherwise the lateral extend (and therefore the

amount of neurons in the model) per visual degree would be prohibitively large (compare to Eq. 5.6). At 5° eccentricity a V1 model covering $2.4^\circ \times 2.4^\circ$ visual degree has a lateral extend of $5 \times 5\,\mathrm{mm}^2$ cortical surface and neurons are positioned on a virtual grid of size $62 \times 62 \times 9$, if one assumes a grid spacing of 80 μm laterally. Vertically, the grid spacing corresponds to 200 μm (see MethodSection 5.2.2.1). The visual field covered by the V1 model is somewhat smaller than the stimulus to avoid boundary effects in the input connectivity. For analogous reasons the LGN neurons were set to cover an intermediate area of $2.8^\circ \times 2.8^\circ$ (77×77 grid).

Having set up the input appropriately we simulated 10 seconds of a typical movie segment and fitted four parameters (see Methods Section 5.2.3): Average input synapse strength of both input streams, the average synaptic weight of the weight distribution for the long-range lateral connections in layer 2/3, and the overall synaptic weight which scales all intracortical synaptic weights relative to the input connections. As outlined in Section 5.2.3, we additionally fitted the average number of units comprised in the multi-units spikes trains from the experimental data. This target data contained spike responses measured in 6 sessions (4 anesthetized macaque monkeys) during about 5 minutes of repeated movie stimulation[1]. As described in the Methods Section 5.2.3, the input synapse strengths were adjusted to the average firing rate increase observed when presenting a movie stimulus relative to spontaneous background. The remaining parameters were varied to yield a close match to spike statistics of the electrophysiological data during a semi-natural movie stimulus. In particular, we fitted the distribution of 2-ISIs, i.e. the distribution of sequential inter-spike interval pairs (see Methods Section 5.2.3). Note that the ISI-distribution is implicitly fitted when using 2-ISI distributions in the optimization.

Figure 5.2 displays the discrepancy of the 2-ISI distribution of models with varied parameters to the distribution derived from the target data. Figure 5.2 A reports that the 2-ISI distribution cannot be matched to a satisfactory degree when one does not extract multi-unit activity from the model ($\lambda_n = 0$, see Methods Section 5.2.3). The resulting "best" ISI-distribution is bimodal not unimodal (Fig. 5.2 B): The network strongly oscillates, when using this "best" parameters. Extending the bounds of the

[1]Note that this target data set contained spike responses to regions of the movie not shown in the simulation, to avoid overfitting the neural dynamics to the particular section of the movie shown to the model.

parameter variation did not change this result (not shown). However, if nearby neurons of the model are pooled to form a single spike train as it is likely to be the case for the experimental data, the fitness landscape resembles a valley with high borders, which has a well defined minimum in the bounds of the parameter variations. Figure 5.2 C shows the result for $\lambda_n = 3$, where the MMD is minimally above its significance threshold (λ_n was varied from 0 to 10). This means that on average 3 neurons are pooled together to form multi-unit spike trains and suggests that joining spike trains of the model to form pseudo multi-unit spike trains is essential for our data.

At the optimal parameters settings the patchy long-range weight scale has to be reduced by a factor of about 2.5 in respect to the average weight reported by Thomson & Bannister (2003) for more local layer 2/3 recurrent connections (having scale 1 in Fig. 5.2 B). The fitted overall weight scale is 110. Although these parameter values show greatest similarity of the 2-ISI distribution and the ISI-distribution matches the target data reasonable well (Fig. 5.2 D), both distributions are not identical. For instance, one notices that longer spike intervals (> 50 ms) are underrepresented (Fig. 5.2 D). Indeed, even for the best fit the MMD statistics is well above threshold (about 50% higher), indicating that 2-ISI distributions derived from the model and the experiments are still significantly different. Choosing different kernel widths or using linear spike intervals did not change this fact considerable. We conjecture that for an improved match of the distributions other parameters, such as individual synaptic weights of neurons, have to be tuned as well. Recall that these parameter are drawn from random distributions and are not constraint by the available experimental data (see Discussion).

Figure 5.3 shows stimulus and response of the network with optimized parameters (Fig. 5.2 B) for a period of 1 second. Network dynamic was rich and laterally structured. Burst-like activity varied in strength and in the amount of participating neurons. The high activity bursts were stimulus induced. Note that the dense spiking activity at second 6.1 in the thalamic input (Fig. 5.3 A) was followed by a strong burst in all layers (Fig. 5.3 C-E). Especially in layer 5 (Fig. 5.3 E), the onset of a burst was laterally displaced, suggesting a traveling wave like activity.

The response pattern is different for individual layers. Layer 4 receives strongest thalamic input and is almost free of bursts. The activity of layer 5 is somewhat reduced in comparison to that in layer 2/3. Bursts are small and short and spatially restricted. In

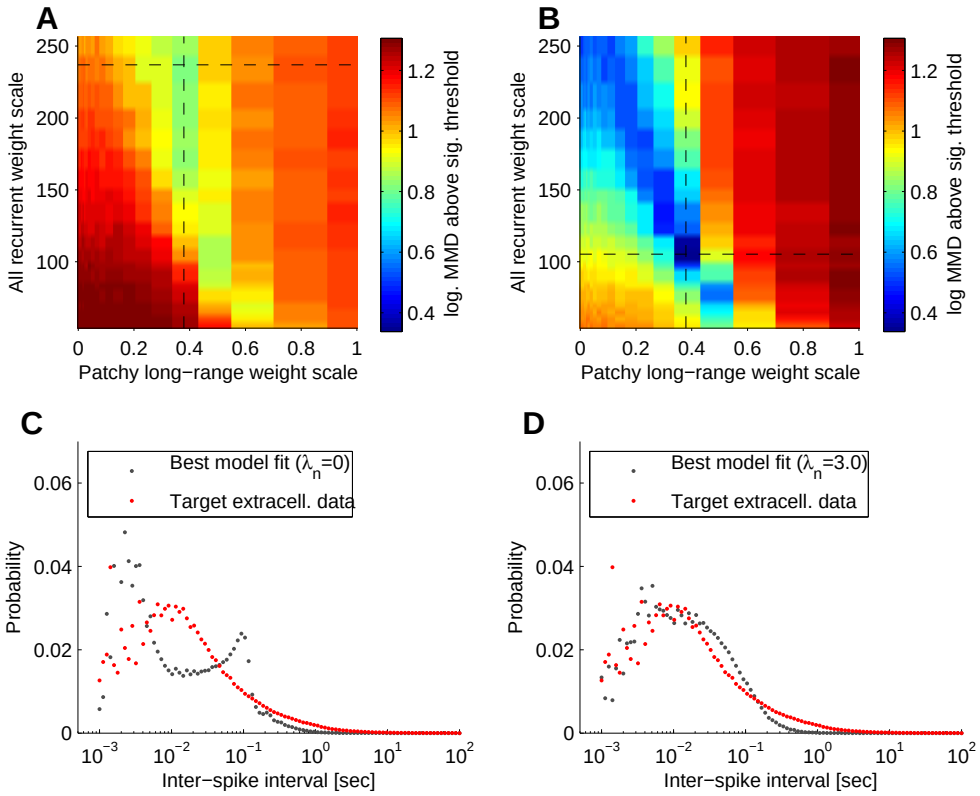


Figure 5.2: Model parameter fit to the 2-ISI distribution of extracellular recorded spikes from V1. **A** and **B**: Optimization criterion Eq. 5.9 for the variation of synaptic weight scale parameters (patchy-long range connections and overall recurrent weight scale). MMD is displayed in color code, in logarithmic units above of the significant threshold (thus a negative value would indicate that the samples derived from model and real data can be assumed to originate from the same underlying 2-ISI distribution with an error of $\alpha = 0.05$). Plots **A** and **C** show the fitness landscape and the resulted ISI-distribution for the optimal parameters when using single neuron spikes trains of the model. Plots **B** and **D** shows the same plots, when one generates pseudo-electrode spike trains and includes the mean number of neurons λ_n pooled as a parameter to be optimized (optimal is here $\lambda_n = 3$, varied from 0 to 10). The best fits are indicated by lines in the upper plots. Plot **C** and **D** show the ISI-distribution of the best fit. Note that the ISI-distribution (which is only implicitly fitted) matches much better the real data when using pseudo electrodes. However, the 2-ISI-distribution is still significantly different from the real data (MMD statistics about 50% above significance threshold). Pseudo electrode position were randomly chosen from all layers.

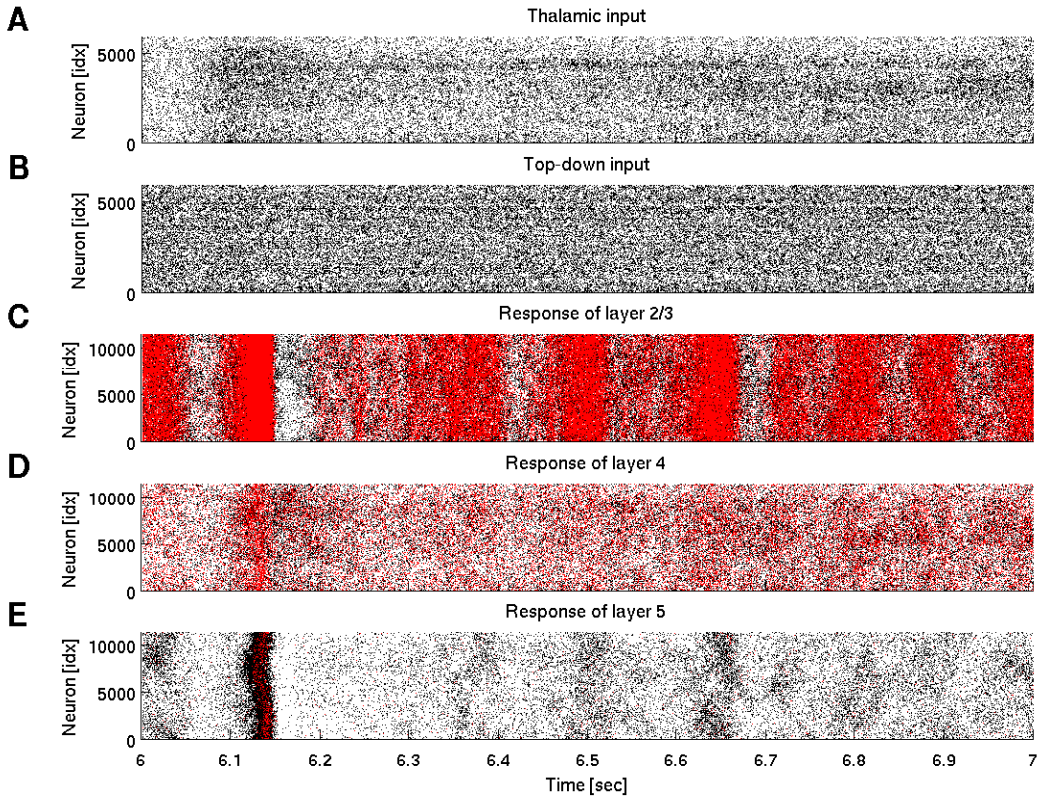


Figure 5.3: Network activity for optimized parameters (see Fig. 5.2 B and D). Plot **A** shows the thalamic input (with spikes of on and off, lagged and non-lagged cells pooled in single spike trains). Neurons on the cuboid grid are ordered linearly for plotting purposes (first vertical axis, then both horizontal axes). That means that highest and lowest index correspond to diametrical opposed corners of the cuboid grid. Due to toroidal boundary conditions these points will most likely be connected, however (Method Section 5.2.2.1). **B**: Top-down input which resembles firing rate distributions (but not in respect to channels correlations and higher order spike statistics) of experimental data from V1. **C**-**E**: Spike trains of all neuron in layer 2/3, 4, and 5, respectively. Inhibitory neurons are plotted in red, excitatory neurons in black. Note that due to the enormous amount of neurons the overall circuit activity appears exaggerated. The different lateral connectivity structure causes characteristic activity patterns in each layer.

contrast, layer 2/3 exhibits higher activity and larger bursts, which often extends over the whole surface of the model. These structural different dynamics are caused by the more numerous lateral connections introduced by the long-range patchy projections within layer 2/3. As the circuit width is 5 mm, long-range projections extend almost across the whole circuit, inducing spatially higher correlated firing pattern. However, because of the preferred connectivity of iso-oriented neurons (compare to Fig. 5.1), activity is likely to be correlated in orientation domains.

5.3.2 Comparison of simulated and experimentally measured spiking statistics

Having optimized the 2-ISI distribution to the experimental data in two general parameters (average weight scales), we compared the resulting firing statistics to that of experimental data in detail. Although our fitting process found parameters where the firing 2-ISI-distributions are similar, the best fit nevertheless exhibited a discrepancy between distributions. Thus firing statistics still be different in the model. Comparing model and experiment using other measures of spiking statistics helps to describe the deviations in more detail, and quantitatively shows which aspects are not captured by the model.

For characterizing the firing statistics in the model we simulated a model repeatedly (60 times) during a typical 20 seconds extract of the movie input using the optimized parameter values from the last section. 1000 pseudo-electrodes were positioned uniformly in the circuit and multi-units spikes trains generated (with optimal $\lambda_n = 3$). These spike trains were compared to all available data from V1. Additionally, the thalamic input stream during movie activity was compared to all available data recorded from dLGN. Thalamic input spike trains were not pooled with neighboring neurons, but the four LGN cell types (on, off, lagged, non-lagged) were combined to form a single spike train.

The results are summarized in Fig. 5.4 and Fig. 5.5. Panel A of Fig. 5.4 displays the ISI-distribution on a logarithmic scale. Simulation and experiment agree quite well, at least for short ISIs; remember that this distribution was implicitly optimized by the parameter fit. As already stated above, the V1 simulation lacked longer spike intervals relative to the experiment. In LGN, simulation and experiments disagree: probability

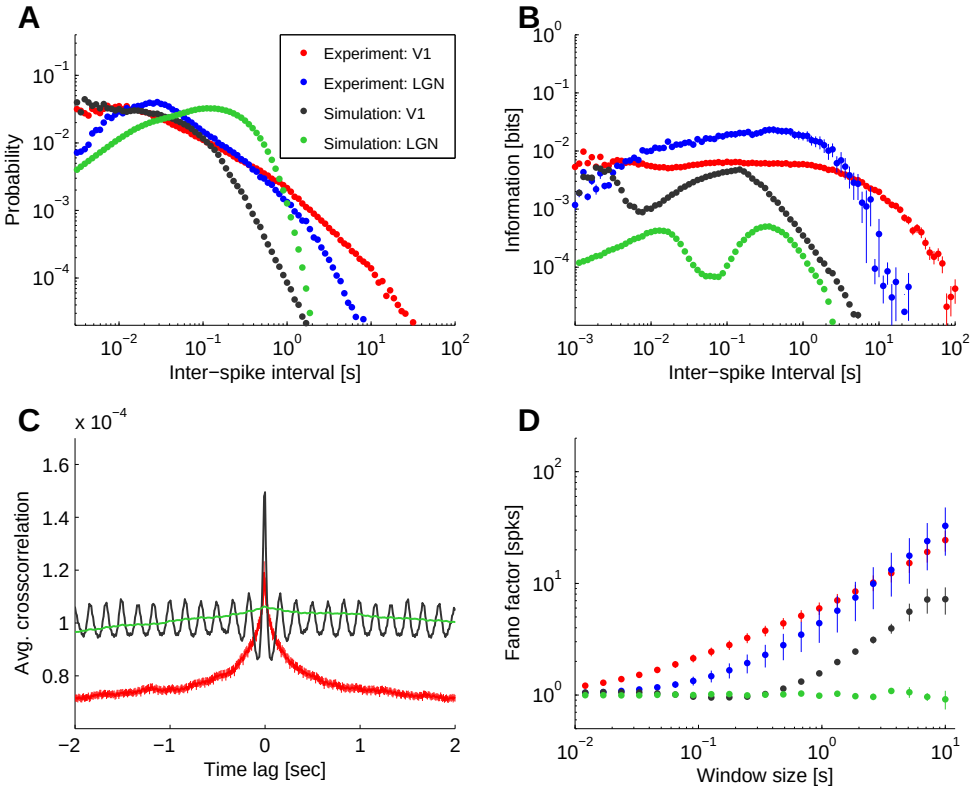


Figure 5.4: Comparison of spike train statistics of electrophysiological recordings from V1 (red) and LGN (blue) of anesthetized macaque to that of the V1 model (gray) and its thalamic input (green). From the V1 model response pseudo-electrode spike trains are generated before comparison ($\lambda_n = 3$). **A**: Inter-spike interval (ISI) distribution estimated on all available data. **B**: Contributions of a given ISI to the dependence of the following ISI (function $J(y)$ of Eq. 5.10)). **C**: Average cross-correlation for simultaneously recorded electrodes (or pseudo-electrode positions in the model). Cross-correlations are expressed in the probability of occurrence of individual spike pairs with fixed lags divided by the mean firing rates of both electrodes. LGN data is omitted because the set of simultaneous recorded electrodes was too small. **D**: Fano factor of spike counts, i.e. the ratio of variance to mean of the spike count within a fixed window, is plotted versus window length. Mean and variance of the spike count are estimated across trials on non-overlapping windows. See Section 5.3.2 for a detailed discussion of the results.

of long ISIs is generally too low in the model. In both areas, LGN and V1, the shape

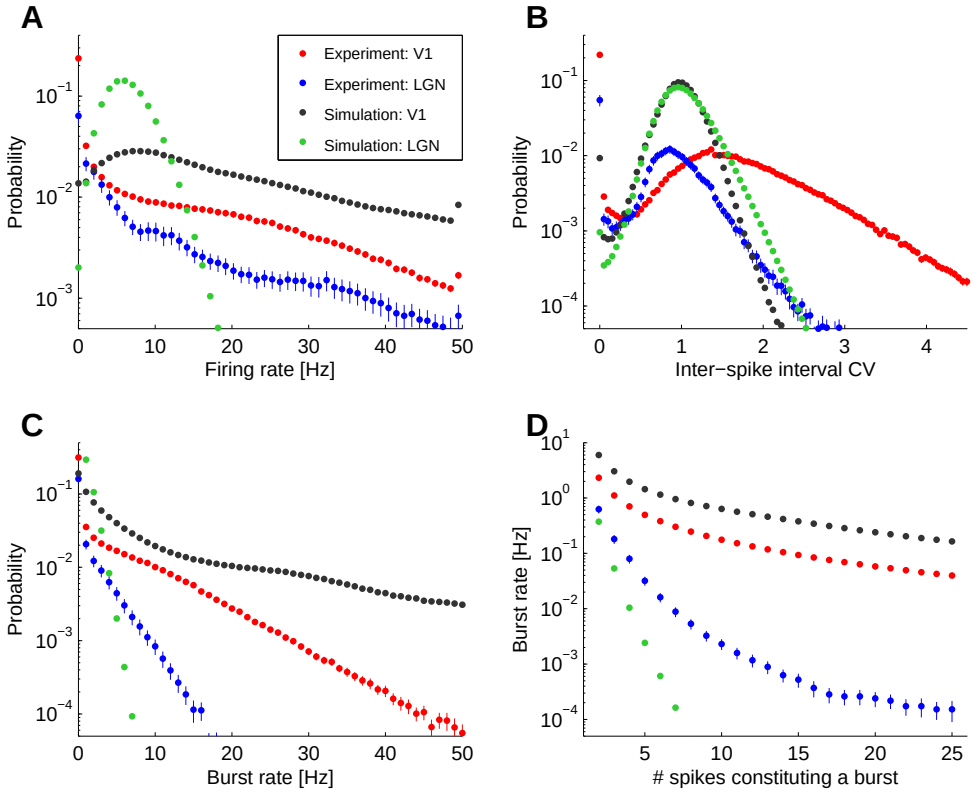


Figure 5.5: Comparison of burst characteristics of experimental data and model. Same data as in Fig. 5.4. Histograms are estimated on overlapping, 2 second lasting time regions and pooled over all available data. **A**: Firing rate distribution. **B**: Inter-spike interval coefficient of variation (ISI-CV) distribution. ISI-CV is defined as the ratio of standard deviation to mean of the ISI-distribution (on each time region). **C**: Burst rate distribution. Burst events are defined as having at least 2 spikes with average ISI of at most 5 ms. **D**: Mean rates of bursts containing at least n spikes with maximal average ISI of 5 ms. See Section 5.3.2 for discussion of the results.

of the distributions from model and experiments agree, however, having a single peak and a monotonous decrease in case of LGN and V1, respectively.

Panel B of Fig. 5.4 summarizes the dependence of one ISI on the following ISI. For homogeneous Poisson spike trains, where spike position are uniformly distributed in time, the length of a given ISI will no tell anything about the length of the following ISI. In neural spike trains, however, this will commonly be the case. The current and

following ISIs can be seen as two random variable X and Y, respectively. A measure for their dependence is the mutual information $I(X,Y) = \sum_x \sum_y p(x,y) \log_2 \frac{p(x,y)}{p(x)p(y)}$ (see Chapter 3 for references). Then the function

$$J(y) = p(y) \sum_x p(x|y) \log_2 \frac{p(x|y)}{p(x)} \tag{5.10}$$

measures the contribution of each ISI y to the overall dependence, since it is $I(X,Y) = \sum_y J(y)$. In other words, if $J(y)$ is high for a particular ISI y this means that the next ISI is well determined by the length of y; one gains information about x. Note that in our case the 2-ISI distributions are symmetric (if one neglects boundary effects), because they are estimated using all possible ISIs of a spike train. We see from Fig. 5.4 B that for experimental data the dependency is quite strong. In V1 in particular, the contribution of ISI to the information is rather constant with ISI duration. The model spike trains seem more random, as the two subsequent ISIs depend less on each other, and in case of the V1 model the dependence is only close to the experiment at the two peaks. These peaks probably arise in the model because of the relatively stereotyped burst length and its slight oscillatory behavior (see below): within a burst it is likely that a short interval is followed by a short one (or a very long one if the bursts happens to end), and a long interval might be preferable followed by a short one within the next burst. In the LGN model the dependence of following ISIs is generally very low, which is to be expected because of the Poissonian rate process generating the spike trains (at least for lower rates). Here the bimodality might arise on the left side due to the regularity inducing gamma renewal process for high firing rates (see Methods Section 5.2.2.2) and for long interval due to the movie content changes. In the experimental data, on the other hand, the relatively constant and high dependence of inter-spike intervals might be caused by slow changing state variables: it is likely that for a given interval duration a similar interval follows.

In Fig. 5.4 C the average cross-correlation between different electrodes is plotted. While electrodes in V1 are correlated for lags up to about 250 ms, the model shows a faster fall-off and a small oscillatory compound (compare also to the network activity in Fig. 5.3 C). The relatively small lateral extend of the model might promote the oscillatory behavior. Similarly, the trial-to-trail variability of model does not reach the variability of the experimental data, although for larger window size the Fano factor at

least increases (Fig. 5.4 D). The LGN model does not match the LGN data: its trial-to-trial variability is not too low for longer times. On the other hand, in the experimental data the variability might be exaggerated because of experimental artifacts. For instance, the electrode might move slightly during a recording session of several hours and thus pick up a different set of neurons in later trials.

In Fig. 5.5 the bursting behavior of experiment and model spike trains are compared. All histograms of Fig. 5.5 are calculated on (overlapping) 2 second regions and pooled over all available electrodes. The firing rate distribution (Fig. 5.5 A) is markedly different for the thalamic input and the LGN. Although the mean firing rate of the LGN model is fitted to the experiments (7 Hz, see Section 5.2.2.2), it is much more likely that there occurs no spike in a random 2 second region in the experimental data than in model spike trains. This is partly an effect of pooling all LGN cell types in the LGN model.

In case of V1, the slope of the tail of the firing distribution is matched well (Fig. 5.5 A), but the probability of low spike rates is too low, similarly to the LGN model. This low probability of long pauses might partly induced by the top-down input model. Because the top-down input is not modulated by the stimulus (which obviously is the case in the brain), it merely acts like a stimulus independent noise source. We set the strength of the top-down input to achieve 2 Hz average firing rate (see Method Section 5.2.3). While this value is realistic as it is estimated from the data, the fact that the rate of each input neuron were *independently* drawn is not. However, lacking precise data about the top-down input correlations we can only state this fact.

Burst rates are generally higher in V1 than in LGN data, which is reproduced in the model (Fig. 5.5 C). However, burst rates are actually higher in the V1 model than in the experiment (see Discussion). The dependence of the rate of the size of bursts is matched qualitatively well for the V1 model (Fig. 5.5 D), whereas long bursts do not occur in our LGN model, although there is some (low) probability that they occur in reality (Fig. 5.5 D). Similarly, high frequency bursts are underrepresented in the LGN model. This can also be seen in Fig. 5.5 C and D where the rate of shorter and longer bursts are compared. The deviations of the LGN model to the data suggests that the simplistic rate-based input model of the LGN is not enough to capture non-Poissonian effects present already in the LGN spike trains.

Figure 5.5 B shows that the high variability of spike intervals in V1 is not captured in the spikes trains from the model. The ISI coefficient of variation (ISI-CV) is higher in reality. For LGN data, however, ISI-CV is lower and peaks near 1, indicating that spike intervals are more regular and more like a Poissonian spike train than in V1. In that respect the LGN data is similar to the LGN model.

In summary, we find that the spike characteristics of the generic V1 and LGN model has similarity to the experimental data but still differs from it in some aspects. In particular, the variability is lower and the changes in firing dynamics over intermediate times is less extreme: the experimental data exhibits both long silent periods and prolonged bursting events. However, besides the differences, there is a general trend that the model statistics resemble the experimental data. It seems that the more sophisticated V1 model is closer to reality than the much simpler rate-based LGN model. Generally, data derived from LGN and V1 is more similar to its simulated counterpart, than experimental data and simulated data to themselves.

5.3.3 Does the model predict the response of experimental data to a movie stimulus?

A model having qualitatively similar spike statistics to experimental data, does not necessarily respond similarly to given stimuli. We wondered to what degree the model predicts the spike trains measured at an electrode in vivo? Figure 5.6 shows the response to repeated trials of a section of a movie stimulus of both an electrode channel and neurons of the model. As described above the movie extract was chosen to include the receptive field of an electrode from one experimental session (channel 7 of "d04nm1"). The center of the circuit is aligned to the center of the receptive field of that electrode. Thus if the model were a reasonable abstraction of the reality, one would assume that firing patterns at a pseudo-electrode in the center of the circuit would be very similar to those observed at this particular electrode channel.

In Fig. 5.6 multiple trials of movie presentation are shown for both the electrode and an pseudo-electrode (with $\lambda_n = 3$) inserted in layer 2/3 of the model (Fig. 5.6 B and Fig. 5.6 D, respectively). The layer position of the electrode is not determined experimentally and electrode tip distance from the dura is not necessarily a reliable indicator (because of the geometry of the cortex). However, the chance that channel 7

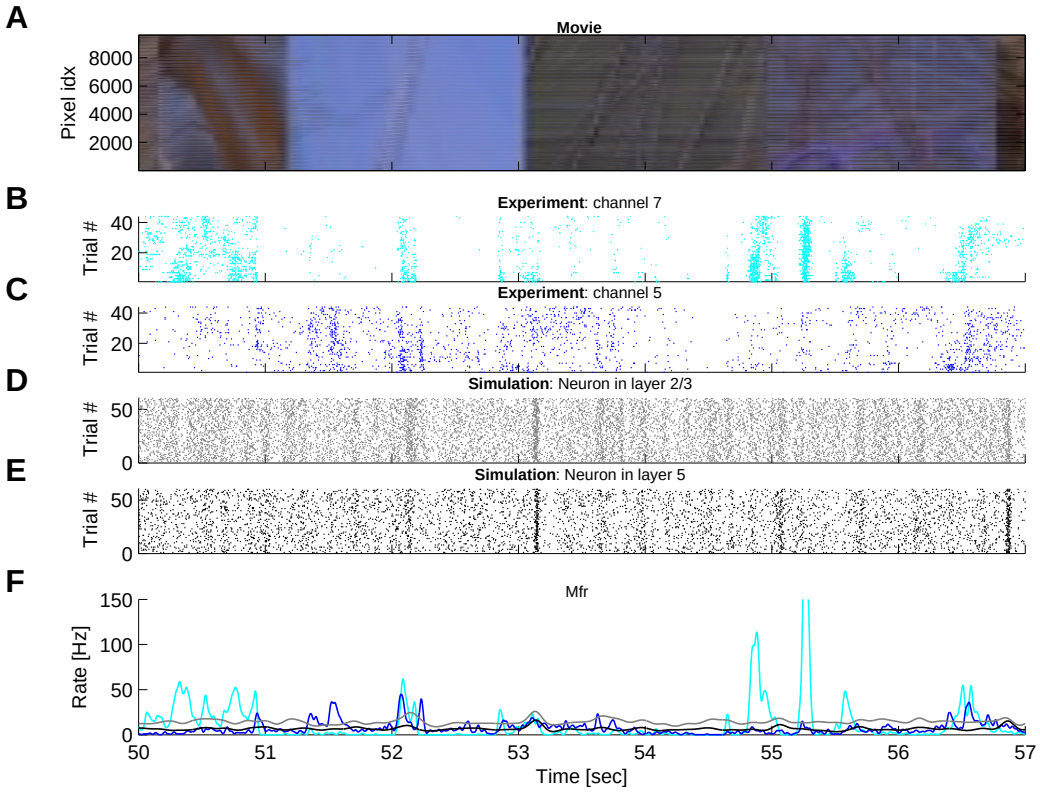


Figure 5.6: Spike trains elicited in response to multiple trials of movie presentations for two experimental electrodes and two pseudo electrodes in the circuit. Movie stimulus the same in experiment and simulation. In simulations a black and white version of the color movie was presented. **A** Movie stimulus in original colors. Movie frames are shown on a 1-dimensional line to visualize the time course (order: pixel columns first, then rows). Plot **B** and **C** show recorded spike trains of two electrodes. The center of the receptive field of electrode in B (channel 7) is used as a reference point to align the center of the circuit to the center of the movie region. The electrode shown in C (channel 5) has overlapping receptive fields with slightly shifted center and is positioned about 1.5 mm further away. Plot **D** and **E** show trials of two pseudo-electrodes generated from the model circuit ($\lambda_n = 3$). They are positioned in the center of the network in layer 2/3 and offset laterally by (simulated) 1.5 mm in layer 5, respectively. Thus their positions correspond to the experimental electrodes in Panel B and Panel C, respectively. Plot **F** shows the mean firing rates of the electrodes in plots B–E using the same color code (smoothed with a 50 ms Gaussian kernel). Note that bursts occur reliably across trials in both, simulation and experiment. Trial-to-trial variability is realistic in simulations. However, correlations of the mean firing rates is not perfect. Nevertheless, at some times the responses of the simulation resembles the experimental observations (e.g. at 52 sec).

recorded in superficial layers is higher since its depth is about 1 mm smaller than a second electrode (channel 5), which is plotted in Fig. 5.6 C. Thus the second electrode is more likely to be positioned in a deeper layer. For comparison we show a pseudo-electrode positioned in layer 5 of the model in Fig. 5.6 E, which also is shifted by 1.5 mm lateral distance as it is the case for the second electrode in the experimental session. Panel A of Fig. 5.6 shows the movie extract which is presented to both model and monkey. In the experiment the movie was shown in color (as depicted), whereas in the model a black and white version was used because only one type of ganglion cells was modeled. The two dimensions of the movie is lined up to one dimension to visualize its temporal evolution.

One can see that the trials of both experiments and simulations show reliable activity to certain parts of the movie. For instance, at second 52, when an object flies through the blue sky in the movie, the response of both model and simulation is a reliable burst at very similar times. On the other hand some prominent and very reliable bursts in the experimental data are not anticipated by the model. At second 55.3 channel 7 fires strongly (Fig. 5.6 B), but in the model no burst occurs in either spike train (Fig. 5.6 D and E).

The mean firing rate of the two experimental electrodes and the pseudo-electrodes from the model are displayed in Fig. 5.6 F. The simulated traces matches at some times the increase seen in one of the electrodes, but in general the correlation is weak. In Fig. 5.7 B the distribution of correlation coefficient between 1000 randomly placed pseudo-electrodes within the circuit and both electrodes of Fig. 5.6 B and C are shown (calculated over 20 seconds stimulus and 200 ms windows). As said, there exist a correlation for most locations in the circuit, but it is relatively low: highest correlation coefficients lie around 0.3.

The trial-to-trial variability is similar in the simulation and the model at least for this movie section. This can be seen when comparing the mean spike counts with the variance over trials (Fig. 5.7 A). The ratio of both values is close to one (for 50 ms windows). Note that we show here only selected electrodes for 20 seconds stimulus duration, whereas all available data was analyzed in Fig. 5.4 D. The overall appearance of spike trials generated by the model is a bit noisier, with less silence in between activity phases (as noted above). Moreover, bursts in the experiment appear to be slightly longer in duration. In the experiments there is a slight drift over trials (trials

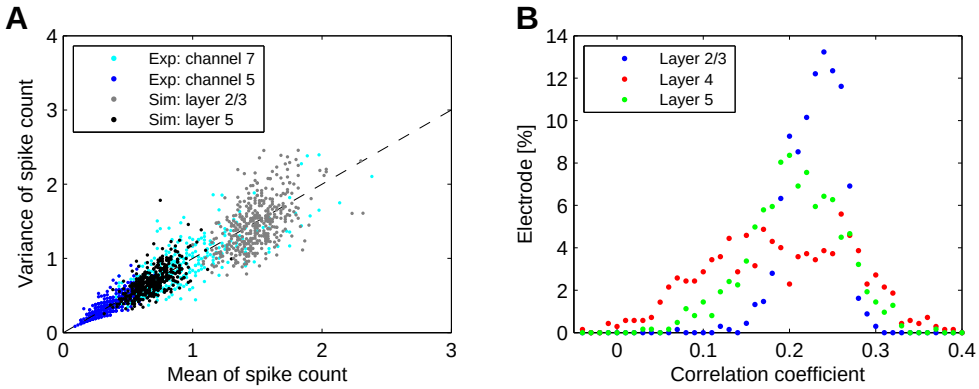


Figure 5.7: Trial-to-trial variability and correlations of responses to movie trials in experiment and simulations (compare to Fig. 5.6). **A**: Mean of spike count across trials on 50 ms windows plotted against variance. Electrodes are identical to that displayed in Fig. 5.6 B-E (but for 20 seconds stimulus duration). **B**: Histogram of correlation coefficients of the two experimental electrodes with 1000 randomly placed pseudo-electrodes from the V1 model, separately plotted for layer position of the pseudo-electrode. Correlation is weak but present in all layers. The mean firing rate was estimated on non-overlapping windows of 200 ms duration.

are number in the order of their recording time), which is probably due to experimental conditions.

In summary, we have applied exemplary a shown that adjusting the network parameters to the 2-ISI distribution of the real data results in firing statistics which resemble the experimentally measured data. However, not all details of the real statistic are captured by the model. This knowledge can be used to systematically improve aspects of the model to achieve a closer match to the data. In a direct comparison of spike trains to the same movie stimulus the correlations in the mean rate remain low. This suggests that the response of the real system is still more complex than our state-of-the art connectionists model.

5.4 Discussion

In this Chapter developed a detailed computer model of several square millimeters area of macaque V1. We have compared it to spike train statistics of electrophysiological

recordings from area V1 and LGN in anesthetized macaque monkey.

The last years have seen several attempts to model large areas of the brain with similar components that we used here, such as inter-laminar connectivity, laterally structured connectivity, realistic input connections, synaptic depression and facilitation, and neurons having one or a few compartments (e.g. Izhikevich & Edelman (2008); Johansson & Lansner (2007); Kremkow *et al.* (2007); Tao *et al.* (2004)). The anatomical detail, e.g. the inclusion of orientation dependent long-range patchy connections, and lateral extend of our V1 model matches recent approaches very well. These models, as well as ours, incorporate many anatomical and physiological details, but they are, of course, still a strong abstraction of reality. For instance, genetic and metabolic maintenance, the electrochemically milieu, blood vessels, or cell signaling systems etc. are completely ignored – just to name a few aspects. However, the computational function of neural systems is thought to arise mainly through communication between neurons using spikes, although other signals might be involved as well (see Chapter 3). Thus it is feasible to try to capture the neural spike train statistics with a model concentrating on neurons and their interactions.

Having a surface area of 25mm^2 the network size of our model is still very small compared to an average of 1200mm^2 for striate cortex of macaque (Essen *et al.* (1984)). In relation to the extend of the long-range connections (SD of 1mm in each lateral direction), our model is of minimal size in that sense that lateral connections are not likely to connect to neuron near its origin via the toroidal boundaries. In principle, thanks to the PCSIM simulation environment[1] used throughout, the network size is easily scalable in the expense of simulation time. Simulating a circuit of 100mm^2 (having a neuron grid of $166 \times 166 \times 9$ at lower eccentricities) for a few seconds, for example, required about two days simulation time on a 16 CPU cores machine, as compared to roughly 6 hours simulation time of the presented model on a dual core machine (for 20 seconds simulation time). Since we simulated the model repetitively for long times (20 seconds and more), we have restricted this study to a relatively small circuit.

With about 35000 neurons and over 4 million synapses our model has still many parameters, which in principle had to be estimated from data. With the limited amount of data available to us, this is impossible. It would take a great amount of collaborative

[1] available from `http://sourceforge.net/projects/pcsim/`

effort to try to fit parameters on the basis of single neurons; although recently the Blue Brain Project started to do just that (Markram (2006)). Thus it is very difficult to constraint such model. In this Chapter we develop a method for adjusting network parameters to achieve realistic firing dynamics in response to a semi-natural movie stimulus.

This method could be used to systematically investigate components of the networks important for achieving certain aspects of realistic dynamics. We exemplified the fitting process in three parameters and achieved a reasonable fit to the ISI-distribution. One could choose different parameters to investigate the importance of components of the network. As optimization algorithm we performed simple grid search, because it is very efficient on a multi-processor system. If more parameters were to be adjusted any gradient descent algorithm, such as adaptive simulated annealing (Ingber (1989)), could readily be applied.

One also could change the optimization criterion for further investigations. For instance, we here exemplary optimized our network model in a few free parameters to the 2-ISI distribution. One could likewise fit more complicated distributions, such as the k-ISI distributions for higher k, or e.g. the distribution of waiting times to the next spikes seen from randomly sampled times t for n neurons simultaneously. The latter distribution could be used to fit correlations between electrodes (and neurons) of multi-electrode recordings. Moreover, the MMD could in principle be used to fit distributions of whole spike trains directly using a kernel for comparing spikes trains (e.g. Shpigelman *et al.* (2003)). In our case, where the model has an enormous degree of freedom, a direct match of spike trains seems unlikely, however.

In comparing the resulting model with statistics of the experimental data, we found that even after having optimizing parameters, the model does not reproduce all subtle aspects of spike trains experimentally observed. In particular, the bursting characteristics of model and experiment are different in that bursts occur more likely and more regularly in the model (Fig. 5.5). Although there might be even an increase in bursting due to anesthesia in the data (see Chapter 2 for a discussion), the mismatch of experiment and simulation is probably due to approximations done within the model.

Recall that the fit included the estimation of the number of neurons pooled to multi-unit spike trains. If the bursting behavior of the neuron model would be devoid of short bursts, this number will be artificially increased to match the ISI-distribution of the

experimental data. We found a mean number of $\lambda_n = 3$ which might be correct, but intuitively seems to be a bit too high, given that we increased the threshold for spike detection to a very high 7 SD of the noise component. Thus the lack of longer inter-spike intervals observed in the simulations might be partly an effect of pooling more spike trains in need of short ISIs to match the target. Another reason for the lack of longer ISIs might the possibly unrealistic noise source from the top-down projection. However, simulating the circuit without top-down input resulted in a less severe, but still similar underrepresentation of longer inter-spike intervals (not shown).

One possibility for the deviation of model and experiment is that the simplified neuron model might misses aspects of the dynamics of real neurons. Although the neuron model can in principle capture a wide variety of firing characteristics (Izhikevich (2006)), and is thus unlikely to fail completely, the mixture of different firing types (e.g. the ratio of regular spiking to intrinsically bursting cells) is generic and might not correspond specifically to V1 of macaques. Therefore in future studies one could attempt to adjust the mixture of firing types by including additional parameters in the ISI-fitting process, which describe the distribution of firing types present in the circuit. However, for this task it would be advantageous to compare spike trains to multi-dimensional polytrode recordings, which allow to sort multi-unit activity to obtain single units spike trains. With single unit spike trains the fitting process would be greatly improved because no pseudo-electrode signals had to be estimated. Additionally, the risk of introducing a sample bias towards very big neurons via the high spike detection threshold would be minimized (Olshausen & Field (2005)).

Moreover, the neuron model lacks complicated dendrite morphology seen in vivo, which are thought to actively and non-linearly shape the response of single neurons (see Yuste & Tank (1996) for a review). Instead the model uses point neurons. In the simulation we saw that the model did not seem capable of reproducing longer-lasting changes in the firing dynamics, e.g. that a neuron is silent for seconds and then bursts reliably (see Fig. 5.6 A for an example). This behavior results in the high ISI-CV in the experiment (Fig. 5.5 B). Dynamics on intermediate time scales might be induced in the cortex by second messengers such as calcium ions. Calcium dynamics is known to be complex, showing waves and spikes, and has pronounced effects on the firing of neurons (see Berridge (1998) for a review). In our model no explicit calcium dynamics, nor that of any other second messenger, are included. Furthermore, since we based the

synapse model on that used by Haeusler & Maass (2007), no slow synaptic currents are included in the model, such as $GABA_B$ and NMDA (Dayan & Abbott (2001)). We expect that a model containing such slow synaptic dynamics will achieve a closer match to the experimentally observed spike train statistics. Analyzing the effects of slow dynamics is a promising direction for future studies.

Besides many anatomical details included in the model, we also neglected some known aspects. For instance, we incorporate only 3 instead of 6 layers and neglected several sublaminar structures existing in macaque V1 (Callaway (1998)). Since connection probabilities are not measured in such detail in macaque as it is done for cat, we instead used data obtained from cat slices of different cortices (Thomson & Bannister (2003)). Presently, there is only one other study describing layer specific connection probability in cat (Binzegger *et al.* (2004)). Although the latter data set is based on the visual cortex, it is derived solely on anatomical consideration, whereas the former data set incorporates functional synaptic weight estimates as well. Since both data sets are from cat and therefore can only serve as a rough approximation to macaque V1, we decided to apply the former because of their advantage having synaptic strength estimated and thus stick to the three layer layout used by Haeusler & Maass (2007). We are aware that particularities of macaque V1 such as the first feedback loop from layer 4C to layer 6 and vice versa suggested by Callaway (1998) is not included. Similarly, the probability distribution for patchy long-range connections in superficial layers, as well as the hard-wired orientation tuning are inspired from cat. Thus, using connection probabilities derived from macaque V1 would most likely lessen the gap of the correspondence of experiment and simulation.

We found that in a direct comparison to experimental data during movie presentation the predictive power of the V1 model was low. Such a direct comparison of a network model to data during complex stimulation has not been attempted so far. The low correlation is to be expected since the synaptic weights are only randomly chosen and unconstrained by the experiment. In future studies one could for example learn the weights to a readout neurons to achieve a high correlation to the experimentally measure neurons.

There are also several other caveats, which explain the relatively low correlation. Despite the approximations and abstraction within the V1 model described above, a main reason might be simplifications in the input model. We found that the statistics

of our rate-based LGN does only very poorly resemble the experimentally data from LGN. Thus, the LGN model might have to be improved to achieve a closer match of the model spike trains to that of experimentally recorded spikes. Besides the LGN model the stimulus was adjusted, too. Since we had to adjust the stimulus size (see Results Section 5.3.1), only about a sixth of the original movie frame area was presented to the model. Even though the reference electrode had a receptive field in the center of this region, neurons might interact with neurons beyond their receptive field. Neither these interacting neurons nor their input were included in the model, because of the size restrictions. Moreover, the color movie had to be transformed to black and white (by simply averaging all color channels). More appropriate would be to incorporate color pathways into the model by including multiple types of color opponent ganglion cells (Sincich & Horton (2005)). In principle, this could be done straightforwardly, although many details about color processing in the visual cortex are unknown (Sincich & Horton (2005)). However, neglecting color information might have a strong effect on the response, because many colors share identical gray scales.

5.5 Conclusion

We have presented a detailed circuit of V1 and its thalamic input stream. Often these models are difficult to adjust to generate reasonable firing dynamics, so that they could be used as basis for further investigations. The method developed and applied here is very helpful in adjusting free parameters to the statistics of experimental measured spikes train. We are confident that the gap in understanding neural systems in terms of neural network models (Olshausen & Field (2005)) will lessen by developing models in close relation to experimental data.

Chapter 6

Conclusion

This dissertation focused on a particular dataset of electrophysiological recordings from macaque monkeys. Machine learning techniques were developed and applied in order to characterize the interdependence between different signals and their role in information processing. Two signals, commonly recorded with electrophysiological electrodes, were analyzed: spiking activity and local field potentials (LFP).

In Chapter 2 their interdependence were analyzed in details while attempting to systematically infer spikes from the local field potentials. This approach did not only visualize concrete spike trains estimated solely on the basis of LFP, it was also possible to pin down the frequency components and oscillation phase features which were most important carriers of the information about spikes, and to investigate the redundancy or independence of individual features. Several findings were quite unexpected. For instance, the relatively high information in low frequency bands, especially at times after spike position. Furthermore, the prediction accuracy as well as the important features did not differ much in case of spontaneous activity or stimulus evoked activity. Both conditions exhibited different spatial decay constant of the prediction performance, however, indicating that during stimulus, neurons are spatially more decorrelated.

Although it was previously known that e.g. γ-power correlates best with spiking activity, the relative contribution and redundancies of different features were quantified. Supervised learning method (support vector machines) are capable of using many features simultaneously to non-linearly predict the target outcome. This method has not been applied to investigate the relationship between LFP and spiking activity before.

The results of this dissertation strongly suggest that local field potentials are an important carrier of information and analyzing LFP in addition to spike trains might be elucidate cortical functions.

This is underlined by the findings of Chapter 3 that LFP might play an active role in information processing. Using information theoretic arguments the gain in information about naturalistic movie scenes is quantified to be over 50%. Remarkably, this improvement is achieved not for simplistic, abstract stimuli but for stimuli which resemble moving natural surroundings. Most notable, the information theoretic approach does not need to know what kind of features of the natural scenes is encoded but instead analyzes any kind of possible features. Although phase-of-firing codes has been suggested earlier for the hippocampal formation (for references see Chapter 3),

this is the first time a phase-of-firing code has been shown to exist for the primary visual cortex. In visual cortex the code is subtle and not immediately accessible by eye. Spikes generally occur more likely at certain phases of the LFP, only a small but crucial jitter, depending reliably on the stimulus content, together with the reliability of LFP phases in response to the stimulus, carries the additional information. In particular, this additional information can be used to disambiguate scenes which illicit similar spike rates.

In state-of-the-art network models the extracellular milieu, and thus a signal like LFP, is commonly neglected and all communications and computational properties are thought to emerge through spiking activity. In Chapter 5 such a network model of a patch of the primary visual cortex was developed. The model incorporates many anatomical and physiological details of the macaque monkey. The laminar organization of neurons, lateral connection profiles as well as orientation tuned input connections are modeled according to anatomical studies. The model shows rich and structured dynamics in response to movie stimuli, including reliable activity bursts and realistic trial-to-trial variability. However, it is evident that temporal structure of responses is still richer in experimental data.

For optimizing parameters of the network model a new method for comparing two multidimensional distributions has been developed in Chapter 4 and applied in Chapter 5. Most remarkably this new method, using the Maximum Mean Discrepancy, can be efficiently computed when relying on kernel spaces and therefore inherits their enormous modularity and capabilities. For instance, distributions on graphs can readily be compared using graph kernels. Moreover, the method performs at least similar to the best existing method. In multidimensional problems it commonly outperforms existing methods, both in accuracy and computational speed.

In Chapter 5 is shown that, although tuned in a few parameters, the firing dynamics of such a model does not capture every detail of the dynamics seen *in vivo*. Although many approximations were made when developing the artificial network model, together with the finding that LFP code for additional information beyond that coded by the spike count alone, it suggests that the restriction on modeling only spiking communications might not be sufficient.

Given the discrepancy between experiment and model, the approximation commonly done in even the most detailed network models, one might conclude that much

remains to be done to finally understand the computational function of neural systems.

References

ADORJAN, P., LEVITT, J.B., LUND, J.S. & OBERMAYER, K. (1999). A model for the intracortical origin of orientation preference and tuning in macaque striate cortex. *Vis Neurosci*, **16**, 303–18. 98

ADRIAN, E. (1928). *The Basis of Sensations*. Norton, New York. 68

AJMONE-MARSAN, C. (1965). Electrical activity of the brain: Slow waves and neuronal activity. *Israel J Med Sci*, **1**, 104–117. 9

ANDERSON, N., HALL, P. & TITTERINGTON, D. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *Journal of Multivariate Analysis*, **50**, 41–54. 88, 91

ARCONES, M. & GINÉ, E. (1992). On the bootstrap of u and v statistics. *The Annals of Statistics*, **20**, 655–674. 89

AREZZO, J., LEGATT, A. & VAUGHAN, H.G.J. (1979). Topography and intracranial sources of somatosensory evoked potentials in the monkey. i. early components. *Electroencephalography & Clinical Neurophysiology*, **46**, 155–172. 9

BARTOS, M., VIDA, I. & JONAS, P. (2007). Synaptic mechanisms of synchronized gamma oscillations in inhibitory interneuron networks. *Nature Review Neuroscience*, **8**, 45–56. 42

BARTSCH, A.P. & VAN HEMMEN, J.L. (2001). Combined Hebbian development of geniculocortical and lateral connectivity in a model of primary visual cortex. *Biol Cybern*, **84**, 41–55. 98

REFERENCES

BEAULIEU, C., KISVARDAY, Z., SOMOGYI, P., CYNADER, M. & COWEY, A. (1992). Quantitative distribution of GABA-immunopositive and -immunonegative neurons and synapses in the monkey striate cortex (area 17). *Cereb Cortex*, **2**, 295–309. 101, 102

BELITSKI, A., GRETTON, A., MAGRI, C., MURAYAMA, Y., MONTEMURRO, M.A., LOGOTHETIS, N.K. & PANZERI, S. (2007). Local field potentials and spiking activity in primary visual cortex convey independent information about natural visual stimuli. *to be submitted to Journal of Neuroscience*. 45

BERRIDGE, M.J. (1998). Neuronal calcium signaling. *Neuron*, **21**, 13–26. 134

BIAU, G. & GYORFI, L. (2005). On the asymptotic properties of a nonparametric l_1-test statistic of homogeneity. *IEEE Transactions on Information Theory*, **51**, 3965–3973. 91

BICKEL, P. (1969). A distribution free version of the smirnov two sample test in the p-variate case. *The Annals of Mathematical Statistics*, **40**, 1–23. 90

BINZEGGER, T., DOUGLAS, R.J. & MARTIN, K.A.C. (2004). A quantitative map of the circuit of cat primary visual cortex. *J Neurosci*, **24**, 8441–53. 109, 134

BISHOP, C.M. (2006). *Pattern recognition and machine learning*. Springer. 14, 15

BLAKE, C.L. & MERZ, C.J. (1998). *UCI repository of machine learning databases*. 95

BLASDEL, G.G., LUND, J.S. & FITZPATRICK, D. (1985). Intrinsic connections of macaque striate cortex: axonal projections of cells outside lamina 4C. *J Neurosci*, **5**, 3350–69. 103

BLUMENFELD, B., BIBITCHKOV, D. & TSODYKS, M. (2006). Neural network model of the primary visual cortex: from functional architecture to lateral connectivity and back. *J Comput Neurosci*, **20**, 219–41. 98

BORGWARDT, K.M., ONG, C.S., SCHONAUER, S., VISHWANATHAN, S.V.N., SMOLA, A.J. & KRIEGEL, H.P. (2005). Protein function prediction via graph kernels. *Bioinformatics*, **21**, i47–i56. 90

BREDFELDT, C.E. & RINGACH, D.L. (2002). Dynamics of spatial frequency tuning in macaque V1. *J Neurosci*, **22**, 1976–84. 109

BROCKMANN, D., BAUER, H.U., RIESENHUBER, M. & GEISEL, T. (1997). SOM-model for the development of oriented receptive fields and orientation maps from non-oriented ON-center OFF-center inputs. In W. Gerstner, A. Germond, M. Hasler & J.D. Nicoud, eds., *Artificial Neural Networks—ICANN '97. 7th International Conference Proceedings*, 207–12, Springer-Verlag, Berlin, Germany. 111

BUCHWALD, J.S. & GROVER, F.S. (1970). Amplitudes of background fast activity characteristic of specific brain sites. *Journal of Neurophysiology*, **33**, 148–159. 9

BUCHWALD, J.S., HALA, E. & SCHRAMM, S. (1965). A comparison of multi-unit activity and eeg activity recorded from the same brain site in chronic cats during behavioral conditioning. *Nature*, **205**, 1012–1014. 9

BULLOCK, T. (1993). Integrative systems research on the brain - resurgence and new opportunities. *Annual Review of Neuroscience*, **16**, 1–15. 69

BURGES, C.J.C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, **2**, 121–167. 14

BUTTS, D., WENG, C., JIN, J., YEH, C., LESICA, N., ALONSO, J. & STANLEY, G. (2007). Temporal precision in the neural code and the timescales of natural vision. *Nature*, **449**, 92–94. 80

BUZAS, P., KOVACS, K., FERECSKO, A.S., BUDD, J.M.L., EYSEL, U.T. & KISVARDAY, Z.F. (2006). Model-based analysis of excitatory lateral connections in the visual cortex. *J Comp Neurol*, **499**, 861–81. 103, 104

BUZSAKI, G. (2002). Theta oscillations in the hippocampus. *Neuron*, **33**, 325–340. 69

BUZSÁKI, G. (2006). *Rhythms of the Brain*. Oxford University Press. 41

BUZSAKI, G. & DRAGUHN, A. (304). Neuronal oscillations in cortical networks. *Science*, 1926–1929. 69

REFERENCES

CALLAWAY, E.M. (1998). Local circuits in primary visual cortex of the macaque monkey. *Annu Rev Neurosci*, **21**, 47–74. 101, 102, 108, 109, 110, 134

CARUANA, R. & JOACHIMS, T. (2004). KDD cup. http://kodiak.cs.cornell.edu/kddcup/index.html. 95

CASELLA, G. & BERGER, R. (2002). *Statistical Inference*. Duxbury, CA: Pacific Grove, 2nd edn. 87

CHANCE, F.S., NELSON, S.B. & ABBOTT, L.F. (1999). Complex cells as cortically amplified simple cells. *Nat Neurosci*, **2**, 277–82. 98

CHAZELLE, B. (2000). A minimum spanning tree algorithm with inverse-ackermann type complexity. *Journal of the ACM*, **47**. 90

COHEN, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement.*, **20**, 37–46. 16

CONWAY, B.R. & LIVINGSTONE, M.S. (2006). Spatial and temporal properties of cone signals in alert macaque primary visual cortex. *J Neurosci*, **26**, 10826–46. 106

COVER, T. & THOMAS, J. (1991). *Elements of Information Theory*. Wiley Interscience. 54, 56

CRONER, L.J. & KAPLAN, E. (1995). Receptive fields of P and M ganglion cells across the primate retina. *Vision Res*, **35**, 7–24. 106

CSICSVARI, J., JAMIESON, B., WISE, K. & BUZSÁKI, G. (2003). Mechanisms of gamma oscillations in Hippocampus of the behaving rat. *Neuron*, **37**, 311–322. 42

DAYAN, P. & ABBOTT, L.F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*. The MIT Press. 134

DE RUYTER VAN STEVENINCK, R., LEWEN, G., STRONG, S., KOBERLE, R. & BIALEK, W. (1997). Reproducibility and variability in neural spike trains. *Science*, **275**, 1805–1808. 54, 55, 68, 76

DESTEXHE, A., CONTRERAS, D. & STERIADE, M. (1999). Spatiotemporal analysis of local field potentials and unit discharges in cat cerebral cortex during natural wake and sleep states. *J. Neurosci.*, **19**, 4595–4608. 44

DESTEXHE, A., CONTRERAS, D. & STERIADE, M. (2001a). LTS cells in cerebral cortex and their role in generating spike-and-wave oscillations. *Neurocomputing*, **38**, 555–563. 44

DESTEXHE, A., RUDOLPH, M., FELLOUS, J.M. & SEJNOWSKI, T.J. (2001b). Fluctuating synaptic conductances recreate in vivo-like activity in neocortical neurons. *Neuroscience*, **107**, 13–24. 102

DONG, D.W. & ATICK, J.J. (1995). Temporal decorrelation: A theory of lagged and nonlagged responses in the lateral geniculate nucleus. *Network: Computation in Neural Systems*, **6**, 159–178. 99, 105, 106, 107

DUDLEY, R.M. (2002). *Real analysis and probability*. Cambridge University Press, Cambridge, UK. 85

ENROTH-CUGELL, C. & ROBSON, J.G. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *J Physiol*, **187**, 517–52. 106

ERNST, U.A., PAWELZIK, K.R., SAHAR-PIKIELNY, C. & TSODYKS, M. (2001). Intracortical origin of visual maps. *Nat Neurosci*, **4**, 431–6. 98

ERWIN, E., OBERMAYER, K. & SCHULTEN, K. (1995). Models of orientation and ocular dominance columns in the visual cortex: a critical comparison. *Neural Comput*, **7**, 425–68. 111

ESSEN, D.C.V., NEWSOME, W.T. & MAUNSELL, J.H. (1984). The visual field representation in striate cortex of the macaque monkey: asymmetries, anisotropies, and individual variability. *Vision Res*, **24**, 429–48. 131

FELLEMAN, D.J. & ESSEN, D.C.V. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*, **1**, 1–47. 110

FISER, J., CHIU, C. & WELIKY, M. (2004). Small modulation of ongoing cortical dynamics by sensory input during natural vision. *Nature*, **431**, 573–578. 45

REFERENCES

FISHER, N. (1993). *Statistical analysis of circular data.*. Cambridge University Press, Cambridge. 53, 72

FORTET, R. & MOURIER, E. (1953). Convergence de la réparation empirique vers la réparation théorique. *Ann. Scient. École Norm. Sup.*, **70**, 266–285. 85

FRIEDMAN, J. & RAFSKY, L. (1979). Multivariate generalizations of the wald-wolfowitz and smirnov two-sample tests. *The Annals of Statistics*, **7**, 697–717. 90

FRIEN, A., ECKHORN, R., BAUER, R., WOELBERN, T. & GABRIEL, A. (2000). Fast oscillations display sharper orientation tuning than slower components of the same recordings in striate cortex of the awake monkey. *European Journal of Neuroscience*, **12**, 1453–1465. 50

FRIES, P., NIKOLIĆ, D. & SINGER, W. (2007). The gamma cycle. *Trends in Neurosciences*, **30**, 309–316. 69, 79

FROMM, G.H. & BOND, H.W. (1964). Slow changes in the electrocorticogram and the activity of cortical neurons. *Electroencephalogrraphy Clinical Neurophysiology*, **17**, 520–3. 9, 43

FROMM, G.H. & BOND, H.W. (1967). The relationship between neuron activity and cortical steady potentials. *Electroencephalography & Clinical Neurophysiology*, **22**, 159–166. 9

GASSER, H.S. & GRUNDFEST, H. (1939). Axon diameters in relation to the spike dimensions and the conduction velocity in mammalian A fibers. *Am J Physiol*, 393–414. 9

GAZERES, N., BORG-GRAHAM, L.J. & FREGNAC, Y. (1998). A phenomenological model of visually evoked spike trains in cat geniculate nonlagged X-cells. *Vis Neurosci*, **15**, 1157–74. 100, 105, 107, 108

GILBERT, C.D., DAS, A., ITO, M., KAPADIA, M. & WESTHEIMER, G. (1996). Spatial integration and cortical dynamics. *Proc Natl Acad Sci U S A*, **93**, 615–22. 99, 103

GIRARD, P., HUPE, J.M. & BULLIER, J. (2001). Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. *J Neurophysiol*, **85**, 1328–31. 105, 110

GOLLISCH, T. & HERZ, A. (2005). Disentangling sub-millisecond processes within an auditory transduction chain. *PLOS Biology*, **3**, 144–154. 68

GRAY, C.M., KONIG, P., ENGEL, A.K. & SINGER, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, **338**, 334–7. 42

GRETTON, A., BORGWARDT, K., RASCH, M.J., SCHÖLKOPF, B. & SMOLA, A.J. (2007). A kernel method for the two-sample-problem. In J. Schölkopf B.; Platt & T. Hofmann, eds., *Advances in Neural Information Processing Systems*, vol. 19, The MIT Press, Cambridge, MA. 86, 88

GRINVALD, A. (1985). Real-time optical mapping of neuronal activity: from single growth cones to the intact mammalian brain. *Ann Rev Neurosci*, **8**, 263–305. 46

GRINVALD, A. & HILDESHEIM, R. (2004). VSDI: A new era in functional imaging of cortical dynamics. *Nature Reviews Neuroscience*, **5**, 874–885. 46

GROVER, F.S. & BUCHWALD, J.S. (1970). Correlation of cell size with amplitude of background fast activity in specific brain nuclei. *Journal of Neurophysiology*, **33**, 160–171. 9

GUPTA, A., WANG, Y. & MARKRAM, H. (2000). Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex. *Science*, **287**, 273–278. 99, 102, 105

HAEUSLER, S. & MAASS, W. (2007). A statistical analysis of information-processing properties of lamina-specific cortical microcircuit models. *Cereb Cortex*, **17**, 149–62. 4, 99, 101, 102, 103, 105, 134

HALL, P. & TAJVIDI, N. (2002). Permutation tests for equality of distributions in high-dimensional settings. *Biometrika*, **89**, 359–374. 91

REFERENCES

HARADA, Y. & TAKAHASHI, T. (1983). The calcium component of the action potential in spinal motoneurones of the rat. *Journal of Physiology*, **335**, 89–100. 69

HARRIS, K. (2005). Neural signatures of cell assembly organization. *Nature Reviews Neuroscience*, **6**, 399–407. 62, 79, 80

HARRIS, K., HENZE, D., HIRASE, H., LEINEKUGEL, X., DRAGOI, G., CZURKO, A. & BUZSAKI, G. (2002). Spike train dynamics predicts theta-related phase precession in hippocampal pyramidal cells. *Nature*, **417**, 738–741. 62, 79

HASENSTAUB, A., SHU, Y., HAIDER, B., KRAUSHAAR, U., DUQUE, A. & MCCORMICK, D.A. (2005). Inhibitory postsynaptic potentials carry synchronized frequency information in active cortical networks. *Neuron*, **47**, 423–435. 43

HENRIE, J.A. & SHAPLEY, R. (2005). LFP power spectra in V1 cortex: the graded effect of stimulus contrast. *J Neurophysiol*, **94**, 479–90. 2, 50

HENZE, D.A., BORHEGYI, Z., CSICSVARI, J., MAMIYA, A., HARRIS, K.D. & BUZSAKI, G. (2000). Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *J Neurophysiol*, **84**, 390–400. 9, 44, 114

HENZE, N. & PENROSE, M. (1999). On the multivariate runs test. *The Annals of Statistics*, **27**, 290–298. 90

HOPFIELD, J. (1995). Pattern-recognition computation using action-potential timing for stimulus representation. *Nature*, **376**, 33–36. 68, 69

HUANG, C.M. & BUCHWALD, J.S. (1977). Interpretation of the vertex short-latency acoustic response: a study of single neurons in the brain stem. *Brain Research*, **137**, 291–303. 9

HUBEL, D.H. & WIESEL, T.N. (1977). Ferrier lecture. Functional architecture of macaque monkey visual cortex. *Proc R Soc Lond B Biol Sci*, **198**, 1–59. 108, 111

HUMPHREY, A.L. & WELLER, R.E. (1988a). Functionally distinct groups of X-cells in the lateral geniculate nucleus of the cat. *J Comp Neurol*, **268**, 429–47. 99

HUMPHREY, A.L. & WELLER, R.E. (1988b). Structural correlates of functionally distinct X-cells in the lateral geniculate nucleus of the cat. *J Comp Neurol*, **268**, 448–68. 99

HUNT, C. (1951). The reflex activity of mammalian small-nerve fibers. *J Physiol (Lond)*, **115**, 456–469. 9

HUXTER, J., BURGESS, N. & O'KEEFE, J. (2003). Independent rate and temporal coding in hippocampal pyramidal cells. *Nature*, **425**, 828–832. 79

INGBER, L. (1989). Very fast simulated re-annealing. *Mathl. Comput. Modelling*, **12**, 967–973. 132

IZHIKEVICH, E.M. (2003). Simple model of spiking neurons. *IEEE Trans Neural Netw*, **14**, 1569–72. 101

IZHIKEVICH, E.M. (2006). *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting (Computational Neuroscience)*. The MIT Press. 101, 133

IZHIKEVICH, E.M. & EDELMAN, G.M. (2008). Large-scale model of mammalian thalamocortical systems. *Proc Natl Acad Sci U S A*, **105**, 3593–8. 131

IZHIKEVICH, E.M., GALLY, J.A. & EDELMAN, G.M. (2004). Spike-timing dynamics of neuronal groups. *Cereb Cortex*, **14**, 933–44. 101

JARVIS, M.R. & MITRA, P.P. (2001). Sampling properties of the spectrum and coherency of sequences of action potentials. *Neural Computation*, **13**, 717–749. 16, 18

JENSEN, O. & LISMAN, J. (2000). Position reconstruction from an ensemble of hippocampal place cells: Contribution of theta phase coding. *Journal of Neurophysiology*, **83**, 2602–2609. 79

JOHANSSON, C. & LANSNER, A. (2007). Towards cortex sized artificial neural systems. *Neural Netw*, **20**, 48–61. 131

JOHNSON, N.L., KOTZ, S., & BALAKRISHNAN, N. (1994). *Continuous Univariate Distributions*, vol. 1. John Wiley and Sons, 2nd edn. 89

REFERENCES

JUERGENS, E., GUETTLER, A. & ECKHORN, R. (1999). Visual stimulation elicits locked and induced gamma oscillations in monkey intracortical- and EEG-potentials, but not in human EEG. *Experimental Brain Research*, **129**, 247–259. 9, 69

KAMONDI, A., ACSADY, L., WANG, X.J. & BUZSAKI, G. (1998). Theta oscillations in somata and dendrites of hippocampal pyramidal cells in vivo: activity-dependent phase-precession of action potentials. *Hippocampus*, **8**, 244–261. 9, 69

KENET, T., BIBITCHKOV, D., TSODYKS, M., GRINVALD, A. & ARIELI, A. (2003). Spontaneously emerging cortical representations of visual attributes. *Nature*, **425**, 954–956. 45

KOCH, C. (2004). *Biophysics of Computation: Information Processing in Single Neurons (Computational Neuroscience)*. Oxford University Press, USA. 112

KOHONEN, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, **43**, 59–69. 111

KREIMAN, G., HUNG, C.P., KRASKOV, A., QUIROGA, R.Q., POGGIO, T. & DICARLO, J.J. (2006). Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. *Neuron*, **49**, 433–45. 41

KREITER, A.K. & SINGER, W. (1992). Oscillatory neuronal responses in the visual cortex of the awake macaque monkey. *European Journal of Neuroscience*, **4**, 369–375. 42

KREMKOW, J., KUMAR, A., ROTTER, S. & AERTSEN, A. (2007). Emergence of population synchrony in a layered network of the cat visual cortex. *Neurocomput.*, **70**, 2069–2073. 131

LEE, H., SIMPSON, G.V., LOGOTHETIS, N.K. & RAINER, G. (2005). Phase locking of single neuron activity to theta oscillations during working memory in monkey extrastriate visual cortex. *Neuron*, **45**, 147–156. 41, 79

LISMAN, J. (2005). The theta/gamma discrete phase code occuring during the hippocampal phase precession may be a more general brain coding scheme. *Hippocampus*, **15**, 913–922. 69

LIU, J. & NEWSOME, W.T. (2006). Local field potential in cortical area MT: stimulus tuning and behavioral correlations. *J Neurosci*, **26**, 7779–90. 41

LOGOTHETIS, N., MERKLE, H., AUGATH, M., TRINATH, T. & UGURBIL, K. (2002). Ultra high-resolution fMRI in monkeys with implanted RF coils. *Neuron*, **35**, 227–42. 10, 48

LOGOTHETIS, N.K. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *J Neurosci*, **23**, 3963–71. 2, 9, 69

LOGOTHETIS, N.K., GUGGENBERGER, H., PELED, S. & PAULS, J. (1999). Functional imaging of the monkey brain. *Nat. Neurosci.*, **2**, 555–62. 49

LOGOTHETIS, N.K., PAULS, J., AUGATH, M., TRINATH, T. & OELTERMANN, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, **412**, 150–157. 2, 46

LUND, J.S., ANGELUCCI, A. & BRESSLOFF, P.C. (2003). Anatomical substrates for functional columns in macaque monkey primary visual cortex. *Cereb Cortex*, **13**, 15–24. 103, 108, 110

MAASS, W., NATSCHLÄGER, T. & MARKRAM, H. (2002). Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, **14**, 2531–2560. 102

MACKAY, D. & MCCULLOCH, W. (1952). The limiting information capacity of a neuronal link. *Bull Math Biophys*, **14**, 127–135. 68

MAINEN, Z.F. & SEJNOWSKI, T.J. (1995). Reliability of spike timing in neocortical neurons. *Science*, **268**, 1503–1506. 46

MARKRAM, H. (2006). The blue brain project. *Nat Rev Neurosci*, **7**, 153–60. 132

MARKRAM, H., WANG, Y. & TSODYKS, M. (1998). Differential signaling via the same axon of neocortical pyramidal neurons. *PNAS*, **95**, 5323–5328. 99, 102

REFERENCES

MARKRAM, H., TOLEDO-RODRIGUEZ, M., WANG, Y., GUPTA, A., SILBERBERG, G. & WU, C. (2004). Interneurons of the neocortical inhibitory system. *Nat Rev Neurosci*, **5**, 793–807. 101

MASTRONARDE, D.N. (1987). Two classes of single-input X-cells in cat lateral geniculate nucleus. I. Receptive-field properties and classification of cells. *J Neurophysiol*, **57**, 357–80. 99

MAUNSELL, J.H., GHOSE, G.M., ASSAD, J.A., MCADAMS, C.J., BOUDREAU, C.E. & NOERAGER, B.D. (1999). Visual response latencies of magnocellular and parvocellular LGN neurons in macaque monkeys. *Vis Neurosci*, **16**, 1–14. 110

MCGUIRE, B.A., GILBERT, C.D., RIVLIN, P.K. & WIESEL, T.N. (1991). Targets of horizontal connections in macaque primary visual cortex. *J Comp Neurol*, **305**, 370–92. 103, 104

MEHTA, M., LEE, A. & WILSON, M. (2002). Role of experience and oscillations in transforming a rate code into a temporal code. *Nature*, **417**, 741–746. 62, 79

MITZDORF, U. (1985). Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena. *Physiol. Rev.*, **65**, 37–99. 9

MITZDORF, U. (1987). Properties of the evoked potential generators: current source-density analysis of visually evoked potentials in the cat cortex. *International Journal of Neuroscience*, **33**, 33–59. 9, 69

MONTEMURRO, M., SENATORE, R. & PANZERI, S. (2007a). A downward biased estimator of spike timing information. *Neurocomputing*, **70**, 1777–1781. 59

MONTEMURRO, M., SENATORE, R. & PANZERI, S. (2007b). Tight data-robust bounds to mutual information combining shuffling and model selection techniques. *Neural Computation*, **19**, 2913–57. 59, 62

MUSICANT, D.R., KUMAR, V. & OZGUR, A. (2003). Optimizing f-measure with support vector machines. In I. Russell & S. Haller, eds., *Proceedings of the Sixteenth International Florida Artificial Intelligence Research Society Conference*, 356–360, AAAI Press. 19

NADASDY, Z., CSICSVARI, J., PENTTONEN, M., HETKE, J., WISE, K. & BUZSAKI, G. (1998). Extracellular recording and analysis of neuronal activity: From single cells to ensembles. In E. H & D. JL, eds., *Neuronal ensembles: strategies from recording and decoding*, 17–55, Wiley-Liss, New York. 9

NATSCHLÄGER, T. & MAASS, W. (2005). Dynamics of information and emergent computation in generic neural microcircuit models. *Neural Netw*, **18**, 1301–8. 36

NELSON, P.G. (1966). Interaction between spinal motoneurons of the cat. *Journal of Neurophysiology*, **29**, 275–287. 9

NUNEZ, P.L. & SHRINIVASAN, R. (2006). *Electric fields of the brain*. Oxford University Press, 2nd edn. 44, 112

OBERMAYER, K. & BLASDEL, G.G. (1993). Geometry of orientation and ocular dominance columns in monkey striate cortex. *J Neurosci*, **13**, 4114–29. 111, 112

OBERMAYER, K., RITTER, H. & SCHULTEN, K. (1990). A principle for the formation of the spatial structure of cortical feature maps. *Proc Natl Acad Sci U S A*, **87**, 8345–9. 111

OBERMAYER, K., BLASDEL, G.G. & SCHULTEN, K. (1992). Statistical-mechanical analysis of self-organization and pattern formation during the development of visual maps. *Phys Rev A*, **45**, 7568–7589. 111, 112

OGAWA, S., TANK, D.W., MENON, R., ELLERMANN, J.M., KIM, S.G., MERKLE, H. & UGURBIL, K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proc Natl Acad Sci U S A*, **89**, 5951–5. 2

O'KUSKY, J. & COLONNIER, M. (1982). Postnatal changes in the number of astrocytes, oligodendrocytes, and microglia in the visual cortex (area 17) of the macaque monkey: a stereological analysis in normal and monocularly deprived animals. *J Comp Neurol*, **210**, 307–15. 102

OLSHAUSEN, B.A. & FIELD, D.J. (2005). How close are we to understanding v1? *Neural Comput*, **17**, 1665–99. 98, 133, 135

REFERENCES

OPTICAN, L. & RICHMOND, B. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. information theoretic analysis. *Journal of Neurophysiology*, **57**, 162–178. 68

OSIPOVA, D., TAKASHIMA, A., OOSTENVELD, R., FERNANDEZ, G., MARIS, E. & JENSEN, O. (2006). Theta and gamma oscillations predict encoding and retrieval of declarative memory. *J Neurosci*, **26**, 7523–31. 41

PANZERI, S. & SCHULTZ, S. (2001). A unified approach to the study of temporal, correlational, and rate coding. *Neural Computation*, **13**, 1311–1349. 57, 58, 68

PANZERI, S. & TREVES, A. (1996). Analytical estimates of limited sampling biases in different information measures. *Network-Computation in Neural Systems*, **7**, 87–107. 58

PANZERI, S., PETERSEN, R., SCHULTZ, S., LEBEDEV, M. & DIAMOND, M. (2001). The role of spike timing in the coding of stimulus location in rat somatosensory cortex. *Neuron*, **29**, 769–777. 58, 80

PANZERI, S., SENATORE, R., MONTEMURRO, M.A., & PETERSEN, R.S. (2007). Correcting for the sampling bias problem in spike train information measures. *J Neurophysiol*, **98**, 1064–1072. 17, 58, 62, 68

PERCIVAL, D.B. & WALDEN, A.T. (2002). *Spectral analyses for physical applications*. Cambridge University Press. 16, 50

PERRY, V.H., OEHLER, R. & COWEY, A. (1984). Retinal ganglion cells that project to the dorsal lateral geniculate nucleus in the macaque monkey. *Neuroscience*, **12**, 1101–23. 105

PETERS, A., PAYNE, B.R. & BUDD, J. (1994). A numerical analysis of the geniculocortical input to striate cortex in the monkey. *Cereb Cortex*, **4**, 215–29. 109

POLA, G., THIELE, A., HOFFMANN, K. & PANZERI, S. (2003). An exact method to quantify the information transmitted by different mechanisms of correlational coding. *Network: Computation in Neural Systems*, **14**, 35–60. 17

QUIROGA, R., NADASDY, Z. & BEN-SHAUL, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Computation*, **16**, 1661–1687. 49, 50

RIEKE, F., WARLAND, D., DE RUYTER VAN STEVENINCK, R. & BIALEK, W. (1999). *Spikes: Exploring the Neural Code*. MIT Press. 46

ROCKLAND, K.S. & VIRGA, A. (1989). Terminal arbors of individual "feedback" axons projecting from area V2 to V1 in the macaque monkey: a study using immunohistochemistry of anterogradely transported Phaseolus vulgaris-leucoagglutinin. *J Comp Neurol*, **285**, 54–72. 110

RODIECK, R.W. (1965). Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Res*, **5**, 583–601. 106

ROSENBAUM, P. (2005). An exact distribution-free test comparing two multivariate distributions based on adjacency. *Journal of the Royal Statistical Society B*, **67**, 515–530. 91

RUBINO, D., ROBBINS, K.A. & HATSOPOULOS, N.G. (2006). Propagating waves mediate information transfer in the motor cortex. *Nature Neuroscience*, **9**. 41

SCHEIN, S.J. & DE MONASTERIO, F.M. (1987). Mapping of retinal and geniculate neurons onto striate cortex of macaque. *J Neurosci*, **7**, 996–1009. 108

SCHERBERGER, H., JARVIS, M.R. & ANDERSEN, R.A. (2005). Cortical local field potential encodes movement intentions in the posterior parietal cortex. *Neuron*, **46**, 347–354. 41

SCHNEIDMAN, E., BIALEK, W. & BERRY, M.J. (2003). Synergy, redundancy, and independence in population codes. *J Neurosci*, **23**, 11539–11553. 17, 18

SCHNUPP, J., HALL, T., KOKELAAR, R. & AHMED, B. (2006). Plasticity of temporal pattern codes for vocalization stimuli in primary auditory cortex. *Journal of Neuroscience*, **26**, 4785–4795. 69

SCHÖLKOPF, B. & SMOLA, A.J. (2002). *Learning with Kernels*. MIT Press, Cambridge, MA. 14

REFERENCES

SERFLING, R. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. 87, 88

SHANNON, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, **27**, 623–656. 54, 74

SHPIGELMAN, L., SINGER, Y., PAZ, R. & VAADIA, E. (2003). Spikernels: Embedding spiking neurons in inner-product spaces. *Advances in Neural Information Processing Systems*. 132

SINCICH, L.C. & HORTON, J.C. (2005). The circuitry of V1 and V2: integration of color, form, and motion. *Annu Rev Neurosci*, **28**, 303–26. 110, 135

STEINWART, I. (2002). On the influence of the kernel on the consistency of support vector machines. *J. Mach. Learn. Res.*, **2**, 67–93. 84, 85

STERIADE, M., MCCORMICK, D.A. & SEJNOWSKI, T.J. (1993). Thalamocortical oscillations in the sleeping and aroused brain. *Science*, **262**, 679–685. 43

STRONG, S., KOBERLE, R., DE RUYTER VAN STEVENINCK, R. & BIALEK, W. (1998). Entropy and information in neural spike trains. *Phys. Rev. Lett.*, **80**, 197–200. 54, 59

TADMOR, Y. & TOLHURST, D.J. (2000). Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes. *Vision Res*, **40**, 3145–57. 106

TAO, L., SHELLEY, M., MCLAUGHLIN, D. & SHAPLEY, R. (2004). An egalitarian network model for the emergence of simple and complex cells in visual cortex. *Proc Natl Acad Sci U S A*, **101**, 366–71. 98, 131

TEICH, A.F. & QIAN, N. (2006). Comparison among some models of orientation selectivity. *J Neurophysiol*, **96**, 404–19. 108

THOMSON, A.M. & BANNISTER, A.P. (2003). Interlaminar connections in the neocortex. *Cereb Cortex*, **13**, 5–14. 99, 101, 103, 104, 105, 113, 115, 119, 134

THOMSON, D.J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE Volume 70*, 1055–1096. 16

TOOTELL, R.B., SILVERMAN, M.S., SWITKES, E. & VALOIS, R.L.D. (1982). Deoxyglucose analysis of retinotopic organization in primate striate cortex. *Science*, **218**, 902–4. 108

TROYER, T.W., KRUKOWSKI, A.E., PRIEBE, N.J. & MILLER, K.D. (1998). Contrast-invariant orientation tuning in cat visual cortex: thalamocortical input tuning and correlation-based intracortical connectivity. *J Neurosci*, **18**, 5908–27. 109

TYLER, C.J., DUNLOP, S.A., LUND, R.D., HARMAN, A.M., DANN, J.F., BEAZLEY, L.D. & LUND, J.S. (1998). Anatomical comparison of the macaque and marsupial visual cortex: common features that may reflect retention of essential cortical elements. *J Comp Neurol*, **400**, 449–68. 102

VAPNIK, V. (1999). *The Nature of Statistical Learning Theory*. Springer. 14

VICTOR, J. & PURPURA, K. (1996). Nature and precision of temporal coding in visual cortex: A metric-space analysis. *Journal of Neurophysiology*, **76**, 1310–1326. 69

VINCENT, J.L., PATEL, G.H., FOX, M.D., SNYDER, A.Z., BAKER, J.T., VAN ESSEN, D.C., ZEMPEL, J.M., SNYDER, L.H., CORBETTA, M. & RAICHLE, M.E. (2007). Intrinsic functional architecture in the anaesthetized monkey brain. *Nature*, **447**, 83–86. 45

WENISCH, O.G., NOLL, J. & VAN HEMMEN, J.L. (2005). Spontaneously emerging direction selectivity maps in visual cortex through STDP. *Biol Cybern*, **93**, 239–47. 98

WIELAARD, D.J., SHELLEY, M., MCLAUGHLIN, D. & SHAPLEY, R. (2001). How simple cells are made in a nonlinear network model of the visual cortex. *J Neurosci*, **21**, 5203–11. 98

YUSTE, R. & TANK, D.W. (1996). Dendritic integration in mammalian neurons, a century after Cajal. *Neuron*, **16**, 701–16. 133