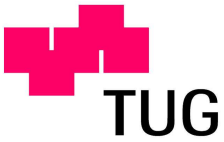
Information processing properties of neocortical microcircuits



Stefan Haeusler

Institute for Theoretical Computer Science

Graz University of Technology

Doctoral thesis

Graz, December 2006

Acknowledgements

I would like to thank Wolfgang Maass, my supervisor, for his guidance and support. His inspiring approach to neuroscience directed me to this exciting research topic.

Additional thanks go to Danko Nikolić for the fruitful collaboration resulting in Chapter 4.

I would like to thank Helmut Pockberger, for his friendly willingness to be the second referee of this thesis.

I am grateful to Thomas Natschlaeger for valuable discussions related to Chapter 2.

Furthermore I would like to thank Ed Callaway, Rodney Douglas, Rolf Koetter, Henry Markram, and Alex Thomson for helpful discussions on research related to Chapter 3.

I also want to express special thanks to Malte Rasch, Robert Legenstein, Michael Pfeiffer, David Sussillo, Helmut Hauser, Prashant Joshi, Stefan Klampfl, Ingrid Preininger and Oliver Friedl for making my stay here so enjoyable.

Finally I am grateful to my wonderful wife for supporting me.

I was partially supported by the Austrian Science Fund FWF and the FACETS project of the European Union.

Abstract

The neocortex has a distinctive modular and laminar structure, that exhibits a characteristic anatomical organization into six cortical layers. One may conjecture that this stereotypical organization is distinguished by specific functional properties that are advantageous for many information processing domains in diverse cortical areas. This thesis investigates the computational operations in the neocortex and how they are related to the underlying stereotypical circuit structure.

It is shown by means of computer simulations of detailed cortical microcircuit models that the biologically realistic lamina-specific synaptic connectivity structure exhibits specific computational advantages over various types of control structures that consist of the same components. Several aspects of the connectivity structures are analyzed for their impact on the information processing capabilities of cortical microcircuit models. It turns out that structures that support noise suppression enhance the computational capabilities of cortical microcircuit models.

Furthermore the analysis of multi-electrode recordings in cat primary visual cortex demonstrates that the temporal dynamics of information is not compatible with simple computational models based on linear operations. The diversity of temporal profiles recorded at different electrodes, the superposition of slowly fading memory from different preceding stimuli, and the fact that linear readout models are sufficient even for nonlinear processing tasks can be seen as support for recently proposed computational models based on nonlinear dynamic systems theory. It is demonstrated that this characteristic features can be replicated by a detailed model for cat primary visual cortex.

Zusammenfassung

Der Neokortex hat eine charakteristische modulare und laminare Struktur, welche eine typische anatomische Organisation in sechs Layern aufweist. Es wird vermutet, dass diese stereotypische Organisation durch spezifische funktionelle Eigenschaften ausgezeichnet ist, welche für viele Informationverarbeitungsgebiete in unterschiedlichen kortikalen Arealen vorteilhaft ist. Diese Arbeit untersucht die Rechenoperationen im Neokortex und wie diese von der zugrunde liegenden stereotypischen Verbindungsstruktur abhängen.

Es wird durch Computersimulationen von detaillierten kortikalen Mikroschaltkreismodellen gezeigt, dass die biologisch realistischen synaptischen Verbindungsstrukturen einen spezifischen Rechenvorteil gegenüber verschiedenen Kontrollstrukturen, welche aus den selben Elementen bestehen, haben. Verschiedene Aspekte der Verbindungsstrukturen werden auf ihre Auswirkungen auf die Informationsverarbeitungsfähigkeiten von kortikalen Mikroschaltkreismodellen analysiert. Es stellt sich heraus, dass Strukturen welche Rauschunterdrückung unterstützen, die Rechenfähigkeiten von kortikalen Mikroschaltkreismodellen verbessern.

Weiters zeigt die Analyse von Mehrelektrodenaufzeichnungen im primären visuellen Kortex der Katze, dass die zeitliche Dynamik von Information nicht mit einfachen Rechenmodellen, welche auf linearen Operationen basieren, vereinbar ist. Die Vielfältigkeit der zeitlichen Profile der Aufzeichnungen verschiedener Elektroden, die Überlagerung der langsam abklingenden Speicherung verschiedener vorangegangener Stimuli und die Tatsache, dass lineare Auslesemodelle sogar für nichtlineare Verarbeitungsaufgaben ausreichend sind, kann als eine Bestätigung der vor kurzem vorgeschlagenen Rechenmodelle basierend auf nichtlinearer dynamischer Systemtheorie angesehen werden. Es wird gezeigt, dass diese charakteristichen Eigenschaften in einem detaillierten Modell für den primären visuellen Kortex der Katze reproduziert werden können.

Contents

Chapter 1

Introduction

The cerebral cortex of mammals has been the subject of intense study for at least four decades. Although much has been discovered about the response properties of individual cortical neurons and about the structure of topographic maps, little is presently known about the computational operations in cortex and how they are achieved by the underlying circuit structure. Even for the extensively studied primary sensory cortices, such as visual cortex, auditory cortex, and somatosensory cortex, the direct relation between circuitry and behavior is still largely unknown.

On the other hand all of these cortical areas have a similar anatomical organization and are more similar to each other than to other brain areas. Cortical neurons seem to obey well-defined rules of connectivity across layers that are characteristic for each layer and have motivated the concept of a *cortical column*. This relatively uniform structure of cerebral cortex suggests that the similarity on the anatomical level extends to the computational level and endows different areas with specific universal functional properties that are advantageous for many information processing domains.

A conceptual framework that could explain and analyze the potentially universal computational capabilities of such microcircuits has been proposed recently by Maass *et al.* (2002)[1], called *liquid state machine*. In this framework a cortical microcircuit is not viewed as an implementation of a single computation, but as

[1]A closely related computational model was discovered by Jäger & Haas (2004) in the context of artificial neural networks.

a more universal computational device that can simultaneously support a large number of different computations. For this computational model the exact operations that are carried out by the cortical microcircuit are less relevant as long as neural readouts can learn to transform the ongoing activity of the cortical microcircuit at any moment in time into the output that is needed for their specific computation. Compared to Turing machines or attractor neural networks this model is more adequate for analyzing parallel real-time computations on time-varying input streams, such as those occurring in generic cognitive information processing tasks, because it does not rely on discrete internal states or require convergence to some stable internal state that are in general hard to control. Within this framework the investigation of the functional properties of cortical microcircuits corresponds to the analysis of the information that is available to neural readouts about results of specific computations.

The aim of this thesis is to gain insight in the functional properties of the neocortex by linking modeling and experimental studies within the framework of the *liquid state machine*. This is achieved by focusing on one hand on computer simulations of detailed microcircuit models in order to relate certain structural features of the neocortical circuitry to specific functional properties. On the other hand a direct experimental test of these computational properties is carried out by analyzing multi-electrode recordings from cat primary visual cortex (area 17). Finally a link between model and experiment is established by comparing certain computational aspects that are found in the experiments to the corresponding results for the simulation of a detailed computer model of a 1.1×1.1 mm^2 patch of cat primary visual cortex complemented by a standard model for retina and LGN.

For the first part of the thesis dedicated to the modeling approach I investigate in Chapter 2 the general applicability of the concept of the *liquid state machine* for computations in neural microcircuits. For this purpose I explore the possible relationship between the high dimensional dynamics of neural circuits and the potential capabilities of neural readouts through computer simulations of generic cortical microcircuits. These neural readouts receive inputs from hundreds of neurons in the circuit and collapse the dynamics of this very high dimensional system to a low dimensional state. A consequence of this collapse is that each readout

neuron can learn to assign equivalence classes of high dimensional microcircuit states that allow invariant readout responses despite the fact that the neural microcircuit may never return to the same state. I analyze the capability of readout neurons to extract clearly structured trajectories from the complex trajectories of transient states of large neural circuits that can exhibit even convergence to virtual attractors. The attractor landscape may be completely different for different readout neurons. They are called virtual because they exist only from the perspective of readout neurons and are not apparent in the high dimensional trajectory. In principle these virtual attractors make it possible to carry out specific computations without changing the dynamics of the microcircuit itself. Furthermore I address the point whether generic neural microcircuits can play a similar role for neural computing as high dimensional kernels for support vector machines, which present a large number of nonlinear combinations of components of the input stream to a readout. This kernel property of neural microcircuits is tested by comparing the performance of a single linear readout to the performance of sophisticated multi-unit classifiers that do not rely on a nonlinear preprocessing stage.

In Chapter 3 I apply the framework of the *liquid state machine* to the analysis of the potential universal functional properties of the neocortex due to its specific anatomical organization. The neocortex has a distinctive modular and laminar organization, that exhibits a characteristic organization into six cortical layers. Sensory regions are organized as arrays of vertical columns with a diameter of about 500 μm that are composed of cells with similar response properties. Each column traverses the six cortical layers and each layer has a unique pattern of inputs, intrinsic connections and outputs. Layer 1 receives input from non-specific thalamus, layer 2/3 from higher cortical regions, layer 5 from specific thalamus and layer 6 receives input from specific thalamic nuclei and multiple brain regions. On the other hand layer 2/3 projects to higher cortical areas and layer 5 provides the main output to lower cortical areas or to subcortical structures, but also projects back to nonspecific thalamus. Finally layer 6 projects to thalamus and multiple brain regions specialized to process different modalities. Additionally all layers are highly interconnected with each other forming a precisely structured microcircuit. Why the neocortex has a laminar design has remained a mystery

from a functional point of view. In contrast to the previous work of Treves (2003) and Raizada & Grossberg (2003) that is based on the assumption of specific computational roles of individual layers of a rather abstract neural network model I approach this question by analyzing the information processing capabilities of a detailed model of a single cortical column with a diameter of about 100 μm. Furthermore I investigate the impact of certain structural features of the lamina-specific connectivity on the computational performance of generic 'neocortical readouts', i.e. of projection neurons in layers 2/3 and layer 5. The purpose of this chapter is to identify certain properties of the connectivity graphs that are most salient for the computational performance of these readouts.

The aim of the second part of this thesis is to investigate experimentally if the computational properties that are predicted by the theory of the *liquid state machine* are supported by multi-electrode recordings from cat area 17. There exists a variety of different computational models for visual processing in the brain that are consistent with most of the available experimental data. Most work on cortical microcircuits is done in the conceptual framework of Hubel & Wiesel (1962) that was developed 40 years ago for cat area 17. This framework postulates a precise feedforward organization of visual processing to explain the formation of specific receptive fields. On the other hand cortical circuits are highly recurrent structures. For instance simple cells in layer 4 of cat receive 95% of their synapses from other cortical cells (Ahmed *et al.* (1994)). Thus notions such as 'feedforward' and 'feedback' are inadequate concepts for the analysis of highly recurrent cortical microcircuits. In the framework of the *liquid state machine* recurrent microcircuits are viewed as kernel-like processors that perform a nonlinear spatio-temporal integration of time-varying inputs thereby providing the computational properties of fading memory and high nonlinearity for visual processing. A direct experimental test of these essential properties for real-time processing of quickly varying visual stimuli has been missing so far for cat primary visual cortex. An example for an experimental study that may be seen as a step in this direction is Hung *et al.* (2005), where it was shown that the firing activity of an unbiased sample of a few neurons in macaque inferior temporal cortex (IT) contains information about a previously shown image, and that this information

lasts for several hundred ms. A similar persistence of information in the cortical entrance stage of visual processing would be somewhat surprising.

In Chapter 4 I focus in particular on the temporal dynamics of information about previously shown visual patterns and how the information about visual inputs arriving at different moments in time is superimposed and merged within the microcircuit. In this context I additionally address the question of how the information about previously shown stimuli is encoded in the neural activity, or equivalently, what kernel-like mappings are used to project this information into the cat primary visual cortex. Up to now it is still a controversial issue as to whether cortical neurons transmit information primarily in their average firing rates or in the precise timing of spikes. Synchronous firing is experimentally observed in the cat superior colliculus (Brecht *et al.* (1999)). Furthermore neurons with overlapping receptive fields in the cat primary visual cortex fire synchronously when activated by common preferred stimuli (Gray *et al.* (1989)). It has been shown for β-lobe neurons in locusts *in vivo* (MacLeod *et al.* (1998)) that precise spike timing in the order of millisecond carries information about input stimuli. Desynchronization of ensembles of projection neurons causes a loss of information in odor-evoked spike trains in single downstream β-lobe neurons. A corresponding study of the information about stimuli that is contained in the precise timing of spikes is missing for the cat primary visual cortex. I therefore focus on the question of how the information about previously shown visual stimuli is encoded in the neural activity in the context of rate coding and precise timing of spikes.

Finally in Chapter 5 a link between modeling and experiment is established by comparing the experimental results obtained in Chapter 4 to the results for a detailed computer model for a 1.1×1.1 mm^2 patch of cat primary visual cortex (area 17). The model is based on the cortical microcircuit model with biologically realistic lamina-specific connectivity presented in Chapter 3 and complemented by a standard model for retina and LGN. In particular the temporal dynamics of information and the importance of the precise timing of spikes for the computational performance of neural readouts is compared to the corresponding results for multi-electrode recordings from cat primary visual cortex.

Chapter 2

Perspectives of the high dimensional dynamics of neural microcircuits from the point of view of low dimensional readouts

Summary

I investigate generic models for cortical microcircuits, i.e. recurrent circuits of integrate-and fire neurons with dynamic synapses. These complex dynamic systems subserve the amazing information processing capabilities of the cortex, but are at the present time very little understood. I analyze the transient dynamics of models for neural microcircuits from the point of view of one or two readout neurons that collapse the high dimensional transient dynamics of a neural circuit into a 1- or 2–dimensional output stream. This stream may for example represent the information that is projected from such circuit to some particular other brain area or actuators. It is shown that simple local learning rules enable a readout neuron to extract from the high dimensional transient dynamics of a recurrent neural circuit quite different low-dimensional projections, that even may contain "virtual attractors" which are not apparent in the high dimensional dynamics of the circuit itself. Furthermore it is demonstrated that the information extraction capabilities of linear readout neurons are boosted by the computational

operations of a sufficiently large preceding neural microcircuit. Hence a generic neural microcircuit may play a similar role for information processing as a kernel for support vector machines in machine learning. I demonstrate that the projection of time-varying inputs into a large recurrent neural circuit enables a linear readout neuron to classify the time-varying circuit inputs with the same power as a complex nonlinear classifiers, such as for example a pool of perceptrons trained by the p-delta-rule, or a feedforward sigmoidal neural net trained by backprop, provided that the size of the recurrent circuit is sufficiently large. At the same time such readout neuron can exploit the stability and speed of learning rules for linear classifiers, thereby overcoming the problems caused by local minima in the error function of nonlinear classifiers. In addition it is demonstrated that pairs of readout neurons can transform the complex trajectory of transient states of a large neural circuit into a simple and clearly structured 2-dimensional trajectory. This 2-dimensional projection of the high-dimensional trajectory can even exhibit convergence to virtual attractors which are not apparent in the high dimensional trajectory.

2.1 Introduction

Computation in biological neural circuits is often modeled by attractor neural networks with low dimensional internal state spaces, and analyzed from the point of view of a human observer with focus on easily discernible features such as convergence to an attractor. However, the dynamics of real neural microcircuits, consisting of a few thousand neurons, represents a trajectory in a very high dimensional dynamical system. Due to its high dimensionality, new phenomena emerge which cannot be observed in the commonly studied 2- or 3-dimensional dynamical systems. Functionally most important are features of the dynamics of neural circuits that can be extracted by readout neurons, i.e. by neurons that receive inputs from hundreds or thousands of neurons in this circuit and transmit low dimensional projections of their transient dynamics to other brain areas, or to actuators. This chapter explores the possible relationship between the high dimensional dynamics of neural circuits and their neural readouts through computer simulations of generic cortical microcircuits.

Recently Maass, Markram & Natschlaeger proposed a general theoretical model, called *liquid state machine* Maass *et al.* (2002), which represents a convenient framework for neural computations in real time for rapidly time varying continuous input functions. It does not require convergence to stable internal states or attractors, since information about past inputs is captured in the perturbations of a high dimensional dynamical system, i.e. in the continuous trajectory of transient internal states. First the input stream is projected into a sufficiently large neural circuit. In general different input streams will cause different trajectories of internal states of the system, i.e., the input streams are separated by the circuit. Secondly a memory-less readout learns to extract salient information from the high dimensional transient states of the circuit. In particular each readout can learn to define its own classes of equivalence of dynamical states within the neural microcircuit, and can then perform its task on novel inputs. Due to this principle of "readout assigned equivalent states of a dynamical system" an invariant readout can be possible despite the fact that the neural microcircuit may never re-visit the same state. Furthermore multiple readouts can be trained to perform different tasks on the same state trajectories of a recurrent neural circuit, thereby enabling parallel real-time computing. Good separation capability of the high dimensional dynamical system for different preceding inputs, in combination with an adequate readout, allows essentially any real-time computation on continuous and bounded time-varying inputs with fading memory to an arbitrary degree of precision. It is shown in Maass *et al.* (2002) that a generic neural microcircuit model tends to have fairly good separation property, due to the biologically realistic diversity of its components and its sparse but recurrent connectivity ("loops within loops"). Adaptivity within the microcircuit itself is not necessary in this context, although it may facilitate the task of the readout for a family of related tasks. This situation is analogous to that of choosing kernels for support vector machines, where there exist general purpose kernels that provide good performance for a large variety of tasks.

Whereas in Maass *et al.* (2002) the potential readout capabilities of *pools* of neurons were explored, I investigate in this chapter the readout capability of *single* integrate-and-fire (I&F) neurons and of pairs of such neurons. It is shown that for a sufficiently large recurrent neural circuit a single neuron as

readout achieves the same classification power for a binary classification task (as specified in section 2.3.2) as sophisticated multi-unit classifiers, such as pools of perceptrons with the p-delta-rule, see Auer *et al.* (2002), voted perceptrons, see Freund & Schapire (1999); Warmuth *et al.* (2002), feedforward sigmoidal neural nets trained by backprop. Hence one may argue that a generic neural microcircuit plays a similar role for neural computing as a high dimensional kernel for support vector machines in machine learning. In addition it is demonstrated that pairs of readout neurons can transform the complex trajectory of transient states of a large neural circuit into a simple and clearly structured 2-dimensional trajectory. This 2-dimensional projection of the high-dimensional trajectory can even exhibit convergence to virtual attractors which are not apparent in the high dimensional trajectory.

2.2 Methods

I carried out computer simulations with a generic recurrent network of I&F neurons as described in Maass *et al.* (2002). The input to the network consisted of spike trains, which diverged to inject current into 30% randomly chosen excitatory neurons. The amplitudes of the input synapses were chosen from a Gaussian distribution, so that each neuron in the recurrent microcircuit received a slightly different input. The input spike trains were generated from randomly generated Poisson spike templates with a frequency of 20 Hz, where each spike in the template was moved by a Gaussian distribution with mean 0 and an SD of 4 ms.

I used randomly connected circuits consisting of I&F neurons, 20% of which were randomly chosen to be inhibitory (see Tsodyks *et al.* (2000)). Unless stated otherwise in the figure legend the circuit size was chosen to be 135 neurons. Parameters of neurons and synapses were chosen in accordance with biological data: membrane time constant 30 ms, absolute refractory period 3 ms (excitatory neurons), 2 ms (inhibitory neurons), threshold 15 mV (for a resting membrane potential assumed to be 0 mV), reset voltage 13.5 mV, constant background current at 13.5 nA, input resistance 1 MΩ.

Connectivity structure: The distribution of connection lengths was chosen to be biologically realistic, with primary local connections and a few longer connec-

tions. More precisely the probability of a synaptic connection from neuron a to neuron b (as well as that of a synaptic connection from neuron b to neuron a) was defined as $C \cdot \exp(-D(a,b)^2/\lambda^2)$, where λ is a parameter which controls both the average number of connections and the average distance between neurons that are synaptically connected (its value was fixed at 1.5 for all simulations reported in this chapter, independent of the size of the network). For the circuits consisting of 135 neurons I assumed that the neurons were located on the integer points of a $15 \times 3 \times 3$ column in space, where $D(a,b)$ is the Euclidean distance between neurons a and b. The neurons of the circuits used for the simulations for Fig. 2.2 and Fig. 2.3 were arranged in $2 \times 2 \times 3$, $3 \times 3 \times 6$, $5 \times 5 \times 4$, $5 \times 5 \times 8$, $7 \times 7 \times 8$, $7 \times 7 \times 12$ and $7 \times 7 \times 16$ columns, whereas the columns of the networks for Fig. 2.6 had the size $3 \times 3 \times 9$, $3 \times 3 \times 11$, $3 \times 3 \times 13$, $3 \times 3 \times 15$, $3 \times 3 \times 18$ and $3 \times 3 \times 21$. Depending on whether a and b were excitatory (E) or inhibitory (I), the value of C was 0.3 (EE), 0.2 (EI), 0.4 (IE), 0.1 (II).

In the case of a synaptic connection from a to b the synaptic dynamics were modeled according to the model proposed in Markram *et al.* (1998), with the synaptic parameters U (use), D (time constant for depression), F (time constant for facilitation) randomly chosen from Gaussian distributions that were based on biological data reported in Gupta *et al.* (2000) and Markram *et al.* (1998). Depending on whether a, b were excitatory (E) or inhibitory (I), the mean values of these three parameters (with D, F expressed in second, s) were chosen to be 0.5, 1.1, 0.05 (EE), 0.05, 0.125, 1.2 (EI), 0.25, 0.7, 0.02 (IE), 0.32, 0.144, 0.06 (II). The scaling parameter A (in nA) was chosen to be 75 (EE), 150 (EI), -47 (IE), -47 (II). In the case of input synapses, the parameter A had a value of 18 nA. The SD of each parameter was chosen to be 50% of its mean (with negative values replaced by values chosen from an appropriate uniform distribution). The time course of postsynaptic currents was modeled by an exponential decay $\exp(-t/\tau_s)$ with $\tau_s = 3$ ms ($\tau_s = 6$ ms) for excitatory (inhibitory) synapses. The transmission delays between liquid neurons were chosen uniformly to be 1.5 ms (EE), and 0.8 for the other connections. For each trial the initial conditions of the circuit were randomly chosen (for each neuron in the circuit the membrane voltage was set at a value drawn from the uniform distribution over the interval [13.5 mV, 15 mV].

I assumed that each readout neuron receives synaptic input from all neurons in the recurrent circuit. The current *liquid state* of the circuit (using the terminology Maass *et al.* (2002)) is defined as the n-dimensional vector of contributions of the n neurons in the circuit to the membrane potential of a generic readout neuron at time t (assuming unit size weights and static synapses for this generic readout neuron). Technically these individual contributions to the membrane potential of a generic readout neuron are the outputs of a low pass filter with a kernel that decays exponentially with a time constant of 30 ms (reflecting the assumed 30 ms membrane time constant of the readout neuron), applied to the spike trains of the n neurons in the recurrent circuit. After training, the weights of a readout neuron have no longer uniform size, and hence each readout neuron defines a different projection of the high dimensional liquid states into one dimension. Therefore strictly speaking a readout neuron does not have full access to the real intrinsic state space of the recurrent network, which consists of the membrane potential of each I&F neuron and the fraction of available synaptic efficacy R and the running value of the utilization of synaptic efficacy u (for terminology see Markram *et al.* (1998)) of each dynamical synapse. The trajectory of the recurrent neural circuit was modeled as a sequence of consecutive liquid states sampled every 20 ms. Each readout neuron defines in general a different projection of this trajectory of liquid states into a 1-dimensional trajectory, and correspondingly each pair of readout neurons defines a different projection of the high dimensional dynamics into 2 dimensions.

2.3 Results

2.3.1 Projecting the input into a larger neural circuit increases the classification power of a readout neuron

When a time-varying input, such as for example a Poisson spike train, is injected as input into a large recurrent circuit of n I&F neurons, it becomes difficult for a human observer to extract information about this input from the resulting dynamics of the circuit. The computer simulations show that in contrast to that, the dynamics of neural circuits becomes easier to classify for a readout neuron

when n is large. This effect is less surprising if one notes that the decision surface which is relevant for the decision whether an I&F neuron fires at some specific time t, can be approximated by a hyperplane in the state space of the dynamical system that models the recurrent circuit of I&F neurons. The number of degrees of freedom of this hyperplane grows with n. Furthermore an empirically well-supported result from statistical learning theory (see for example Vapnik (1998)) implies that the discrimination power of a perceptron (or hyperplane) increases when the inputs that need to be classified are first projected nonlinearly into a sufficiently high-dimensional space. This effect has been demonstrated by Jäger (2001) in the context of artificial neural networks. I would like to argue that it may also play an important role in biological neural computation, and may contribute to the large and seemingly universal computational power of recurrent neural microcircuits. It may have received little attention so far in the analysis of neural computation, because it is not observable in small circuits. It also can usually not be observed in larger models for neural circuits if their architecture has been engineered by the modeler for a particular task, since that often entails that the dynamics of that circuit is restricted to a lower dimensional subspace of its state space. For this reason I focus on the dynamics of generic models for recurrent neural microcircuits that reflect biological data in their connection statistics, and which have not been engineered for a particular purpose.

As a benchmark test for memory retrieval and pattern classification by neural circuits I considered the task illustrated in Fig. 2.1. Four Poisson spike trains were randomly generated and fixed as spike pattern templates. More precisely, two such patterns, templates 1 and 2, were fixed for the time interval from 0 to 250 ms, and two other ones, templates 3 and 4, for the second interval from 250 to 500 ms. Input spike trains over 500 ms were randomly composed of noise variations (Gaussian distribution with mean 0, SD 4 ms) of one of the templates 1, 2 in their first half followed by noisy variations of one of the templates 3, 4 in their second half. I defined that an input spike train belongs to class 1(2) if its first half was generated from the template spike train 1 (2), no matter whether its second half had been generated from. A readout neuron was required to carry out a classification at time t = 500 ms, after a noisy variation of one of the 2 spike pattern templates 3 or 4 had been sent into the circuit and had "overwritten" the

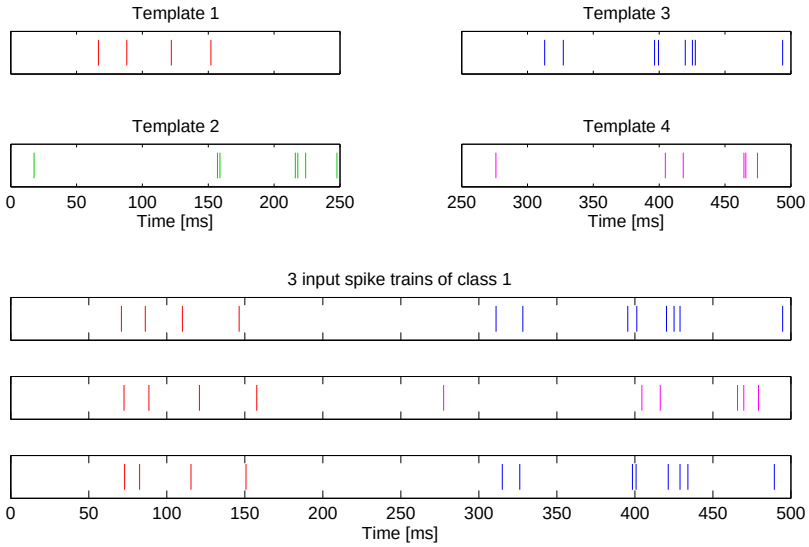


Figure 2.1: Input distribution for the classification task. For each experiment four Poisson spike trains were randomly generated and fixed as spike pattern templates 1-4. Input spike trains over 500 ms were randomly composed of noise variations (Gaussian distribution with mean 0, SD 4 ms) of one of the templates 1, 2 in their first half followed by noisy variations of one of the templates 3, 4 in their second half. An input spike train was defined to belong to class 1(2) if its first half was generated from the template spike train 1 (2), no matter whether its second half had been generated from. Three typical noisy variations of class 1 are shown in the lower part of the figure.

transient dynamics caused by the first pattern templates. One can therefore view the earlier pattern as one that sets the "context" for the second one, and it may be important from the functional point of view to recover this "context" at a later point in time. This classification task is relatively difficult, since it requires the integration of information over a time interval (and from a temporal distance) of 250 ms, which is fairly large compared to the membrane time constant of a single neuron (30 ms). For this discrimination task only the weights of the synapses of the readout neuron were adapted, thus leaving the dynamics of the recurrent neural circuits unspecialized, potentially providing unbiased input to myriads of other readout neurons that are specialized to extract other information about the input to the circuit (see Fig. 8 in Maass *et al.* (2002)). Several standard

algorithms for single-unit and multi-unit neural classifiers were applied: pools of perceptrons with the p-delta-rule, see Auer *et al.* (2002), voted perceptrons, see Freund & Schapire (1999); Warmuth *et al.* (2002), and backprop for feedforward sigmoidal neural nets.

Panels A, B and C in Fig. 2.2 show that each type of readout achieves a better performance when spike trains generated from input distributions as described above are injected into a larger recurrent circuit. Furthermore it can be seen that even for single neuron readouts (panels A, B) the classification error approaches 0 when the size of the recurrent circuit grows[1]. This effect is reminiscent of a frequently exploited effect in machine learning (more precisely in support vector machines and other kernel based methods). There one projects the given data first nonlinearly into a very high dimensional space. Within this high dimensional space the projections of the original data from different classes usually become linearly (or nearly linearly) separable, see Vapnik (1998). But an essential difference to kernel-based methods in machine learning is that there the projection into a high dimensional space is not carried out explicitly, whereas in the neural model the nonlinear projection of the input stream into the high dimensional state set of the circuit may be viewed as the essential computational operation of the generic neural microcircuit model.

For a human observer the liquid states at time $t = 500$ ms that result from input spike trains from the classes 1 and 2 look indistinguishable, like two sets of state vectors that are drawn from the uniform distribution over the state set. However readout neurons can be trained to recognize their inherent structural similarities, and are therefore able to classify also novel examples drawn from these classes. In order to demonstrate that their performance, which improves for larger sizes of the recurrent circuit, is due to this hidden structural similarity, and not to other scaling effects, the experiment were repeated with the same number of states drawn from a uniform distribution over the state set of the corresponding recurrent circuits (with a random assignment of class labels). Panel D of Fig. 2.2 shows that the readout can still be trained to classify states in the training set

[1]Further work is needed to explore when exactly this occurs. It appears to depend both on values of parameters of the circuit (e.g. λ) and on the type of inputs and the number of training examples.

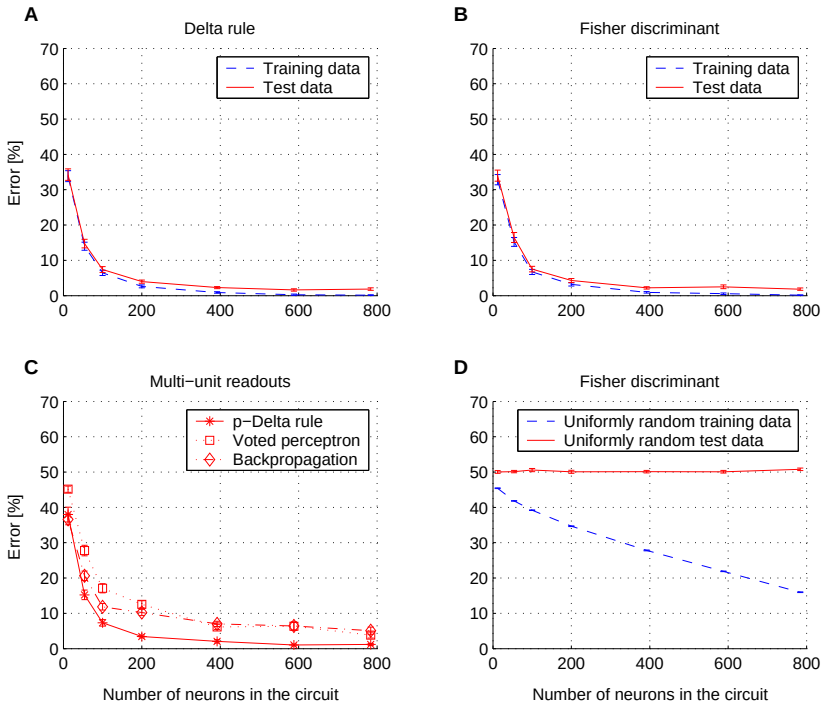


Figure 2.2: Demonstration that a larger recurrent neural circuit increases the classification power of a readout neuron. **A:** The readout was to trained to carry out a classification of the previously injected input time series as described in Fig. 1 by means of the internal state of the recurrent network at time $t = 500$ ms (after the complete spike train had been injected into recurrent circuit). The solid and dashed lines show for training and test inputs the error rates of a hyperplane (= perceptron) in the n-dimensional state space of the recurrent circuit, trained with the well-known delta learning rule (see Hertz *et al.* (1991)), as a function of the size n of the recurrent circuit of I&F neurons. For each value of n 100 randomly drawn recurrent circuits were generated. For each circuit a new input distribution was fixed and 800 input spike trains from this distribution were chosen for training, whereas 80 novel examples were used for testing. The error rates represent the averages over the performance of different circuits (error bars indicate SEM). **B:** Corresponding results for the classification task applying the Fisher-discriminant algorithm (see Duda *et al.* (2001)) instead of the delta learning rule to optimize the weights of a perceptron. The error rates were almost the same for this algorithm (whereas its computing time was much shorter).

Figure 2.2 (continued): **C:** Corresponding results for the classification task applying multi-unit readouts instead of perceptrons: p-delta learning rule (using an array of 10 perceptrons see Auer *et al.* (2002)), backpropagation (applied to 10 feedforward sigmoidal neurons), and voted perceptrons (using an array of 10 perceptrons) implemented as outlined in Warmuth *et al.* (2002). The error rates were almost the same as for perceptrons. **D:** Results of a control experiment. In order to show that a better readout performance for larger recurrent circuits can only be achieved if the complexity of the input stays constant (or grows at a lower rate than the circuit size), the same algorithms as in panel B were applied to compute an optimal separating hyperplane for randomly chosen internal states of the recurrent circuit. The training set consisted of 800 randomly drawn state vectors, drawn from the uniform distribution over all state vectors (with randomly assigned class labels) instead of 800 liquid states at time $t = 500$ ms that resulted from injecting input spike trains from a fixed distribution into these circuits. The performance on these training sets decreased with the circuit size. But in the case of such randomly labeled state sets there was no generalization possible, hence the performance on test data from the same distribution (uniform distribution over all state vectors) yielded an error of 50%.

with an error of less than 50% (but much larger than for the previously considered classes). Furthermore this error on the training set decreases with the circuit size. But in the case of such randomly labeled state sets there is no generalization possible, hence the performance on test data from the same distribution yields an error of 50%.

The test that was applied here provides a generally applicable method for quantifying the characteristic inherent similarity of states within each of two classes A, B of liquid states, even in cases where this inherent similarity of states can not be detected by a human observer. This method proposes to compare the classification performance of readouts that were trained to classify states from these two classes with that of the same type of readouts (using the same training algorithm) trained to classify states from two classes C, D of the same size, whose elements were drawn from the uniform distribution over the state space (with randomly assigned class labels). The classification performance on these two other classes will in general also be better than random guessing, since the readout can store information about these particular two sets C and D in its

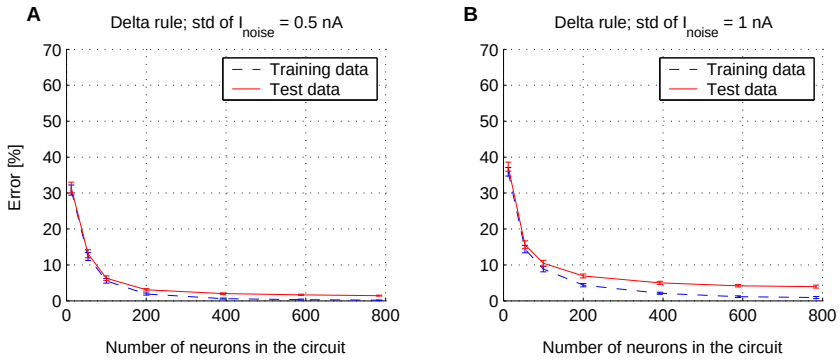


Figure 2.3: Results for the classification task of Fig. 2A with two levels of noise added to the input currents independently for all neurons in the circuit. For each simulation time step Gaussian noise with mean 0 and standard deviation 0.5 nA and 1 nA respectively was injected into the neurons. The classification error increased for the higher noise level, whereas the results showed little effect for the lower noise level.

weights. Hence its classification performance will improve with the number of weights in the readout, and hence with the size of the circuit. However if there is some structural similarity among states within one of the two classes A, B, a trained readout neuron achieves a much higher performance for classification of states from these two classes. The difference in error rates achieved for the classification of states from the two original classes A, B and the two classes C, D quantifies how much common structure the readout can extract from each of the two original classes A, B of liquid states.

In order to test the robustness of the classification capabilities of a readout neuron with regard to additional noise in the circuit I added Gaussian noise with mean 0 and standard deviation 0.5 nA and 1 nA respectively to the input currents of each neuron in the circuit at each simulation time step. This corresponds to Gaussian noise on the membrane potentials with a std of 1/52 (1/26) of the voltage difference between the threshold and the membrane potential for a constant background current of 13.5 nA (reset voltage). As illustrated in Fig. 2.3 the classification error increased for the higher noise level, whereas the results for the lower noise level were comparable to those without this extra noise.

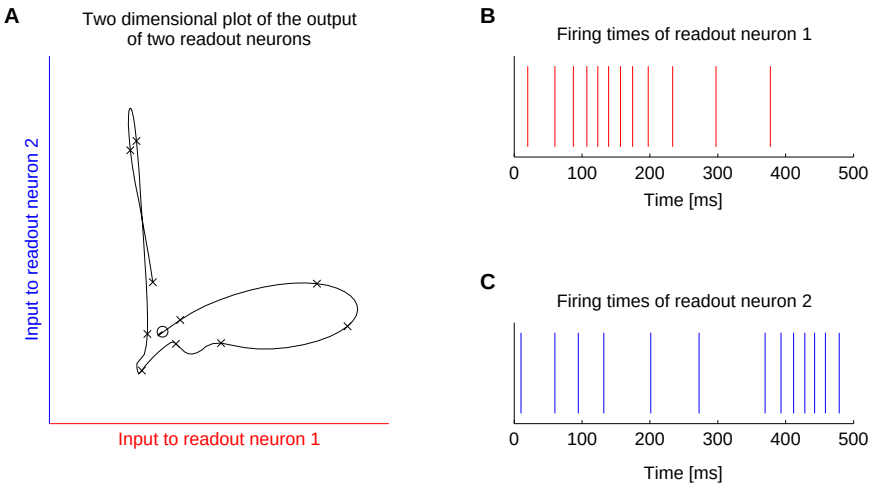


Figure 2.4: **A:** A Typical 2-dimensional projection of the high dimensional trajectory of liquid states in a recurrent neural circuit, represented by the combined synaptic inputs from the neurons in this circuit to 2 readout neurons. The beginning of the trajectory is marked by a circle, with crosses on the curve at every 50 ms interval. **B, C:** Resulting spike trains of these 2 readout neurons.

2.3.2 Finding structure in complex high dimensional trajectories of circuit dynamics

If one tracks the n-dimensional trajectory defined by the states of a recurrent circuit of n I&F neurons as a function of time, this trajectory is likely to resemble Brownian motion. However, from the point of view of readout neurons the same trajectory may have a simple and clear structure, and even converge to a "virtual attractor". The input to the recurrent circuit consists in the following always of 8 spike trains that are simultaneously injected into the circuit.

I focus on the information that *pairs* of readout neurons can extract from the high dimensional trajectory of liquid states of a recurrent neural circuit. The time course of the synaptic input to such readout neurons[1] is plotted as a curve in the plane, such as illustrated in Fig. 2.4A. The approximate structure of this 2-dimensional curve is captured by the resulting spike trains of these 2 readout

[1]More precisely: the total contribution of the neurons in the recurrent circuit to the membrane potential of these two readout neurons, with an assumed membrane time constant of 30 ms.

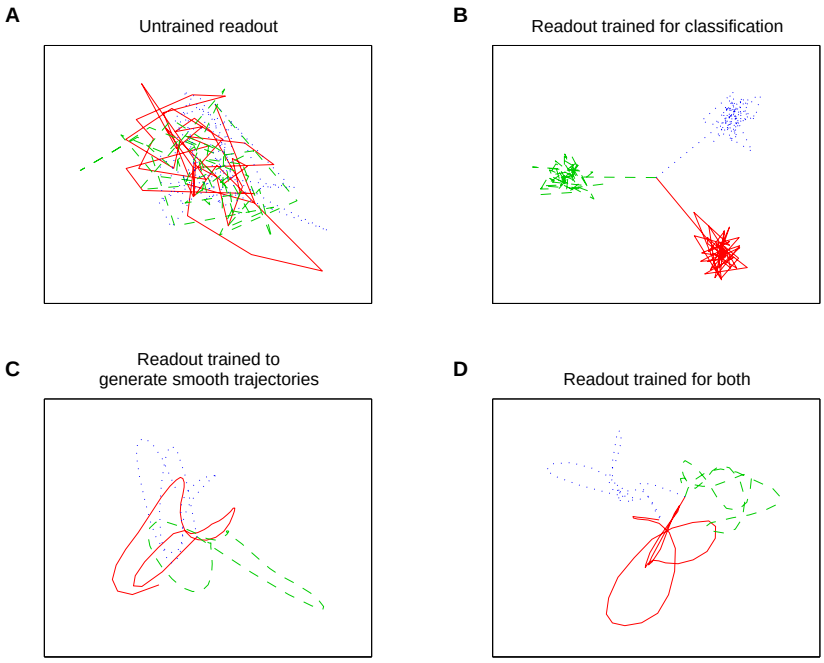


Figure 2.5: Three trajectories of the synaptic input to 2 readout neurons (indicated by solid, dashed, and dotted lines) caused by three different input Poisson spike trains to the recurrent circuit during a time interval of 1 s. For randomly assigned synaptic weights these trajectories resemble Brownian motion as shown in panel **A**. For other values of the weights the same trajectories may appear well separated in space as illustrated in panel **B**, or smooth as displayed in panel **C**. For the calculations of the weights in both cases the well-known Fisher discriminant algorithm was applied for panel B directly to the points of the trajectories and for panel C to their second time derivatives. A combination of both perspectives, enabling both classifications through clear spatial separation and smooth tracking of the trajectory (see panel D), can be achieved by applying the Fisher discriminant to the union of the two point sets used for panels B and C.

neurons (see Fig. 2.4B, C), and can therefore be transmitted to other neural circuits.

In Fig. 2.5 the input to two readout neurons is plotted in this way for three different time varying inputs to a recurrent circuit consisting of 135 I&F neurons. Depending on the choice of synaptic weights for these readout neurons, these 2-

dimensional trajectories may look like Brownian motion (Fig. 2.5A), trajectories that move fast into different attractor basins (Fig. 2.5B), smooth trajectories with a characteristic dynamical structure (Fig. 2.5C), or trajectories that are smooth and move into different attractor basins (Fig. 2.5D). The first output (Fig. 2.5A) is the typical result if the weights of the two readout neurons are randomly chosen. If one chooses the weights of these two readout neurons according to the Fisher discriminant Duda *et al.* (2001), the resulting curves move for these three time-varying inputs to three different "virtual attractor basins". With a slight variation of the Fisher discriminant (apply it to the sets of 2^{nd} order derivatives with regard to time for points on these three trajectories, not to the sets of points on these trajectories) one gets the outputs of the two readout neurons that define the curves shown in Fig. 2.5C. These curves are quite smooth and exhibit a clear temporal evolution. By applying the Fisher discriminant to the union of points on the trajectories and their second derivatives with regard to time, one gets responses of the two readout neurons that combine the effects of Fig. 2.5B and C: they move on smooth curves to different attractor basin (shown in Fig. 2.5A). Although the Fisher discriminant is usually only viewed as a global optimization procedure, the resulting setting of the weights of the two readout neurons can also be approximated by an incremental learning algorithm: the MSE-algorithm, see section 5.8.2. in Duda *et al.* (2001), which is local and unsupervised, and therefore not unrealistic from the biological point of view. For the simulations reported in this chapter I used the exact implementation of the Fisher discriminant. The results of Fig. 2.5 show that the low dimensional trajectory extracted by 2 readout neurons from a fairly large neural circuit may have little visible structural similarity with the high dimensional trajectory of transient states of that circuit, and my even move to "virtual attractors" that are not apparent from the high dimensional trajectory.

Analogously as with the classification task considered for Fig. 2.2, there exists an interesting scaling law, which prevents the observation of these effects in 2- or 3-dimensional dynamical systems, or in small models for neural circuits. The capability of pairs of readout neurons to extract smooth 2-dimensional trajectories from the very complex trajectory of firing activity in a recurrent circuit of I&F neurons increases with the number of neurons in this circuit (Fig. 2.6). The same

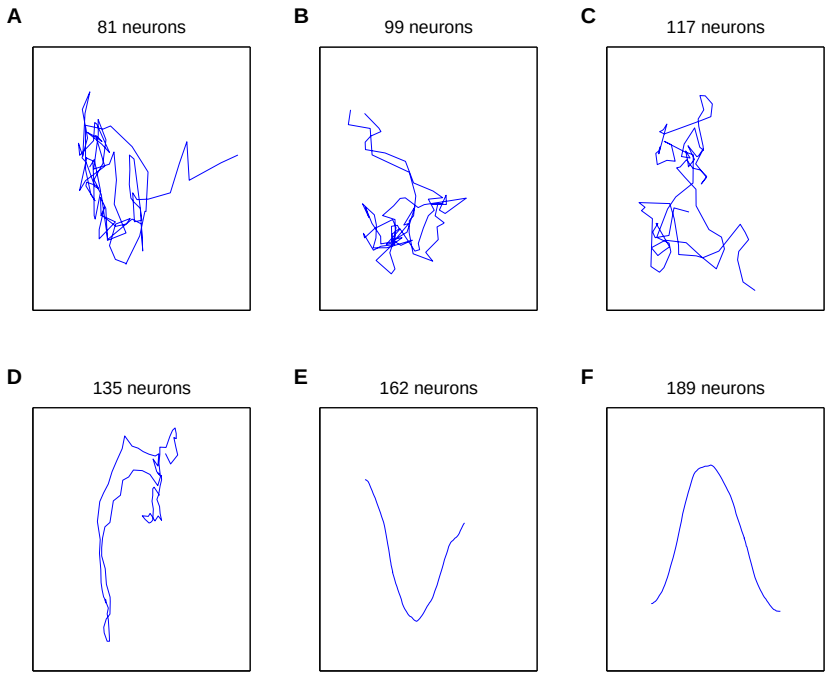


Figure 2.6: Dependence of the capability of 2 readout neurons to transform a trajectory of a recurrent circuit of n I&F neurons into a smooth low dimensional projection on the number n of neurons in the circuit. The input consisted of eight parallel injected random Poisson spike trains over a time interval of 2 s. In each case the same optimization method as for Fig. 2.5C was applied to the weights of the 2 readout neurons.

input spike train was injected into recurrent neural circuits of varying size, and for each recurrent circuit a pair of readout neurons was optimized (as for Fig. 2.5C) to generate a smooth 2-dimensional trajectory of their synaptic input. In order to show that the increased smoothness of the projections depends on the size of the network and is not due to a slow down of its dynamic the spike rasters with the responses of 10 randomly chosen neurons of each recurrent circuit to input as used for the simulations for Fig. 2.6 are illustrated in Fig. 2.7.

Fig. 2.8 shows that the smooth large scale structure that pairs of trained readout neurons can extract from a complex dynamics of high dimensional circuits, may generalize to input where one of the eight simultaneously injected spike trains was replaced by a random Poisson spike train at each trial. This

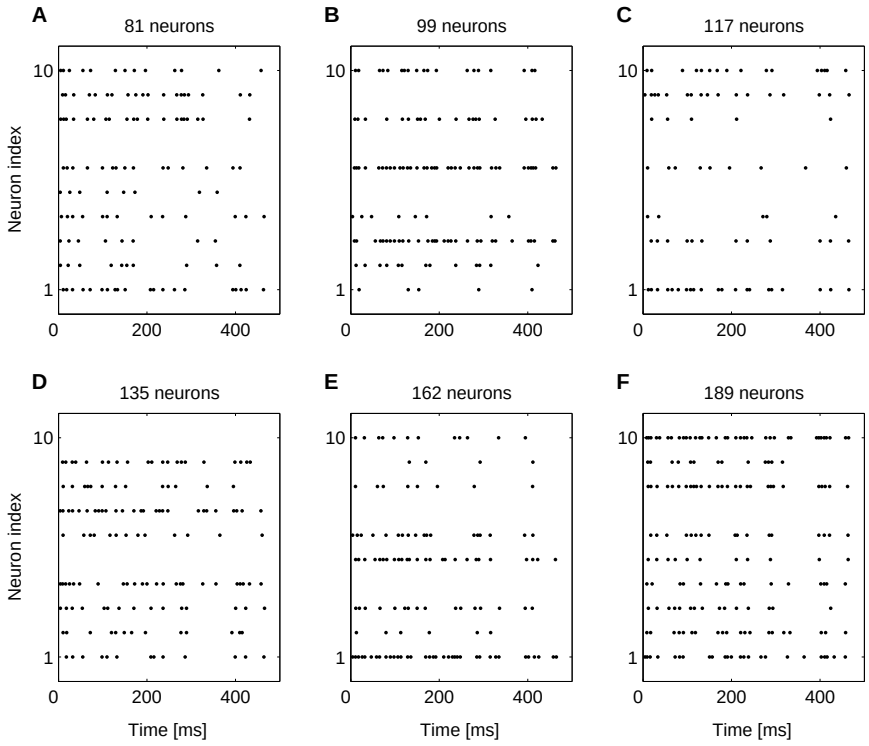


Figure 2.7: Spike rasters with the responses of 10 randomly chosen neurons from the recurrent circuits of n I&F neurons, for a circuit input as used for the simulation for Fig. 2.6. The level of activity of the these neurons does not diminish with circuit size. This shows that the increased smoothness of the low dimensional projections of the circuit trajectories shown in Fig. 2.6 is not the result of diminishing firing activity in the larger circuits.

additional input may represent independent spatio-temporal information about the environment received from other areas in the neocortex. Two readout neurons were trained to respond with smooth trajectories of similar shape to the quite diverse high dimensional dynamics caused by this random spike train. In other words the readout was trained to assign equivalence classes in the internal state space, which contained all possible internal states at time point t that could result from different previously injected random spike trains. After training the readout neurons transformed the trajectories of liquid states that resulted from input composed of the same seven fixed spike trains but a previously not seen random spike train into closely related 2-dimensional projections (panels B-E).

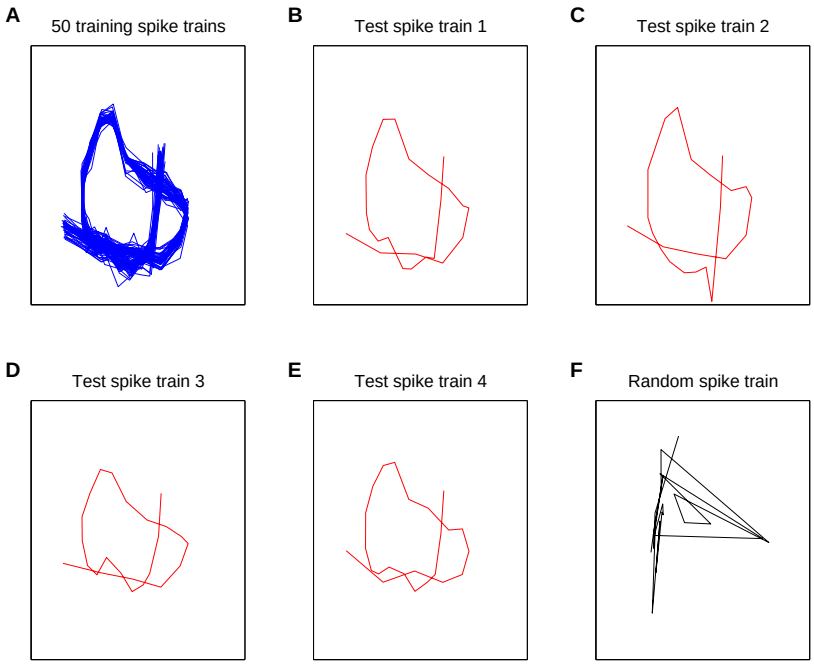


Figure 2.8: **A:** The smooth large scale structure that pairs of trained readout neurons can extract from a complex dynamics of high dimensional circuits can generalize to novel input. One of eight fixed random input spike trains of length 0.5 s that were simultaneously injected into a recurrent circuit of 135 I&F neurons was replaced by a novel random Poisson spike train at each trial. The weights of the two readout neurons were chosen (similarly as in Fig. 2.5C) in such a way that the trajectories have about the same shape from the point of view of these readout neurons. This generalized to previously not shown variations of the random input spike train (**B - E**). The trajectory caused by entirely different input to the recurrent circuit (with the same firing rate) induced a completely different temporal evolution of membrane potentials in the readout (**F**), what shows that the response of the two readout neurons was still highly selective.

However their response is still highly selective and the trajectory of liquid states caused by entirely different input to the recurrent circuit (with the same firing rate) induces a completely different temporal evolution of membrane potentials in the readout (panel F).

2.3.3 Different readouts may create diverse virtual attractor landscapes

I showed that different readout neurons can be trained to extract a diverse set of features from the same high dimensional neural activity of a sufficiently large recurrent circuit of I&F neurons. Hence for a specific information processing task it may not be necessary to manipulate this high dimensional trajectory itself. Instead, the number of degrees of freedom for readout neurons are chosen so large, that they can be trained to extract individualized smooth paths on virtual attractor landscapes from the same high dimensional circuit dynamics.

These attractors are called virtual, because they are not real attractors of the underlying dynamics but just look like attractors from the perspectives of certain low dimensional projections. Furthermore these attractors are transient, i.e. they represent temporary attractors formed within the transient behavior of the system. In other words: they are sets of states that attract certain trajectories during a certain time segment, but not permanently. Nevertheless, they may represent the result of a computation for a low dimensional readout. Hence the presence of virtual attractors makes it in principle possible to carry out particular computations needed by specific readouts without changing the dynamics of the recurrent circuit itself (thereby leaving it ready to serve as analog memory for other readouts with completely different tasks). The remarkable flexibility that remains when just the low-dimensional readouts are adapted for specific computational tasks is demonstrated in Fig. 2.9 for 3 different pairs of readout neurons. For all three panels of Fig. 2.9 the inputs to the recurrent circuit (and the resulting circuit dynamics) are identical. However the temporal evolution of the readout responses has a different large-scale structure for each of the three pairs of readout neurons. Inputs to the recurrent circuit were three different spike trains a, b, c. The first pair of readout neurons was just trained to separate the trajectories resulting from these 3 inputs by smooth responses (panel A). The second pair was trained in addition to create a virtual common attractor for patterns a and b, but not for c (panel B). The third pair of readout neurons was trained to move the responses for patterns a and c to a common attractor, while keeping the trajectory for pattern b away from this attractor. Altogether Fig. 2.9

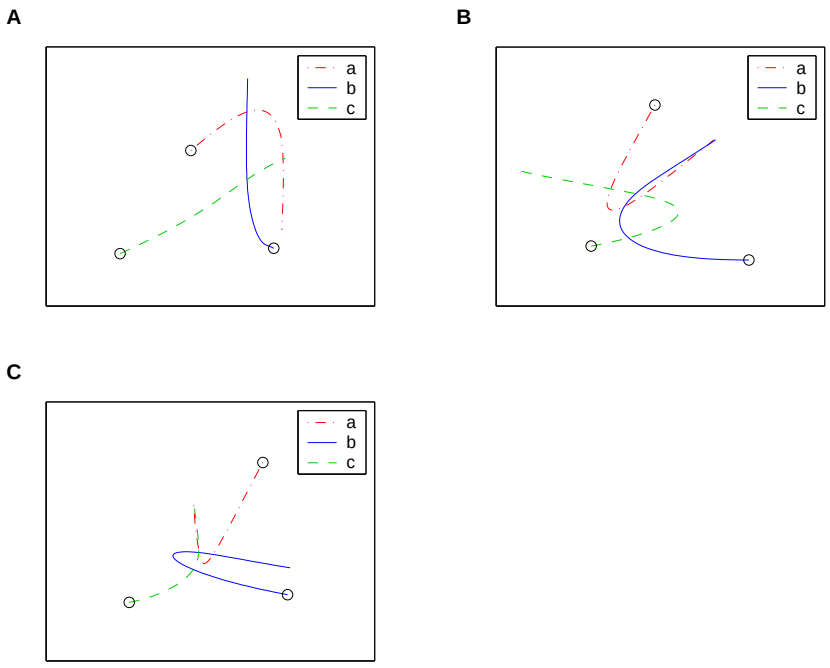


Figure 2.9: Different pairs of readout neurons may create different virtual attractor landscapes. **A:** Two readout neurons were trained to respond to the 3 trajectories of liquid states caused by the injection of 3 different input spike trains during a time interval of 1 s to a recurrent circuit of 375 I&F neurons (in a $5 \times 5 \times 15$ column) with smooth well-separated responses, like in Fig. 2.5A (beginnings of the 2 dimensional response-trajectories marked by circles). **B, C:** Responses of 2 other pairs of readout neurons to the same 3 trajectories of liquid states as in panel A. For panel B a pair of readout neurons were trained to move only for inputs a and b to a common attractor. For panel C another pair of readout neurons was trained to move only for inputs a and c a common attractor.

suggests an alternative to modeling neural dynamics by low-dimensional attractor neural networks: While the internal dynamics of a generic high-dimensional neural microcircuit may be extremely complex, different pools of readout neurons may use this high dimensional dynamics as a universal source of online information, and can be trained to extract low dimensional trajectories that move on virtual attractor landscapes. Since these virtual attractor landscapes may differ from readout to readout, one arrives in this way at a possible scheme for parallel

real-time processing with the help of high dimensional dynamical systems.

2.4 Discussion

Many tools and concepts that have been developed for the investigation of dynamical systems are very useful for analyzing low dimensional autonomous dynamical systems. However new effects have to be taken into account when one analyzes high dimensional dynamical systems such as those implemented by cortical microcircuits. Typically these systems are constantly bombarded with inputs from sensory neurons and other neural circuits, hence they are non-autonomous. Furthermore these systems have to compute in real-time, and therefore need to retrieve information for their computational tasks from trajectories of transient states of the circuit. Since the high dimensionality of the neural dynamics increases the capability of a readout neuron to select and represent specific components of the information, the high dimensional trajectory need not be engineered for a specific task. Rather, different readout neurons can extract completely different aspects for their specific task. In fact, from their point of view the high dimensional dynamics may even appear to move towards well-defined attractor basins, but this virtual attractor landscape may be a completely different one for each readout neuron.

Another beneficial aspect of the high dimensionality of the dynamics of neural microcircuits is the resulting boosting of the classification power of single readout neurons, which has been demonstrated in this chapter. This implies that very simple and robust learning algorithms, which cannot get stuck in local minima, can be used to train these readouts. The effects exhibited in this chapter may help to provide challenges and ideas for the development of a new theory of dynamical systems that is adequate for high dimensional non-autonomous systems with diverse components, and can therefore be used to analyze real-time computing in neural microcircuits. In particular I will apply the framework of the *liquid state machine* in the next chapter to the analysis of the potential universal functional properties of the neocortex. I will investigate the impact of the lamina-specific synaptic connectivity in stereotypical cortical microcircuits on the computational

performance and address the question of how certain structural features of the neocortical circuitry relate to specific functional properties.

Chapter 3

A statistical analysis of information processing properties of lamina-specific cortical microcircuit models

Summary

A major challenge for computational neuroscience is to understand the computational function of lamina-specific synaptic connection patterns in stereotypical cortical microcircuits. Previous work on this problem had focused on hypothesized specific computational roles of individual layers and connections between layers, and had tested these hypotheses through simulations of abstract neural network models. I approach this problem by studying instead the dynamical system defined by more realistic cortical microcircuit models as a whole, and by investigating the influence which its laminar structure has on the transmission and fusion of information within this dynamical system. The circuit models that I examine consist of Hodgkin-Huxley neurons with dynamic synapses, based on detailed data from Markram et al. *(1998); Thomson* et al. *(2002) and Gupta* et al. *(2000). I investigate to what extent this cortical microcircuit template supports the accumulation and fusion of information contained in generic spike inputs into layer 4 and layers 2/3, and how well it makes this information accessible to projection*

neurons in layers 2/3 and layer 5. I exhibit specific computational advantages of such data-based lamina-specific cortical microcircuit model by comparing its performance with various types of control models that have the same components and the same global statistics of neurons and synaptic connections, but are missing the lamina-specific structure of real cortical microcircuits. I conclude that computer simulations of detailed lamina-specific cortical microcircuit models provide new insight into computational consequences of anatomical and physiological data.

3.1 Introduction

The neocortex is composed of neurons in different laminae that form precisely structured microcircuits. In spite of numerous differences depending on age, cortical area, and species, many properties of these microcircuits are stereotypical, suggesting that neocortical microcircuits are variations of a common microcircuit template (Douglas & Martin (2004); Douglas *et al.* (1995); Kalisman *et al.* (2005); Mountcastle (1998); Nelson (2002); Silberberg *et al.* (2002); White (1989)). One may conjecture that such microcircuit template is distinguished by specific functional properties, which enable it to subserve the enormous computational and cognitive capabilities of the brain in a more efficient way than, for example, a randomly connected circuit with the same number of neurons and synapses. The potential computational function of laminar circuit structure has already been addressed in numerous articles, see for example Douglas & Martin (2004); Raizada & Grossberg (2003); Treves (2003) and the references in these recent publications. Treves (2003) and Raizada & Grossberg (2003) investigated specific hypotheses regarding the computational role of lamina-specific structure, and have supported these hypotheses through computer simulations of rather abstract models for neural circuits.

The results of Treves (2003) show that in this more abstract setting the laminar circuit structure yields a small advantage regarding the separation of two types of information: at which horizontal location of the circuit input has been injected ("where" information) and information about the particular pattern which had been injected there ("what" information). I complement this analysis by taking a closer look at the temporal dynamics of information at a single horizontal

location, more precisely within a single column with a diameter of about 100 μm. I find that from this perspective the computational advantage of laminar circuits is substantially larger: around 30% (depending on the specific type of information processing task), rather than just 10% as observed in Treves (2003).

The stereotypical cortical microcircuit is a highly recurrent circuit that involves numerous superimposed positive and negative feedback loops Douglas & Martin (2004). Most methods that have been developed in engineering sciences in order to design and analyze such recurrent circuits focus on the system behavior of the recurrent circuit as a whole, since it has turned out to be not feasible to understand the emergent dynamics of nonlinear recurrent circuits merely on the basis of specifications of their components. This systems-perspective of stereotypical cortical microcircuits had first been emphasized by Douglas *et al.* (1995), and had led to their definition of an abstract model of a "canonical microcircuit". It was demonstrated in Douglas *et al.* (1995) that this systems-perspective provides a new way of understanding the role of inhibition in cortical microcircuits, in particular the way in which relatively small changes of inhibitory feedback may cause large changes in the gain of the system. However the temporal dynamics of neuronal and synaptic activity had not been taken into account in these early models. Also much fewer data on stereotypical connection patterns were available at that time. Furthermore no attempt had previously been made to analyze with rigorous statistical methods emergent information processing capabilities of the resulting detailed microcircuit models. The goal of this chapter is to close this gap.

I investigate the information processing capabilities of detailed microcircuit models based on data from Thomson *et al.* (2002) on lamina-specific connection probabilities and connection strengths between excitatory and inhibitory neurons in layers 2/3, 4, 5, and on 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. The analysis is based on the assumption that stereotypical cortical microcircuits have some "universal" computational capabilities, and can carry out quite different computations in diverse cortical areas. Consequently it concentrates on the generic information processing capability to hold

and fuse information contained in Poisson input spike trains from two different sources (modeling thalamic or cortical feedforward input into layer 4, and lateral or top-down input into layers 2/3). In addition I examined the capability of such circuit models to carry out linear and nonlinear computations on time-varying firing rates of these two afferent input streams. In order to avoid – necessarily quite biased – assumptions about the neuronal encoding of the results of such computations, I analyzed the information which is available about the results of such computations to the generic "neural users", i.e., to pyramidal neurons in layers 2/3 (which typically project to higher cortical areas) and to pyramidal neurons in layer 5 (which typically project to lower cortical areas or to subcortical structures, but also project for example from V1 back to nonspecific thalamus , i.e. to the intralaminar and midline nuclei that do not receive direct primary sensory input, and through this relay to higher cortical areas, see Callaway (2004)).

In contrast to the model in Maass *et al.* (2002) (see Destexhe & Marder (2004) for a discussion) I have not used simply linear regression to estimate the information available to such readout neurons, whose output is modeled by a weighted sum of postsynaptic potentials (with an exponential decay time constant of 15 ms) in response to spikes from presynaptic neurons. Rather, I added here the constraint that the contribution of an excitatory (inhibitory) presynaptic neuron needs to have a positive (negative) weight in such weighted sum. In addition I took into account that a readout neuron in layers 2/3 or layer 5 only receives synaptic inputs from a rather small subset of neurons in the microcircuit according to the data of Thomson *et al.* (2002) (which imply that in a circuit of 560 neurons a neuron in layers 2/3 has on average 84 presynaptic neurons, and a neuron in layer 5 has on average 109 presynaptic neurons, see Fig. 3.1). But as in the earlier model the parameters of synapses were not modified within the circuit for specific computational tasks, only the weights of synaptic connections to such symbolic readout neurons in layers 2/3 and 5 (which were not modeled to be part of the circuit, in the sense that they did not project back into the circuit). [1]

[1]This simplification was made in this chapter for pragmatic reasons, since first results on the case with feedback Maass *et al.* (2006) suggest that it requires a separate analysis.

3.2 Methods

The currently most complete set of data on connection probabilities and efficacies of synaptic connections between 6 specific populations of neurons in cortical microcircuits (excitatory and inhibitory neurons in layers 2/3, 4, 5) has been assembled in Thomson *et al.* (2002). Intracellular recordings with sharp electrodes from 998 pairs of identified neurons were made to assemble these data. 679 paired recordings were made from somatosensory, motor and visual areas of adult rats, and 319 from visual areas in adult cats. The sampling was made randomly within a lateral spread of $50-100\,\mu m$ Thomson (2005). For those pairs where both data from rat and from cat are given in Thomson *et al.* (2002), I took the data from rat (see Fig. 3.1). Only for pairs of neurons within layer 4 no data from rat are given in Thomson *et al.* (2002), hence the corresponding data in Fig. 3.1 are from cat. [1]

The short term dynamics of cortical synapses (i.e., their specific mixture of paired pulse depression and paired pulse facilitation) is known to depend on the type of the presynaptic and postsynaptic neuron (see for example Gupta *et al.* (2000); Markram *et al.* (1998); Thomson (2003)). I modeled this short term synaptic dynamics according to the model proposed in Markram *et al.* (1998), with synaptic parameters U, D, F. The model predicts the amplitude A_k of the PSP for the k^{th} spike in a spike train with interspike intervals $\Delta_1, \Delta_2, \ldots, \Delta_{k-1}$ through the recursive equations

$$\begin{aligned} A_k &= w \cdot u_k \cdot R_k \\ u_k &= U + u_{k-1}(1-U)exp(-\Delta_{k-1}/F) \\ R_k &= 1 + (R_{k-1} - u_{k-1}R_{k-1} - 1)exp(-\Delta_{k-1}/D) \end{aligned} \tag{3.1}$$

with hidden dynamic variables $u \in [0,1]$ and $R \in [0,1]$ whose initial values for the first spike are $u_1 = U$ and $R_1 = 1$ (see Maass & Markram (2002) for

[1] Some of the pairings were rarely observed and the corresponding entries suffer from small sample size (see Thomson *et al.* (2002) for details). Also very small neurons in rat may have been missed Thomson (2005). In addition it is possible that in some cortical microcircuits connections exist between pairs of neurons for which no connections were reported in Thomson *et al.* (2002) (see for example Dantzker & Callaway (2000) for the case of connections to inhibitory neurons in layers 2/3).

a justification of this version of the equations, which corrects a small error in Markram *et al.* (1998)). The deterministic synapse model is designed to model the average sum of postsynaptic responses resulting from the concerted action of multiple stochastic synaptic release sites. Results reported in section 3 show that the inclusion of short term synaptic plasticity has a significant impact on the information processing capability of the circuit models. [1]

The parameters U, D, and F were chosen in the computer model from Gaussian distributions that reflect data reported in Markram *et al.* (1998) and Gupta *et al.* (2000) for each type of connection (note that the parameter U is according to Markram *et al.* (1998) largely determined by the initial release probability of the synaptic release sites involved). Depending on whether the input was excitatory (E) or inhibitory (I), the mean values of these three parameters U, D, F (with D, F expressed in seconds) were chosen to have the mean values that were reported in these articles (see Table 3.1). The standard deviation SD of each parameter was chosen to be 50% of its mean (with negative values replaced by values chosen from an uniform distribution between 0 and two times the mean).

The microcircuit models that I examined consisted of three layers, with 30%, 20% and 50% of the neurons assigned to layers 2/3, layer 4 and layer 5, respectively. Each layer consisted of a population of excitatory neurons and a population of inhibitory neurons with a ratio of 4:1. Synaptic connections between the neurons in any pair of the resulting 6 populations were randomly generated in accordance with the empirical data from Table 3.1 and Fig. 3.1. Most circuits that were simulated consisted of 560 neurons. The mean number of presynaptic neurons from a neuron in such circuit was then 76, yielding altogether an average of 42594 synapses in the circuit.

[1] Long term synaptic plasticity within the simulated circuit was not included in this study for pragmatic reasons because of the additional complex issues involved, but will be addressed in subsequent studies.

[1] Connections from L2/3-I to L5-E are reported in Thomson *et al.* (2002), but are discussed only qualitatively. Hence the entry for connections from L2/3-I to L5-E (marked by a question mark) is only an extrapolation. The same applies to connections from L4-I to L2/3-I. No data on the amplitudes of IPSPs from L5-I to L5-I are given in Thomson *et al.* (2002), hence the corresponding entry is just a guess.

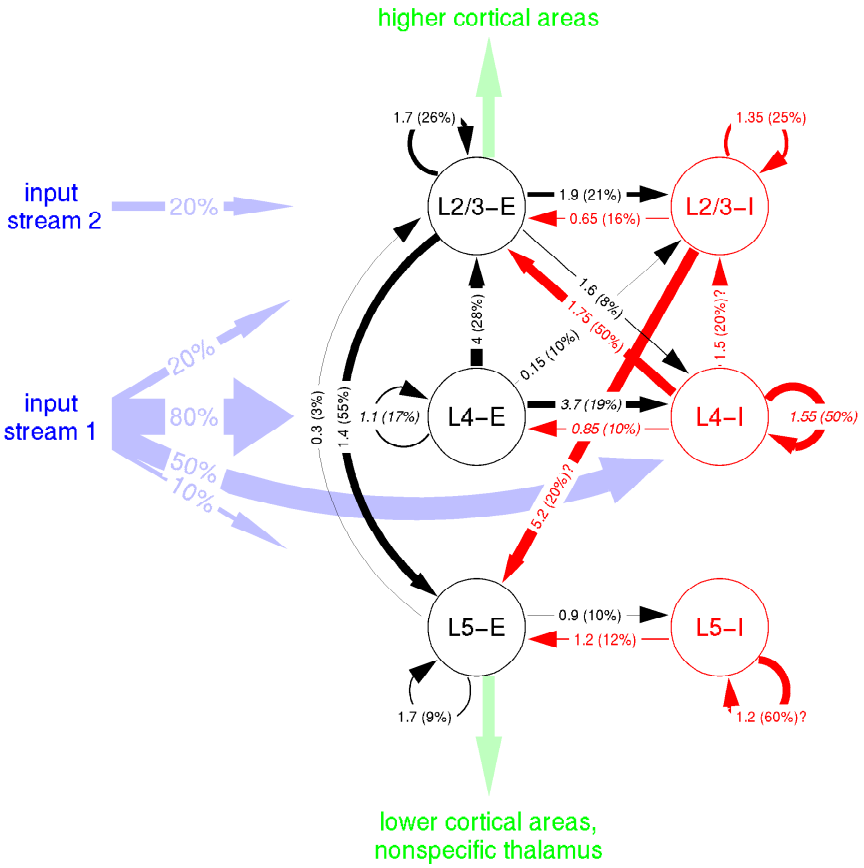


Figure 3.1: Cortical microcircuit template. Numbers at arrows denote connection strengths (mean amplitude of postsynaptic potentials, PSPs, measured at soma in mV) and connection probabilities (in parentheses) according to Thomson *et al.* (2002), for connections between cortical neurons in 3 different layers, each consisting of an excitatory (E) and an inhibitory (I) population, with an estimated maximal horizontal distance of up to 100 μm. Most of the data are from rat cortex, except for interconnections in layer 4 (italic), which are from cat.[1]Percentages at input streams denote connection probabilities for input neurons used in the simulations. In addition each neuron receives background noise reflecting the synaptic inputs from a large number of more distal neurons (see Methods).

As models for excitatory and inhibitory neurons I chose conductance based single compartment Hodgkin-Huxley neuron models with passive and active properties modeled according to Destexhe *et al.* (2001) and Destexhe & Pare (1999).

from/to	E	I
E	**0.5, 1.1, 0.05**	**0.05, 0.125, 1.2**
I	**0.25, 0.7, 0.02**	**0.32, 0.144, 0.06**

Table 3.1: Synaptic parameters that scale the short term dynamics of synapses according to the type (excitatory or inhibitory) of the pre- and postsynaptic neuron: Mean values of U, D, F according to Markram *et al.* (1998) and Gupta *et al.* (2000).

In accordance with experimental data on neocortical and hippocampal pyramidal neurons (Hoffman *et al.* (1997); Magee *et al.* (1998); Magee & Johnston (1995); Stuart & Sakmann (1994)) the active currents comprise a voltage dependent Na^+ current (Traub & Miles (1991)) and a delayed rectifier K^+ current (Traub & Miles (1991)). For excitatory neurons a non-inactivating K^+ current (Mainen *et al.* (1995)) responsible for spike frequency adaptation was included in the model. The peak conductance densities for the Na^+ current and delayed rectifier K^+ current were chosen to be $500pS/\mu m^2$ and $100pS/\mu m^2$ respectively, and the peak conductance density for the non-inactivating K^+ current was chosen to be $5pS/\mu m^2$. The membrane area of the neuron was set to be 34636 μm^2 as in Destexhe *et al.* (2001). For each simulation the initial conditions of each neuron, i.e. the membrane voltage at time $t = 0$, were drawn randomly (uniform distribution) from the interval [-70, -60] mV.

A cortical neuron receives synaptic inputs not only from immediately adjacent neurons (which were modeled explicitly in the computer model), but also smaller background input currents from a large number of more distal neurons. In fact, intracellular recordings in awake animals suggest that neocortical neurons are subject to an intense bombardment with background synaptic inputs, causing a depolarization of the membrane potential and a lower input resistance commonly referred to as 'high conductance state' (for a review see Destexhe *et al.* (2003)). This was reflected in the computer model by background input currents that were injected into each neuron (in addition to explicitly modeled synaptic inputs from afferent connections and from neurons within the circuit). The conductances of these background currents were modeled according to Destexhe *et al.* (2001) as a one-variable stochastic process similar to an Ornstein-Uhlenbeck

process with mean $g_e = 0.012\mu$S and $g_i = 0.057\mu$S, variance $\sigma_e = 0.003\mu$S and $\sigma_i = 0.0066\mu$S, and time constants $\tau_e = 2.7$ ms and $\tau_i = 10.5$ ms, where the indices e/i refer to excitatory and inhibitory background input conductances, respectively. According to Destexhe *et al.* (2001) this model captures the spectral and amplitude characteristics of the input conductances of a detailed biophysical model of a neocortical pyramidal cell that was matched to intracellular recordings in cat parietal cortex *in vivo*. Furthermore the ratio of the average contributions of excitatory and inhibitory background conductances was chosen to be 5 in accordance with experimental studies during sensory responses (Anderson *et al.* (2000); Borg-Graham *et al.* (1998); Hirsch *et al.* (1998)).

The maximum conductances of the synapses were chosen from a Gaussian distribution with a SD of 70% of its mean (with negative values replaced by values chosen from an uniform distribution between 0 and two times the mean). The mean maximum conductances of the synapses were chosen to reproduce the mean amplitude of PSPs given in Fig. 3.1 at the resting membrane potential (in the presence of synaptic background activity).

Two afferent input streams, each consisting of either 4 or 40 spike trains (i.e., 4 or 40 input channels), were injected into the circuit. Each of the channels of the first input stream (representing thalamic, or feedforward cortical input) was injected mainly into layer 4, i.e. to 50% of its inhibitory neurons and 80% of its excitatory neurons, but also into 20% of the excitatory neurons in layer 2/3 and 10% of the excitatory neurons in layer 5 (all randomly chosen). [1] The average number of inputs converging to an excitatory neuron in layer 4 is therefore 3.2 or 32. [2] This is roughly in the range suggested by experimental measurements of the

[1] This input distribution reflects qualitatively the evidence cited in ch. III of White (1989) that "thalamocortical afferents to layer 4 synapse not only with layer 4 nonpyramidal neurons, but also with a wide variety of both pyramidal and nonpyramidal neuronal types whose cell bodies occur throughout layers 2 to 6".

[2] Computer simulations suggest that smaller connection probabilities from external input neurons can be chosen if the amplitudes of resulting PSPs are scaled up accordingly. For the case of 40 input channels I carried out simulations with lower input connectivity for input stream 1 while keeping the product of PSP amplitude and connection probability constant. The results about performance differences between data-based circuits and amorphous control circuits (see Table 3.2) are largely invariant to these changes, even if the connection probabilities for external input neurons are scaled down to $1/5^{th}$ of the previously given values.

variability of EPSPs in simple cells of cat visual cortex with varying levels of LGN stimulation (Ferster (1987)) and cross-correlation experiments between monosynaptically linked cells of the LGN and cat visual cortex (Tanaka (1983)), which suggest that at least 10 LGN cells provide input to each simple cell. The mean conductance of input synapses was chosen to generate a PSP with a mean amplitude of 1.9 mV at the resting membrane potential (in the presence of synaptic background activity). This value corresponds to the lower bound of the estimate of geniculate input to a single neuron in layer 4 of adult cats given in Chung & Ferster (1998). It was multiplied in the simulations with a scaling parameter S_{I1} that reflects the biologically unrealistic number of input neurons in these simulations (see discussion below). Each of the channels of the second afferent input stream was injected into 20% of the excitatory neurons in layers 2/3 (also with a mean amplitude of 1.9 mV, multiplied with another scaling parameter S_{I2}).

Altogether there remain 3 parameters for which values have to be chosen in order to arrive at functional computer models of cortical microcircuits. These parameters S_{RW}, S_{I1}, S_{I2} scale (in the form of multiplicative factors) the amplitudes of PSPs for all synaptic connections within the circuit ("recurrent weights"), the amplitudes of PSPs from input stream 1, and the amplitudes of EPSPs from input stream 2. They have to be chosen in such a way that they account for the difference in scale between the simulated microcircuits and biological cortical microcircuits. Values for these 3 parameters cannot be read off from the previously mentioned data, and one has to suspect that adequate values depend also on the species, on the specific cortical microcircuit *in vivo* that one wants to model, on the current state of various homeostatic processes, on the current behavioral state (including attention) of the organism, and on the intensity of the current afferent input.

The parameter S_{I1} was chosen so that the afferent input stream 1 (consisting of 40 Poisson spike trains at 20 Hz) caused (without input stream 2 and without recurrent connections, i.e. S_{RW} set to 0) an average firing rate of 15 Hz in layer 4. The parameter S_{I2} was analogously chosen so that the afferent input stream 2 (generated like input stream 1) caused an average firing rate of 10 Hz in layers 2/3. In either case only one of the two input streams was activated. With this procedure I obtained $S_{I1} = 14$ and $S_{I2} = 33$. For simulations with input streams

consisting of 4 Poisson spike trains these values were multiplied by 10. The input synapses were chosen to be static, i.e. the synaptic parameters were set to $U = 1$, $D = 0$, and $F = 0$, and their maximum conductances were chosen from a Gaussian distribution with a SD of 70% of its mean (with negative values replaced by values chosen from an uniform distribution between 0 and two times the mean).

The parameter S_{RW} accounts for the average number of synaptic inputs to a neuron from other neurons in the circuit (apart from the globally modeled background synaptic input, see above), hence for the difference in circuit size between the simulated microcircuit models and a real cortical microcircuit. It turned out that a value of 60000/(number of neurons in the simulated circuit) for S_{RW} produced in layer 5 of the simulated circuit for the standard values of S_{I1} and S_{I2} a realistic low but significant firing activity of 8.5 Hz (see Fig. 3.2), hence this value was used as standard value for S_{RW}. This value scales the average number 76 of presynaptic neurons in a circuit of 560 neurons up to 107 times that value, yielding thereby an average of 8132 presynaptic neurons. This number is consistent with the estimates for the total number of synapses on a neuron given in Binzegger *et al.* (2004), that range from 2981 to 13075 for different cell types in cat visual cortex. Some additional synaptic input was modeled by background synaptic input (see above).

These standard values of the parameters S_{RW}, S_{I1}, S_{I2} were used throughout the computer experiments, except for the results reported in Fig. 3.10 and 3.11, where I analyzed the impact of these parameters on the reported results. In this case I simulated circuit models with randomly chosen values from independent uniform distributions over the interval $[0.1 \times \text{standard value}, 3.1 \times \text{standard value}]$ for all three parameters.

The computer simulations examined how much information about each preceding temporal segment (of length 30 ms) of each of the two input streams was accessible to a hypothetical projection neuron in layers 2/3 and to a hypothetical

[1]The population firing rates of different layers are somewhat, but not totally correlated. The maximum correlation coefficient between the population firing rates of layer 2/3 and layer 4, layer 2/3 and layer 5, and layer 4 and layer 5 (for $t > 150$ ms, 1 ms bin size and arbitrary lag) is 0.62, 0.65 and 0.56, respectively.

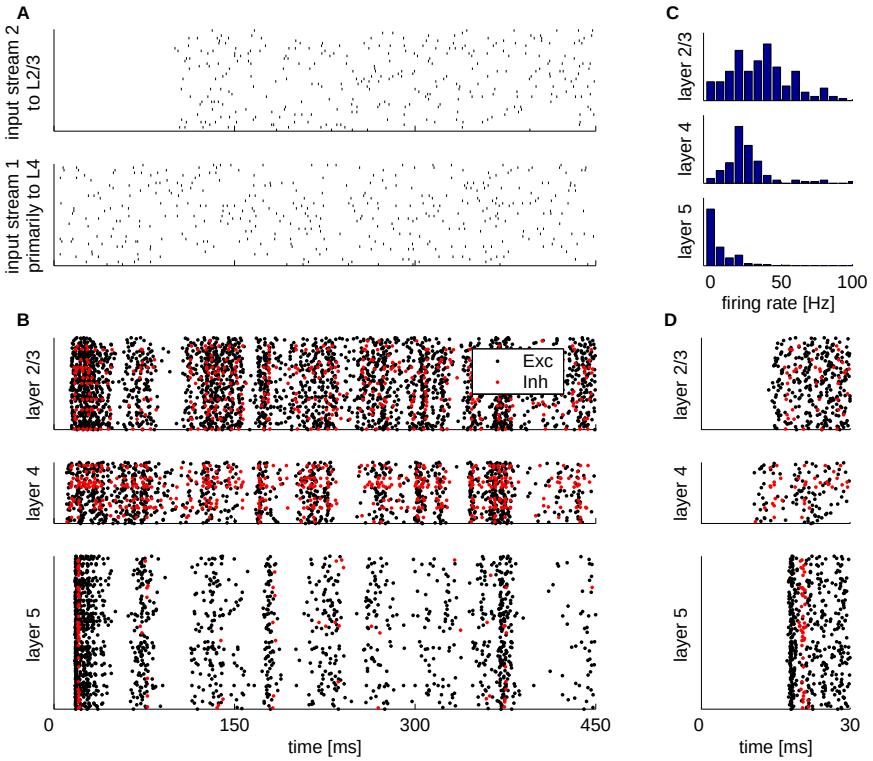


Figure 3.2: **A:** Two input streams consisting each of 40 Poisson spike trains (the input to layers 2/3 starts here 100 ms later). **B:** Spike raster of data-based cortical microcircuit model (consisting of 560 neurons) for the two input streams shown in A. The vertical dimension is scaled according to the number of neurons in each layer. Spikes of inhibitory neurons are indicated in red. [1]**C:** Distribution of firing rates in B (after onset of input into layers 2/3 and layer 4), showing an automatically emerging exponential distribution of firing rates in layer 5. Mean firing rate: 8.5 Hz. **D:** Enlargement of the initial time segment, showing a spread of excitation from layer 4 to superficial and deep layers that qualitatively matches data from Armstrong-James *et al.* (1992).

projection neuron in layer 5. The excitatory and inhibitory presynaptic neurons for such a hypothetical readout neuron were randomly chosen in the same manner as for any other excitatory neuron in that layer (i.e., according to Fig. 3.1), but no synaptic connections from a readout neuron back into the circuit were included. This amounted to an average of 84 presynaptic neurons for a readout neuron in layers 2/3, and 109 presynaptic neurons for a readout neuron in layer 5. The

weights of synaptic connections from these presynaptic neurons were optimized for specific tasks. In contrast to the simulations discussed in Maass *et al.* (2002), the resulting number of inputs to such readout neuron was much smaller than the circuit size. The projection or readout neurons themselves were modeled as linear neurons, i.e., their output was a weighted sum of low pass filtered spikes (exponential decay with a time constant of 15 ms, modeling the time constants of synaptic receptors and membrane of a readout neuron). Care was taken to make sure that weights from excitatory (inhibitory) presynaptic neurons could not become negative (positive). For this purpose I used the linear least squares method with non-negativity constraints Lawson & Hanson (1974) to optimize the weights for a particular task. This is in contrast to the linear regression which was used in Maass *et al.* (2002). For each training or test example, which consisted of an input and a target value for the readout neuron, a simulation of the microcircuit model was performed. Each input for the readout neuron was generated by collecting the low pass filtered version of the presynaptic spike trains to the readout neuron at time point $t = 450$ ms. Each corresponding target value was calculated for the various tasks described below (see Results). In order to correctly apply the linear least squares method with non-negativity constraints the spike trains of inhibitory (excitatory) neurons were convolved with negative (positive) exponential kernels and the corresponding readout weights were multiplied by -1 (+1) after training. For classification tasks the linear readout neuron was trained to output the class labels, i.e. 0 or 1, whereas a classification was obtained by thresholding the output at 0.5 (analogous to the firing threshold of a real cortical neuron). This algorithm yields a weight vector $< w_1, ..., w_n >$ with $w_i \geq 0$ if the i^{th} presynaptic neuron of the readout is excitatory, and $w_i \leq 0$ if the i^{th} presynaptic neuron is inhibitory. Within these (linear) constraints this restricted form of linear regression minimizes the error of the readout on the training examples. [1] This typically resulted in an assignment of weight 0 (corresponding to a silent synapse in a biological circuit) to about 2/3 of these synapses. Hence a typical readout neuron had less than 40 nonzero weights, and therefore a much smaller

[1] In MATLAB one can execute this optimization algorithm through the command LSQNONNEG.

capability to extract information in comparison with the model considered in Maass *et al.* (2002).

For each computer simulation at least 10 circuits were generated. For the experiments shown in Fig. 3.5 and Fig. 3.8 20 circuits were used. New spike templates were generated each time when a new circuit was drawn, in order to avoid accidental dependencies on properties of specific spike templates. For the training of the readout neurons 1500 simulations were performed with randomly drawn Poisson inputs over 450 ms. Furthermore 300 simulations with new randomly generated inputs were used for testing. The error bars in the figures denote standard errors. All performance results in this chapter (except for some diagnostic results reported in Fig. 3.8, see legend) are for test inputs that had not been used for the training of readouts, and freshly generated random initial conditions and background noise for all neurons in the circuit.

All simulations were carried out with the CSIM software Natschläger *et al.* (2003) in combination with MATLAB.

3.3 Results

Injection of two input streams consisting of Poisson spike trains into layer 4 and into layers 2/3 of the microcircuit model produced a response (see Fig. 3.2) whose successive onset in different layers qualitatively matches data on cortical microcircuits *in vivo* (Armstrong-James *et al.* (1992)). In addition the firing rates in layer 5 automatically acquire a biologically realistic exponential distribution (see e.g. Amit & Brunel (1997); Baddeley *et al.* (1997); v. Vreeswijk & Sompolinsky (1998, 1996)). [1] Fig. 3.3 gives an impression of the fairly large trial to trial variability of firing activity within the circuit for the same spike input patterns, which resulted from jitter in the spike input (top row) and internal noise (bottom row) due to the injection of randomly varying background input currents to all neurons in order to model in-vivo conditions (see Methods). [2] Hence the

[1] The distribution of firing rates in layers 2/3 and 4 reflects the typical rate distribution of Poisson spike trains that was induced by the Poisson input to these layers.

[2] In addition, for all subsequently considered computational tasks independently chosen spike patterns had been previously injected as afferent inputs, causing a fairly large variance of

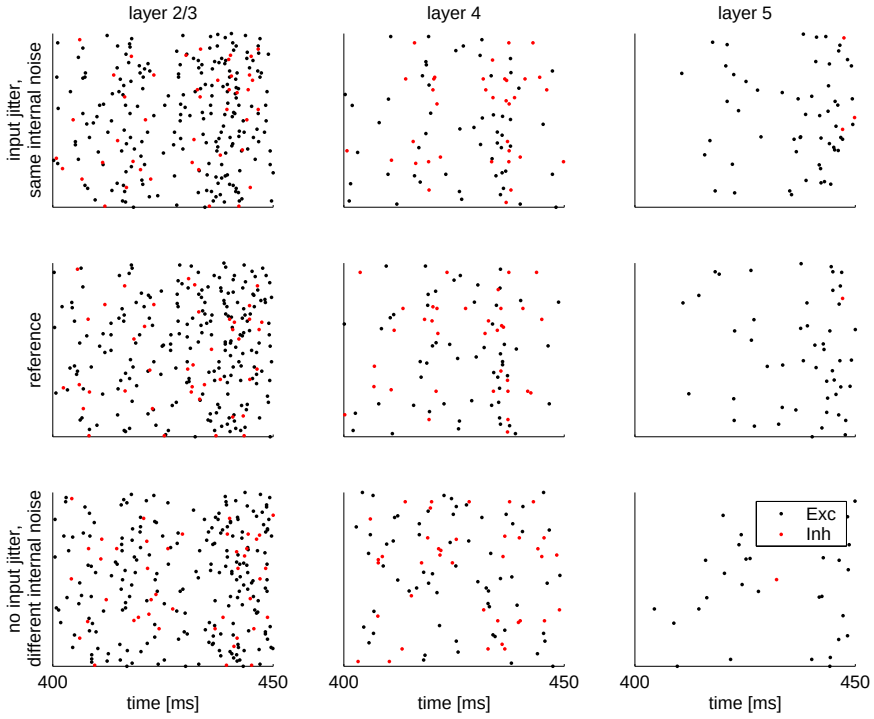


Figure 3.3: Impact of temporal jitter of input spikes (Gaussian distribution with mean 0 and SD 1 ms) and background noise of neurons in the circuit on the circuit response (see Methods). The rows in the middle and at the top show the spike rasters resulting from two trials with identical background noise and with input spike patterns that were identical except for their temporal jitter. The bottom row shows how much the spike raster for a trial with novel background noise and identical input spike patterns differs from that for the trial shown in the middle row. This illustrates that the simulated circuits (which were subject to both sources of noise) reflect qualitatively the commonly observed large trial-to-trial variability of neural responses *in vivo* to repetitions of the same stimulus.

simulated circuits reflect qualitatively the commonly observed large trial-to-trial variability of neural responses *in vivo* to repetitions of the same stimulus.

I tested these microcircuit models on a variety of generic information processing tasks that are likely to be related to actual computational tasks of neural microcircuits in cortex:

initial states of dynamic synapses.

- classification of spike patterns in either of the two afferent input streams (requiring invariance to noise and to spike input from the other input stream)
- temporal integration of information contained in such spike patterns
- fusion of information from spike patterns in both input streams in a nonlinear fashion (related to "binding" tasks)
- real-time computations on the firing rates from both input streams.

For information processing tasks with spike patterns I randomly generated spike pattern templates consisting of 30 ms segments of 40 Poisson spike trains at 20 Hz (see Fig. 3.4). More precisely the spike trains of each of the two input streams were of length 450 ms and consisted of 15 consecutive time segments of length 30 ms. For each segment 2 spike pattern templates were generated randomly. For the actual input one of the two templates of each time segment was chosen randomly (with equal probability) and a noisy variation of it, where each spike was moved by an amount drawn from a Gaussian distribution with mean 0 and SD 1 ms (see the panel on the right hand side of Fig. 3.4), was injected into the circuit. Such temporal jitter in the spike input causes significant changes in the circuit response (see Fig. 3.3), and it is a nontrivial task for readout neurons to classify spike patterns in spite of this fairly large trial-to-trial variability of the circuit response. I also tested retroactive classification of preceding spike patterns that had been injected 30 ms before, and hence were "overwritten" by independently chosen subsequent spike patterns. Furthermore a nonlinear XOR computation on spike patterns in the two concurrent input streams was examined in order to test the capability of the circuit to extract and combine information from both input streams in a nonlinear manner. The task is to compute the exclusive-or (XOR) [1] of the two bits that represent the labels of the two templates from which the most recent spike patterns in the two input streams had been generated (for example, its target output is 1 for both time segments for the input shown on the right hand side of Fig. 3.4). Note that this computation

[1] The XOR outputs 1 if exactly one of its two input bits has value 1, it outputs 0 if the input bits are 00 or 11.

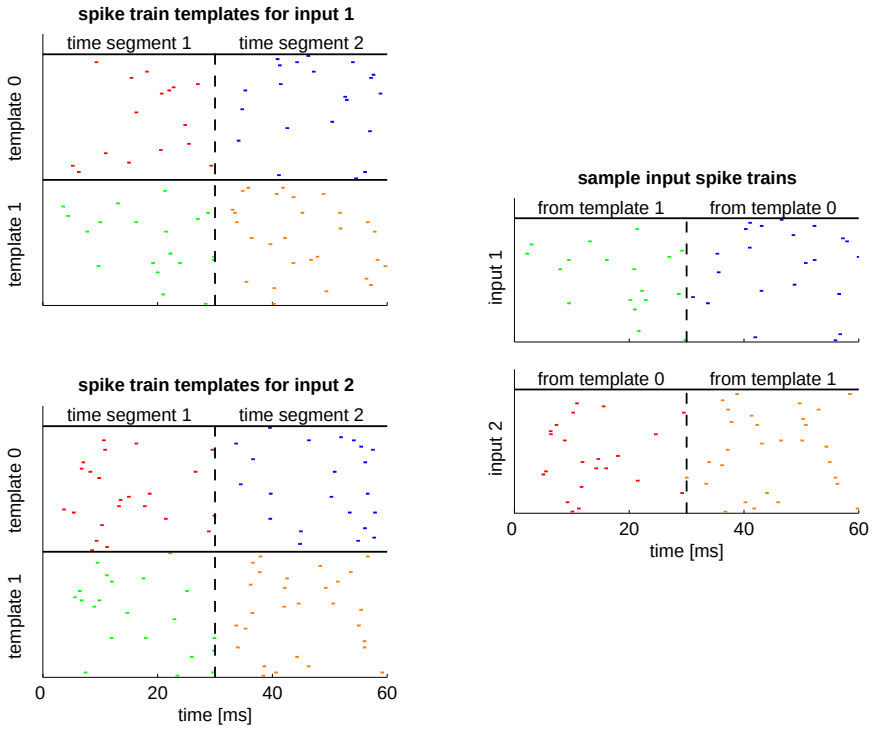


Figure 3.4: Input distributions for the spike pattern classification and exclusive-or (XOR) tasks. The task is to compute the XOR of the two bits that represent the labels of the two templates from which the most recent spike patterns in the two input streams had been generated, i.e. classify as 1 if their template labels are different and 0 otherwise. The spike trains of each of the two input streams were of length 450 ms and consisted of 15 time segments of length 30 ms. For each segment 2 templates were generated randomly (40 Poisson spike trains at 20 Hz). The actual spike trains of each input of length 450 ms used for training or testing were generated by choosing for each segment one of the two previously chosen associated templates, and then generating a jittered version by moving each spike by an amount drawn from a Gaussian distribution with mean 0 and a SD 1 ms (a sample is shown in the panel on the right hand side).

involves a nonlinear "binding" operation on spike patterns, since it has to give a low output value if and only if either noisy versions of the spike templates with label 1 appeared both in input streams 1 and 2, or if noisy version of the spike templates with label 0 appeared both in input stream 1 and 2.

In addition I analyzed linear and nonlinear computations on time-varying firing rates of the two input streams. The spike trains of each of the two input

streams were of length 450 ms and consisted of 15 consecutive time segments of length 30 ms. For each input stream and each time segment 4 Poisson spike trains were generated with a randomly chosen frequency between 15 Hz and 25 Hz. The actual firing rates used for the computations on these input firing rates were calculated from these spike trains with a sliding window of 15 ms width. I used input streams consisting of just 4 spike trains for these tests since the performance of both data-based circuits and control circuits was quite low if input rates were represented by 40 spike trains.

The emergent computational properties of data-based microcircuit models are recorded in Fig. 3.5 (green bars). The performance of the trained readouts for test inputs (which are generated from the same distribution as the training examples, but not shown during training) was measured for all binary classification tasks by the kappa coefficient, which ranges over [-1,1], and assumes a value ≥ 0 if and only if the resulting classification of test examples makes fewer errors than random guessing. [1] For tasks that require an analog output value the performance of the trained readout was measured on test examples by its correlation coefficient with the analog target output. The accuracy of computations achieved by trained readout neurons from microcircuit models with a data-based laminar structure is compared with the accuracy achieved by trained readout neurons from control circuits (red bars in Fig. 3.5) whose data-based laminar connectivity structure has been scrambled by replacing the source and target neurons of each synaptic connection within the circuit by randomly drawn neurons of the same type, i.e. excitatory or inhibitory neurons, under the constraint that no synaptic connection occurs twice (I will usually refer to these circuit models as amorphous circuits). Note that this procedure does not change the total number of synapses, the synapse type alignment with regard to pre- and postsynaptic neuron type, the global distributions of synaptic weights or other synaptic parameters, or the sets of neurons which receive afferent inputs or provide output to readout neurons. The connectivity structure of amorphous circuits is (apart from different

[1] The kappa coefficient measures the percentage of agreement between two classes expected beyond that of chance and is defined as (P_o - P_c)/(1 - P_c), where P_o is the observed agreement and P_c is the chance agreement. Thus for classification into 2 equally often occurring classes one has $P_C = 0.5$.

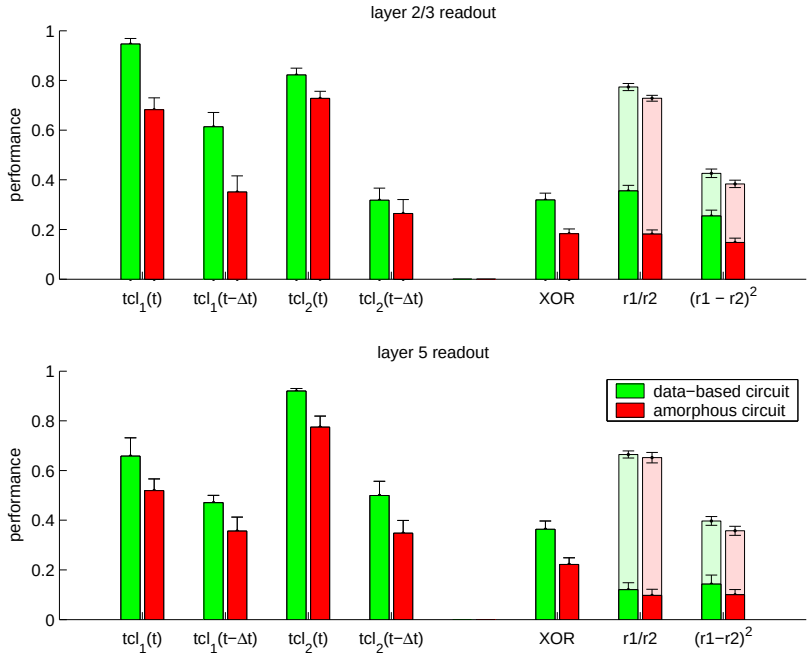


Figure 3.5: Performance of trained linear readout neurons in layers 2/3 and layer 5 (see Methods) for various classification tasks on spike patterns and computations performed on the rates of the two input streams, both for data-based laminar microcircuit models (green bars) and for control circuits where the laminar structure had been scrambled (red bars). $\mathrm{tcl}_{1/2}(\mathrm{t})$ denotes retroactive classification of noisy spike patterns (inputs were generated as shown in Fig. 3.4) in input streams 1 or 2 that were injected during the preceding time interval [t-30ms,t] into two classes according to the template from which each spike pattern had been generated. $\mathrm{tcl}_{1/2}(t - \Delta t)$ refers to the more difficult task to classify at time t the spike pattern before the last one that had been injected during the time interval $[t - 60\ ms, t - 30\ ms]$. For XOR classification the task is to compute at time $t = 450$ ms the XOR of the template labels (0 or 1) of both input streams injected during the preceding time segment [420 ms, 450 ms]. On the right hand side the performance results for real-time computations on the time-varying firing rates $r_1(t)$ of input stream 1 and $r_2(t)$ of input stream 2 (both consisting of 4 Poisson spike trains with independently varying firing rates in the two input streams). The light bars show performance results for the two target functions $r_1(t)/r_2(t)$ and $(r_1(t) - r_2(t))^2$, and the bold bars for the performance on the nonlinear components of these real-time computations at any time t

connection probabilities between and within the populations of excitatory and inhibitory neurons) identical with that of the graphs studied in classical random graph theory Bollobas (1985) (with the four connection probabilities for these two populations taken from the data-based circuits).

Fig. 3.5 shows that data-based circuits perform significantly better for the majority of the considered information processing tasks, except for the rate tasks for a readout neuron in layer 5 and the task $tcl_2(t - \Delta t)$ for a readout neuron in layer 2/3 (for which the performance increase was not significant). In particular, potential projection neurons in layers 2/3 and layer 5 have in a data-based laminar circuit better access to the information contained in the current and preceding spike patterns from either afferent input stream. The results show that potential readout neurons can classify spike patterns from either afferent input stream independently from the simultaneously injected spike pattern in the other input stream (and independently from the fairly high trial-to-trial variability shown in Fig. 3.3). One interesting detail can be observed for the two tasks involving computations on firing rates. Here the performance of data-based and control circuits is about the same (see light bars in Fig. 3.5), but readouts from layer 2/3 perform for data-based circuits significantly better on the nonlinear component of these computations (see bold bars in front of the light bars in Fig. 3.5). [1]

The actual performance achieved by trained readouts from microcircuit models depends on the size of the circuit (theoretical results predict that it will automatically improve when the circuit size increases Maass *et al.* (2002)). This is demonstrated in Fig. 3.6 for one of the computational tasks considered in Fig. 3.5 (XOR of labels of spike patterns from the two afferent input streams), both for data-based and for control circuits. Fig. 3.6 also shows that the superior performance of data-based circuits does not depend on the circuit size. The performance improvement of circuits consisting of 1000 neurons compared to circuits consisting of 160 neurons is somewhat smaller for rate tasks. For instance the performance of a readout neuron in layer 2/3 or layer 5 trained for the two rate

[1] This nonlinear component of the target functions r_1/r_2 and $(r_1 - r_2)^2$ resulted by subtracting from these functions an (for the considered distribution of r_1, r_2) optimally fitted linear function.

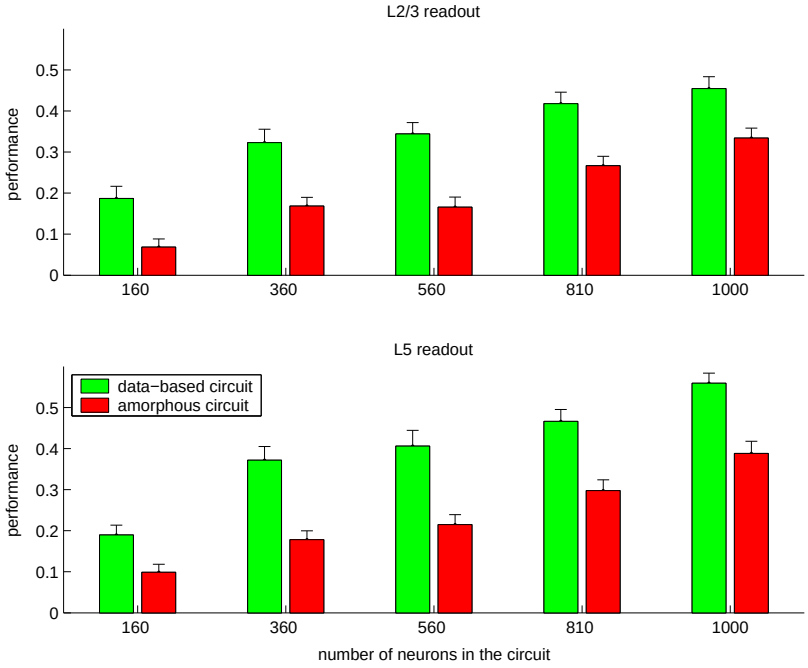


Figure 3.6: Performance (see Methods) of projection neurons in layers 2/3 and layer 5 for different circuit sizes, with and without a data-based laminar structure, for the computation of the XOR task. These results show that the superior performance of data-based circuits does not depend on the circuit size.

tasks r_1/r_2 and $(r_1 - r_2)^2$ increases on average by only 25% for data-based circuits and 20% for amorphous circuits.

The preceding results show that microcircuits with a data-based laminar structure have superior computational capabilities for a large variety of computational tasks. This raises the question *why* this is the case. I approach this question from two different perspectives. I first examine which aspects of the data-based circuit structure are essential for their superior performance. Obviously the procedure for generating amorphous circuits destroys not only the laminated structure of data-based circuits, but also other structural properties such as the distribution of degrees of nodes in the underlying connectivity graph, and its cluster structure. I therefore introduce three additional types of control circuits in order to analyze the impact of specific structural features on the performance. Secondly I exhibit a characteristic feature of the internal dynamic of these different circuit types

that is correlated with their computational performance.

I first studied the computational impact of small-world properties of data-based circuits. Small-world networks have been characterized in Watts & Strogatz (1998) through two properties. They have a higher clustering coefficient (measured by the proportion of immediate neighbors of nodes in the graph that are connected by a link) than amorphous circuits, while maintaining a comparable average shortest path length. [1] Data-based cortical microcircuit models have in fact small-world properties according to this definition, since their clustering coefficient (that has a value of 36%) is 38% higher than in amorphous circuits, while their average shortest path length is about the same (1.75 links). [2] In order to decide whether these small-world properties are sufficient for inducing the superior computational properties of data-based circuits, I generated control circuits that have the same size, clustering coefficient, and average shortest path length as data-based circuits by the spatial growth algorithm described in Kaiser & Hilgetag (2004) (with parameters $\alpha = 4$, $\beta = 1.32$ and 560 nodes). Subsequently these undirected graphs were converted to directed graphs by randomly replacing each edge with a synapse (that is randomly oriented) or a reciprocal synaptic connection, with a probability so that the total number of synaptic connections and reciprocal synaptic connections is identical to the corresponding number for data-based circuits. [3]

For the third type of control circuit I considered circuits that have the same distributions of in- and outdegrees for neurons as data-based circuits. The in- and outdegree of a neuron is defined as the total number of incoming and outgoing synaptic connections, respectively.

For this purpose I generated data-based circuits and subsequently exchanged the target neurons of randomly chosen pairs of synapses with pre- and postsynaptic neurons of the same category (excitatory or inhibitory), until the laminar specific connectivity structure disappeared (no exchange was carried out if either

[1] Note that both properties refer to the structure of the underlying undirected graph, where directed edges are replaced by undirected links.

[2] The long-range cortical connectivities in the cat and macaque monkey brain have clustering coefficients of 55% and 46%, respectively, as reported in Hilgetag *et al.* (2000).

[3] It should be noted that this procedure does not reproduce the same fraction of synapse types as for data-based circuits and amorphous circuits.

tasks/circuits	amorphous	small-world	degree-controlled	degree-controlled w/o input or output specificity	random short term synaptic dynamics	static synapses
memory	32.6	41.6	12.0	35.8	48.3	65.7
non-linear	36.9	11.3	-2.3	4.6	40.1	38.7
other	12.2	5.3	-0.6	5.6	14.1	6.9
all	25.0	15.4	1.6	12.0	30.4	30.6

Table 3.2: Average percentage of performance decrease compared to data-based circuits (averaged over tasks and readout types) for 7 types of control circuits (defined in the text) and the tasks defined in the legend for Figure 3.5. The memory tasks are $tcl_1(t-\Delta t), tcl_2(t-\Delta t)$, the non-linear tasks are XOR and computations on the purely nonlinear components of r_1/r_2 and $(r_1 - r_2)^2$ and other tasks are the remaining tasks considered in Figure 3.5. Only degree-controlled circuits achieve better performance than data-based circuits for some tasks.

of the two resulting new connections existed already). This circuit type also has small-world properties, but the average cluster coefficient was smaller than for data-based circuits (only 27% higher than in amorphous networks). I refer to this circuit type as degree-controlled circuits.

Degree-controlled circuits preserve the distribution of degrees among neurons that receive external input or provide input to a readout neuron. Therefore a control circuits was added, referred to as degree controlled circuits without input or output specificity, by randomly exchanging neurons in different layers of degree-controlled circuits. The degree distributions of neurons for all 5 types of circuits are shown in Figure 3.7.

An important structural feature of all circuit types considered until now is the alignment of synapse type with regard to pre- and postsynaptic neuron type according to Table 3.1. In order to analyze the impact of the alignment of dynamic synapses on the performance I randomly exchanged the synaptic parameter triplets, i.e. U, D and F, that define the short term plasticity between all synapses. In the last type of control circuits all dynamic synapses were replaced by static synapses (with weights rescaled so that the mean firing rate in layer 5 stayed fixed).

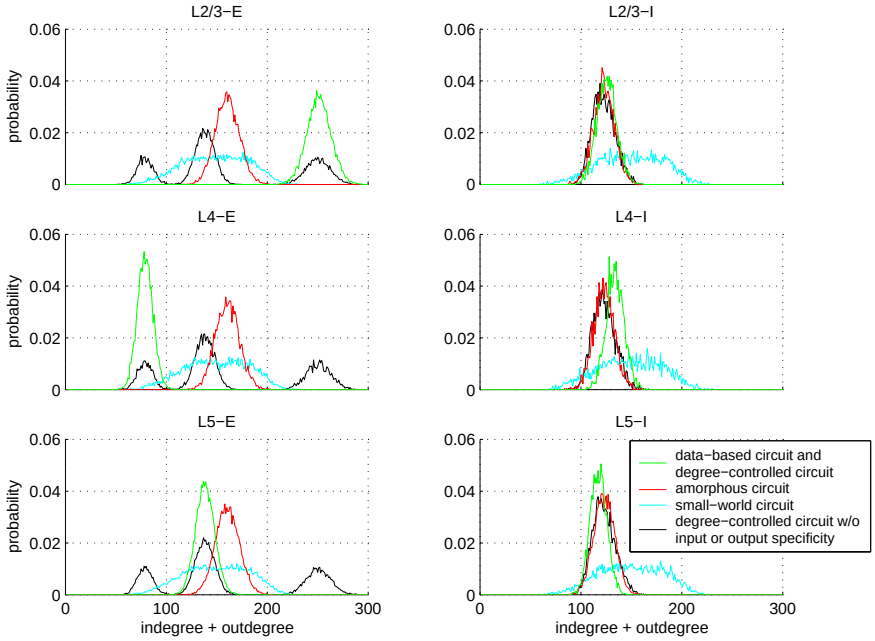


Figure 3.7: Degree distributions of neurons in different populations for 5 types of circuits (averaged over 100 circuits). All circuits have degree distributions that can be better approximated by sums of Gaussians than by a power law of the form $P(d) \propto d^{-\gamma}$ for neurons of degree d and a positive constant γ (i.e. the circuits are not scale-free). The correlation coefficients for least square fits for sums of Gaussians and power law distributions are > 0.96 and < 0.08, respectively. Thus none of these circuits are scale-free, which shows that their difference in performance cannot be explained on the basis of this concept. The computational analysis (see Table 3.2) implies that the varying locations of peaks for different layers of data-based circuits are essential for their superior computational performance.

A summary of the performance of all 7 different types of control circuits is shown in Table 3.2. The small world property increases the performance of amorphous circuits to some extent, but a more important structural feature is the degree distribution defined by data-based circuits. If this degree distribution matches the degree distribution of data-based circuits for each single layer, and therefore matches also the specific input and output topology of data-based circuits, the average performance is comparable to the performance of data-based circuits. Table 3.2 also shows (see column 5) that a data-based assignment of

synapse types (according to Table 3.1) is essential for good computational performance. The last column shows that circuits with static synapses also have inferior computational properties.

In order to elucidate the relationship between inherent properties of the circuit dynamics and computational performance I studied one fundamental – but relatively simple – information processing task in more detail: retroactive classification of spike patterns into two classes, in spite of noise. More precisely, the task was to classify the 2×4 input spike trains generated from two templates (as in Fig. 3.4) into two classes, in spite of a subsequent waiting period of 100 ms (during which identical spike trains were injected in either case), and in spite of widely different initial conditions (caused by different preceding spike inputs) and the relatively high internal noise that models bombardment with unrelated background synaptic input in the 'high conductance state' (compare the middle and bottom row of Fig. 3.3 to see stochastic changes in the spike response caused by the latter). This task tests the capability of circuits to maintain information about spike patterns that had been injected more than 100 ms ago. This information is reduced by noise resulting from inherent noise in neurons and varying initial conditions ("in-class variance"). The solid green lines in Fig. 3.8 show that readout neurons in layers 2/3 and layer 5 of data-based circuits can learn quite fast from relatively few training examples to guess which of the two fixed spike patterns had previously been injected. A comparison with the red lines show that their error on new examples of this task (test error) is significantly smaller than that of readout neurons in amorphous control circuits. Furthermore this advantage is not reduced when more training examples become available. The superiority of readouts from data-based circuits results both from a better fit to the training data (dashed curves in Fig. 3.8), and from a smaller generalization error (= distance between solid and dashed curve). [1]

A more intrinsic explanation for the better computational performance of data-based laminar circuits is provided by the theory of computations in dynamical systems (see Legenstein & Maass (2005) for a review). Fig. 3.9 shows that a

[1] Note that all types of circuits have for a smaller number of training examples a smaller error on the training set but a larger error on the test set due to the well-known overfitting effect that is studied extensively in statistical learning theory Vapnik (1998).

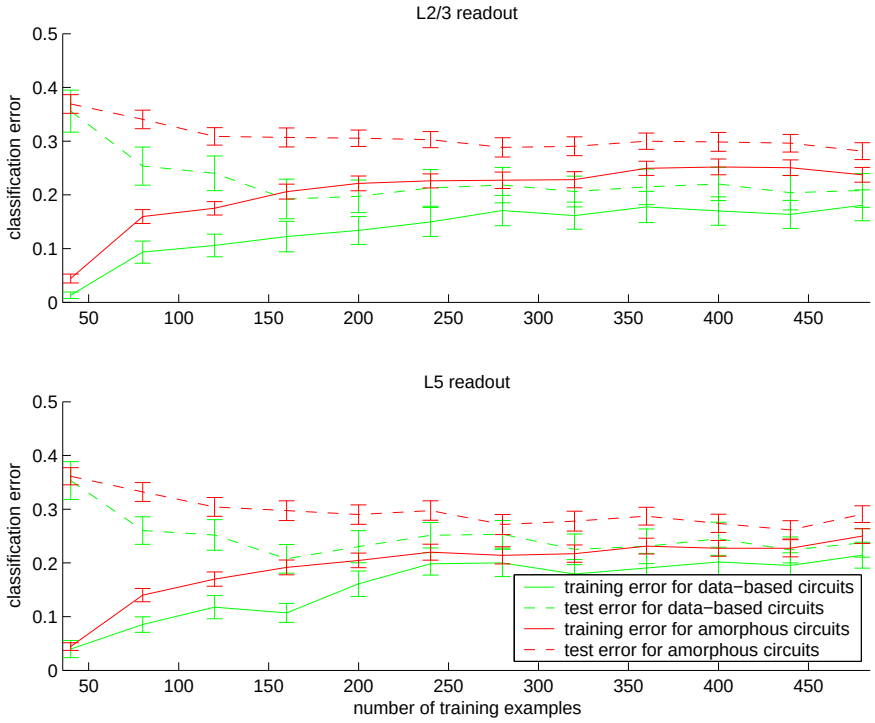


Figure 3.8: Training and testing error of readouts from data-based and amorphous microcircuit models as function of the size of the training set. The task was retroactive classification of two randomly created spike patterns of length 100 ms (consisting of 4 Poisson spike trains at 20 Hz) after identical input of length 100 ms was subsequently injected into the circuit, in spite of varying initial conditions (caused by independently generated preceding Poisson spike inputs of the same type) and noisy background input currents as before. The top panel shows the performance of a potential readout neuron in layers 2/3 with 84 presynaptic neurons, trained by sign-constrained linear regression (see Methods). The bottom panel depicts the performance for a potential readout neuron in layer 5 with 109 presynaptic neurons. Both types of readouts perform better for data-based laminar circuit, both on the training set (with new random drawings of initial conditions and background noise) and on the testing set. This holds for all sizes of the training set that were considered.

data-based circuit works in a substantially less chaotic regime than an amorphous control circuit. Its sensitivity to tiny differences in initial conditions is also less than in the other 3 types of control circuits that preserve selected aspects of the

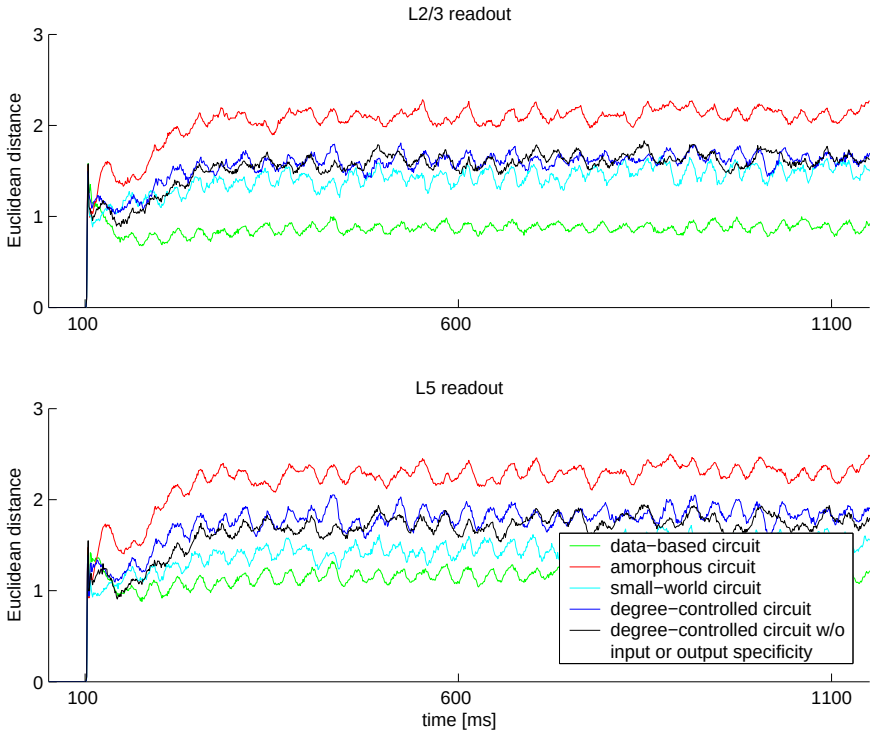


Figure 3.9: Euclidean distances in trajectories of circuit states (more precisely: of input vectors to readout neurons in layers 2/3 and in layer 5) resulting from moving a single input spike (at 100 ms) by 0.5 ms. Shown is the average from simulations of 400 randomly generated circuits with different initial conditions and background noise. Initial conditions and internal noise were chosen to be identical in both trials of each simulation, as in standard tests for estimating the Lyapunov exponent of (deterministic) dynamical systems, see Legenstein & Maass (2005). The curves show lasting differences in Euclidean distance between circuit states which are about twice as large for amorphous circuits, thereby indicating a more chaotic dynamics than in laminar circuits with the same number of neurons and synapses.

data-based network structure. A less chaotic dynamics implies better generalization capability to new inputs for many different types of dynamical systems. This observation is of interest because one might assume that the number of synapses per neuron is the essential parameter that determines the amount of chaos in the circuit. But all circuits for which results are plotted in Fig. 3.9 have the same number of synapses.

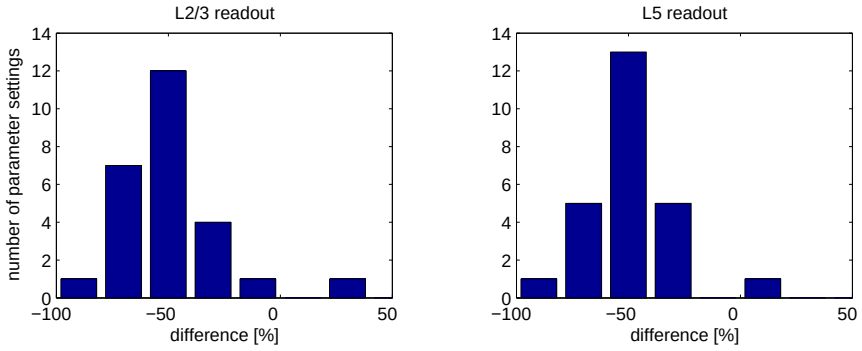


Figure 3.10: Data-based circuit structure reduces the impact of noise. For each of two circuit types, i.e. for data-based and amorphous circuits, 50 circuits were generated and 500 identical inputs as generated for the task in Fig. 3.8 were injected into each of them. The difference in the variance of circuit responses (averaged over the 50 circuits) was evaluated from the perspective of readout neurons. More precisely: The variance of the input to readout neurons, after injecting 500 times the same input into the circuit, was analyzed by how much this variance was reduced for the data-based circuit structure (expressed as percentage of change in comparison with amorphous control circuits). The smaller variance for data-based circuits shows that their dynamics is less influenced by internal noise and different initial conditions, thereby providing better generalization capabilities of trained readouts. This experiment was repeated for 30 randomly chosen settings of the scaling parameters S_{RW}, S_{I1}, S_{I2} (see Methods), and this figure shows for how many of these parameter settings a specific noise-reduction was achieved for data-based circuits. The mean in-class variance for all parameter settings was 0.58, SD 0.29, for data-based circuits. The percentages for 5 of these circuits were not entered into this plot since their in-class variance was below 0.01, resulting from the fact that they hardly responded to the input. The percentage of change of in-class variance for the standard setting of the scaling parameters was -38 for layer 2/3 readouts and -41 for layer 5 readouts. These results show that the noise reduction capability of the data-based circuits was not an accidental property of the standard setting of the scaling parameters.

The dashed curves in Fig. 3.8 show that another reason for the better computational performance of data-based circuit models is that the synaptic weights of their readout neurons can be better fitted to the training data. This fact can be explained in terms of the in-class variance of high-dimensional circuit states

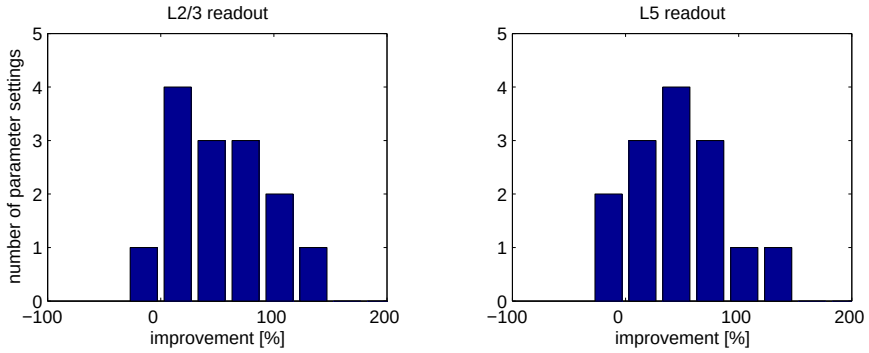


Figure 3.11: Percentage of improvement in performance of readouts from laminar circuits for different values of the 3 scaling parameters, for the same XOR task as discussed in Fig. 3.4 and 3.5 (but with just 2 times 4 input spike trains). The percentage of improvement was measured for 16 randomly drawn settings of the scaling parameters S_{RW}, S_{I1}, S_{I2}. Two of these settings yielded extremely low performance for both data-based and amorphous circuits (below 0.08, hence below the SD of the performance data for all 16 parameter settings), and were therefore omitted from the plot. The improvement in performance for the standard setting of these parameters was 74% for layer 2/3 readouts and 64% for layer 5 readouts. This suggests that a laminar circuit has a superior computational capability for most parameter settings, hence for a wide variety of stimulus intensities and regulatory states of neural systems *in vivo*.

caused by varying initial conditions and internal noise (for repeated trials with the same spike input to the circuit). The correlation between this in-class variance on training data and the classification error of trained readouts on test data for the task considered in Fig. 3.8 was 0.80 for readouts from layer 2/3 and 0.72 for readouts from layer 5 for data-based laminar circuits.

Fig. 3.10 shows that for amorphous control circuits this in-class variance is generally larger. Furthermore Fig. 3.10 shows that this noise-suppressing feature of the dynamics in data-based laminar circuits is not an accidental property of the fixed setting of the 3 parameters S_{RW}, S_{I1}, S_{I2} (which scale the weights of recurrent synaptic connections, the amplitudes of input stream 1, and the amplitudes of input stream 2) that I used for the simulations reported so far (see Methods). Fig. 3.10 shows that this noise-suppressing feature appears also for all other (randomly chosen) settings of these parameters that were tested.

Fig. 3.11 shows that also the superior computational performance of data-based laminar microcircuit models is not an accidental consequence of a particular choice of these 3 parameters, but holds for most of their potential values. This fact is demonstrated here for the XOR on spike patterns that was already previously discussed. [1] This suggests that a laminar circuit has for some tasks superior computational capability for a fairly large variation of dynamic regimes. This is of interest because different behavioral states, different states of homeostatic processes, or different input intensities may give rise to a variety of different dynamic regimes of cortical microcircuits.

3.4 Discussion

I demonstrated that data-based laminar connectivity structure enhances the information processing capabilities of cortical microcircuit models. In particular I have shown that such data-based circuit model can accumulate, hold, and fuse information contained in two afferent spike input streams. It should be noted that the computations that were analyzed in the computer experiments were biologically realistic real-time computations on dynamically varying input streams, rather than static computations on batch inputs that are usually considered in modeling studies. In contrast to the circuit models from Buonomano & Merzenich (1995) and Maass *et al.* (2002), the circuit models that were analyzed in this chapter not only have a biologically realistic laminar structure, but also consist of Hodgkin-Huxley type neurons (with additional background input based on data which are conjectured to be representative for the 'high conductance state' of cortical circuits *in vivo* Destexhe *et al.* (2003)). In addition the simulations discussed in this chapter took a substantially larger trial-to-trial variability into account. Furthermore information was not extracted from all neurons as in Buonomano & Merzenich (1995) and Maass *et al.* (2002), but from a much smaller subset of neurons that represents the typical set of presynaptic neurons

[1] The SD of the performance of readouts from control circuits for different parameter settings was 0.15 for readouts from layer 2/3 and 0.18 for readouts from layer 5. The performance improvement for data-based laminar circuits was somewhat correlated with the performance (cc. 0.16 for layer 2/3, 0.59 for layer 5 readouts)

for a projection neuron in layers 2/3 or 5. In addition the extraction of information by such projection neurons was for the first time subjected to the constraint that the signs of weights of incoming synapses cannot be chosen arbitrarily in a biological circuit, but are determined by the type (excitatory or inhibitory) of each presynaptic neuron. Although this means that not the full power of linear regression (or of the perceptron learning rule) can be used for optimizing such more realistic readouts, I show that even under these biologically more realistic conditions a substantial amount of information can be extracted by projection neurons in layers 2/3 or layer 5. Furthermore the results in Fig. 3.6 show that their performance increases with circuit size, making it reasonable to conjecture that almost perfect performance will be achieved by a circuit model which is sufficiently large so that the number of presynaptic neurons approaches realistic values of a few thousand.

I demonstrated in Fig. 3.5 that data-based laminar microcircuit models perform significantly better than control circuits (which are lacking the laminar structures but are otherwise identical with regard to their components and overall connection statistics) for a wide variety of fundamental information processing tasks. This superiority holds for most settings of the parameters which scale the global strength of afferent inputs and of recurrent connections, corresponding to a wide variety of stimulus intensities and regulatory states of neural systems *in vivo* (Fig. 3.11). I also analyzed which aspect of the connectivity structure of data-based laminar circuits is responsible for their better computational performance. I arrived (on the basis of the results reported in Table 3.2) at the conclusion that their particular distribution of degrees of nodes (relative to circuit inputs and projection neurons) is primarily responsible, more so than the small world property of data-based circuits. I propose that this computational superiority of laminar circuits can be understood in terms of the properties of the dynamical system which is defined by such microcircuit models: I have shown in Fig. 3.9 and 3.10 that the dynamics of laminar circuits is somewhat less influenced by internal noise and noise in the input, thereby providing better generalization capabilities of trained readouts, and a better fit to training data because of the reduced variance in circuit responses.

In Chapter 4 I continue by investigating experimentally if the computational properties of the cortical microcircuit model as explored in this chapter, in particular fading memory and nonlinear computing on inputs from different time frames, are supported by multi-electrode recordings in cat primary visual cortex. This second part of the thesis is therefore dedicated to a direct experimental test of this essential properties predicted by the theory of the *liquid state machine*.

Chapter 4

Temporal dynamics of information carried by ensembles of neurons in the primary visual cortex

Summary

I use multi-electrode recordings from cat primary visual cortex to investigate whether a simple linear classifier can extract information about previously presented stimuli. I find that information is extractable and that it even lasts for several hundred milliseconds after the stimulus has been removed. In a fast sequence of stimulus presentations, information about both new and old stimuli is superimposed and can be is extracted simultaneously. Nonlinear relations between successively presented pattern in a stimulus sequence can be extracted suggesting highly nonlinear properties of cortical representations. It is shown that the precise timing of spikes in the order of milliseconds carries information about the visual stimuli that is not contained in rate responses.

4.1 Introduction

It has recently been argued that fundamental aspects regarding the organization of computations in visual cortex are still unknown Olshausen & Field (2006). In fact, there exist completely different computational models for the processing of visual inputs in the brain that are consistent with most of the available experimental data. On one end of the spectrum one finds models which propose that specific computational operations are carried out by specialized neural circuits (often seen as being organized through precisely structured maps of columns, where each column is specialized for carrying out a specific computational operation, such as for example the application of a Gabor filter to a particular patch in the image). Complex vision tasks, such as visual object recognition, are then modeled by a subsequent hierarchy of other processing layers that carry out further specific computational operations (Serre *et al.* (2005)). At the opposite end of the spectrum one finds models which emphasize that the visual system (in fact, the whole brain) is a recurrent neural network (or more abstractly: a dynamical system), whose organizational structure is incompatible with such isolated sequential applications of specific computational operations. In between both extremes one finds models that explicitly take feedback connections between cortical areas (and between cortex and thalamus) into account, but assign specific computational roles to such feedback, such as for example input completion or "predicting away".

One particular prediction of models that emphasize the aspect that the visual system is a recurrent circuit concerns the time course of information about visual inputs in primary visual cortex. Computer simulations of sparsely connected recurrent circuits of neuron-like processing units Buonomano & Merzenich (1995); Jäger & Haas (2004); Maass *et al.* (2002), as well as theoretical analysis of related circuits with simpler (linear) processing units Jäger (2002); White *et al.* (2004) make the following prediction. Inputs to such circuits leave a trace in the circuit activity that is in general difficult to "decode" with the naked eye, but which can be read out quite easily by suitably trained linear or nonlinear classifiers ("readouts"). In fact, if such readout receives direct inputs from a sufficiently large set of neurons in the circuit, even very simple linear classifiers suffice in many

cases. The validity of this prediction for the time course of visual information in a higher cortical area (area IT, i.e. inferior temporal cortex) has recently been verified by Hung *et al.* (2005). They showed (through a series of single-unit recordings) that a hypothetical linear readout that gets as input the spike counts (for the last 50 ms) from different neurons (recorded in different trials, but with the same visual input), can recover from these spike counts the information which image had been shown to the monkey several hundred ms ago.

If such slowly fading information about a previous stimulus would also occur when a rapid sequence of different images is shown (especially if it can be demonstrated for primary visual cortex, rather than for the highest visual area IT), it would pose an obstacle for computational models that postulate a precisely organized sequence of processing steps (i.e., a pipelined computational organization) in the visual cortex, where each processing layer works at any moment on a single "frame" of an input movie. If such superposition of information from different frames can be demonstrated in primary visual cortex, it would also pose an obstacle for models that propose a precisely structured feedback from higher visual areas (that completes or explains each frame in terms of higher level concepts), since the temporal dispersion of information would make it very hard to bring bottom-up sensory information and top-down interpretation together in a coherent manner for any single frame. On the other hand, such superposition of information from subsequent frames of online inputs is an essential ingredient for models that emphasize the dynamical system aspect of the recurrently connected visual system as a whole. Furthermore, it was shown in Natschläger & Maass (2005) to be a direct consequence of generic recurrent circuits of neurons. This recurrent circuitry is viewed in the context of the Liquid State Machine (LSM) model of Maass *et al.* (2002) as an analogue for a kernel of a support vector machine (SVM) in machine learning Schölkopf & Smola (2002); Vapnik (1998), whose computational task is to superimpose and mix information from different time frames and different sources (such as various sensory modalities, internal states, or even motor plans) in such a way, that a large number of projection neurons can independently learn to extract specific aspects of such fused information that are needed by the particular brain area to which they project.

The exact computational operations that are carried out by neurons and microcircuits are less relevant for this computational model, as long as a number of large scale properties for the simultaneous response of large ensembles of neurons (more precisely, for those ensembles of neurons that are synaptically connected to projection neurons in various cortical areas) are met. Such desirable properties are for example: diversity of responses of individual neurons, and high mutual information between their collective response and salient components in the recent sensory input stream.

I report in this chapter results from a direct experimental test of this prediction. I analyzed the superposition of information from different frames of visual input in sets of neurons in primary visual cortex through the following type of experiment. The experimental data was provided by the Department of Neurophysiology of the Max-Planck-Institute for Brain Research in Frankfurt (Main), Germany. Rapid sequences of complex visual patterns (large white letters with a black background, see Fig. 4.1A) were shown to anesthetized cats, and the resulting spiking activity in primary visual cortex (area 17) was recorded with multi-unit recordings.

I analyzed the information that is contained in the spikes from about 100 simultaneously recorded neurons from the perspective of a hypothetical projection neuron, i.e. the spike train from each recorded neuron was low-pass-filtered (in order to mimic the time course of postsynaptic potentials in a projection neuron that are caused by these spikes, see Fig. 4.1D). A linear combination of these traces (with weights that remained fixed after training) was formed. It was examined whether the weights could be chosen in such a way, that a comparison of the current value of the weighted sum of these traces with a fixed threshold (corresponding to the firing threshold of a hypothetical projection neuron) provided a classification of an earlier visual input that had been shown several hundred ms ago. This was examined in particular in cases where another frame had been shown during the intervening time interval.

4.2 Methods

4.2.1 Experiments

The experimental recordings were planned and performed by Danko Nikolić from the Department of Neurophysiology of the Max-Planck-Institute for Brain Research in Frankfurt (Main), Germany. In four cats anesthesia was induced with ketamine and maintained with a mixture of 70% N_2O and 30% O_2 and with halothane (0.4-0.6%). The cats were paralysed with pancuronium bromide applied intravenously (Pancuronium, Organon, 0.15 mg $kg^{-1}h^{-1}$). Multi-unit activity (MUA) was recorded from area 17 and by using multiple silicon-based 16-channel probe (organized in a 4×4 spatial matrix) which were supplied by the Center for Neural Communication Technology at the University of Michigan (Michigan probes). The inter-contact distances were 200 μm (0.3-0.5 MΩ impedance at 1000 Hz). Signals were amplified 1000$\times$ and, to extract unit activity, were filtered between 500 Hz and 3.5 kHz. Digital sampling was made with 32 kHz frequency and the waveforms of threshold-detected action potentials were stored for an off-line spike sorting procedure. The probes were inserted approximately perpendicular to the surface of the cortex, allowing us to record simultaneously from neurons at different cortical layers and at different columns. This setup resulted with a cluster of overlapping receptive fields (RF), all RFs being covered by the stimuli (see Fig. 4.1A) (more details on recording techniques can be found in Schneider & Nikolić (2006); Schneider *et al.* (2006)).

Stimuli were presented binocularly on a 21" computer monitor (HITACHI CM813ET, 100 Hz refresh rate) and by using the software for visual stimulation ActiveSTIM
(www.ActiveSTIM.com). Binocular fusion of the two eyes was achieved by mapping the borders of the respective RFs and then by aligning the optical axes with an adjustable prism placed in front of one eye. The stimuli consisted of single white letter with elementary features suitable for area 17 and spanning approximately 5° of visual angle. The stimuli were presented on a black background for a brief period of time. Fig. 4.1A illustrates the spatial relation between the constellation of RFs and the stimulus in one of the experimental setups. In each

stimulation condition either a single letter was presented or a sequence of up to three letters. For presentations of single letters the letters A and D each presented for 100 ms were used.

Stimulus sequences were made with letters A, B, C, D, and E. I compared either the responses across the sequences ABC, DBC, and ADC (cat 1 and 4) or across sequences ABE, CBE, ADE, and CDE (cats 2 and 3). Each member of a sequence was presented for 100 ms and the blank delay-period separating the presentation of letters lasted also 100 ms. Each stimulation condition (single letter or a sequence) was presented 50 times for cat 1, 150 times for cat 2 and cat 3 and 100 times for cat 4. The order of presentation was randomized across the stimulation condition. Example raster plots of responses to two different sets of stimuli can be seen in Fig. 4.1B.

For the analysis of the neuronal responses to single letter presentations I discarded datasets with a maximum mean firing rate (averaged over electrodes and trials) below 50 Hz. For the analysis of the responses to sequences of letters only datasets with a maximum mean firing rate above 80 Hz were used.

4.2.2 Data analysis

The typical spike trains prior to the application of spike-sorting procedures are illustrated in Fig. 4.1B. All datasets showed high trial-to-trial variability, with an average fano factor of about 8. If all the single units that resulted from the spike-sorting procedure included into the analysis, this resulted in too sparse data representations and hence in overfitting. Therefore only units with mean firing rates $\geq$10 Hz were used and single units with less frequent firings were pooled into multi-unit signals. These methods resulted in datasets with 66 to 124 simultaneously recorded units for further analysis.

The recorded spike times were convolved with an exponential kernel with a decay time constant of $\tau = 20$ ms (the performance of classifiers is similar when one uses instead spike counts for the last 20 or 30 ms, see Suppl. Fig. A.1). A linear classifier was trained to discriminate between pairs of stimuli on the basis of the convolved spike trains at time points $t \in \{0, 10, ..., 700\}$ ms after stimulus onset (using only the vectors of 66 to 124 values of the convolved time series

at time t). I refer to this classifier as R_t. A second type of classifier, which I refer to as R_{int}, was trained to carry out such classification simultaneously for all time points $t \in \{150, 160, ..., 450\}$ ms (see Fig. 4.9). If not otherwise stated, the results for type R_t classifiers are reported. A linear classifier applied to the convolved spike data (i.e., an equivalent to low-pass-filtering) can be interpreted as an integrate-and-fire (I&F) neuron with synaptic inputs modeled as Dirac delta functions. The time constant of 20 ms reflects the temporal properties of synaptic receptors and of the membrane. A classification is obtained due to the firing threshold of the I&F neuron.

Linear classifiers were trained with linear-kernel support vector machines (SVMs). In section 4.3.3 I also discuss performance results for nonlinear SVMs with quadratic and radial basis function (RBF) kernels (see Schölkopf & Smola (2002); Vapnik (1998)). The parameter C of the SVMs was chosen to be 50 for recordings from cat 1 and cat 4 and 10 for recordings from cat 2 and cat 3. The parameter σ^2 of RBF kernels was set to 0.5 in order to achieve optimal test performance. The classification performance was always estimated with 10-fold cross validation, in which the number of positive and negative examples was balanced for the training and for the test set. All the reported performance data are for the test set. Error bars in the figures denote the standard error of the test error of all 10 cross validation runs.

4.3 Results

4.3.1 Persistent information about preceding visual stimuli

The spike trains from a randomly sampled set of neurons in primary visual cortex carry for several hundred ms information about previously shown visual stimuli, even for anesthetized cats (Fig. 4.2). Furthermore this information can be extracted by simple linear classifiers (applied to low pass filtered spike trains, see Methods), hence potentially also by a suitably trained projection neuron.

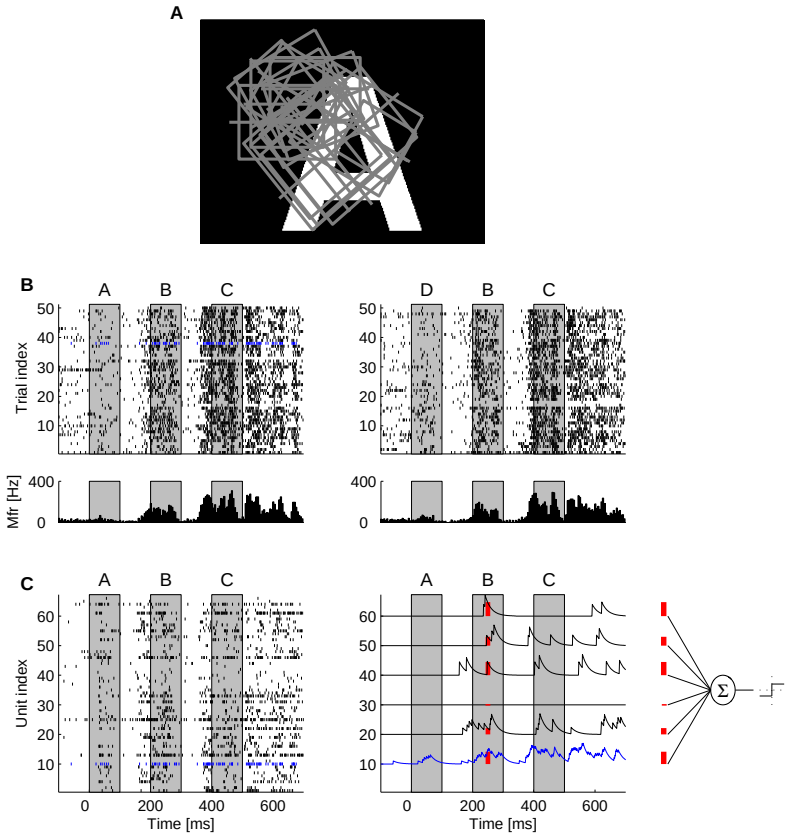


Figure 4.1: Experimental setup and illustration of the chosen method for data analysis. **A:** An example of a visual stimulus in relation to the constellation of receptive fields (gray rectangles) from one Michigan probe. **B: Upper part:** Spike times recorded from one neuron across 50 stimulus presentations and for two stimulus sequences (ABC and DBC). In this and in all other figures the gray boxes indicate the periods during which the letter-stimuli were visible on the screen. **Lower part:** Peri-stimulus time histogram for the responses of this neuron. Mfr: mean firing rate. **C:** Spike trains from 66 identified neurons for one trial (trial number 38, see blue trace in B) are shown on the left. The spikes from neuron number 10, for which the responses in all trials are shown in B, are plotted in blue. Each spike train is convolved with an exponential kernel (i.e., low pass filtered), as shown for the spike train from 6 selected neurons on the right hand side. The values of the resulting traces at a particular time point t (shown in red for $t = 250$ ms, for just 6 of the 66 neurons), but from all 66 neurons, constitute the vector of numbers to which a classifier (in most cases: a linear classifier) is applied in the data analysis. The target output of the classifier is in this case 1, since A (instead of D) had been shown in this trial as the first stimulus in the sequence.

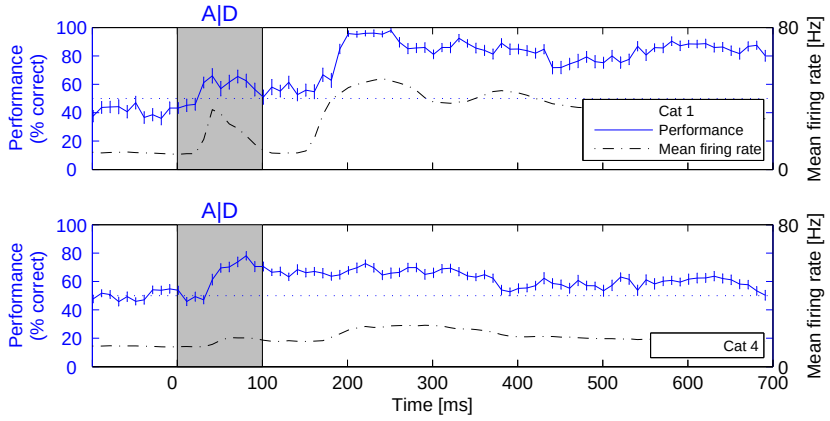


Figure 4.2: Persistence of information in primary visual cortex for a single visual stimulus (for 2 cats). In one cat (lower panel) the firing rates remained rather low, and hence the available information was lower. Information was extracted by linear classifiers R_t. Performance results shown are for test data.

4.3.2 Superposition of information about different sequentially presented stimuli

I also find, that when sequences of visual stimuli were presented, neuronal activity continued to carry information about the first stimulus during and after the presentation of the second stimulus (Fig. 4.3). However, this does not imply that information about the second visual stimulus did not enter primary visual cortex. Rather, information about the second visual stimulus can be recovered (by another linear classifier) from the firing activity of the same ensemble of neurons (Fig. 4.4 A,B). In fact, information about this second stimulus persists also, even during and after the presentation of a third stimulus. This implies, that the firing activity of the recorded ensemble of neurons carries during some time period simultaneously information about two sequentially presented stimuli (Fig. 4.4B). Consequentially, also results of computations that essentially involve information from both stimuli can be extracted from the same ensemble of neurons by a linear classifier (Fig. 4.4C).

Moreover linear readouts can extract information about results of purely non-

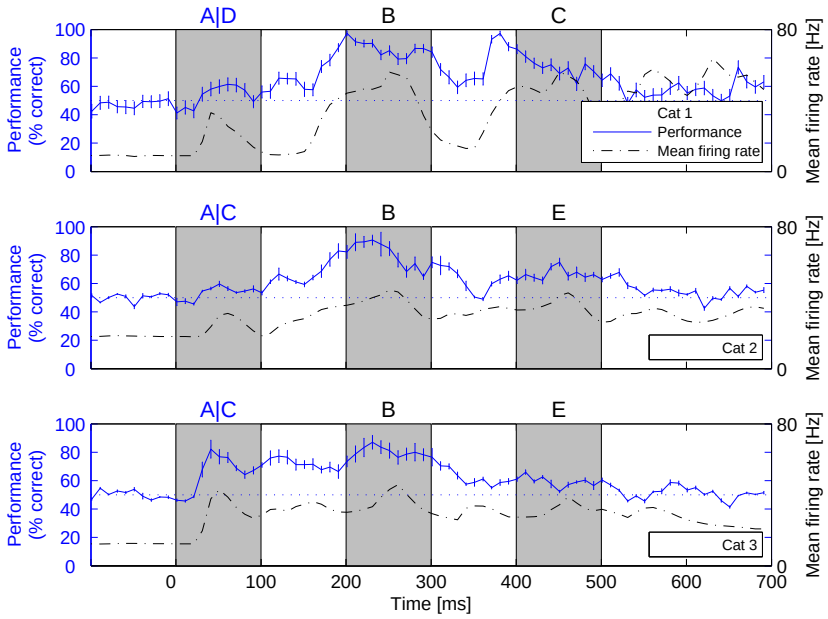


Figure 4.3: Persistence of information about the first stimulus in spite of subsequent further stimuli. Classifiers R_t were trained to identify the first letter in the sequences ABC vs. DBC in one experiment (cat 1) and sequences ABE vs. CBE in other two experiments (cats 2 and 3). In all cases, the performance reached its maximum shortly before or during the presentation of the second letter. In one case (cat 1) information about the first letter remained present even during the presentation of the third letter.

linear computations[1] on visual stimuli. This indicates nonlinear processing of visual inputs in the visual pathway up to the level of the primary visual cortex. The reason for this is that in case of only linear processing in the visual pathway the output of a linear readout in the primary visual cortex is necessarily a linear function of the visual input. Clearly the output of this linear readout has a vanishing correlation coefficient with any purely nonlinear function of the visual input. Therefore the nonvanishing correlation coefficient with the purely nonlinear XOR function of the 2 bits encoded by the 2 choices $A|C$ and $B|D$ as visual inputs shown in Fig. 4.4C indicates nonlinear processing of visual inputs in the visual pathway (see Suppl. Theorem 1).

[1] Functions are considered as purely nonlinear if for a given input distribution the correlation coefficient between its output and the output of any linear nonconstant function vanishes.

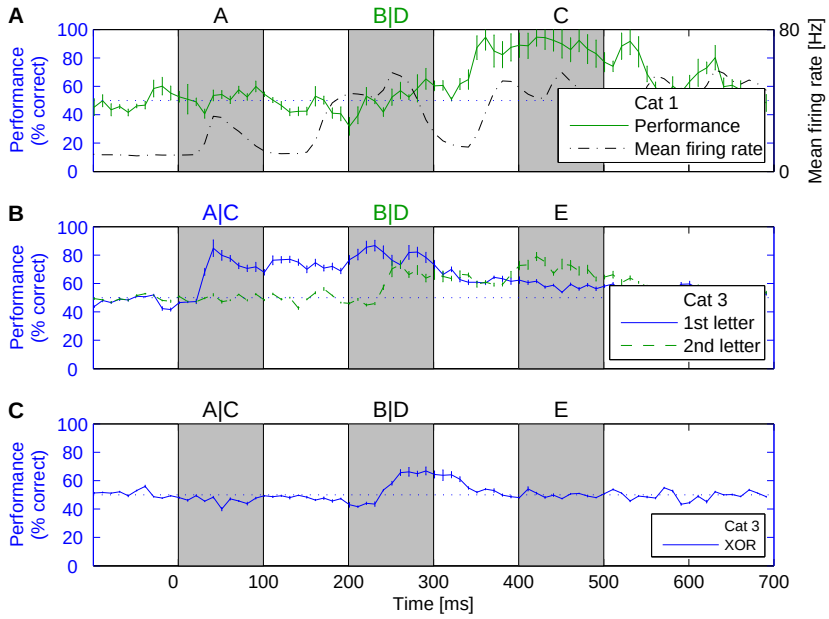


Figure 4.4: Simultaneous persistence of information about the second stimulus (in the same neurons). **A:** Performance of classifiers R_t trained to identify the second letter in the sequences ABC and ADC. Similarly to the results in Fig. 4.3, the performance is still significant during the presentation of the third stimulus. **B:** Simultaneously available information about two different letters of a sequence. Two classifiers identified either the first or the second letter of the following four sequences: ABE, CBE, ADE, and CDE . **C:** Linear classifiers R_t were trained to compute the XOR function of the 2 bits encoded by the 2 choices $A|C$ and $B|D$ (solid line).

4.3.3 Information contained in firing rates

Additional data analysis provides insight into the specific way how information was encoded by the ensembles of neurons from which was recorded. The dash-dotted curves in Fig. 4.2 - 4.4 indicated already that the performance of each classifier was correlated with the time-varying mean firing rate (with the mean taken over all trials and all recording electrodes, but not over time). For example, the correlation coefficients were 0.45, 0.68, 0.66 for the 3 panels of Fig. 4.3. The performance of the classifier was also positively correlated with the difference in the mean firing rates in response to a different stimulus at the beginning of the sequence (see Fig. 4.5). However the corresponding correlation coefficients were

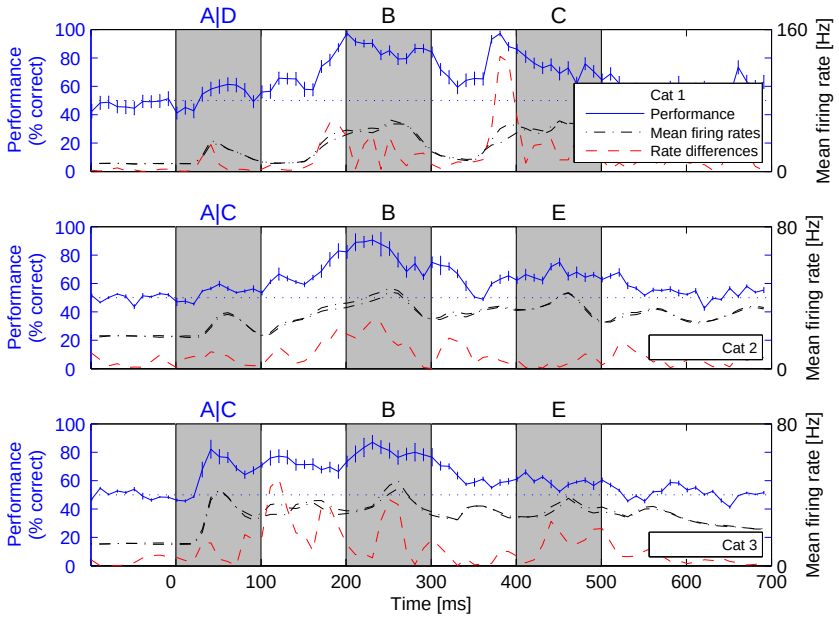


Figure 4.5: The relation between classifier's performance and i) the mean firing rates (dash-dotted lines) and ii) the difference in the mean firing rates between two stimulation conditions (8 fold magnified, dashed lines). The results are for the same data as in Fig. 4.3. The mean firing activity of all neurons from which we recorded exhibits some class specific differences.

somewhat smaller (0.37, 0.42, 0.46).

The weights of the linear classifiers R_t that were trained (independently for each t) to classify from the low pass filtered spike train at a specific time t the stimulus that had been shown during the time interval [0, 100] ms were not time-invariant. However one can also train classifiers R_{int} to carry out such classification for any time t within some interval int. For an interval int of length 100 ms a linear classifier R_{int} achieves almost the same performance as the classifier R_t (see Fig. 4.6A). Fig. 4.6C shows the performance of an even more invariant classifier for the whole time interval $int = [150, 450]$ ms. As expected, the performance of such very general classifier R_{int} is for most time points t lower than that of the more specialized linear classifiers R_t. Nevertheless, Fig. 4.4C shows that a substantial amount of information about a preceding stimulus is contained in the spike responses in a time-invariant manner for a 300 ms long

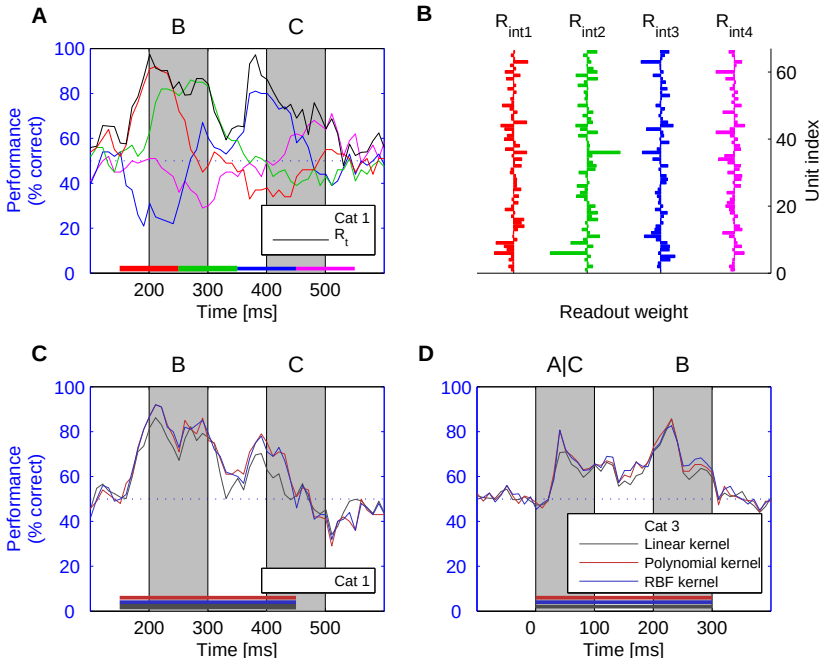


Figure 4.6: Performance of classifiers R_{int} that were trained to extract information at any time point within some interval *int*. **A:** Performance of 4 linear classifiers trained for disjoint intervals of length 100 ms, marked by a bar at the bottom in the same color as the corresponding performance curve (for test data). Each classifier reaches almost the same performance as the more specialized classifiers R_t for t within the interval *int*, but its performance drops quite fast outside of *int*. **B:** Weight vectors of the 4 classifiers from A. **C, D:** Performance of more invariant classifiers that classify the first letter for all time points t within an interval *int* of length 300 ms (for 2 cats). Performance is lower than that of more specialized classifiers, but still significant.

time interval that starts 50 ms after the offset of the first stimulus, and extends through and beyond the time interval when the second stimulus was shown. This results imply, that the persistent information that is contained in the spike trains about the first stimulus cannot be explained in any simple way through neuronal adaptation (for example through a reduced response to the second stimulus if a neuron had already been activated by the first stimulus).

It turned out that a nonlinear classifier (SVM with polynomial or RBF kernel) could not achieve a significant performance improvement for any classifier R_t that was specialized for a single time point t (relative to stimulus onset). However

Fig. 4.6C shows that a SVM with a quadratic kernel performs significantly better than a SVM with a linear kernel for the time-invariant classification task of the classifier R_{int} for $int = [150,\ 450]$ ms. A plot of the class-specificity of the firing rates of the most informative neurons for this task (Fig. 4.8) shows, that some correlations between these firing rates produce in fact for certain time periods within the interval $[150,\ 450]$ ms significant class-specific differences, that complement differences of individual firing rates which occur at other time points. Remarkably, it turns out that additional nonlinear computations (besides pairwise correlations) do not increase the performance of a classifier R_{int} for $int = [150,\ 450]$ ms, since a SVM with an RBF kernel does not perform better then a classifier with a quadratic kernel (see Fig. 4.6C).

A nonlinear classifier only provided a performance improvement for time-invariant classification in the case where the response to the first stimulus was perturbed by a second stimulus (see Fig. 4.7).

It turned out that a relatively small subset of neurons contained already most of the information that the nonlinear classifier R_{int} for $int = [150,\ 450]$ ms could extract from the whole ensemble of neurons (see Fig. 4.8A).

4.3.4 Information contained in precise spike times

I also tested whether precise spike times in primary visual cortex carried information about the visual stimulus that had been shown several hundred ms ago. For that purpose the recorded data was perturbed by jittering the spike times. More precisely, each spike was moved by an amount that was drawn from a Gaussian distribution with mean 0 and some standard deviation SD. The results in Fig. 4.9 show that even values of a few ms for SD reduce the performance of classifiers R_t (and classifiers R_{int}) significantly. This also holds true if one first shuffles spike trains between different trials (see Fig. 4.9B). This finding is particularly remarkable in view of the high trial-to-trial variability of neuronal responses in these experiments (shown in Fig. 4.10).

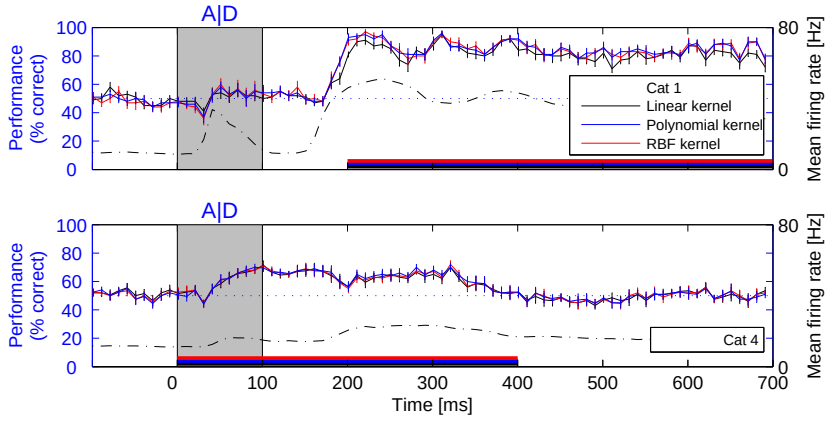


Figure 4.7: Performance of more invariant classifiers R_{int} that were trained to classify a single visual stimulus for all time points t within an interval int of length 500 ms and 400 ms for cat 1 and cat 4, respectively. The performance is comparable to that of the more specialized classifiers R_t in Fig. 4.2. The performance of nonlinear classifiers is for this task not significantly better. This is in contrast to the case where a second intervening stimulus had been shown (see Fig. 4.6C). In the case where only a single visual stimulus was shown, linear classifiers R_{int} are able to extract information about this stimulus for several hundred ms.

4.4 Discussion

I have shown for the presented experimental data that even in the primary visual cortex of anesthetized cats the information about preceding visual stimuli lasts for several hundred ms. These data are consistent with those of Hung *et al.* (2005), who had demonstrated such persistence of information about a single stimulus in a higher cortical area (area IT). It is somewhat surprising that a similar persistence of information can be found in the cortical entrance stage for visual information, which might be seen as being substantially more input-driven rather than percept-driven. Furthermore, the data show that information from different frames of visual input sequences is superimposed, and information about completely different frames is transmitted simultaneously by spikes from the same ensemble of neurons in primary visual cortex. These findings, in conjunction with the observed diversity of individual neural responses (see Fig. 4.8B), support computational models for visual processing that view the highly recurrent neuronal

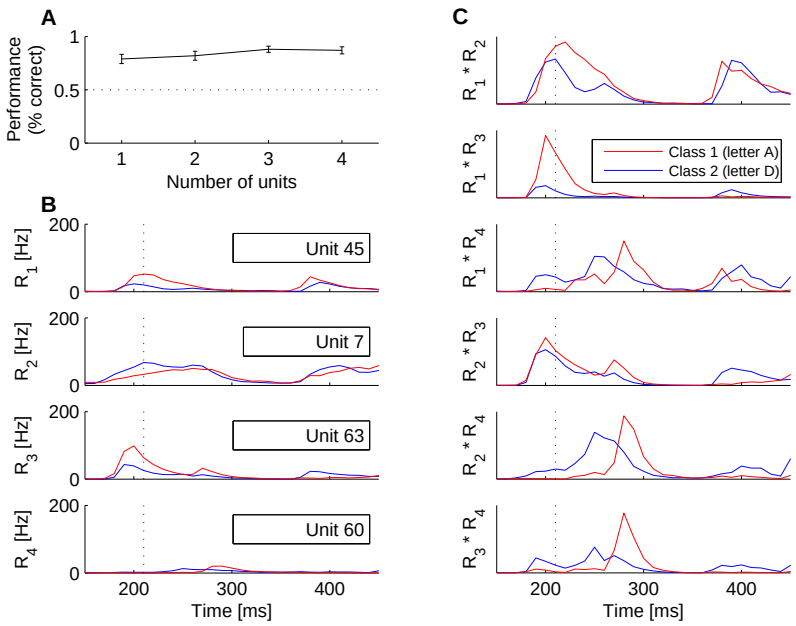


Figure 4.8: Analysis of contributions of individual neurons to the performance of a classifier R_{int} for $int = [150, 450]$ ms (SVM with quadratic kernel from Fig. 4.6C). **A:** Performance is already maximal with very few neurons. **B:** Average firing rates of the 4 most informative neurons 210 ms after stimulus onset for the two classes (defined by the first stimulus A or D) marked in red (for A) and blue (for D). **C:** Class dependent correlations exhibit marked differences at various time points.

networks in the visual cortex as kernel-like processors that fuse information from different time slices in such a way, that numerous different classifications and conclusions can be extracted by projection neurons that receive synaptic inputs from a large number of neurons within this recurrent circuit. Such superposition of information is also a prediction of preceding computer simulations of generic cortical microcircuits (Natschläger & Maass (2005)). On the other hand, these experimental data are harder to explain on the basis of computational models that view the visual cortex as a more orderly processing pipeline, where a sequence of precisely structured processing steps is applied sequentially to each frame in a sequence of visual inputs. The experimental data are also hard to explain on the basis of computational models that postulate precisely structured and well-defined computational roles for feedback connections to primary visual

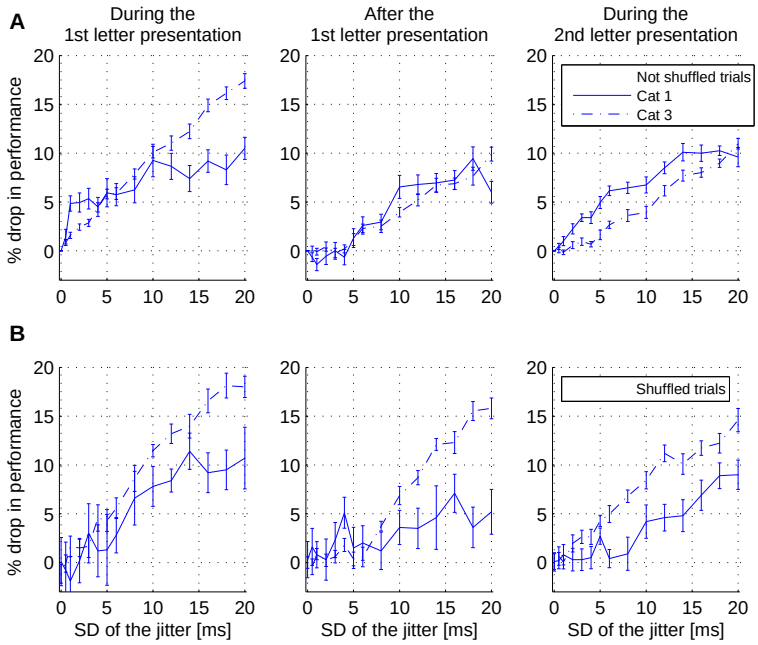


Figure 4.9: Drop in performance for the classifiers R_t in Fig. 4.3 due to the Gaussian jitter of spike times. **A:** Drop in performance computed for three points in time according to the original peaks in performance in Fig. 4.3. For cat 1, these peaks were $t \in \{60, 120, 200\}$ ms and for cat 3, $t \in \{40, 120, 230\}$ ms. The performance drops for these three points in time are shown in the three panels, respectively, and in the order left to right. A standard deviation (SD) of the jitter of only a few milliseconds decreased the classification performance significantly. **B:** Drop in performance for spike data that was shuffled randomly across trials. The drop in performance was computed at the performance peaks at $t \in \{40, 130, 220\}$ ms, for cat 1, and $t \in \{40, 120, 250\}$ ms, for cat 3. Jittering the spike times of trial shuffled data resulted in a similar performance drop when compared to un-shuffled data.

cortex, such as input completion. These models require a more precise temporal organization of information about subsequent frames of visual inputs, in order to make a comparison of bottom-up visual input with top-down interpretations possible. Superposition of information from several frames seems to make input completion for any single frame impossible.

I used methods for extracting information from simultaneously recorded ensembles of neurons that mimic the computational capabilities of hypothetical

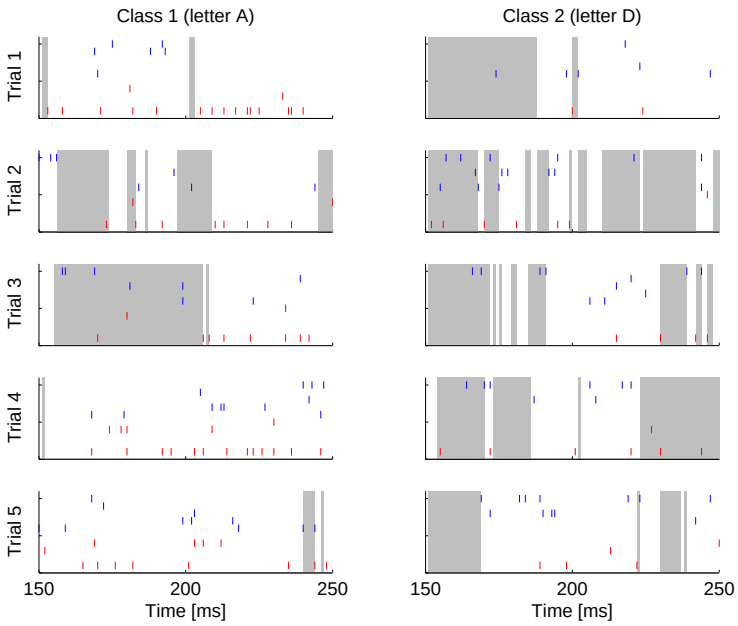


Figure 4.10: Trial-to-trial variability of neuronal responses and classification errors of a classifier R_{int} for $int = [150,\ 450]$ ms (SVM with linear kernel from Fig. 4.6C) during the time interval $[150,\ 250]$ ms for 10 test trials (five for each of the two classes defined by the first stimulus A or D). Shown are the spike times for the 10 units with the 5 largest positive weights (marked in blue) and the 5 largest negative weights (marked in red). Gray shading indicates time periods during which the corresponding trial was misclassified by the classifier.

projection neurons, thereby showing that this information is not only present in the spike trains from this ensemble of neurons, but could in principle also be extracted and used by the brain. But further experiments in awake and behaving animals are needed in order to test whether this information is actually used by the animal. Another open problem arises from the observation that in some cases the performance of classifiers that also had access to correlations between the firing of individual neurons performed better than simple linear classifiers. Experiments with larger numbers of electrodes will have to determine whether such correlations are still needed when one records from a number of neurons that approaches the member of presynaptic neurons for a biological neuron.

A surprising fact that emerged from the data analysis was, that the informa-

tion about visual stimuli that were shown (and had disappeared) several hundred ms ago, was not only contained in the firing rates, but also in the precise timing of spikes. A random jittering of the actually recorded spike trains by just a few ms reduced the amount of information (even if the readout classifier has previously been trained on such artificially jittered spike trains). This result suggests that spike timing in a fairly large ensemble of neurons in primary visual cortex contains extra information over fairly long periods of time, in addition to the information contained in firing rates.

In Chapter 5 these findings are compared to the results for a detailed computer model for a 1.1×1.1 mm^2 patch of cat primary visual cortex (area 17). In contrast to the microcircuit model in Chapter 3 that was intended to model a generic neocortical micro-column this model also incorporates lateral short range connectivity as found in cat. Additionally it is complemented by standard models for retina and LGN that replicate the statistical structure of cortical inputs in a biologically realistic way. I will focus in particular on the temporal dynamics of information about previously shown stimuli and how this information is encoded in the neural activity.

Chapter 5

Comparison of a detailed model for cat primary visual cortex with experimental recordings

Summary

A circuit model for a small patch of cat primary visual cortex (area 17) is presented, that is based on detailed anatomical and physiological data and complemented by a standard model for the retina and the LGN. The temporal dynamics of neural activity and information in the model is compared with multi-electrode recordings from cat primary visual cortex. It is shown that in the model information from different frames of visual input sequences is superimposed and information about completely different frames is simultaneously available to a neural readout. Additionally it is demonstrated that the information about preceding stimuli is not only contained in slowly varying rates, but also in the precise timing of spikes with a temporal resolution of a few milliseconds. Both of these findings closely resemble the results for recordings in primary visual cortex of cat. Finally it is shown that the nonlinear properties of cortical representations observed in experiments can be sufficiently explained in terms of the nonlinear computations performed locally by the recurrent circuitry.

5.1 Introduction

Processing of visual information in the primary visual cortex (V1) has been a subject of extensive research. Nevertheless, its computational role and the mechanisms underlying it remain poorly understood. 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)). In Schwabe & Obermayer (2005) a common principle of optimal encoding of sensory information was hypothesized that resulted in changes in the stimulus tuning functions which relate to experimentally observed changes during attentional modulations and perceptual learning.

This study combines detailed computer modeling with neurobiological experiments in order to arrive at a more accurate model for primary visual cortex with respect to neuronal firing statistics and temporal dynamics of information. The model consists of a small patch of area 17 and is based on the cortical microcircuit model described in Chapter 3 that implements experimental data data from Thomson *et al.* (2002) on lamina-specific connection probabilities and connection strengths between excitatory and inhibitory neurons in layers 2/3, 4, 5, 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.

In contrast to Chapter 3, where I concentrated on the generic information processing capability to hold and fuse information contained in generic Poisson input spike trains, I here explicitly model 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 I implemented 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. Physiological studies show that in response to natural time-varying stimuli (movies) neurons in the cat LGN remove temporal correlations at the time scale of 50 - 300 ms and have a power spectrum that is whitened in the frequency range of 3-15 Hz (Dan *et al.* (1996)). Furthermore this model accounts for lagged and non-lagged cells, which have been observed experimentally (Humphrey & Weller (1988a,b); Mastronarde (1987)).

The temporal dynamics of information about previously shown stimuli is compared between the V1 model and recordings from cat primary visual cortex (area 17). For this purpose the experimental setup used for the analysis of the multi-electrode recordings in Chapter 4 is replicated for the V1 model. Additional to identical stimuli, i.e. sequences of white letters, the readouts were modeled as virtual electrodes that are arranged in a realistic manner with respect to position and number of neurons (units) from which they record.

5.2 Methods

5.2.1 Retina and LGN

I modeled a realistically dense lattice of LGN X-cells, that dominate central cat V1 physiology Ferster (1990). The model consisted of 10000 LGN cells that were laid out on a $50 \times 50 \times 4$ grid. Anatomical and physiological studies have shown that within the X pathway one LGN cell receives retinal input from very few, i.e. about 1-3, ganglion cells (Hamos *et al.* (1987); Mastronarde (1992); Robson (1993)). Therefore each LGN X-cell in the model received input from a single retinal ganglion X-cell and each retinal ganglion X-cell projected to four LGN X-cells (Troyer *et al.* (1998); Worgotter & Koch (1991)). At 5^o of eccentricity the resulting 50×50 grid of retinal ganglion X-cells corresponds to a patch of 1.61×1.61 mm^2 retina (Peichl & Wassle (1979)), which covers a visual field of $7.1^o \times 7.1^o$ (Bishop *et al.* (1962)).

The LGN cells closely followed the standard design of Dong & Atick (1995); Troyer *et al.* (1998) (see also Teich & Qian (2006)) and were modeled as a spatio-

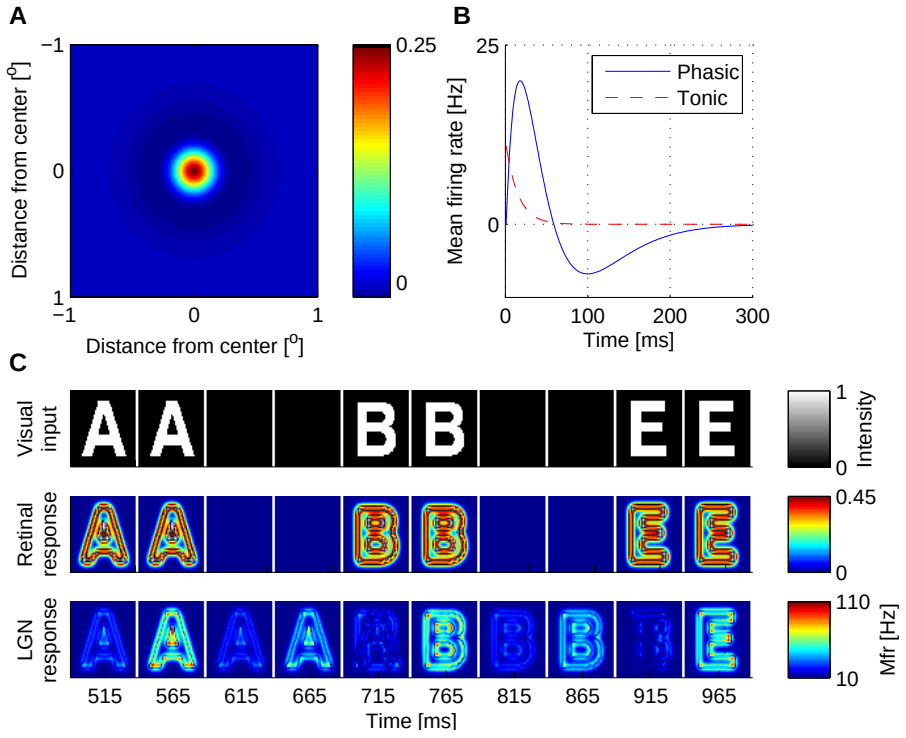


Figure 5.1: Model for the retina and the LGN. **A:** Spatial retinal kernels in the shape of a 'Mexican hat'. **B:** Temporal LGN kernels used to convolve the retinal response consisting of a temporal phasic response and a weaker longer lasting tonic response. The negative response of the phasic component is converted to positive firing rates by the lagged response of the LGN model. The total response to a flash lasts about 200 ms. **C:** Temporal evolution of the retinal and the LGN response to a visual stimulus for consecutive points in time separated by 50 ms (from left to right). Shown is the absolute value of the filter output that corresponds to the sum of the responses of on-center and off-center cells for the retinal model and lagged and non-lagged cells for the LGN model. Retinal and LGN cells are located on a horizontal 50×50 grid covering a visual field of $7.1^o \times 7.1^o$.

temporal filter bank with non-linearities. The filter bank converted time varying input signals which were given by the photoreceptor activity on the retina into firing rates of LGN neurons. From these firing rates spike trains were generated as LGN output. I extended this standard model by two features. Additional to the short transient high-frequency phasic discharge of LGN cells (in response to

a small light spot) I implemented a longer lasting tonic discharge with temporal dynamics taken from Gazeres *et al.* (1998). Secondly, I accounted for the difference in firing statistics for the initial phasic response and the tonic response by implementing the switching gamma renewal process described in Gazeres *et al.* (1998) as spike generation mechanism. In particular the initial phasic response of LGN cells is more regular than the tonic response, which could be caused by an increased contribution of bursts. Finally, no mechanism for luminance and contrast gain control was implemented because current models (Bonin *et al.* (2005); Troyer *et al.* (1998)) are only applicable to drifting gratings with fixed contrast through time and not to sequences of letters.

In detail each of the 50×50 retinal cells had a center-surround spatial receptive field that was modeled by a difference of Gaussians

$$h_{spatial} = G_{\sigma_{r+}}(x_r, y_r) - G_{\sigma_{r-}}(x_r, y_r) \tag{5.1}$$

$$G_{\sigma_{r+/-}}(x_r, y_r) = \frac{1}{2\pi\sigma_{r+/-}^2} \exp\left(-\frac{x_r^2 + y_r^2}{2\sigma_{r+/-}^2}\right) \tag{5.2}$$

where the parameters $\sigma_{r+} = 0.11^o$ and $\sigma_{r-} = 0.33^o$ were taken from Gazeres *et al.* (1998). This filter resembles spatial decorrelation properties of the retina Dong & Atick (1995). Positive retinal output values were assigned to on-center cells and negative values to off-center cells.

The retinal output was filtered by the LGN model using the temporal kernels illustrated in Fig. 5.1B. The output of the LGN model consisted of the sum of two components. A phasic component that was modeled according to Dong & Atick (1995) by

$$k_{phasic}(t) = k_0 t(1 - \pi\omega_c t)\exp(-2\pi\omega_c t), \quad t \geq 0$$

with a characteristic noise cutoff $\omega_c = 5.5$, and a tonic component that was implemented according to Gazeres *et al.* (1998) with $k_{tonic}(t) = k_1 \exp(-t/\tau),\ t \geq 0$ and $\tau = 15$ ms. The parameters k_0 and k_1 were chosen so that the maximum phasic response was 3 times larger that the tonic response and yielded a value of 160 Hz in response to an optimal centered spot (Gazeres *et al.* (1998)). Negative LGN output values were assigned to lagged cells and positive values were assigned

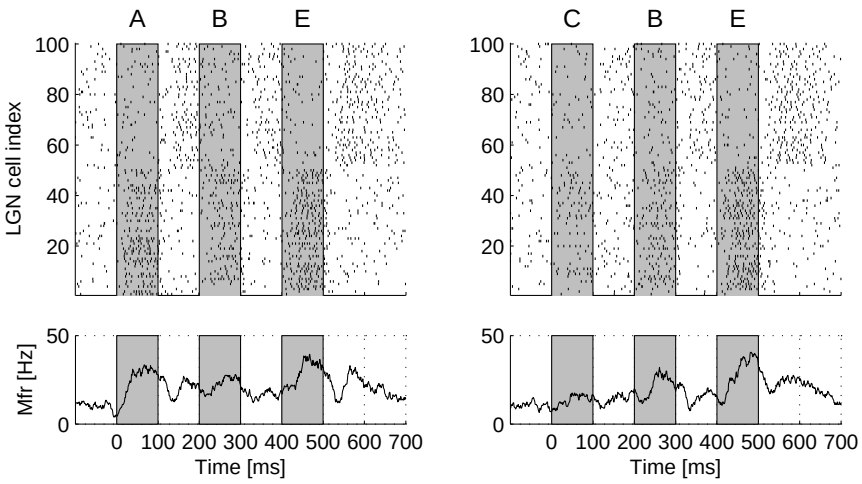


Figure 5.2: LGN output spike trains for the two sequences ABE and CBE. **A:** Spike trains of 100 LGN cell with firing rates drawn from a gamma renewal process switching at a frequency of 30 Hz between regular spiking and a Poisson process for spontaneous discharge. The first (second) half of the LGN cells correspond to lagged (non-lagged) cells. In this and in all other figures the gray boxes indicate the time periods during which the letter-stimuli were visible on the screen. **B:** Mean firing rate (Mfr) of the LGN cells shown in B calculated with a sliding window of length 20 ms.

to non-lagged cells. The spontaneous discharge frequency of LGN cells was chosen to be 10 Hz (Gazeres *et al.* (1998)).

Therefore each visual input to the retina caused four different responses, i.e. that of any combination of non-lagged or lagged cells in the LGN model with on-center or off-center cells in the retinal model. The four LGN cells corresponding to the four different responses to a single visual input were aligned along the z-axis of the $50 \times 50 \times 4$ grid. The preprocessing stages of the retinal and the LGN model are illustrated in Fig. 5.1C.

The firing rates of LGN cells were converted to spike trains by means of a switching gamma renewal process introduced by Gazeres *et al.* (1998) for modeling spontaneous and visually evoked spike trains within the same mathematical framework. The switching frequency between regular spiking during phasic responses ($r = 5$) and a Poisson process for spontaneous discharge ($r = 1$) was chosen to be 30 Hz. An example of LGN output spike trains is shown in Fig. 5.2.

5.2.2 Cortical model

The model for cat primary visual cortex was based on the model described in Chapter 3 and consisted of three layers, with 30%, 20% and 50% of the neurons assigned to layers 2/3, layer 4 and layer 5, respectively. Each layer consisted of a population of excitatory neurons and a population of inhibitory neurons with a ratio of 4:1. Connection probabilities and connection strengths between 6 specific populations of neurons (excitatory and inhibitory neurons on layer 2/3, 4 and 5) were chosen according to experimental data assembled in Thomson *et al.* (2002). 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)). Neurons were modeled as conductance based single compartment Hodgkin-Huxley neurons (Destexhe & Pare (1999); Destexhe *et al.* (2001)). In addition each neuron received synaptic background noise reflecting the bombardment of synaptic inputs from a large number of more distal neurons which cause a depolarization of the membrane potential and a lower input resistance commonly referred to as 'high conductance state' (Destexhe *et al.* (2001)). Two afferent input streams were injected into the circuit. The first input stream (representing LGN input) was injected mainly into layer 4, but also to a lesser extent into layers 2/3 and layer 5. The second afferent input stream (modeling top-down input) was injected into layer 2/3.

In contrast to the microcircuit model in Chapter 3 that was intended to model a generic neocortical micro-column, I here adapted this model for the specific application to a 1.1×1.1 mm^2 patch of cat area 17. The model consisted of 4840 neurons located on a $22 \times 22 \times 10$ grid with a lattice spacing of 50 μm. This roughly corresponds to the size of a vertically oriented pyramidal cell module in cat area 17 that is centered around a cluster of apical dendrites from layer 5 pyramidal cells (Peters & Yilmaz (1993)). In cat each of this pyramidal cell modules contains approximately 203 neurons.

In accordance with biological data short range lateral connectivity is isotropic and nonspecific and inhibition acts on a shorter length scale than excitation (Callaway (1998); Callaway & Wiser (1996); Fitzpatrick *et al.* (1985); Lund (1987)).

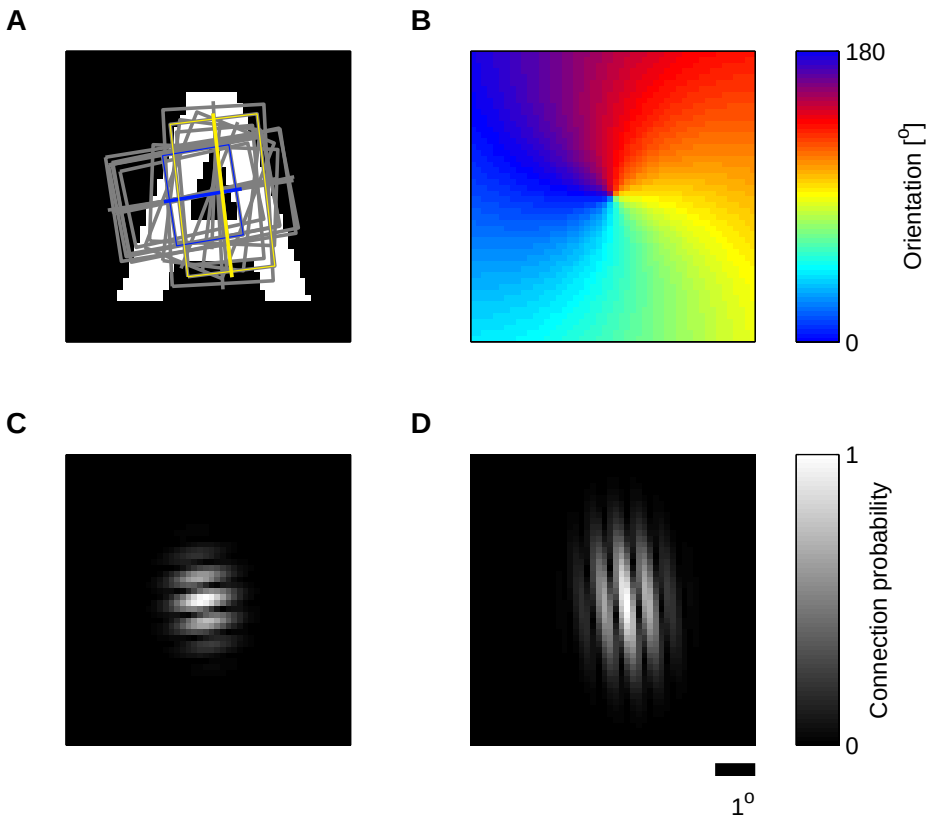


Figure 5.3: Cortical receptive fields. **A:** Receptive field contours for 12 neurons located in layer 4 of the microcircuit model. The center and the orientation of the receptive field of each neuron is determined by a visuotopic map and an orientation map, respectively. **B:** Orientation map laid out in the shape of a pinwheel. The orientation of a receptive field is chosen according to the location of its center in the visual field. **C:** Receptive field of an inhibitory neuron associated with the blue contour shown in panel A. The probability of a connection between the inhibitory neuron and a LGN cell that receive input from a certain position on the retina is given by the absolute value of a Gabor function. Negative and positive values of the Gabor function represent connections to LGN on-center cells and LGN off-center cells, respectively. **D:** Receptive field of an excitatory neuron associated with the yellow contour shown in panel A.

The lateral connection probabilities were modeled by a Gaussian

$$c = c_0 \exp\left(-\frac{x_c^2 + y_c^2}{2\sigma_c^2}\right)$$

where σ_c was chosen to be 200 μm and 100 μm for excitatory and inhibitory connections, respectively, and x_c and y_c specified the distances between two cortical cells in the two horizontal directions. The value for the excitatory short range connectivity corresponds roughly to the 260 μm measured by Buzas *et al.* (2006) for cat area 17. For each synaptic connection the peak of the Gaussian c_0 was set to the corresponding value used in Chapter 3 for connections between neurons within a lateral distance of 100 μm.

5.2.3 Cortical receptive fields

The LGN inputs were injected mainly into layer 4, i.e. to 50% of its inhibitory neurons and 80% of its excitatory neurons, but also into 20% of the excitatory neurons in layer 2/3 and 10% of the excitatory neurons in layer 5. The connection probability between a LGN cell and a V1 cell was modeled according to Troyer *et al.* (1998) by a two-dimensional Gabor function with parameters taken from Teich & Qian (2006). For vertically oriented V1 cells the Gabor function is defined by

$$g(x_{lc}, y_{lc}) = \exp\left(-\frac{x_{lc}^2}{2\sigma_{x_{lc}}} - \frac{y_{lc}^2}{2\sigma_{y_{lc}}}\right) \times \cos(2\pi f_g x_{lc} + \phi),$$

where f_g and ϕ determined the spatial frequency and spatial phase, respectively, and $\sigma_{x_{lc}}$ and $\sigma_{y_{lc}}$ set the number of subregions and the aspect ratio of the receptive field. The Gabor function was properly rotated around the center in order to obtain receptive fields with orientation θ.

The distances between a LGN cell and a cortical cell in the two horizontal directions, i.e. x_{lc} and y_{lc}, were calculated after the location of the LGN cell was topographically mapped onto the cortex. According to Buzas *et al.* (2003) the visuotopic map of area 17 is locally smooth and independent of the orientation map. The visuotopic mapping was implemented by aligning the centers of the LGN grid and the cortical grid in the horizontal directions. Subsequently the LGN grid was rescaled so that a shift of 1^0 in the visual field corresponds to a shift of 0.9 mm in the cortex (Troyer *et al.* (1998)). The orientation preference of V1 cells of the model was laid out as a pinwheel as illustrated in Fig. 5.3B.

All connections from the LGN model to the microcircuit model were excitatory. Positive and negative values of the Gabor function represented connections

from on-center LGN cells and off-center LGN cells, respectively (regardless if the LGN cells were lagged or non-lagged). The absolute values of the Gabor function represented connection probabilities. All connection probabilities were additionally rescaled by a single global scaling factor so that a layer 4 excitatory cell had on average 32 LGN inputs. This is the same number as chosen in Chapter 3.

The width and the length of a receptive field were defined by the number of subregions and the aspect ration. The number of subregions was given by the ratio of the width of the Gaussian to the width of a half-cycle of the sinusoid. The aspect ratio was defined by the ratio of the length of the Gaussian to the width of a half-cycle of the sinusoid. The width and the length of the Gaussian were defined by the 5% value of the peak of the Gaussian along the short and the long axes.

Orientation tuning was implemented according to the inhibition-dominant recurrent model proposed in McLaughlin *et al.* (2000); Tao *et al.* (2004). In contrast to feedforward models for orientation selectivity of cortical cells recurrent models assume only a weak feedforward orientation bias by the mechanism of Hubel and Wiesel and a subsequent sharpening of the tuning by phase-insensitive intracortical connections (Carandini & Ringach (1997); Douglas *et al.* (1995); Somers *et al.* (1995)). In the inhibition-dominant recurrent model this tuning is achieved by a more broadly tuned feedforward LGN input to inhibitory cells than to excitatory cells to produce an effective Mexican hat. According to Teich & Qian (2006) aspect ratios of 4.54 and 2.0 were used for excitatory and inhibitory cells, respectively, and 2.65 subregions were chosen for both cell types.

The layout of spatial frequency maps in cat area 17 remains controversial. Several descriptions of the organization of spatial frequency preference have been proposed (Everson *et al.* (1998); Hubener *et al.* (1997); Issa *et al.* (2000); Shoham *et al.* (1997); Tolhurst & Thompson (1982); Tootell *et al.* (1981)) that share the common finding that spatial frequency varies systematically within a hypercolumn. In contrast Sirovich & Uglesich (2004) showed that a nonspecific component biased some of these results and argued that there is little evidence for spatial frequency columns on the scale of cortical orientation columns. Therefore the spatial frequency for each receptive field was drawn randomly from a Gaussian distribution with mean 0.8 cycles/^{o} and standard deviation 0.1 cycles/^{o}. This

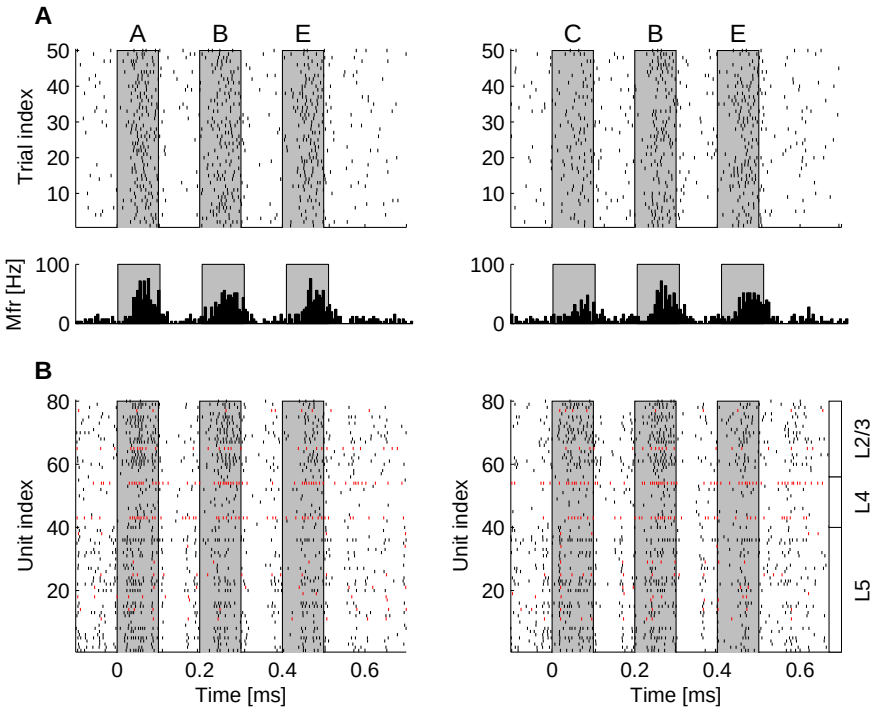


Figure 5.4: **A: Upper part:** Spike times recorded from neuron number 76 across 50 stimulus presentations for the two stimulus sequences ABE and CBE. **Lower part:** Peri-stimulus time histogram for the responses of this neuron. Mfr: mean firing rate. **B:** Spike trains from 80 selected neurons located around the positions of virtual electrodes for trial number 5. Spike trains of excitatory and inhibitory neurons are indicated in black and red, respectively.

corresponds roughly to the mean preferred spatial frequency of cortical cells in cat at 5^o of eccentricity (Movshon *et al.* (1978)). The spatial phase preference of cells varies widely from cortical cell to cortical cell and does not cluster with any other receptive field property (DeAngelis *et al.* (1999)). In agreement with these results the spatial phase was drawn randomly from a uniform distribution over the interval $[0, 2\pi]$.

The second input stream modeling top-down input consisted of 50×50 Poisson inputs with firing rates between 10Hz and 15Hz that were injected into layer 2/3 (see Chapter 3 for details). The same visuotopic mapping as for input stream 1 was applied. Instead of the Gabor connectivity function a uniform connection

probability was chosen. The maximum extent of lateral connections was scaled such that a cortical neuron could receive input from at most 40 LGN neurons. The connection probability was set to the corresponding values used in Chapter 3 [1]. Because of the visuotopic mapping the inputs to two different cortical cells are less correlated than in Chapter 3 and represent a stronger noise source. Therefore the input strength of input stream 2 was scaled by a factor of 1/2.

5.2.4 Classification task

The stimuli consisted of sequences of white letters, i.e. ABE, CBE, ADE and CDE, of size 50×50 that were presented on a black background. The letters covered approximately 5^o of the visual field. Each letter of a sequence was presented for 100 ms. Intervals between the presentations of subsequent letters lasted 100 ms.

In order to replicate the results for the multi-electrode recordings reported in Chapter 4 a virtual 4×4 electrode array centered in the microcircuit model was simulated. This was done be recording the spiking activity of the 80 closest neurons around the 16 virtual electrodes. Adjacent electrodes were separated horizontally by a distance of 200μm. Because the layers from which the real electrodes recorded are unknown the virtual electrodes recorded from all neurons along the vertical axis of the model (covering all layers). The spike raster of neurons recorded by a virtual electrode array is shown in Fig. 5.4.

The recorded spike times of neurons were convolved with an exponential kernel with a decay time constant of $\tau = 20$ ms. A linear support vector machine (SVM) was trained to discriminate between pairs of stimuli on the basis of the convolved spike trains at time points $t \in \{0, 10, ..., 700\}$ ms with respect to stimulus onset. The parameter C of the linear SVM was chosen to be 5.

Each stimulus sequence was presented 50 times. The classification performance was estimated with 5-fold cross validation and averaged over 5 circuits. All reported results are for test data. Error bars indicate the standard error of the test error of all 5 cross validation runs.

[1] With this setup the number of inputs injected into a V1 neuron is identical to the corresponding number of inputs obtained for the model in Chapter 3

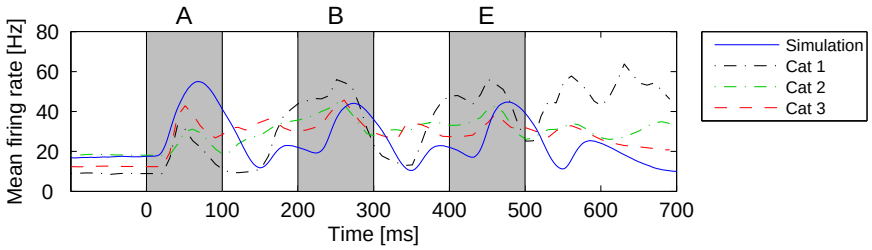


Figure 5.5: Mean firing rates of the microcircuit model and multi-electrode recordings in cat primary visual cortex for a similar experiment as in Chapter 4. The mean firing rates were averaged over all trials for the sequence ABE. The temporal evolution of the mean firing rate varies for different cats but all curves show characteristic activity peaks during and approximately 80 ms after the presentation of letters. The activity peaks of the microcircuit model due to the lagged and non-lagged responses of LGN cells are in good agreement with the experimental data from cat 3 with a correlation coefficient of 0.53.

5.3 Results

The time course of the mean firing rate of the V1 model shows characteristic activity peaks corresponding to the lagged and non-lagged responses of LGN cells to the visual stimuli (Fig. 5.5). For a short time interval, during the presentation of the second and the third letter of a sequence, lagged and non-lagged LGN cells respond simultaneously. The temporal evolution of the mean firing rate of multi-electrode recordings in cat primary visual cortex for a similar experiment (see Chapter 4) varies for different cats. However, all curves show the same characteristic activity peaks as also observed for the V1 model.

Fig. 5.6A and B show that classifiers can extract substantial information about the first and the second letter in a sequence even after the letters were removed and new letters were shown. Note that the information about the previously shown letter is available although the neuronal rate responses decrease for a short period of time to the level of spontaneous activity. This is even more prominent during the presentation of the third letter (Fig. 5.6B) than during the presentation of the second letter (Fig. 5.6A) because the two options B or D for the second time slot represent an additional noise source for the classifier (trained to identify

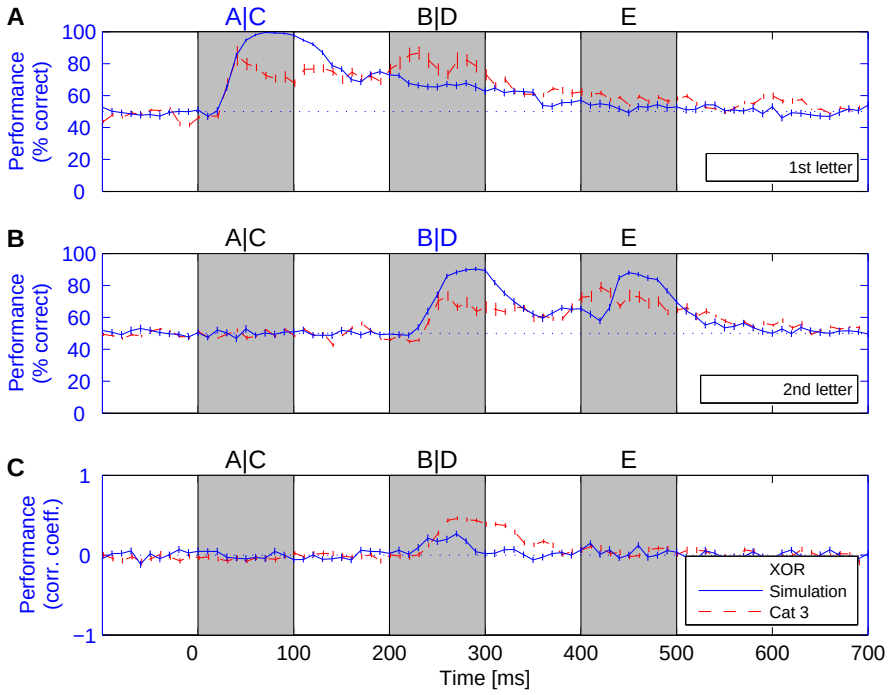


Figure 5.6: Comparison of the temporal evolution of information about previously shown stimuli for the microcircuit model and for a similar experiment in cat primary visual cortex (Chapter 4, cat 3). The classifiers were trained to discriminate at time t between pairs of stimuli on the basis of the convolved spike trains recored with a real and a virtual electrode for the cat and the microcircuit model, respectively. **A:** Performance of classifiers trained to identify at time t the first letter in a sequence. **B:** Performance of classifiers trained to identify at time t the second letter in a sequence. **C:** A linear classifiers was trained to compute at time t the XOR function of the 2 bits encoded by the 2 choices A or C and B or D. The results for the microcircuit model and for the recordings from cat show similar performance peaks due to the lagged and non-lagged responses of LGN cells. The results for the XOR task suggest that the performance is primarily limited by the available information about each of its input bits.

the first letter in a sequence).

The temporal dynamics of information shows similar performance peaks for the V1 model when compared to experimental data from cat (Chapter 4, cat 3). The time points of the performance peaks match with the corresponding time points of the activity peaks for the mean firing rate. The results for cat show

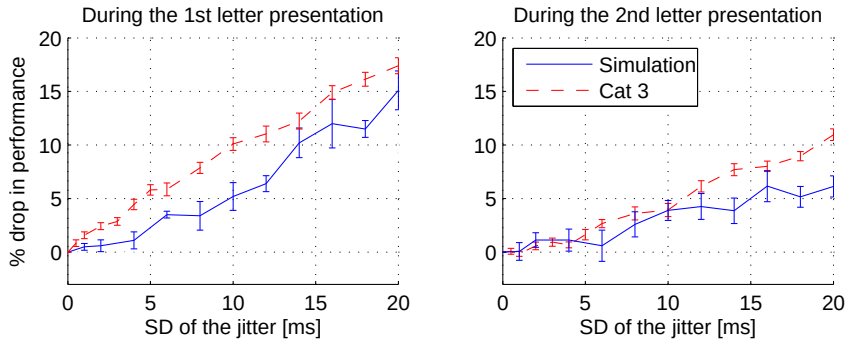


Figure 5.7: Drop in performance for the classifiers in Fig. 5.6A due to adding Gaussian jitter to the spike times. The drop in performance was computed for two points in time according to the performance peaks during the first and the second letter presentation. For the V1 model these peaks were at $t \in \{70, 200\}$ ms and for cat at $t \in \{40, 230\}$ ms. For the V1 model and the cat a standard deviation (SD) of only a few milliseconds decreases the classification performance significantly.

somewhat better temporal integration properties. The performance of a classifier trained to classify the first letter in a sequence is less affected by the appearance of the second letter. Nevertheless, the total time interval during which information about previously shown stimuli is available is comparable between the V1 model and the experimental results for cat.

After the presentation of the second letter, the spiking activity contains simultaneously high information about both of the sequentially presented letters as also observed in the experimental data. Virtual electrode recordings from neurons in the V1 model contain information about the exclusive OR (XOR) of the 2 bits that are encoded through the choice of the first two letters, as also observed experimentally (see Fig. 5.6C). This suggests that the information that is present in the neural activity about the result of this nonlinear XOR computation is primarily limited by the available information about each of its input bits. This information is mainly contained in the lagged and non-lagged responses of LGN cells.

I further analyzed how the information is encoded in the neuronal activity. The ability of classifiers to distinguish between two stimuli is positively correlated

with the time-varying mean firing rate (averaged over all trials and all recorded electrodes, but not over time). The correlation coefficient of the performance of a classifier trained to classify the first letter in sequence and the mean firing rate is 0.45. This value is somewhat lower than the corresponding value of 0.66 for cat. The correlation coefficient of the performance of the classifier and the difference in the mean firing rates caused by the two stimuli classes has a value of 0.71 and is larger than for cat (0.46).

Similar to the experimental results the precise timing of spikes carries information about the stimulus identity. Perturbing the recorded spike times by an amount drawn from a Gaussian distribution with mean 0 and a standard deviation (SD) of a few ms decreases the performance of the classifier significantly. (Fig. 5.7). The classification performance dropped even for time points 200 ms after the stimulus onset. The similarity between the results for the simulation and the experiment suggests that the amount of information contained in the precise timing of spikes is not due to a specifically engineered neural coding in the brain but rather a general consequence of computing in recurrent neural microcircuits.

5.4 Discussion

I have shown that a detailed computer model of a small patch of cat primary visual cortex that implements experimental data on lamina-specific connection probabilities and connection strengths between neurons in three different layers replicates the characteristic features of the temporal dynamics of information found in experiments (Chapter 4).

The information about previously shown stimuli contained in the neural activity of the microcircuit is not erased by the presentation of a new stimulus but it is possible to extract information about multiple stimuli simultaneously. Furthermore a linear classifiers applied to the spike times recorded with simulated electrodes can extract information about results of nonlinear computations on visual stimuli in the same order of magnitude as real electrodes from cat primary visual cortex. Thus the nonlinear properties of cortical representations observed in experiments can be sufficiently explained in terms of the nonlinear computations performed locally by the recurrent synaptic connections. The results for

the XOR task suggest that the performance is primarily limited by the available information about each of its input bits.

Finally the results also agree with the experimental results obtained in Chapter 4 in indicating that information is not only contained in rate responses, but that the precise timing of spikes with a temporal resolution of a few milliseconds carries additional information about the stimulus. The results for the V1 model show a similar performance drop when compared to the corresponding results for cat. This suggests that computer simulations represent an adequate testbed for the further analysis of this effect.

Chapter 6

Conclusions

The preceding chapters have demonstrated how the conceptual framework of the *liquid state machine* can be applied to gain further understanding of the computational properties of the neocortical circuitry.

In Chapter 2 I investigated the high dimensional transient dynamics of neural microcircuit models from the point of view of one or two readout neurons. It was shown that pairs of readout neurons can extract clearly structured low-dimensional projections of these complex transient states that can even exhibit convergence to virtual attractors which are not observable in the high dimensional trajectories. These virtual attractor landscapes may differ for each readout neuron making it in principle possible to carry out particular computations needed by specific readouts without changing the dynamics of the recurrent circuit itself. Furthermore it was demonstrated that sufficiently large neural microcircuits can play a similar role for information processing as kernels for support vector machines in machine learning and enable linear readout neurons to classify time-varying inputs with the same power as complex nonlinear classifiers. Linear readouts have the advantage that they can be trained by simple learning algorithms that cannot get stuck in local minima of the error function.

In Chapter 3 the specific computational properties of detailed cortical microcircuit models with biologically realistic lamina-specific synaptic connection patterns were explored. It was shown that these cortical microcircuit models exhibit specific computational advantages over various types of control circuits that have the same components and the same global statistics of synaptic connections,

but are missing the lamina-specific structure of real cortical microcircuits. This superiority holds for various values of the global strength of recurrent connections and afferent input streams, corresponding to different regulatory states of neural systems *in vivo* and stimulus intensities. Furthermore these information processing capabilities increase with circuit size suggesting that almost perfect performance will be achieved for biologically realistic circuit sizes. It was shown that the connectivity graphs defined by this cortical microcircuit template have small-world properties, but it turned out that other properties of the connectivity graphs are more salient for the computational performance. In particular the distribution of the total number of input and output synapses to a neuron relative to circuit inputs and projection neurons is primarily responsible for the better computational performance. It was shown that these properties of the connectivity graphs support noise suppression and thereby enhance the computational capabilities of the cortical microcircuit model.

In Chapter 4 an experimental test of the predictions of the computational model *liquid state machine* was carried out for visual processing in primary visual cortex of anesthetized cats. Some quite surprising facts about the temporal evolution of the information about visual stimuli that can be recovered from multi-unit recordings were demonstrated. First, this information lasts several hundred ms, and information from subsequently presented stimuli is superimposed. Second, this information about preceding stimuli is not only contained in slowly varying rates, but also in the precise timing of spikes. Altogether the experimental paradigm that has been explored differs in essential aspects from most experiments that aim at providing information about computational processes in primary visual cortex: The data consisted of recordings from a randomly chosen subset of neurons, rather than from a subset of neurons that were selected because they responded in a predictable manner. Furthermore a sequence of visual stimuli was presented, rather than a single one, which is arguably closer to natural vision than a single static image. This setup provides an opportunity to analyze the fusion of information from several preceding time segments. Finally, it turned out that almost all information that could be extracted in this way from multi-unit spike activity by state-of-the-art nonlinear classifiers, could also be extracted by linear classifier. Therefore it is not impossible that the same information, and in

fact substantially more information from several thousand presynaptic neurons instead of the roughly hundred neurons from which was recorded, is available to projection neurons that read out information from primary visual cortex. These experiments demonstrated that the temporal dynamics of information in primary visual cortex is not compatible with simple models based on linear filters. The exhibited superposition of slowly fading memory from different preceding stimuli, the diversity of temporal profiles recorded at different electrodes, and the fact that linear readout models sufficed even for nonlinear processing tasks can be seen as support for recently proposed computational models Jäger & Haas (2004); Maass *et al.* (2002) based on nonlinear dynamic systems theory.

Finally, in Chapter 5 it was demonstrated that a detailed computer model of a small patch of cat primary visual cortex that is based on the realistic cortical microcircuit model presented in Chapter 3 exhibits the characteristic features of the temporal dynamics of information found experimentally in Chapter 4. The results agree with respect to the capability of a readout to extract simultaneously information about different preceding stimuli. Furthermore linear readouts can retrieve information about nonlinear relations between these stimuli indicating nonlinear fusion of information due to the recurrent synaptic connections of the microcircuit model. Finally, it was shown the precise timing of spikes with a temporal precision of a few milliseconds carries additional information about previously shown stimuli that is not contained in rate responses.

Apparently the neural circuit models considered in this thesis represent the most detailed data-based cortical microcircuit models whose information processing capabilities have been analyzed so far. The results of this analysis show that it is possible to exhibit through extensive computer simulations specific computational consequences of their circuitry, thereby creating a link between detailed anatomical and neurophysiological data and their likely computational consequences.

Appendix A

Appendix

Supplementary Theorem 1 *The correlation coefficient between the XOR function* $f_{XOR} : \{\mathbf{s}_A, \mathbf{s}_C\} \times \{\mathbf{s}_B, \mathbf{s}_D\} \to \mathbb{R}$ *of the visual stimuli* $\mathbf{s}_A, \mathbf{s}_B, \mathbf{s}_C, \mathbf{s}_D \in \mathbb{R}^m$ *with* $m \in \mathbb{N}$ *defined by*

$$\begin{aligned} f_{XOR}(\mathbf{s}_A, \mathbf{s}_B) &= f_{XOR}(\mathbf{s}_C, \mathbf{s}_D) = 0 \\ f_{XOR}(\mathbf{s}_A, \mathbf{s}_D) &= f_{XOR}(\mathbf{s}_C, \mathbf{s}_B) = 1 \end{aligned}$$

and any nonconstant linear function $f_L : \{\mathbf{s}_A, \mathbf{s}_C\} \times \{\mathbf{s}_B, \mathbf{s}_D\} \to \mathbb{R}$ *is zero.*

Proof: The correlation coefficient between f_{XOR} and f_L is defined by

$$CC = \frac{\mathrm{COV}(f_{XOR}, f_L)}{\mathrm{VAR}(f_{XOR})\ \mathrm{VAR}(f_L)}$$

where COV and VAR denote the covariance and the variance, respectively. Any linear function $f_L : \{\mathbf{s}_A, \mathbf{s}_C\} \times \{\mathbf{s}_B, \mathbf{s}_D\} \to \mathbb{R}$ can be written as

$$f_L = \mathbf{w}_1 \mathbf{s}_1 + \mathbf{w}_2 \mathbf{s}_2 + b$$

with $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^m$, $b \in \mathbb{R}$ and $\mathbf{s}_1 \in \{\mathbf{s}_A, \mathbf{s}_C\}$, $\mathbf{s}_2 \in \{\mathbf{s}_B, \mathbf{s}_D\}$. Furthermore

$$\begin{aligned}
\mathrm{COV}(f_{XOR}, f_L) &= \sum_{\mathbf{s}_1,\mathbf{s}_2} f_{XOR}(\mathbf{s}_1,\mathbf{s}_2) f_L(\mathbf{s}_1,\mathbf{s}_2)/4 \\
&\quad - \sum_{\mathbf{s}_1,\mathbf{s}_2} f_{XOR}(\mathbf{s}_1,\mathbf{s}_2)/4 \cdot \sum_{\mathbf{s}_1,\mathbf{s}_2} f_L(\mathbf{s}_1,\mathbf{s}_2)/4 \\
&= 0 \cdot (\mathbf{w}_1\mathbf{s}_A + \mathbf{w}_2\mathbf{s}_B + b)/4 \\
&\quad + 1 \cdot (\mathbf{w}_1\mathbf{s}_A + \mathbf{w}_2\mathbf{s}_D + b)/4 \\
&\quad + 1 \cdot (\mathbf{w}_1\mathbf{s}_C + \mathbf{w}_2\mathbf{s}_B + b)/4 \\
&\quad + 0 \cdot (\mathbf{w}_1\mathbf{s}_C + \mathbf{w}_2\mathbf{s}_D + b)/4 \\
&\quad - (\mathbf{w}_1\mathbf{s}_A + \mathbf{w}_2\mathbf{s}_B + \mathbf{w}_1\mathbf{s}_C + \mathbf{w}_2\mathbf{s}_D + 2b)/4 \\
&= 0.
\end{aligned}$$

The variance of nonconstant f_L is larger than 0 and thus $CC = 0$. ∎

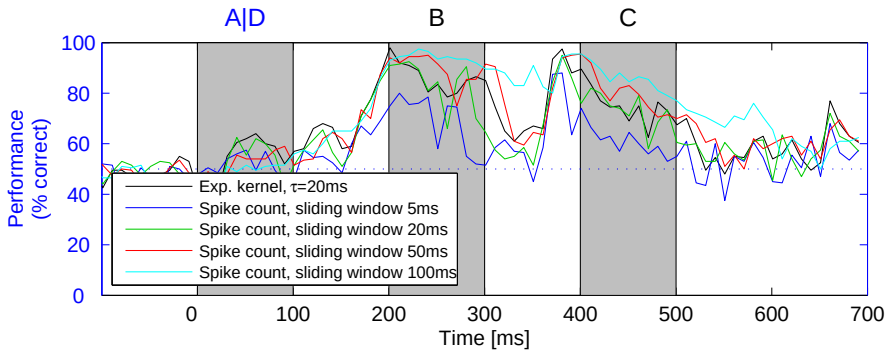


Figure A.1: Comparison of the performance of linear classifiers R_t trained to discriminate between pairs of stimuli on the basis of convolved spike trains (exponential kernel with a decay time constant of $\tau = 20$ ms) or spike counts during the last 20, 50 or 100 ms (see Chapter 4.2.2). The performance of classifiers trained on convolved spike trains is similar to the performance of classifiers trained on spike counts during the last 20 ms.

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. 82

Ahmed, B., Anderson, J.C., Douglas, R.J., Martin, K.A. & Nelson, J.C. (1994). Polyneuronal innervation of spiny stellate neurons in cat visual cortex. *J Comp Neurol*, **341**, 39–49. 4

Amit, D.J. & Brunel, N. (1997). Model of global spontaneous activity and local structured activity during delay periods in the cerebral cortex. *Cerebral Cortex*, **7**, 237–252. 42

Anderson, J., Lampl, I., Reichova, I., Carandini, M. & Ferster, D. (2000). Stimulus dependence of two-state fluctuations of membrane potential in cat visual cortex. *Nature Neuroscience*, **3**, 617–621. 37

Armstrong-James, M., Fox, K. & Das-Gupta, A. (1992). Flow of excitation within rat barrel cortex on striking a single vibrissa. *J Neurophysiol.*, **68**, 1345–1358. 40, 42

Auer, P., Burgsteiner, H. & Maass, W. (2002). Reducing communication for distributed learning in neural networks. In J.R. Dorronsoro, ed., *Proc. of the International Conference on Artificial Neural Networks – ICANN 2002*, vol. 2415 of *Lecture Notes in Computer Science*, 123–128, Springer. 10, 15, 17

Baddeley, R., Abbott, L.F., Booth, M.C., Sengpiel, F., Freeman, T., Wakeman, E.A. & Rolls, E.T. (1997). Responses of neurons in primary

and inferior temporal visual cortices to natural scenes. *Proc R Soc Lond B Biol Sci*, **264**, 1775–1783. 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. 82

Binzegger, T., Douglas, R.J. & Martin, K.A. (2004). A quantitative map of the circuit of cat primary visual cortex. *J. Neurosci.*, **24**, 8441–8453. 39

Bishop, P.O., Kozak, W. & Vakkur, G.J. (1962). Some quantitative aspects of the cat's eye: axis and plane of reference, visual field co-ordinates and optics. *J Physiol*, **163**, 466–502. 83

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. 82

Bollobas, B. (1985). *Random Graphs*. Academic Press, London. 48

Bonin, V., Mante, V. & Carandini, M. (2005). The suppressive field of neurons in lateral geniculate nucleus. *J Neurosci*, **25**, 10844–10856. 85

Borg-Graham, L.J., Monier, C. & Fregnac, Y. (1998). Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature*, **393**, 369–373. 37

Brecht, M., Singer, W. & Engel, A.K. (1999). Patterns of synchronization in the superior colliculus of anesthetized cats. *J Neurosci*, **19**, 3567–3579. 5

Buonomano, D.V. & Merzenich, M.M. (1995). Temporal information transformed into a spatial code by a neural network with realistic properties. *Science*, **267**, 1028–1030. 58, 62

Buzas, P., Volgushev, M., Eysel, U.T. & Kisvarday, Z.F. (2003). Independence of visuotopic representation and orientation map in the visual cortex of the cat. *Eur J Neurosci*, **18**, 957–968. 89

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. *submitted for publication*. 89

Callaway, E.M. (1998). Local circuits in primary visual cortex of the macaque monkey. *Annu Rev Neurosci*, **21**, 47–74. 87

Callaway, E.M. (2004). Feedforward, feedback and inhibitory connections in primate visual cortex. *Neural Networks*, **17**, 625–632. 32

Callaway, E.M. & Wiser, A.K. (1996). Contributions of individual layer 2-5 spiny neurons to local circuits in macaque primary visual cortex. *Vis Neurosci*, **13**, 907–922. 87

Carandini, M. & Ringach, D.L. (1997). Predictions of a recurrent model of orientation selectivity. *Vision Res*, **37**, 3061–3071. 90

Chance, F.S., Nelson, S.B. & Abbott, L.F. (1999). Complex cells as cortically amplified simple cells. *Nat Neurosci*, **2**, 277–82. 82

Chung, S. & Ferster, D. (1998). Strength and orientation tuning of the thalamic input to simple cells revealed by electrically evoked cortical suppression. *Neuron*, **20**, 1177–1189. 38

Dan, Y., Atick, J. & Reid, R. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory. *J. Neurosci.*, **15**, 3351–62. 83

Dantzker, J.L. & Callaway, E.M. (2000). Laminar sources of synaptic input to cortical inhibitory interneurons and pyramidal neurons. *Nature Neuroscience*, **3**, 701–707. 33

DeAngelis, G.C., Ghose, G.M., Ohzawa, I. & Freeman, R.D. (1999). Functional micro-organization of primary visual cortex: receptive field analysis of nearby neurons. *J Neurosci*, **19**, 4046–4064. 91

Destexhe, A. & Marder, E. (2004). Plasticity in single neuron and circuit computations. *Nature*, **431**, 789–795. 32

Destexhe, A. & Pare, D. (1999). Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. *J. Neurophysiol.*, **81**, 1531–1547. 35, 87

Destexhe, A., Rudolph, M., Fellous, J.M. & Sejnowski, T.J. (2001). Fluctuating synaptic conductances recreate in vivo-like activity in neocortical neurons. *Neuroscience*, **107**, 13–24. 35, 36, 37, 87

Destexhe, A., Rudolph, M. & Pare, D. (2003). The high-conductance state of neocortical neurons in vivo. *Nat. Rev. Neurosci.*, **4**, 739–751. 36, 58

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. 82, 83, 85

Douglas, R.J. & Martin, K.A. (2004). Neuronal circuits of the neocortex. *Annu. Rev. Neurosci.*, **27**, 419–451. 30, 31

Douglas, R.J., Koch, C., Mahowald, M., Martin, K. & Suarez, H. (1995). Recurrent excitation in neocortical circuits. *Science*, **269**, 981–985. 30, 31, 90

Duda, R.O., Hart, P.E. & Storck, D.G. (2001). *Pattern Classification*. John Wiley, New York, 2nd edn. 16, 21

Ernst, U.A., Pawelzik, R., K, Sahar-Pikielny, C. & Tsodyks, M. (2001). Intracortical origin of visual maps. *Nat Neurosci*, **4**, 431–6. 82

Everson, R.M., Prashanth, A.K., Gabbay, M., Knight, B.W., Sirovich, L. & Kaplan, E. (1998). Representation of spatial frequency and orientation in the visual cortex. *Proc Natl Acad Sci U S A*, **95**, 8334–8338. 90

Ferster, D. (1987). Origin of orientation-selective epsps in simple cells of cat visual cortex. *Journal of Neuroscience*, **7(6)**, 1780–1791. 38

Ferster, D. (1990). X- and Y-mediated synaptic potentials in neurons of areas 17 and 18 of cat visual cortex. *Vis Neurosci*, **4**, 115–133. 83

Fitzpatrick, D., Lund, J.S. & Blasdel, G.G. (1985). Intrinsic connections of macaque striate cortex: afferent and efferent connections of lamina 4C. *J Neurosci*, **5**, 3329–3349. 87

Freund, Y. & Schapire, R.E. (1999). Large margin classification using the perceptron algorithm. *Machine Learning*, **37**, 277–296. 10, 15

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**, 1157–1174. 85, 86

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**, 334–337. 5

Gupta, A., Wang, Y. & Markram, H. (2000). Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex. *Science*, **287**, 273–278. 11, 29, 31, 33, 34, 36, 82, 87

Hamos, J.E., Van Horn, S.C., Raczkowski, D. & Sherman, S.M. (1987). Synaptic circuits involving an individual retinogeniculate axon in the cat. *J Comp Neurol*, **259**, 165–192. 83

Hertz, J., Krogh, A. & Palmer, R.G. (1991). *Introduction to the Theory of Neural Computation*. Addison-Wesley (Redwood City, CA). 16

Hilgetag, C., Burns, G., O'Neill, M. & Scannell, J. (2000). Anatomical connectivity defines the organization of clusters of cortical areas in the macaque monkey and the cat. *Philos Trans R Soc Lond B Biol Sci.*, **355**, 91–110. 50

Hirsch, J.A., Alonso, J.M., Reid, R.C. & Martinez, L.M. (1998). Synaptic integration in striate cortical simple cells. *J. Neurosci.*, **18**, 9517–9528. 37

Hoffman, D.A., Magee, J.C., Colbert, C.M. & Johnston, D. (1997). K+ channel regulation of signal propagation in dendrites of hippocampal pyramidal neurons. *Nature*, **387**, 869–875. 36

Hubel, D.H. & Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol*, **160**, 106–154. 4

Hubener, M., Shoham, D., Grinvald, A. & Bonhoeffer, T. (1997). Spatial relationships among three columnar systems in cat area 17. *J Neurosci*, **17**, 9270–9284. 90

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. 83

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. 83

Hung, C.P., Kreiman, G., Poggio, T. & DiCarlo, J.J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, **310**, 863–866. 4, 63, 75

Issa, N.P., Trepel, C. & Stryker, M.P. (2000). Spatial frequency maps in cat visual cortex. *J Neurosci*, **20**, 8504–8514. 90

Jäger, H. (2001). The "echo state" approach to analyzing and training recurrent neural networks. GMD Report 148, German National Research Center for Information Technology. 13

Jäger, H. (2002). Short term memory in echo state networks. GMD Report 152, German National Research Center for Information Technology. 62

Jäger, H. & Haas, H. (2004). Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. *Science*, **304**, 78–80. 1, 62, 101

Kaiser, M. & Hilgetag, C. (2004). Spatial growth of real-world networks. *Phys. Rev. E*, **69**, 036103. 50

Kalisman, N., Silberberg, G. & Markram, H. (2005). The neocortical microcircuit as a tabula rasa. *Proc Natl Acad Sci*, **102**, 880–885. 30

Lawson, C.L. & Hanson, R.J. (1974). Chapter 23. In *Solving Least Squares Problems*, 161, Prentice-Hall. 41

Legenstein, R. & Maass, W. (2005). What makes a dynamical system computationally powerful? In S. Haykin, J.C. Principe, T. Sejnowski & J. McWhirter, eds., *New Directions in Statistical Signal Processing: From Systems to Brain*, MIT Press, to appear, http://www.igi.tugraz.at/maass/psfiles/165.pdf. 53, 55

Lund, J.S. (1987). Local circuit neurons of macaque monkey striate cortex: I. Neurons of laminae 4C and 5A. *J Comp Neurol*, **257**, 60–92. 87

Maass, W. & Markram, H. (2002). Synapses as dynamic memory buffers. *Neural Networks*, **15**, 155–161. 33

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. 1, 9, 10, 12, 14, 32, 41, 42, 48, 58, 62, 63, 87, 101

Maass, W., Joshi, P. & Sontag, E.D. (2006). Principles of real-time computing with feedback applied to cortical microcircuit models. In Y. Weiss, B. Schölkopf & J. Platt, eds., *Advances in Neural Information Processing Systems*, vol. 18, 835–842, MIT Press. 32

MacLeod, K., Backer, A. & Laurent, G. (1998). Who reads temporal information contained across synchronized and oscillatory spike trains? *Nature*, **395**, 693–698. 5

Magee, J., Hoffman, D., Colbert, C. & Johnston, D. (1998). Electrical and calcium signaling in dendrites of hippocampal pyramidal neurons. *Annu. Rev. Physiol.*, **60**, 327–346. 36

Magee, J.C. & Johnston, D. (1995). Characterization of single voltage-gated Na+ and Ca2+ channels in apical dendrites of rat CA1 pyramidal neurons. *J. Physiol.*, **487 (Pt 1)**, 67–90. 36

Mainen, Z.T., Joerges, J., Huguenard, J.R. & Sejnowski, T.J. (1995). A model of spike initiation in neocortical pyramidal neurons. *Neuron*, **15**, 1427–1439. 36

Markram, H., Wang, Y. & Tsodyks, M. (1998). Differential signaling via the same axon of neocortical pyramidal neurons. *PNAS*, **95**, 5323–5328. 11, 12, 29, 31, 33, 34, 36, 82, 87

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. 83

Mastronarde, D.N. (1992). Nonlagged relay cells and interneurons in the cat lateral geniculate nucleus: receptive-field properties and retinal inputs. *Vis Neurosci*, **8**, 407–441. 83

McLaughlin, D., Shapley, R., Shelley, M. & Wielaard, D.J. (2000). A neuronal network model of macaque primary visual cortex (V1): orientation selectivity and dynamics in the input layer 4Calpha. *Proc Natl Acad Sci U S A*, **97**, 8087–8092. 90

Mountcastle, V.B. (1998). *Perceptual Neuroscience: The Cerebral Cortex*. Harvard University Press, Cambridge. 30

Movshon, J.A., Thompson, I.D. & Tolhurst, D.J. (1978). Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *J Physiol*, **283**, 101–120. 91

Natschläger, T. & Maass, W. (2005). Dynamics of information and emergent computation in generic neural microcircuit models. *Neural Networks*, **18**, 1301–1308. 63, 76

Natschläger, T., Markram, H. & Maass, W. (2003). Computer models and analysis tools for neural microcircuits. In R. Kötter, ed., *Neuroscience Databases. A Practical Guide*, chap. 9, 123–138, Kluwer Academic Publishers (Boston). 42

Nelson, S. (2002). Cortical microcircuits: diverse or canonical? *Neuron*, **36**, 19–27. 30

Olshausen, B.A. & Field, D.J. (2006). What is the other 85% of V1 doing? In T.J. Sejnowski & L. van Hemmen, eds., *Problems in Systems Neuroscience*, 182–211, Oxford Univ. Press. 62

Peichl, L. & Wassle, H. (1979). Size, scatter and coverage of ganglion cell receptive field centres in the cat retina. *J Physiol*, **291**, 117–141. 83

Peters, A. & Yilmaz, E. (1993). Neuronal organization in area 17 of cat visual cortex. *Cereb Cortex*, **3**, 49–68. 87

Raizada, R.D. & Grossberg, S. (2003). Towards a theory of the laminar architecture of cerebral cortex: computational clues from the visual system. *Cerebral Cortex*, **13**, 100–113. 4, 30

Robson, J.A. (1993). Qualitative and quantitative analyses of the patterns of retinal input to neurons in the dorsal lateral geniculate nucleus of the cat. *J Comp Neurol*, **334**, 324–336. 83

Schneider, G. & Nikolić, D. (2006). Detection and assessment of near-zero delays in neuronal spiking activity. *J. Neurosci Methods*, **152**, 97–106. 65

Schneider, G., Havenith, M.N. & Nikolić, D. (2006). Spatio-temporal structure in large neuronal networks detected from cross correlation. *Neural Computation*, ***in press***. 65

Schölkopf, B. & Smola, A.J. (2002). *Learning with Kernels*. MIT Press, Cambridge, MA. 63, 67

Schwabe, L. & Obermayer, K. (2005). Adaptivity of tuning functions in a generic recurrent network model of a cortical hypercolumn. *J Neurosci*, **25**, 3323–32. 82

Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G. & Poggio, T. (2005). A theory of object recognition: computations and circuits

in the feedforward path of the ventral stream in primate visual cortex. *MIT, Cambridge, MA*, **AI Memo 2005-036**. 62

Shoham, D., Hubener, M., Schulze, S., Grinvald, A. & Bonhoeffer, T. (1997). Spatio-temporal frequency domains and their relation to cytochrome oxidase staining in cat visual cortex. *Nature*, **385**, 529–533. 90

Silberberg, G., Gupta, A. & Markram, H. (2002). Stereotypy in neocortical microcircuits. *Trends Neurosci.*, **25**, 227–230. 30

Sirovich, L. & Uglesich, R. (2004). The organization of orientation and spatial frequency in primary visual cortex. *Proc Natl Acad Sci U S A*, **101**, 16941–16946. 90

Somers, D.C., Nelson, S.B. & Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *J Neurosci*, **15**, 5448–5465. 90

Stuart, G.J. & Sakmann, B. (1994). Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. *Nature*, **367**, 69–72. 36

Tanaka, K. (1983). Cross-correlation analysis of geniculostriate neuronal relationships in cats. *J. Neurophysiol.*, **49**, 1303–1318. 38

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. USA*, **101**, 366–371. 82, 90

Teich, A.F. & Qian, N. (2006). Comparison among some models of orientation selectivity. *J Neurophysiol*, **96**, 404–419. 83, 89, 90

Thomson, A.M. (2003). Presynaptic frequency- and pattern-dependent filtering. *J. of Comp. Neurosci.*, **15**, 159–202. 33

Thomson, A.M. (2005). Personal communication. 33

Thomson, A.M., West, D.C., Wang, Y. & Bannister, A.P. (2002). Synaptic connections and small circuits involving excitatory and inhibitory neurons in layers 2 - 5 of adult rat and cat neocortex: triple intracellular recordings and biocytin labelling in vitro. *Cerebral Cortex*, **12**, 936–953. 29, 31, 32, 33, 34, 35, 82, 87

Tolhurst, D.J. & Thompson, I.D. (1982). Organization of neurones preferring similar spatial frequencies in cat striate cortex. *Exp Brain Res*, **48**, 217–227. 90

Tootell, R.B., Silverman, M.S. & De Valois, R.L. (1981). Spatial frequency columns in primary visual cortex. *Science*, **214**, 813–815. 90

Traub, R.D. & Miles, R. (1991). *Neuronal Networks of the Hippocampus*. Cambridge Univ. Press, Cambridge, UK. 36

Treves, A. (2003). Computational constraints that may have favoured the lamination of sensory cortex. *J. Computational Neuroscience*, **14**, 271–282. 4, 30, 31

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**, 5908–5927. 83, 85, 89

Tsodyks, M., Uziel, A. & Markram, H. (2000). Synchrony generation in recurrent networks with frequency-dependent synapses. *J. Neuroscience*, **20**. 10

v. Vreeswijk, C. & Sompolinsky, H. (1998). Chaotic balanced state in a model of cortical circuits. *Neural Comput.*, **10**, 1321–1371. 42

v. Vreeswijk, C.A. & Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, **274**, 1724–1726. 42

Vapnik, V.N. (1998). *Statistical Learning Theory*. John Wiley (New York). 13, 15, 53, 63, 67

Warmuth, M.K., Rätsch, G., Mathieson, M., Liao, J. & Lemmen, C. (2002). Active learning in the drug discovery process. In T.G. Dietterich, S. Becker & Z. Ghahramani, eds., *Advances in Neural Information Processing Systems (NIPS)*, vol. 14, in press. 10, 15, 17

Watts, D.J. & Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature*, **393**, 440 – 442. 50

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. 82

White, E.L. (1989). *Cortical Circuits*. Birkhaeuser, Boston. 30, 37

White, O.L., Lee, D.D. & Sompolinsky, H. (2004). Short-term memory in orthogonal neural networks. *Phys. Rev. Letters*, **92**, 148102. 62

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. 82

Worgotter, F. & Koch, C. (1991). A detailed model of the primary visual pathway in the cat: comparison of afferent excitatory and intracortical inhibitory connection schemes for orientation selectivity. *J Neurosci*, **11**, 1959–1979. 83