

# Physics of Neural Networks

E. Domany J.L. van Hemmen  
K. Schulten (Eds.)

## Models of Neural Networks II

Temporal Aspects of Coding and Information Processing in Biological Systems



Springer-Verlag

# *Physics of Neural Networks*

---

## **Series Editors:**

E. Domany J. L. van Hemmen K. Schulten

## **Advisory Board:**

H. Axelrad  
R. Eckmiller  
J. A. Hertz  
J. J. Hopfield  
P. I. M. Johannesma  
D. Sherrington  
M. A. Virasoro

# *Physics of Neural Networks*

---

*Models of Neural Networks*

E. Domany, J.L. van Hemmen, K. Schulten (Eds.)

*Neural Networks: An Introduction*

B. Müller, J. Reinhardt

*Models of Neural Networks II: Temporal Aspects of Coding and Information*

*Processing in Biological Systems*

E. Domany, J.F. van Hemmen, K. Schulten (Eds.)

E. Domany J. L. van Hemmen  
K. Schulten (Eds.)

# Models of Neural Networks II

Temporal Aspects of Coding and Information  
Processing in Biological Systems

With 90 Figures



Springer-Verlag  
New York Berlin Heidelberg  
London Paris Tokyo  
Hong Kong Barcelona  
Budapest

*Series and Volume Editors:*

**Professor Eytan Domany**  
Department of Electronics  
Weizmann Institute of Science  
76100 Rehovot  
Israel

**Professor Dr. J. Leo van Hemmen**  
Institut für Theoretische Physik  
Technische Universität München  
D-85H7 Garching bei München  
Germany

**Professor Klaus Schulten**  
Department of Physics  
and Beckman Institute  
University of Illinois  
Urbana, IL 61801  
USA

Library of Congress Cataloging-in-Publication Data  
Models of neural networks II/[edited by] Eytan Domany,  
J. Leo van Hemmen, Klaus Schulten.

p. cm.

Includes bibliographical references and index.

ISBN-13:978-1-4612-8736-0 e-ISBN-13:978-1-4612-4320-5  
DOI: 10.107/978-1-4612-4320-5

1. Neural networks (Computer science) I. Hemmen, J. L. van (Jan Leonard), 1947-. II. Domany, Eytan, 1947-. III. Schulten, K. (Klaus) IV. Title: Models of neural networks 2. V. Title: Models of neural networks three.

QA76.87.M58 1994

006.3—dc20

94-28645

Printed on acid-free paper.

© 1994 Springer-Verlag New York, Inc.

Reprint of the original edition 1994

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Production managed by Natalie Johnson; manufacturing supervised by Genieve Shaw.  
Photocomposed copy provided from the editors' LaTeX file.

9 8 7 6 5 4 3 2 1

ISBN 0-387-94362-5 Springer-Verlag New York Berlin Heidelberg  
ISBN 3-540-94362-5 Springer-Verlag Berlin Heidelberg New York

# Preface

Since the appearance of Vol. 1 of *Models of Neural Networks* in 1991, the theory of neural nets has focused on two paradigms: information coding through coherent firing of the neurons and functional feedback.

Information coding through coherent neuronal firing exploits time as a cardinal degree of freedom. This capacity of a neural network rests on the fact that the neuronal action potential is a short, say 1 ms, spike, localized in space and time. Spatial as well as temporal correlations of activity may represent different states of a network. In particular, temporal correlations of activity may express that neurons process the same “object” of, for example, a visual scene by spiking at the very same time. The traditional description of a neural network through a firing rate, the famous S-shaped curve, presupposes a wide time window of, say, at least 100 ms. It thus fails to exploit the capacity to “bind” sets of coherently firing neurons for the purpose of both scene segmentation and figure-ground segregation.

Feedback is a dominant feature of the structural organization of the brain. Recurrent neural networks have been studied extensively in the physical literature, starting with the ground breaking work of John Hopfield (1982). Functional feedback arises between the specialized areas of the brain involved, for instance, in vision when a visual scene generates a picture on the retina, which is transmitted to the lateral geniculate body (LGN), the primary visual cortex, and then to areas with “higher” functions. This sequence looks like a feed-forward structure, but appearances are deceiving, for there are equally strong recurrent signals. One wonders what they are good for and how they influence or regulate coherent spiking. Their role is explained in various contributions to this volume, which provides an in-depth analysis of the two paradigms. The reader can enjoy a detailed discussion of salient features such as coherent oscillations and their detection, associative binding and segregation, Hebbian learning, and sensory computations in the visual and olfactory cortex.

Each volume of *Models of Neural Networks* begins with a longer paper that puts together some theoretical foundations. Here the introductory chapter, authored by Gerstner and van Hemmen, is devoted to coding and information processing in neural networks and concentrates on the fundamental notions that will be used, or treated, in the papers to follow.

More than 10 years ago Christoph von der Malsburg wrote the meanwhile classical paper “The correlation theory of brain function.” For a long time this paper was available only as an internal report of the Max-Planck Institute for Biophysical Chemistry in Göttingen, Germany, and is here

made available to a wide audience. The reader may verify that notions which seemed novel 10 years ago still are equally novel at present.

The paper “Firing rates and well-timed events in the cerebral cortex” by Moshe Abeles does exactly what its title announces. In particular, Abeles puts forward cogent arguments that the firing rate by itself does not suffice to describe neuronal firing. Wolf Singer presents a careful analysis of “The role of synchrony in neocortical processing and synaptic plasticity” and in so doing explains what coherent firing is good for. This essay is the more interesting since he focuses on the relation between coherence — or synchrony — and oscillatory behavior of spiking on a global, extensive scale.

This connection is taken up by Ritz et al. in their paper “Associative binding and segregation in a network of spiking neurons.” Here one finds a synthesis of scene segmentation and binding in the associative sense of pattern completion in a network where neural coding is by spikes only. Moreover, a novel argument is presented to show that a hierarchical structure with feed-forward and feedback connections may play a dominant role in context sensitive binding. We consider this an explicit example of functional feedback as a “higher” area provides the context to data presented to several “lower” areas.

Coherent oscillations were known in the olfactory system long before they were discovered in the visual cortex. Zhaoping Li describes her work with John Hopfield in the paper “Modeling the sensory computations of the olfactory bulb.” She shows that here too it is possible to describe both odor recognition and segmentation by the very same model.

Until now we have used the notions “coherence” and “oscillation” in a loose sense. One may ask: How can one attain the goal of “Detecting coherence in neuronal data?” Precisely this is explained by Klaus Pawelzik in his paper with the above title. He presents a powerful information-theoretic algorithm in detail and illustrates his arguments by analyzing real data. This is important not only for the experimentalist but also for the theoretician who wants to verify whether his model exhibits some kind of coherence and, if so, what kind of agreement with experiment is to be expected.

As is suggested by several papers in this volume, there seems to be a close connection between coherence and synaptic plasticity; see, for example, the essay by Singer (Secs. 13 and 14) and Chap. 1 by Gerstner and van Hemmen. Synaptic plasticity itself, a fascinating subject, is expounded by Brown and Chattarji in their paper “Hebbian synaptic plasticity.” By now long-term depression is appreciated as an essential element of the learning process or, as Willshaw aptly phrased it, “What goes up must come down.” On the other hand, Hebb’s main idea, correlated activity of the pre- and postsynaptic neuron, has been shown to be a necessary condition for the induction of long-term potentiation but the appropriate time window of synchrony has not been determined unambiguously yet. A small time window in the millisecond range would allow to learn, store, and retrieve

spatio-temporal spike patterns, as has been pointed out by Singer and implemented by the Hebbian algorithm of Gerstner et al. Whether or not such a small time window may exist is still to be shown experimentally.

A case study of functional feedback or, as they call it, reentry is provided by Sporns, Tononi, and Edelman in the essay “Reentry and dynamical interactions of cortical networks.” Through a detailed numerical simulation these authors analyze the problem of how neural activity in the visual cortex is integrated given its functional organization in the different areas. In a sense, in this chapter the various parts of a large puzzle are put together and integrated so as to give a functional architecture. This integration, then, is sure to be the subject of a future volume of *Models of Neural Networks*.

*The Editors*

# Contents

Preface	v
Contributors	xv
<b>1. Coding and Information Processing in Neural Networks</b>	<b>1</b>
<i>Wulfram Gerstner and J. Leo van Hemmen</i>	
1.1 Description of Neural Activity . . . . .	1
1.1.1 Spikes, Rates, and Neural Assemblies . . . . .	2
1.2 Oscillator Models . . . . .	4
1.2.1 The Kuramoto Model . . . . .	6
1.2.2 Oscillator Models in Action . . . . .	23
1.3 Spiking Neurons . . . . .	35
1.3.1 Hodgkin–Huxley Model . . . . .	35
1.3.2 Integrate-and-Fire Model . . . . .	38
1.3.3 Spike Response Model . . . . .	39
1.4 A Network of Spiking Neurons . . . . .	47
1.4.1 Stationary Solutions — Incoherent Firing . . . . .	55
1.4.2 Coherent Firing — Oscillatory Solutions . . . . .	68
1.5 Hebbian Learning of Spatio-Temporal Spike Patterns . . . . .	72
1.5.1 Synaptic Organization of a Model Network . . . . .	73
1.5.2 Hebbian Learning . . . . .	74
1.5.3 Low-Activity Random Patterns . . . . .	78
1.5.4 Discussion . . . . .	81
1.6 Summary and Conclusions . . . . .	83
References . . . . .	85
<b>2. The Correlation Theory of Brain Function</b>	<b>95</b>
<i>Christoph von der Malsburg</i>	
Foreword . . . . .	95
2.1 Introduction . . . . .	96
2.2 Conventional Brain Theory . . . . .	96
2.2.1 Localization Theory . . . . .	96
2.2.2 The Problem of Nervous Integration . . . . .	98
2.2.3 Proposed Solutions . . . . .	100
2.3 The Correlation Theory of Brain Function . . . . .	104
2.3.1 Modifications of Conventional Theory . . . . .	104
2.3.2 Elementary Discussion . . . . .	106

2.3.3	Network Structures . . . . .	107
2.3.4	Applications of Correlation Theory . . . . .	112
2.4	Discussion . . . . .	115
2.4.1	The Text Analogy . . . . .	115
2.4.2	The Bandwidth Problem . . . . .	116
2.4.3	Facing Experiment . . . . .	117
2.4.4	Conclusions . . . . .	118
	References . . . . .	118
<b>3.</b>	<b>Firing Rates and Well-Timed Events in the Cerebral Cortex</b>	<b>121</b>
	<i>M. Abeles</i>	
3.1	Measuring the Activity of Nerve Cells . . . . .	121
3.2	Rate Functions and Stationary Point Processes . . . . .	124
3.3	Rate Functions for Nonstationary Point Processes . . . . .	129
3.4	Rate Functions and Singular Events . . . . .	132
	References . . . . .	138
<b>4.</b>	<b>The Role of Synchrony in Neocortical Processing and Synaptic Plasticity</b>	<b>141</b>
	<i>Wolf Singer</i>	
4.1	Introduction . . . . .	141
4.2	Pattern Processing and the Binding Problem . . . . .	143
4.3	Evidence for Dynamic Interactions Between Spatially Distributed Neurons . . . . .	147
4.4	Stimulus-Dependent Changes of Synchronization Probability . . . . .	149
4.5	Synchronization Between Areas . . . . .	151
4.6	The Synchronizing Connections . . . . .	151
4.7	Experience-Dependent Modifications of Synchronizing Connections and Synchronization Probabilities . . . . .	155
4.8	Correlation Between Perceptual Deficits and Response Synchronization in Strabismic Amblyopia . . . . .	156
4.9	The Relation Between Synchrony and Oscillations . . . . .	157
4.10	Rhythm Generating Mechanisms . . . . .	159
4.11	The Duration of Coherent States . . . . .	160
4.12	Synchronization and Attention . . . . .	162
4.13	The Role of Synchrony in Synaptic Plasticity . . . . .	163
4.14	The Role of Oscillations in Synaptic Plasticity . . . . .	164
4.15	Outlook . . . . .	166
4.16	Concluding Remarks . . . . .	167
	References . . . . .	167

<b>5. Associative Binding and Segregation in a Network of Spiking Neurons</b>	<b>175</b>
<i>Raphael Ritz, Wulfram Gerstner, and J. Leo van Hemmen</i>	
5.1 Introduction . . . . .	175
5.2 Spike Response Model . . . . .	177
5.2.1 Starting Point . . . . .	177
5.2.2 Neurons and Synapses . . . . .	178
5.3 Theory of Locking . . . . .	183
5.3.1 Equation of Motion . . . . .	183
5.3.2 Stationary States . . . . .	184
5.3.3 Oscillatory States . . . . .	187
5.3.4 A Locking Condition . . . . .	189
5.3.5 Weak Locking . . . . .	190
5.4 Simulation Results . . . . .	191
5.4.1 Three Scenarios . . . . .	191
5.4.2 Interpretation . . . . .	194
5.4.3 Spike Raster . . . . .	196
5.4.4 Synchronization Between Two Hemispheres . . . . .	198
5.5 Application to Binding and Segmentation . . . . .	201
5.5.1 Feature Linking . . . . .	201
5.5.2 Pattern Segmentation . . . . .	201
5.5.3 Switching Between Patterns . . . . .	204
5.6 Context Sensitive Binding in a Layered Network with Feedback . . . . .	206
5.7 Discussion . . . . .	209
5.8 Conclusions . . . . .	214
References . . . . .	214
<b>6. Modeling the Sensory Computations of the Olfactory Bulb</b>	<b>221</b>
<i>Zhaoping Li</i>	
6.1 Introduction . . . . .	221
6.2 Anatomical and Physiological Background . . . . .	223
6.3 Modeling the Neural Oscillations in the Olfactory Bulb . . . . .	225
6.3.1 General Model Structure . . . . .	226
6.3.2 The Olfactory Bulb as a Group of Coupled Nonlinear Oscillators . . . . .	227
6.3.3 Explanation of Bulbar Activities . . . . .	230
6.4 A Model of Odor Recognition and Segmentation in the Olfactory Bulb . . . . .	232
6.4.1 Emergence of Oscillations Detects Odors: Patterns of Oscillations Code the Odor Identity and Strength . . . . .	232
6.4.2 Odor Segmentation in the Olfactory Bulb — Olfactory Adaptation . . . . .	235

6.4.3	Olfactory Psychophysics — Cross-Adaptation, Sensitivity Enhancement, and Cross-Enhancement . . . . .	237
6.5	A Model of Odor Segmentation Through Odor Fluctuation Analysis . . . . .	238
6.5.1	A Different Olfactory Environment and a Different Task . . . . .	238
6.5.2	Odor Segmentation in an Adaptive Network . . . . .	239
6.6	Discussion . . . . .	244
	References . . . . .	248
<b>7.</b>	<b>Detecting Coherence in Neuronal Data</b>	<b>253</b>
	<i>Klaus Pawelzik</i>	
7.1	Introduction . . . . .	254
7.1.1	Correlations and Statistical Dependency . . . . .	258
7.2	Time Resolved Detection of Coherence . . . . .	261
7.2.1	Statistical Dependency of Continuous Signals . . . . .	263
7.2.2	Detection of Coherent Patterns of Activities . . . . .	266
7.3	Memory and Switching in Local Field Potentials from Cat Visual Cortex . . . . .	268
7.3.1	Temporal Coherence of Local Field Potentials . . . . .	268
7.3.2	Spatial Coherence of Local Field Potentials . . . . .	271
7.4	A Model-Dependent Approach . . . . .	272
7.4.1	Renewal Dynamics . . . . .	273
7.4.2	Switching in the Assembly Dynamics . . . . .	273
7.4.3	Network States are Hidden . . . . .	274
7.4.4	Model Identification from Multiunit Activities . . . . .	275
7.5	Memory and Switching in Multiunit Activities from Cat Visual Cortex . . . . .	276
7.5.1	Identifying the Local Pool Dynamics . . . . .	276
7.5.2	Testing the Assembly Hypothesis . . . . .	276
7.6	Reconstruction of Synchronous Network States . . . . .	277
7.7	Summary . . . . .	279
	References . . . . .	283
<b>8.</b>	<b>Hebbian Synaptic Plasticity: Evolution of the Contemporary Concept</b>	<b>287</b>
	<i>Thomas H. Brown and Sumantra Chattarji</i>	
8.1	Concept of a Hebbian Synapse . . . . .	287
8.1.1	Contemporary Concept of a Hebbian Synaptic Modification . . . . .	288
8.2	Experimental Evidence for Hebbian Synaptic Mechanisms .	289
8.2.1	Induction of Hippocampal LTP . . . . .	290
8.3	Biophysical Models of LTP Induction . . . . .	290
8.3.1	Second-Generation Spine Model . . . . .	296
8.4	Bidirectional Regulation of Synaptic Strength . . . . .	298

8.4.1	Theoretical Representations of Generalized Hebbian Modifications . . . . .	299
8.4.2	Experimental Evidence for Long-Term Synaptic Depression . . . . .	302
8.5	Interaction Between Dendritic Signaling and Hebbian Learning . . . . .	306
	References . . . . .	307
<b>9.</b>	<b>Reentry and Dynamical Interactions of Cortical Networks</b>	<b>315</b>
	<i>Olaf Sporns, Giulio Tononi, and Gerald M. Edelman</i>	
9.1	Introduction . . . . .	315
9.2	Models of Cortical Integration . . . . .	317
9.2.1	Dynamic Behavior of Single Neuronal Groups . . . . .	317
9.2.2	Coupled Neuronal Groups . . . . .	320
9.2.3	Figure-Ground Segregation . . . . .	325
9.2.4	Cooperative Interactions Among Multiple Cortical Areas . . . . .	331
9.3	Summary and Conclusion . . . . .	335
9.3.1	Cortical Integration at Multiple Levels . . . . .	336
9.3.2	Temporal Dynamics . . . . .	336
9.3.3	Correlations and Behavior . . . . .	336
9.3.4	Functions of Reentry . . . . .	337
	References . . . . .	337
	<b>Index</b>	<b>343</b>

# Contributors

MOSHE ABELES Department of Physiology, Hadassa Medical School, Hebrew University, Jerusalem, Israel

THOMAS H. BROWN Department of Psychology and Cellular and Molecular Physiology and Yale Center for Theoretical and Applied Neuroscience, Yale University, New Haven, CT 06520, USA

SUMANTRA CHATTARJI Department of Psychology and Cellular and Molecular Physiology and Yale Center for Theoretical and Applied Neuroscience, Yale University, New Haven, CT 06520, USA

GERALD M. EDELMAN The Neurosciences Institute, 3377 North Torrey Pines Court, La Jolla, CA 92037, USA

WULFRAM GERSTNER Institut für Theoretische Physik, Technische Universität München, D-85747 Garching bei München, Germany

J. LEO VAN HEMMEN Institut für Theoretische Physik, Technische Universität München, D-85747 Garching bei München, Germany

ZHAOPING LI Rockefeller University, Box 272, 1230 York Avenue, New York, NY 10021, USA

CRISTOPH VON DER MALSBURG Institut für Neuroinformatik, Ruhr-Universität Bochum, Postfach 102185, D-44780 Bochum, Germany

KLAUS PAWELZIK Institut für Theoretische Physik, Universität Frankfurt, Robert-Mayer-Straße 8-10, D-60325 Frankfurt am Main, Germany

RAPHAEL RITZ Institut für Theoretische Physik, Technische Universität München, D-85747 Garching bei München, Germany

WOLF SINGER Max-Planck-Institut für Hirnforschung, Deutscheschordenstraße 46, D-60528 Frankfurt am Main, Germany

OLAF SPORNS The Neurosciences Institute, 3377 North Torrey Pines Court, La Jolla, CA 92037, USA

GIULIO TONONI The Neurosciences Institute, 3377 North Torrey Pines Court, La Jolla, CA 92037, USA

# 1

# Coding and Information Processing in Neural Networks

Wulfram Gerstner and J. Leo van Hemmen<sup>1</sup>

with 29 figures

**Synopsis.** This paper reviews some central notions of the theoretical biophysics of neural networks, viz., information coding through coherent firing of the neurons and spatio-temporal spike patterns. After an introduction to the neural coding problem we first turn to oscillator models and analyze their dynamics in terms of a Lyapunov function. The rest of the paper is devoted to spiking neurons, a pulse code. We review the current neuron models, introduce a new and more flexible one, the spike response model (SRM), and verify that it offers a realistic description of neuronal behavior. The corresponding spike statistics is considered as well. For a network of SRM neurons we present an analytic solution of its dynamics, analyze the possible asymptotic states, and check their stability. Special attention is given to coherent oscillations. Finally we show that Hebbian learning also works for low activity spatio-temporal spike patterns. The models which we study always describe globally connected networks and, thus, have a high degree of feedback. We only touch upon functional feedback, that is, feedback between areas that have different tasks. Information processing in conjunction with functional feedback is treated explicitly in a companion paper [94].

## 1.1 Description of Neural Activity

In the study of neural networks a key question is: How should we describe neural activity? Is a rate code sufficient or do we need a much higher temporal resolution so that single spikes become important and pulse or interval code is more appropriate? This kind of problem is what the introductory section is devoted to.

---

<sup>1</sup>Physik-Department der TU München, D-85747 Garching bei München, Germany.

### 1.1.1 SPIKES, RATES, AND NEURAL ASSEMBLIES

The brain consists of an enormous number of neurons with a high degree of connectivity of approximately 10,000 synapses per neuron. If the activity of a single neuron is recorded by an electrode near the soma or the axon, we find a *spike train* consisting of several short pulses—the so-called action potentials—separated by intervals of variable length. If several neurons are recorded at the same time—as has been done in recent multielectrode recordings [81,106]—we find a complex spatio-temporal spike pattern where different neurons fire at different times. When looking at such a spike pattern we may wonder what it means. Also, as a theoretician, one should ask how such an activity pattern can be described. Both questions are directly related to the problem of *neural code*: How is the information on the environment encoded in a neuronal signal?

To be specific, we want to ask the following questions. Can the activity of a *single* neuron be related uniquely to a specific feature of the stimulus (single “grandmother” cell code) or do we always need to consider *ensembles* of equivalent neurons (ensemble or activity code). If there are ensembles, are they local or rather distributed all over the network? If the ensembles are distributed, how can we find neurons that belong together within the same “assembly”? On the other hand, if single neurons are important, we must know which quantity of the neuronal signal contains the relevant information. Is it enough to consider the time-averaged firing rate (*rate code*) as defined by the mean number of action potentials in a given time interval, say 100 or 500 ms? Or should we consider single spikes and the intervals between them (*pulse* or *interval code*)? These questions address open problems in the theory of neural networks and, depending on the experimental system under consideration, allow different answers.

In studies on sensory neurons it has been shown repeatedly that the *mean firing rate* is directly related to the stimulus intensity. Similarly, in motor control the muscle response is related to the firing rate of the neuronal signal. These results which date back to Adrian [10] have led to the idea that the (time-averaged) mean firing rate contains all relevant information, independent of the location of the neuron and the task to be done. This seems to be an implicit assumption in most experiments up to now—despite the fact that a *pulse* or *interval code* would in principle allow to convey much more information [4,27,93,121].

It is only recently that the time structure of neuronal signals has been considered in more detail. Optican and Richmond [101] could show that a time resolution of 20 – 30 ms is necessary for a correct interpretation of neuronal signals in the primate inferior cortex. Dinse et al. [26] investigated the temporal structure of receptive fields in primate visual cortex and found changes on a time scale of 10 – 30 ms. These results are an indication that the mean firing rate which presupposes an averaging interval of at least 100 ms is not sufficient to describe neuronal activity.

Based on a careful evaluation of experimental data on the visual system of the fly, de Ruyter van Steveninck and Bialek could show that *single spikes of a single neuron* contain important information on the stimulus [111]. Moreover, it is possible to reconstruct the stimulus, if the complete spike train is known [16], whereas this is obviously impossible if only the mean firing rate is given. A similar approach has been taken by Eckhorn et al. [29], who showed that single spikes of neurons in the primate visual cortex allow one to guess the spatial and temporal properties of a visual stimulus within the last 100 ms before the spike. In auditory cortex it is known that interaural time differences in the ms range are used by some animals for source localization [78]. These results can be seen as a direct proof that single spikes or, equivalently, interspike intervals can be meaningful in some experimental situations.

Nevertheless, most models of a neural network are restricted to a firing rate description. A single neuron is usually reduced to a simple nonlinear unit where the input (stimulus or driving current) is connected to the output (mean firing rate) via a sigmoidal gain function [70]; for a review, see [61] and [63]. Thus the time structure of neuronal signals is completely disregarded.

A different class of network models deals with ensemble averaged activities only [132]. This allows one to describe variable activity at a good time resolution, but the spatial resolution is limited due to ensemble averaging. Such an approach has become quite popular to describe the phenomena of collective oscillations in the cortex [77,114,116,130]. It is based on the idea that neurons with similar properties are situated next to each other in columns across the cortex [72]. If the structure and connectivity within one column is neglected, each column defines an ensemble (or “pool”) of equivalent neurons. Several ensembles can be connected so as to form a network of units which are described by a single variable, the ensemble activity. Note that such an approach presupposes that only ensemble averaged quantities are relevant, an assumption which may, or may not, be satisfied.

Since the question of neural code is still open we should make no a priori assumption in this respect when designing a model network. For this reason we prefer to start with a network of *single spiking neurons*. This is the approach which is taken in this chapter. It will turn out that there are several distinct types of collective state in such a model network. First there are stationary states of constant activity where all active neurons fire *incoherently*. These states which can be completely characterized by mean firing rates correspond to the macroscopic states studied in traditional models of neural networks. Second, there are *coherent* oscillatory states where active neurons tend to fire in synchrony. This is a new class of states beyond the traditional approach. In order to understand the principles of synchronization leading to coherent firing we study an abstract model of “locking” phenomena in Sec. 1.2 of this chapter. Finally, there are complex spatio-

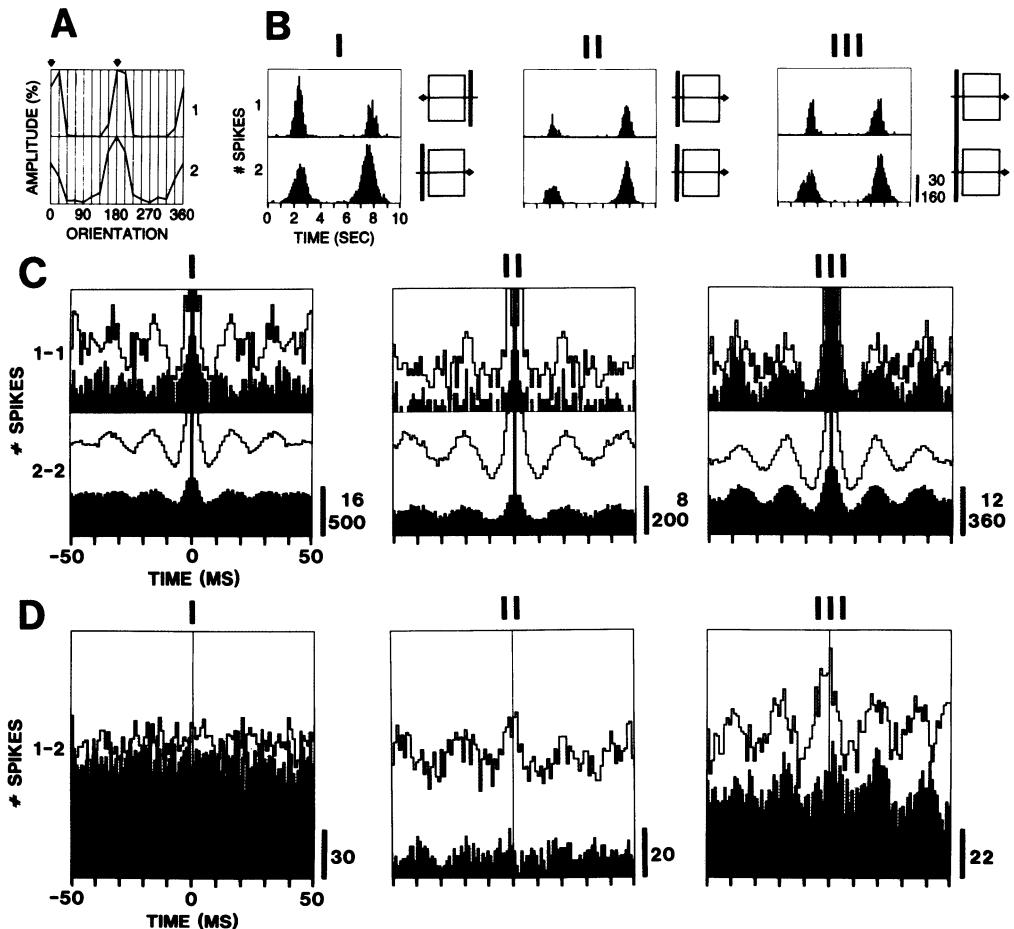
temporal patterns. In this case the full spatial and temporal information on firing times of different neurons is relevant. How spatio-temporal patterns can be learned and retrieved is discussed in Sec. 1.5 of this chapter.

Let us now turn to the problem of collective oscillations. This is an area which has attracted a lot of attention recently after oscillatory behavior has been found in the visual cortex of cat and monkey [8,28,51,52,118]. For a long time, oscillations have also been known to exist in the olfactory system [25,41]. We will start with a simple toy model of oscillatory processes and proceed with more realistic models of these and other phenomena afterwards.

## 1.2 Oscillator Models

Spikes are localized in space and time. As compared to rate coding we have an extra degree of freedom: time. A simple example of a temporal structure is the situation where some neurons in a network fire synchronously. In this case, those neurons which spike *coherently* can, for instance, signal that together they characterize a single object. If coherent spiking is cyclically repeated, the multiunit activity shows an oscillatory structure and those neurons which participate in the collective oscillation exhibit a classical *phase locking*. A typical experiment showing this coherent behavior is the following [8,28,51,52,118]. A long light bar (Fig. 1.1(b), III) moves across two receptive fields,  $\Delta_1$  and  $\Delta_2$ , which are spatially separated from each other. The response of the system is a coherent oscillation, that is, a synchronous spiking that repeats itself more or less periodically and also leads to an oscillatory structure in the *cross-correlation* between  $\Delta_1$  and  $\Delta_2$ . If the light bar is cut into two parts with a hole in between (Fig 1.1(b), II), then the cross-correlation is diminished. If the two light bars move in opposite directions (Fig. 1.1(b), I), then there is no correlation at all, in agreement with the intuitive interpretation that now there are two distinct objects; cf. Fig. 1.1(d). The autocorrelograms always exhibit a pronounced oscillation indicating that we may, and do, have a coherent oscillation in each of the regions separately.

In the following sections we will explore the theoretical context which is needed to describe this new degree of freedom. Here we concentrate on a so-called oscillator description of a neuron. It is simple, shows the gist of some of the locking phenomena seen in experiment but the price we pay is that it is not very realistic. In short, it has didactic qualities which are worthwhile to be spelt out. We first turn to the Kuramoto model proper (Sec. 1.2.1) that underlies most of the early theoretical work on phase locking in neural networks. Then we analyze the performance of various oscillator models in Sec. 1.2.2.



**Fig. 1.1. Experimental evidence for the idea that long-range coherence reflects global stimulus properties.** In the present case, the coherence shows up as an oscillatory correlation. (a) Orientation tuning curves of neuronal responses recorded from two electrodes {1,2} in the primary visual cortex of the cat. The electrodes are separated by 7 mm and show a preference for vertical light bars ( $0^\circ$  and  $180^\circ$  at both recording sites). A tuning curve tells us how well a neuron responds to a stimulus of a specified direction. (b) Poststimulus time histograms (PSTH) of the responses recorded at each site for each of three different but typical stimulus conditions: (I) two light bars moving in opposite directions, (II) two *disjoint* light bars moving in the same direction, and (III) one connected light bar moving across both receptive fields. A schematic diagram of the receptive fields locations and the stimulus configurations used is displayed to the right of each PSTH. (c) Auto-correlograms and (d) cross-correlograms computed for the neuronal responses at the sites 1 and 2 for each of the three stimulus conditions (I – III); except for I, the second direction of stimulus movement is shown with unfilled bars. The numbers on the vertical calibration correspond to the number of coincident spikes. The cross-correlation is nonexistent in I, weak in II, and strong in III. In all three cases, the *auto*-correlograms exhibit a pronounced oscillation. Taken from [117].

### 1.2.1 THE KURAMOTO MODEL

Van der Pol's mathematical description of frequency locking [103] has initiated a lasting theoretical interest in all kinds of locking phenomena that occur in nonlinear oscillators. Originally, a single or only a few oscillators were involved; cf. in particular van der Pol's seminal work on an LC circuit with triode and forcing term and its follow-up. The interest in the *collective* behavior of a large assembly of nonlinear oscillators is more recent. It has been stimulated greatly by a model proposed by Kuramoto [85–88], who assumed  $N$  oscillators coupled “all-to-all” and described by a phase  $\phi_i$ ,  $1 \leq i \leq N$ , with

$$\dot{\phi}_i = \omega_i - \frac{K}{N} \sum_{j=1}^N \sin(\phi_i - \phi_j) . \quad (1.1)$$

Here  $\dot{\phi} = d\phi/dt$ ,  $K \geq 0$ , and the frequencies  $\omega_i$  are independent, identically distributed random variables. The underlying probability measure on the reals is denoted by  $\mu$ . In contrast to an extensive part of the existing literature but in agreement with practical requirements, its support is supposed to be contained in a *bounded* interval. Hence no  $\omega$  is ever found outside this interval.

In the above framework, a neuron is reduced to a single phase. What's nice is that the oscillators are allowed to be different in that each one has its own eigenfrequency  $\omega_i$ ; just take  $K = 0$  in Eq. (1.1) to see what “eigenfrequency” means. For the moment we only note that determining this eigenfrequency may be a problem. Spikes have gone completely. We may, however, *define* a spike by, say,  $\phi_i$  being in a small interval around zero. We will meet other interpretations [38, 115, 119, 120] as we proceed. The connectivity in the cortex is high, of the order  $10^4$ . Thus the all-to-all coupling makes sense. Cortical areas also contain quite a few neurons so it is sensible that the model concentrates on the collective behavior of the network. Mathematically, this means that we are interested in the limit  $N \rightarrow \infty$ .

Two things have to be constantly borne in mind, though. First, a neuron's activity is an all-or-none process whereas the oscillators à la Eq. (1.1) are just the opposite. Furthermore, they feel each other all the time whereas real neurons notice each other *only if and when they spike*. So, in contrast to the oscillator phase description, both the input and the output of a biological neuron are all-or-none phenomena. In spite of that, the simplicity of the model is seductive and it allows a remarkably profound understanding of locking phenomena.

The Kuramoto model has been studied extensively<sup>2</sup> during recent years.

---

<sup>2</sup>A nice, annotated, bibliography of the literature up to 1986 has been provided by Kopell [79].

We refer in particular to the beautiful work of Ermentrout and Kopell [33, 35–38], and Strogatz, Mirollo, and Matthews [95–97, 123–125]. There appears to exist a critical  $K_c$  such that for  $K > K_c$  the system is in a *phase-locked* state characterized by  $\dot{\phi}_i = \dot{\phi}$  for all  $i = 1, \dots, N$ , whereas no such state exists for  $K < K_c$ . Instead, one then encounters a partially coherent state or, as we will see below when studying discrete frequency distributions, a state which is not coherent at all.

Though the model (1.1) looks quite simple, appearances are deceiving. Here we prove phase locking by exhibiting a Lyapunov function  $\mathcal{H}$ . The phase-locked state itself is known since long [85–87] but the present stability proof is new [62]. We show that a phase-locked state is a minimum of  $\mathcal{H}$  and thus (quite) unique and (asymptotically) stable. Furthermore, we are able to offer a physical interpretation of this state and its stability.

In the following subsections we introduce our Lyapunov function  $\mathcal{H}$ , determine the phase-locked state as its minimum, and study the associated fixed-point equations. The Lyapunov function is not specific to the Kuramoto model, which is of mean-field or infinite-range type. It holds for *any* model with symmetric interactions of finite instead of infinite range. For the Kuramoto model, the Lyapunov function  $\mathcal{H}$  represents the energy of an *XY* ferromagnet of strength  $K$  in a random field induced by the random distribution of the frequencies  $\omega_i$ . As such, the system is *frustrated* in that the *XY* ferromagnet likes to have all spins parallel, that is, all  $\phi_i$  equal and thus perfectly phase locked, whereas the random field  $(\omega_i - \langle \omega \rangle)$  tries to break the ferromagnetic order. As  $K$  decreases, the model exhibits a phase transition at  $K_c$ : For  $K > K_c$  the ferromagnet wins and the system is totally phase locked whereas for  $K < K_c$  the random field takes over. We estimate both  $K_c$  and the range of the order parameter  $r$  that describes the macroscopic extent of the phase locking. We study several examples, face the question what happens, if we are given a *discrete* frequency distribution and some *but not all* of the oscillators can lock, and finally discuss the salient differences between the present, more general type of model and the “generic” one with absolutely continuous, symmetric distributions such as the Gaussian and the Lorentzian [35, 37, 85–88, 123–125]. It will turn out that the latter type of model does not behave in a truly generic way since the *partial* locking [87, 124], which shows up for  $K$  in an open interval just below  $K_c$ , is absent in models with a discrete distribution. In other words, different universality classes exist.

### Lyapunov function

The dynamics (1) allows an interesting *sum rule*. We add the  $\dot{\phi}_i$ , divide by  $N$ , and find

$$\frac{d}{dt} \left( N^{-1} \sum_{i=1}^N \dot{\phi}_i \right) = \left( N^{-1} \sum_{i=1}^N \omega_i \right) - \frac{K}{N^2} \sum_{i,j} \sin(\phi_i - \phi_j) . \quad (1.2)$$

Since the sine is an odd function the second sum on the right vanishes (interchange  $i$  and  $j$ ) and we are left with

$$\frac{d}{dt} \left( N^{-1} \sum_{i=1}^N \phi_i \right) = N^{-1} \sum_{i=1}^N \omega_i \equiv \langle \omega \rangle . \quad (1.3)$$

As  $N \rightarrow \infty$  the quantity  $\langle \omega \rangle$  is a nonrandom number and equals the mean  $\int d\mu(\omega) \omega$  by the strong law of large numbers [89]. If, then, we have phase locking defined by  $\dot{\phi}_i = \dot{\phi}$ ,  $1 \leq i \leq N$ , we are bound to find  $\dot{\phi} = \langle \omega \rangle$ .

We now introduce new variables  $\varphi_i$  defined by

$$\varphi_i = \phi_i - \Omega t , \quad (1.4)$$

where  $\Omega$  is at our disposal. In terms of the  $\varphi_i$  the equations of motion (1.1) reappear in the form

$$\dot{\varphi}_i = (\omega_i - \Omega) - \frac{K}{N} \sum_j \sin(\varphi_i - \varphi_j) . \quad (1.5)$$

Suppose for a moment that there was no randomness so that  $\omega_i = \omega$  for  $1 \leq i \leq N$ . If we choose  $\Omega = \omega$ , we then arrive at a simple gradient dynamics [86],

$$\dot{\varphi}_i = - \frac{K}{N} \sum_j \sin(\varphi_i - \varphi_j) = - \frac{\partial \mathcal{H}}{\partial \varphi_i} \Rightarrow \dot{\varphi} = - \nabla \mathcal{H} \quad (1.6)$$

with

$$\mathcal{H} = - \frac{K}{2N} \sum_{i,j} \cos(\varphi_i - \varphi_j) = - \frac{K}{2N} \sum_{i,j} \mathbf{S}_i \cdot \mathbf{S}_j . \quad (1.7)$$

$\mathcal{H}$  is the Hamiltonian of an  $XY$  ferromagnet. The spins  $\mathbf{S}_i$  are unit vectors and the dynamics (1.6) is a gradient dynamics with  $\mathcal{H}$  as a Lyapunov function, viz.,

$$\dot{\mathcal{H}} = \sum_i \frac{\partial \mathcal{H}}{\partial \varphi_i} \dot{\varphi}_i = - \sum_i \left( \frac{\partial \mathcal{H}}{\partial \varphi_i} \right)^2 = - \|\nabla \mathcal{H}\|^2 \leq 0 . \quad (1.8)$$

The inequality in Eq. (1.8) is strict, unless we reach a minimum of  $\mathcal{H}$ , where  $\nabla \mathcal{H} = 0$ . Since  $\mathbf{S}_i \cdot \mathbf{S}_j \leq 1$ , a minimum of Eq. (1.7) is reached as soon as  $\mathbf{S}_i \cdot \mathbf{S}_j = 1$  for all  $i$  and  $j$ , that is, when all spins are parallel. Therefore, asymptotically  $\varphi_i(t) \rightarrow \varphi_\infty$  for all  $i$  and we obtain a perfect phase locking. In terms of the original variables we have  $\phi_i(t) = \omega t + \varphi_\infty$ ,  $1 \leq i \leq N$ .

Is the minimum for  $\mathcal{H}$  unique? No, not quite. Due to Eq. (1.7) we can write

$$\mathcal{H} = - \frac{1}{2} KN \left( N^{-1} \sum_{i=1}^N \mathbf{S}_i \right)^2 , \quad (1.9)$$

which is evidently invariant under a *uniform* rotation of all the  $S_i$  or, in terms of the  $\varphi_i$ 's, under the transformation  $\varphi_i \rightarrow \varphi_i + \alpha$ ,  $1 \leq i \leq N$ . In terms of the stability matrix, all its eigenvalues are strictly negative, except for one, which vanishes. In this way the rotational invariance of Eq. (1.9) is taken care of. A minimum is stable and, orthogonally to this direction, asymptotically stable [67]. We now turn to the case of a *nondegenerate* distribution of the  $\omega_i$ .

Also for Eq. (1.5) a Lyapunov function exists. We can, and will, define the angles  $\varphi_i \bmod 2\pi$ . Equation (1.5) tells us quite explicitly that there is no harm in doing so. Then

$$\mathcal{H} = -\frac{K}{2N} \sum_{i,j} \cos(\varphi_i - \varphi_j) - \sum_i (\omega_i - \Omega) \varphi_i \quad (1.10)$$

induces a gradient dynamics for Eq. (1.5). Hence  $\mathcal{H}$  is a Lyapunov function and the dynamics (1.5) converges to a minimum of  $\mathcal{H}$ , if it exists. One might object that restricting  $\varphi \bmod 2\pi$  to  $[-\pi, \pi]$  is an artefact. If we start far away from a minimum, some of the  $\varphi_i$  may hit the border and jump from  $-\pi$  to  $\pi$  or conversely. That does change the second term on the right in Eq. (1.10). No jumping occurs, however, if a minimum of  $\mathcal{H}$  can be localized *in the interior* of  $[-\pi, \pi]^N$ . In a suitable neighborhood, the system then converges to the minimum and we even have asymptotic stability. It is also plain that the idea which has led us to Eq. (1.10) is equally valid, if the mean-field interaction  $K/2N$  is replaced by a finite-range interaction  $J_{ij}$ . Since the modifications of the arguments below are straightforward they will not be spelled out.

An extremum of  $\mathcal{H}$  is characterized by  $\nabla \mathcal{H} = 0$ , that is, by the fixed-point equation

$$0 = (\omega_i - \Omega) - \frac{K}{N} \sum_j \sin(\varphi_i - \varphi_j) \quad (1.11)$$

for  $1 \leq i \leq N$ . Summing over  $i$  we obtain

$$\Omega = N^{-1} \sum_{i=1}^N \omega_i = \langle \omega \rangle \quad . \quad (1.12)$$

This determines  $\Omega$  and is consistent with the observation following the sum rule (1.3). For a finite-range interaction [100,112,113], exactly the same argument holds, including the sum rule, if the  $J_{ij}$  are symmetric, that is,  $J_{ij} = J_{ji}$ . We now continue with the Kuramoto model.

Let us denote the difference between  $\omega_i$  and  $\Omega = \langle \omega \rangle$  by  $\Delta(\omega) = \omega_i - \langle \omega \rangle$ . To solve Eq. (1.11), viz.,

$$\Delta(\omega_i) = \frac{K}{N} \sum_j \sin(\varphi_i - \varphi_j) \quad , \quad (1.13)$$

we introduce an order parameter  $r$  and an associated variable  $\psi$  through [85,87,88]

$$r e^{i\psi} = N^{-1} \sum_{j=1}^N e^{i\varphi_j} . \quad (1.14)$$

The right-hand side being a convex combination of complex numbers in the convex unit disk  $r \exp(i\psi)$  is in the unit disk itself and  $0 \leq r \leq 1$ . Using Eq. (1.14) and  $\sin(x) = [\exp(ix) - \exp(-ix)]/2i$  we rewrite Eq. (1.6):

$$\begin{aligned} \dot{\varphi}_i &= \Delta(\omega_i) - \frac{K}{2Ni} \sum_{j=1}^N \left[ e^{i(\varphi_i - \varphi_j)} - e^{-i(\varphi_i - \varphi_j)} \right] \\ &= \Delta(\omega_i) - \frac{K}{2i} \left[ e^{i(\varphi_i - \psi)} - e^{-i(\varphi_i - \psi)} \right] \end{aligned}$$

so that

$$\dot{\varphi}_i = \Delta(\omega_i) - Kr \sin(\varphi_i - \psi) . \quad (1.15)$$

This equation explicitly tells us that  $\dot{\varphi}_i$  is governed by both  $\Delta(\omega_i)$  and the collective variables  $r$  and  $\psi$ . The fixed-point equation (1.11) now assumes the simple form

$$\Delta(\omega_i) = Kr \sin(\varphi_i - \psi) . \quad (1.16)$$

It is basic to all that follows.

For the sake of simplicity we suppose that the  $\omega_i$  assume only finitely many values  $\{\omega\}$  with probabilities  $\{p(\omega)\}$ . We then can introduce [56–58] sublattices  $I(\omega) = \{i; \omega_i = \omega\}$  consisting of all  $i$  with  $\omega_i = \omega$ . By the strong law of large numbers [89] the size  $|I(\omega)|$  of the sublattice  $I(\omega)$  is given by  $|I(\omega)| \sim p(\omega)N$  as  $N \rightarrow \infty$ .

On a particular sublattice  $I(\omega)$  all  $\Delta(\omega_i)$  have the same value  $\Delta(\omega)$ . According to Eq. (1.16) all  $\varphi_i$  also assume the very same value  $\varphi(\omega)$  given by

$$\varphi(\omega) - \psi = \arcsin[\Delta(\omega)/Kr] . \quad (1.17)$$

We will verify shortly that the order parameter  $r$  is such that  $|\Delta(\omega)/Kr| \leq 1$ , and  $\arcsin(0)$  should be 0 and not  $\pi$ . By Eqs. (1.14) and (1.17) we have

$$r e^{i\psi} = \sum_{\{\omega\}} p(\omega) \exp \left\{ i \left[ \psi + \arcsin \left( \frac{\Delta(\omega)}{Kr} \right) \right] \right\} \quad (1.18)$$

and since  $r \geq 0$ :

$$r = \sum_{\{\omega\}} p(\omega) \cos \left[ \arcsin \left( \frac{\Delta(\omega)}{Kr} \right) \right] . \quad (1.19)$$

If  $\Delta(\omega) \equiv 0$ , then the system is in a perfectly locked state ( $\psi = \varphi_\infty$ ) associated with an energy minimum of the  $XY$  ferromagnet and  $r = \sum_{\{\omega\}} p(\omega) = 1$

is a *stable* solution. That is,  $\arcsin(0)$  has to vanish, and

$$\cos \left\{ \arcsin \left[ \frac{\Delta(\omega)}{Kr} \right] \right\} = \left[ 1 - \left( \frac{\Delta(\omega)}{Kr} \right)^2 \right]^{1/2} . \quad (1.20)$$

Combining this with Eq. (1.19) we arrive at the fixed-point equation

$$r = \sum_{\{\omega\}} p(\omega) \left[ 1 - \left( \frac{\Delta(\omega)}{Kr} \right)^2 \right]^{1/2} . \quad (1.21)$$

By construction, a solution  $r$  is bound to be such that  $|\Delta(\omega)/Kr| \leq 1$ . For  $\Delta(\omega) \equiv 0$  (or  $K = \infty$ ) Eq. (1.21) has the solution  $r = 1$ . For  $|\Delta(\omega)|$  small (or  $K$  large but finite) we then also obtain a solution by the implicit function theorem.

For given  $\Delta(\omega)$  one may wonder how small  $K$  can be chosen ( $K \geq K_c$ ) and what is the nature of the transition at  $K_c$  where  $r$  ceases to exist as a solution of Eq. (1.21). Putting  $x = (Kr)^2 \in [0, K^2]$  we can rewrite Eq. (1.21) in the form

$$K^{-1} x = \sum_{\{\omega\}} p(\omega) [x - \Delta^2(\omega)]^{1/2} \equiv \vartheta(x) , \quad (1.22)$$

where  $\vartheta(x)$  is defined for  $x \geq \Delta_m^2 = \sup_{\omega} \Delta^2(\omega)$ . On its domain,  $\vartheta$  is a convex combination of concave functions and thus *concave* itself. Moreover, Eq. (1.22) tells us that, as we decrease  $K$ , there exists a critical  $K_c$  such that we find a (stable) solution for  $K > K_c$  and no solution for  $K < K_c$ . Hence there exists *no* global phase locking for  $K < K_c$ . Some examples will be presented later on. The critical  $K_c$  itself can be obtained from Eq. (1.22),

$$K_c^{-1} = \sup_{x \geq \Delta_m^2} \sum_{\{\omega\}} p(\omega) [x - \Delta^2(\omega)]^{1/2} x^{-1} . \quad (1.23)$$

The restriction  $x \geq \Delta_m^2$  is irrelevant for a discrete distribution, since in that case the function  $\vartheta$  is strictly concave and such that the maximum in Eq. (1.23) is assumed for  $x > \Delta_m^2$ .

Stepping back for an overview, we now want to interpret the Lyapunov function (1.10) as a Hamiltonian so that a minimum of  $\mathcal{H}$  represents a ground state of a physical spin system. The first term on the right is the energy of an  $XY$  ferromagnet of mean-field type with coupling strength  $K$ . This term aims at keeping *all* spins parallel. The second term represents a kind of *random field* with strength  $(\omega_i - \langle \omega \rangle)$  and mean  $\langle \omega - \langle \omega \rangle \rangle = 0$ . To

minimize the energy, the second term wants to make the  $\varphi_i$  with  $\omega_i - \langle \omega \rangle > 0$  positive and those with  $\omega_i - \langle \omega \rangle < 0$  negative — as positive and negative as possible. So the two terms counteract each other. There is *frustration*. For large  $K$  the ferromagnet wins in that the system is phase locked, the sublattices are homogeneous, and, as  $K$  decreases, their phases  $\varphi(\omega)$  rotate away slowly from a single fixed direction, a minimal energy configuration of the  $XY$  ferromagnet. This is brought out clearly by Eq. (1.17). The minimum of  $\mathcal{H}$  performs a kind of “unfolding” in the phase space  $[-\pi, \pi]^N$  — like a multiwinged butterfly, the phases being the wings. Herewith the stability of the phase-locked state obtains a natural explanation. As  $K$  decreases further and reaches  $K_c$  the random field takes over. No solution to Eq. (1.16) and no global phase locking exist beyond  $K_c$  anymore.

It may be advantageous to picture the transition at  $K_c$ . To this end we take  $N = 2$  in Eq. (1.10), with  $\omega_1 - \Omega = -1/2$  and  $\omega_2 - \Omega = 1/2$ , and use the third dimension to plot  $\mathcal{H}$ ; see Fig. 1.2.  $\mathcal{H}$  is a function of  $\varphi_1 - \varphi_2$  and, thus, rotationally invariant. The dynamically relevant direction is orthogonal to  $\varphi_1 - \varphi_2 = \text{constant}$ . This can be seen from the gradient dynamics induced by Eqs. (1.8) and (1.10). For  $K > K_c = 1$ , the Lyapunov function has ripples and the system always gets stuck in a minimum of  $\mathcal{H} = -(K/2) \cos(\varphi_1 - \varphi_2) + (\varphi_1 - \varphi_2)/2$ . For  $K < K_c$ , there is no ripple and, hence, no locking. In the following three sections we will estimate  $K_c$  and the order parameter  $r$ , consider some examples, and study the stability of the phase-locked state in more detail.

### Estimating $K_c$ and $r_c$

One of the main problems is estimating  $r_c$  and  $K_c$ , the critical values of  $r$  and  $K$ . We will do so in the limit  $N \rightarrow \infty$ . For large  $x$ , we can estimate the right-hand side of Eq. (1.23) by modifying an argument of Ermentrout’s [33]. We do this for a general probability measure  $\mu$  and write  $K_c = \sup_x \Theta(x)$ , where  $x \geq \Delta_m^2$  and

$$\Theta(x) = x^{-1} \int d\mu(\omega) \sqrt{x - \Delta^2(\omega)} = x^{-1} \vartheta(x) . \quad (1.24)$$

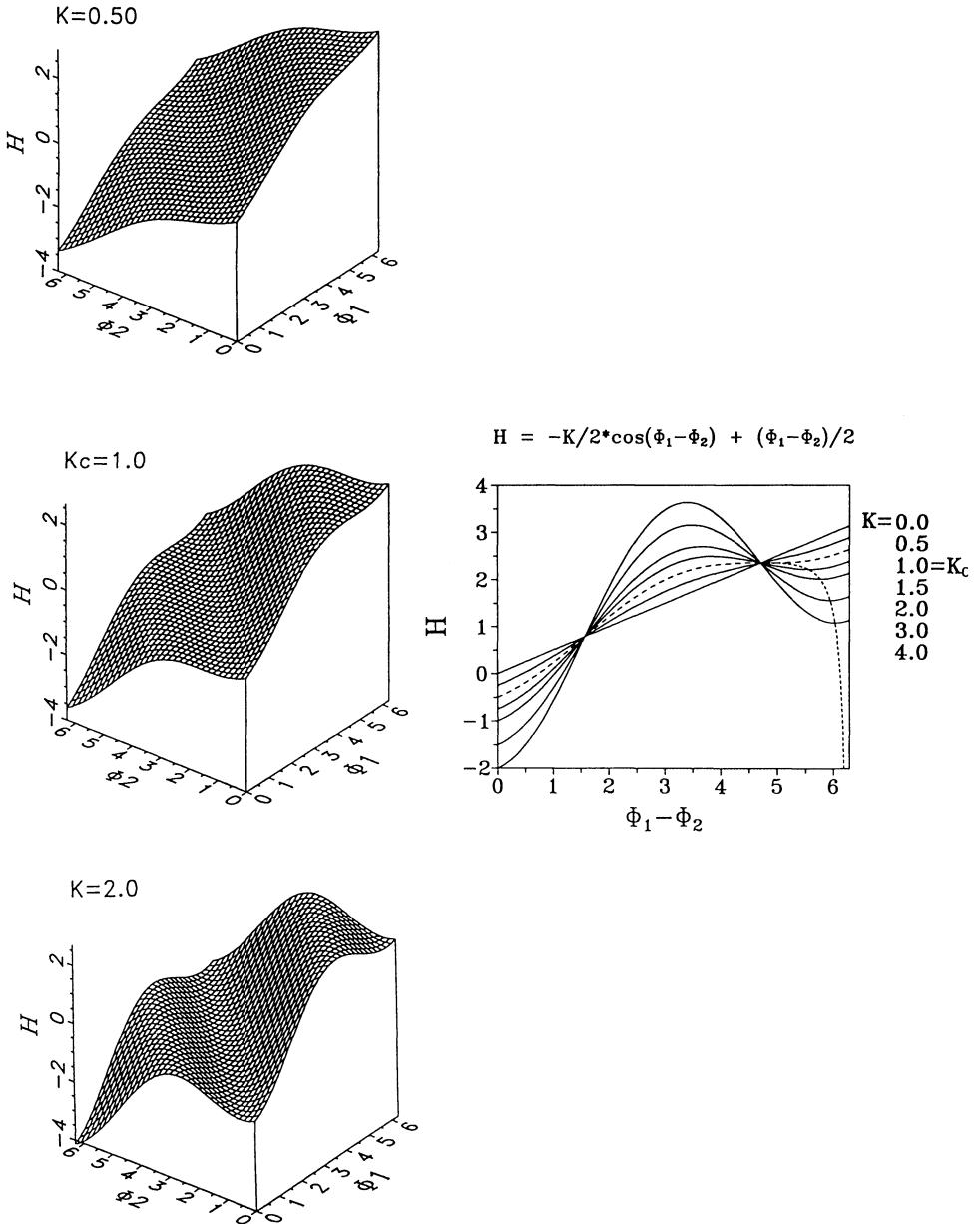
Computing the derivative  $\Theta'$  we find

$$\Theta'(x) = \frac{1}{2x^2} \int d\mu(\omega) \frac{2\Delta^2(\omega) - x}{\sqrt{x - \Delta^2(\omega)}} . \quad (1.25)$$

Hence  $\Theta'(x) < 0$  and  $\Theta(x)$  is decreasing for  $x = (Kr)^2$  beyond  $2\Delta_m^2$ . Thus we obtain the estimate

$$\Delta_m \leq (Kr)_c \leq \sqrt{2} \Delta_m . \quad (1.26)$$

We now turn to lower bounds for  $K_c$  and  $r$  separately.



**Fig. 1.2.** The Lyapunov function  $\mathcal{H}$ . Left column. For  $N = 2$ ,  $\mathcal{H}$  has been plotted as a function of  $\varphi_1$  and  $\varphi_2$  on  $[0, 2\pi] \times [0, 2\pi]$  for various values of  $K$ . The frequencies are  $\omega_1 - \Omega = -1/2$  and  $\omega_2 - \Omega = 1/2$ . The motion of the system is a gradient dynamics which evolves in a plane perpendicular to the line  $\varphi_1 = \varphi_2$ . The intersection with  $\mathcal{H}$  contains the trajectory. For  $K > K_c$ , two ripples occur in the surface and the system gets phase locked in the minimum of one of them. An intersection with the upper ripple (right-hand corner, with  $0 \leq \varphi_1 - \varphi_2 \leq 2\pi$ ) has been plotted in the right column. The dashed line on the right going downwards indicates the location of the minima as  $K$  varies. Taken from [62].

The square root being a concave function we apply Jensen's inequality [110] to Eq. (1.22) and find

$$K^{-1}x \leq \left\{ \int d\mu(\omega) [x - \Delta^2(\omega)] \right\}^{1/2} = [x - \langle \Delta^2(\omega) \rangle]^{1/2} , \quad (1.27)$$

and after squaring this,

$$K^{-2}x^2 - x + \langle \Delta^2(\omega) \rangle \leq 0 . \quad (1.28)$$

The condition (1.28) can be realized only if the discriminant is positive, that is,  $K^2 \geq 4\langle \Delta^2(\omega) \rangle$ . Thus we find

$$K_c \geq 2\sqrt{\langle \Delta^2(\omega) \rangle} . \quad (1.29)$$

We cannot do better since the inequality (1.29) becomes an equality in case  $p(\omega_1) = p(\omega_2) = 1/2$ , as we will see shortly.

To derive a lower bound for  $r_c$  we start with Eq. (1.26), viz.,  $\Delta_m \leq (Kr)_c$ . Though this inequality does not look optimal, it actually is. In the next section we will see that it is saturated by the uniform distribution. If so, we now need a lower bound for  $K_c^{-1}$ . To this end we combine Eqs. (1.23) and (1.26), restrict  $x$  to the interval  $[\Delta_m^2, 2\Delta_m^2]$ , and evaluate the right-hand side of Eq. (1.23) at  $x = 2\Delta_m^2$  so as to get

$$\begin{aligned} K_c^{-1} &\geq \sum_{\{\omega\}} p(\omega) [2\Delta_m^2 - \Delta^2(\omega)]^{1/2} / 2\Delta_m^2 \\ &= \sum_{\{\omega\}} p(\omega) \left[ 2 - \left( \frac{\Delta(\omega)}{\Delta_m} \right)^2 \right]^{1/2} / 2\Delta_m \geq (2\Delta_m)^{-1} . \end{aligned} \quad (1.30)$$

Thus we arrive at the extremely simple inequality

$$r_c \geq \Delta_m / K_c \geq 1/2 . \quad (1.31)$$

It tells us explicitly that a continuous transition from the phase-locked to a nonlocked state with vanishing  $r$  is to be excluded. Note that in obtaining Eq. (1.31) we have not made any special assumption concerning the probability distribution of the frequencies  $\omega_i$ . Neither do we assert that  $r$  must vanish for  $K < K_c$ . There is just no global phase locking.

The inequality (1.31) also provides us with an upper bound for  $K_c$  in that  $K_c \leq 2\Delta_m$ . In case  $p(\omega_1) = p(\omega_2) = 1/2$ , this upper bound and the lower bound (1.29) coincide, so that the upper bound is optimal as well.

### Examples: Discrete and continuous distributions

The simplest nontrivial distribution is the one with two frequencies,  $\omega_1$  and  $\omega_2 > \omega_1$ , and probabilities,  $p(\omega_2) = p$  and  $p(\omega_1) = 1 - p$ . Then  $\Delta(\omega_1) = -(\omega_2 - \omega_1)p \leq 0$  and  $\Delta(\omega_2) = (\omega_2 - \omega_1)(1 - p) \geq 0$  while

$$\langle \Delta^2(\omega) \rangle = p(1 - p)(\omega_2 - \omega_1)^2 . \quad (1.32)$$

$p = 1/2$ , so we get

$$\langle \Delta^2(\omega) \rangle^{1/2} = \frac{1}{2} (\omega_2 - \omega_1) = |\Delta(\omega)| = \Delta_m . \quad (1.33)$$

The fixed-point equation (1.22) takes the form

$$K^{-1}x = (1-p)\sqrt{x - [p(\omega_2 - \omega_1)]^2} + p\sqrt{x - [(1-p)(\omega_2 - \omega_1)]^2} \quad (1.34)$$

with  $x = (Kr)^2 \geq \Delta_m^2$  and  $\Delta_m = \max \{(\omega_2 - \omega_1)p, (\omega_2 - \omega_1)(1-p)\}$ . Whatever  $p$ , there is a remarkably simple expression for the phase difference,

$$\varphi(\omega_2) - \varphi(\omega_1) = \arcsin[(\omega_2 - \omega_1)/K] , \quad (1.35)$$

which can be proven, for example, by combining an addition formula for two arcsines with Eq. (1.21).

We now return to the case  $p = 1/2$ . Then Eq. (1.34) can be squared so as to give

$$K^{-2}x^2 - x + [(\omega_2 - \omega_1)/2]^2 = 0 . \quad (1.36)$$

We get a positive solution to Eq. (1.36) as long as its discriminant is positive, that is,

$$K \geq K_c = |\omega_2 - \omega_1| . \quad (1.37)$$

At  $K_c$  we find  $x(K_c) = K_c^2/2$  so that  $r(K_c) = 1/\sqrt{2}$ . Taking into account both Eqs. (1.37) and (1.33) one easily verifies that inequality (1.29) has been turned into an equality; in short, it is optimal. Combining Eqs. (1.35) and (1.37) we get that in this particular case and at  $K_c$ :

$$\varphi_{\max} - \varphi_{\min} = \arcsin(1) = \pi/2 . \quad (1.38)$$

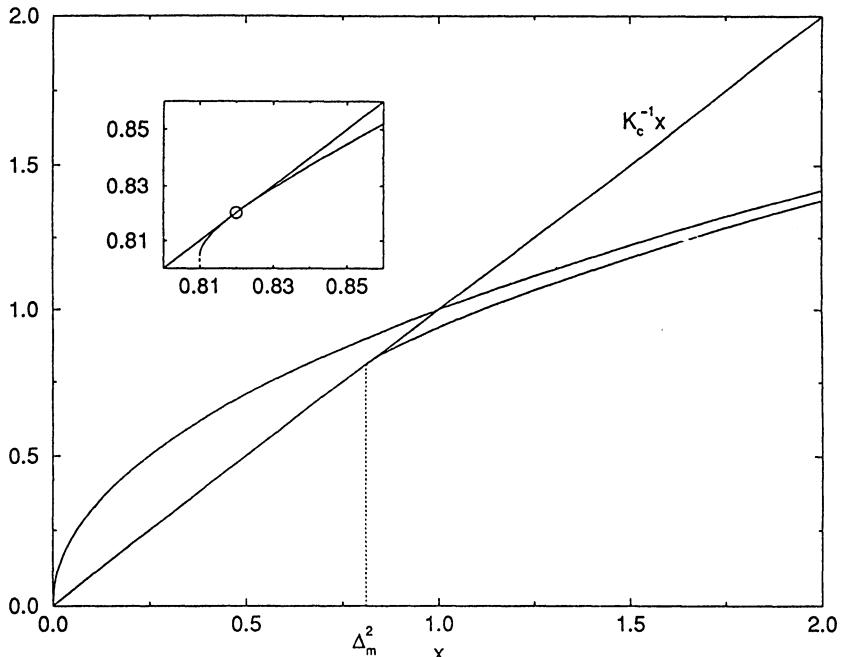
The limit  $p \rightarrow 0$  can also be handled analytically. Here  $\Delta(\omega_2) \gg |\Delta(\omega_1)|$  and Eq. (1.34) may be approximated by

$$K^{-1}x = (1-p)\sqrt{x} + p[x - (\omega_2 - \omega_1)^2]^{1/2} \quad (1.39)$$

with  $x \geq (\omega_2 - \omega_1)^2$ . Figure 1.3 shows that in this limit  $K_c = |\omega_2 - \omega_1|$  once more so that  $r_c = 1$ , as is to be expected. We explicitly see that the side condition  $x \geq \Delta_m^2$  is harmless. Equation (1.35) implies that here too Eq. (1.38) holds. Note that the limit  $p \rightarrow 0$  is different from the case  $p = 0$ . The latter has perfect locking *whatever*  $K$ .

Furthermore, the above results for  $p = 0, 1/2$ , and 1 (due to the symmetry  $p \rightarrow (1-p)$ ) suggest that Eqs. (1.37) and (1.38) hold for any  $p$ . That is indeed the case.

**Proposition:** For the bimodal distribution with  $p(\omega_2) = p$  and  $p(\omega_1) = 1-p$  we have  $K_c = |\omega_2 - \omega_1|$ , whatever  $p$ . Moreover, at  $K_c$ , the phases



**Fig. 1.3.** Graphical solution to the fixed-point equation (1.34) with  $\omega_2 - \omega_1 = 1$ . The lower curve and the vertical dashed line represent the case  $p = 0.1$ . The inset shows why here the side condition  $x \geq \Delta_m^2$  is irrelevant. Note that the limit  $p \rightarrow 0$ , that is, the square root with  $x \geq (\omega_2 - \omega_1)^2$ , differs from the case  $p = 0$ , viz., the square root with  $x \geq 0$ .

of the two sublattices belonging to  $\omega_1$  and  $\omega_2$  are orthogonal, that is, Eq. (1.38) holds.

**Proof:** There is no harm in taking  $\omega_2 > \omega_1$ . Turning to Eq. (1.22), we note that in the present case  $\vartheta(x)$  starts with a square-root singularity, is monotonically increasing and strictly concave for  $x \geq \Delta_m^2$ , that the side condition is irrelevant when we apply Eq. (1.23), and that the maximum is unique — as is exemplified by Fig. 1.3. Putting the derivative of the right-hand side of Eq. (1.23) equal to zero, we state as a *fait accompli* that the unique  $x(p)$  maximizing Eq. (1.23) equals

$$x(p) = \left[ p^2 + (1-p)^2 \right] (\omega_2 - \omega_1)^2 .$$

A little algebra then suffices to verify  $K_c = \omega_2 - \omega_1$  and, taking advantage of Eq. (1.35), we find Eq. (1.38). 2

One might guess that Eq. (1.38) is generally true. Our final example shows that this is not the case. The uniform distribution on  $[-1, 1]$  is a

favorite of the literature. Its fixed-point equation is ( $x \geq \Delta_m = 1$ )

$$K^{-1}x = \frac{1}{2} \int_{-1}^1 d\omega (x - \omega^2)^{1/2} . \quad (1.40)$$

The integral can be done exactly and

$$K_c^{-1} = \sup_{x \geq 1} \frac{1}{2} [x^{-1} \sqrt{x-1} + \arcsin(1/\sqrt{x})] . \quad (1.41)$$

One either applies an argument of Ermentrout's [33] to Eq. (1.40) or checks explicitly that the right-hand side of Eq. (1.41) assumes its maximum at  $x = (Kr)^2 = 1$  so that  $K_c = 4/\pi$ . In addition, Eq. (1.17) implies  $\varphi_{\max} - \varphi_{\min} = \pi$ .

### Stability

The existence of a Lyapunov function guarantees a rather strong form of stability of its minima, which are fixed points of the equation of motion (1.6). Here we will not delve into a formal analysis. Instead we refer the reader to the literature [62] and just ask what can be said beforehand. To this end, we return to the  $XY$  ferromagnet (1.7). Due to the gradient dynamics (1.6), the system relaxes to a minimum of  $\mathcal{H}$  which is characterized by  $\varphi_i = \varphi_\infty$  for  $1 \leq i \leq N$ . Is it unique? No, as we have seen, it is not. The ground state of Eq. (1.7) is rotationally invariant and it remains so, if we add the random field to  $\mathcal{H}$  so as to arrive at Eq. (1.10). The reason is that a uniform rotation through  $\alpha$  produces an extra term:

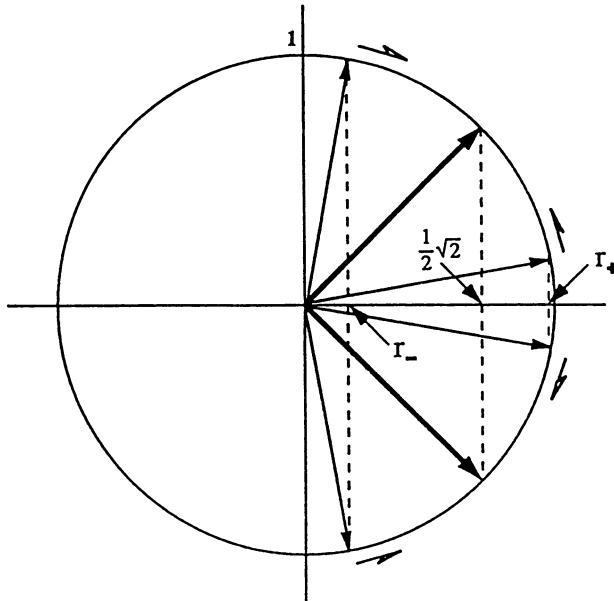
$$\frac{\alpha}{N} \sum_{i=1}^N (\omega_i - \langle \omega \rangle) = 0 ,$$

which vanishes by the very definition (1.3) of  $\langle \omega \rangle$ . Thus we expect, and find, a permanent eigenvalue zero belonging to the eigenvector  $\mathbf{1} = (1, 1, \dots, 1)$  of the Jacobian matrix at a fixed point of the equations of motion (1.6). In passing we note that a gradient dynamics always evolves in a space *orthogonal* to  $\mathbf{1}$ , as is brought out clearly by Fig. 1.2.

It may be clarifying to consider a simple example explicitly, viz., the case  $p(\omega_1) = p(\omega_2) = 1/2$ . The fixed-point equation (1.36) has two roots which for small  $\omega_2 - \omega_1 > 0$  (or large  $K$ ) lead to the following two values for the order parameter  $r$ :

$$r_+ = 1 - \frac{1}{2} \left( \frac{\omega_2 - \omega_1}{K} \right)^2 , \quad r_- = \frac{\omega_2 - \omega_1}{2K} . \quad (1.42)$$

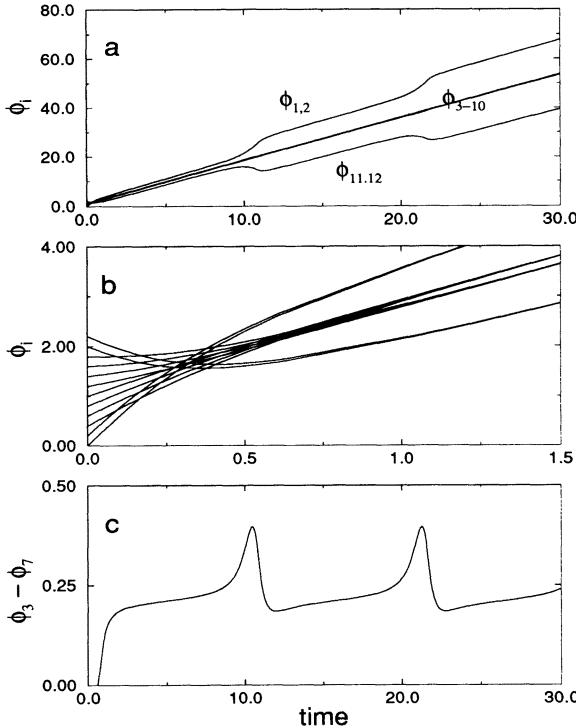
The phases of the states corresponding to  $r_+$  and  $r_-$  have been indicated in Fig. 1.4. They allow a simple interpretation. In the limit  $(\omega_2 - \omega_1)/K \rightarrow 0$ , the first corresponds to *all* spins parallel. It is a stable ground state.



**Fig. 1.4.** Interpreting the stationary points of  $\mathcal{H}$ . For large  $K$ , a system with  $p(\omega_1) = p(\omega_2) = 1/2$  gives rise to two stationary points of the Lyapunov function  $\mathcal{H}$  with order parameter values  $r_+$  and  $r_-$ . The former corresponds to a minimum of  $\mathcal{H}$  and is stable whereas the latter corresponds to a maximum of  $\mathcal{H}$  and is unstable. The phases are given by the angles with the positive horizontal axis and  $r_\pm$  is a convex combination of the  $r$  values of the two sublattices with weight  $1/2$ . So  $r_+ \approx 1$  and  $r_- \approx 0$  and the two sublattices have their block spins nearly parallel or antiparallel. As  $K$  decreases, the two upper and lower phases approach each other and they meet at  $K_c$  at angles  $\varphi_{\max} = \pi/4$  and  $\varphi_{\min} = -\pi/4$ . In other words, at  $K_c$  they merge,  $\varphi_{\max} - \varphi_{\min} = \pi/2$ . Taken from [62].

The second has the spins on both  $I(\omega_1)$  and  $I(\omega_2)$  parallel but  $\varphi(\omega_2) - \varphi(\omega_1) \approx \pi$ , that is, the sublattices have their spins *antiparallel* and the total magnetization vanishes. This configuration corresponds to an energy *maximum* of the *XY* ferromagnet. It is a stationary point ( $\nabla \mathcal{H} = 0$ ) but evidently an *unstable* one; cf. Eq. (1.9). As  $(\omega_2 - \omega_1)/K$  increases,  $\varphi(\omega_2) - \varphi(\omega_1)$  increases for the stable configuration and decreases for the unstable one. The phases of the stable and the unstable configurations meet each other — at least in this case — at  $\varphi = \pi/4$  and  $\varphi = -\pi/4$ , respectively. That is, they meet at  $K_c$ . As they merge the phase-locked state disappears.

If the interaction between the oscillators is no longer all-to-all, as in the Kuramoto model, but local, say nearest-neighbor, then many (local) minima may exist. This already holds for the relatively simple *XY* ferromagnet



**Fig. 1.5. Partial phase locking for discrete distributions.** For  $K < K_c$ , “partial” phase locking exists in the sense that we have a total phase locking on each of the sublattices  $I(\omega)$  but *not* between them. (a,b) Solution of Eq. (1.1) for  $N = 12$  oscillators with  $\omega_1 = \omega_2 = 4$ ,  $\omega_3 = \dots = \omega_6 = 2$ ,  $\omega_7 = \dots = \omega_{10} = 1.5$ , and  $\omega_{11} = \omega_{12} = -0.5$ , corresponding to probabilities  $1/6$ ,  $1/3$ ,  $1/3$ , and  $1/6$ , respectively. Here  $K = 3.0 < K_c = 3.08$ . The initial condition is the homogeneous distribution  $\phi_{12}(0) > \phi_{11}(0) > \dots > \phi_2(0) > \phi_1(0)$ . Asymptotically (not shown here),  $\phi_1$  and  $\phi_2$ ,  $\phi_3$  to  $\phi_6$ ,  $\phi_7$  to  $\phi_{10}$ ,  $\phi_{11}$  and  $\phi_{12}$  merge sublatticewise, as is already suggested by b. However, c exhibits  $\phi_3 - \phi_7$  for large times and shows that the sublattices  $I(1.5)$  and  $I(2.0)$ , though exhibiting something like a partial phase locking when  $\phi_3 - \phi_7$  is pretty flat, do *not* lock exactly. Interestingly, the large humps in c occur in concert with those of  $\phi_{1,2}$  and  $\phi_{11,12}$  in a.

that emerges, if all frequencies are identical — that is, as in Eq. (1.6). Here we have one global minimum, viz., the state with all spins parallel. Even in a one-dimensional ring [34] there is already a second, local, minimum given by  $\varphi_i = \alpha + 2\pi i/N$  where  $N$  is the length of the ring and  $\alpha$  refers to the degeneracy due to a uniform rotation. Note, however, that the Hamiltonian (1.9) of the Kuramoto model has only a *single* minimum.

### Partial phase locking

Phase locking is apparently the rule, if the  $\omega_i$  do not scatter too much. A natural question then is: what happens when some but not all of the oscillators can lock? For example, for  $K = 3$  we take  $\omega_i = -0.5, 1.5, 2.0$ , and 4.0 while the  $p(\omega_i)$  equal  $1/6, 1/3, 1/3$ , and  $1/6$ , respectively. By the fixed-point equation (1.22) we obtain  $K < K_c = 3.08$ , which slightly exceeds  $K = 3$ . The sublattices  $I(1.5)$  and  $I(2.0)$  could lock, at least in principle since  $K > 0.5$ , whereas  $I(-0.5)$  and  $I(4.0)$  have to stay apart, unlocked. If so, one might think that the frequency common to the sublattices  $I(1.5)$  and  $I(2.0)$  would be 1.75. This is not true, due to the exact sum rule (1.3). Neither do they lock exactly, nor is their “common frequency” the appropriately weighted mean of the sublattice frequencies; cf. Figs. 1.5a and 1.5c. Moreover, in numerical simulations it turns out that asymptotically, as  $t \rightarrow \infty$ , all phases  $\varphi_i(t)$  on a single sublattice  $I(\omega)$  approach the *same* limit  $\varphi(\omega; t)$ ; cf. Fig. 1.5b. Hence we end up with a *reduced* dynamics:

$$\dot{\varphi}(\omega) = \Delta(\omega) - K \sum_{\{\omega'\}} p(\omega') \sin[\varphi(\omega) - \varphi(\omega')] \quad (1.43)$$

obeying the exact sum rule

$$\frac{d}{dt} \left( \sum_{\{\omega\}} p(\omega) \varphi(\omega) \right) = 0 \quad . \quad (1.44)$$

In view of the Lyapunov function (1.10) the reduction is easily understood. Though  $K$  is less than  $K_c$  and, thus, a stationary point of  $\mathcal{H}$  cannot be found, the ferromagnetic interaction is at least *minimized on the sublattices*, if there the spins are parallel, that is,  $\varphi_i(t) = \varphi(\omega; t)$  for all  $i \in I(\omega)$ . So it is fair to call this a *sublattice phase locking*. The “minimizing path” itself depends on the distribution of the  $\omega$ ’s. Moreover, since the “partial” phase locking of the sublattices that in principle could lock is *not* an exact one, a rigorous but simple description of the system’s behavior for  $K < K_c$  is hard to imagine — except for Eq. (1.43).

It may be well to contrast the present results with those obtained for more “generic” models [33,35,37,85–88,123–125] that have an absolutely continuous frequency distribution with a symmetric and one-humped density function, such as the Gaussian and the Lorentzian, and ask whether their behavior is truly generic. In this type of model one has [87,124], as  $K$  decreases from infinity, *two* transitions: One at  $K_c$  where the random field takes over partially in that the system is only partially phase locked, and another one at  $K_{pc} < K_c$  where also the partially locked state disappears. For  $K < K_{pc}$  the system behaves truly incoherently. Partial phase locking means that oscillators with frequencies near the center of the distribution

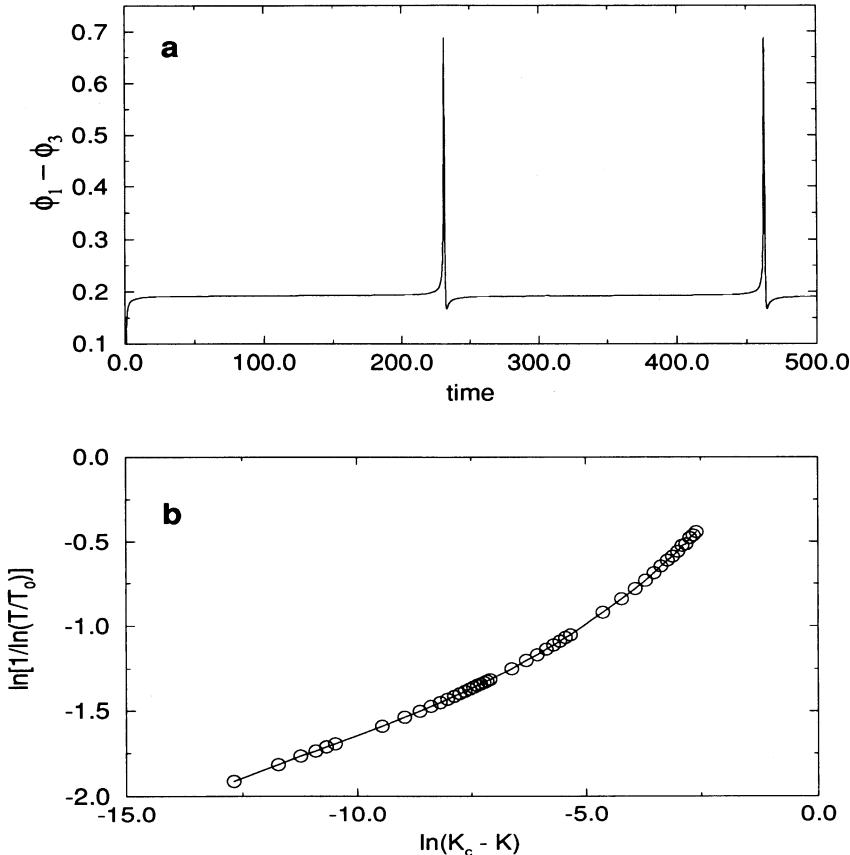
remain locked whereas outlying oscillators are desynchronized. In passing we note that a uniform distribution on a bounded interval has  $K_{pc} = K_c$ .

If partial locking were generic, a discrete frequency distribution would show a similar behavior. For example, let us return to the system of the previous section (Fig. 1.5) that has four frequencies,  $\omega_i = -0.5, 1.5, 2$ , and 4 with respective probabilities  $1/6, 1/3, 1/3$ , and  $1/6$  — a symmetric, one-humped distribution. Without  $\omega_i = -0.5$  and 4, the system would phase lock for  $K > 0.5$  (cf. the Examples) and we therefore expect that, if the partial-locking behavior were generic, then for  $K = 3 < K_c = 3.08$  the sublattices belonging to  $\omega_i = 1.5$  and 2, which are also nearest to  $\langle \omega \rangle = 1.75$ , would lock together as well. As one sees in Fig. 1.5c they do not. We have verified that they lock nowhere below  $K_c$ .

Even more can be said. We have also analyzed the behavior as  $K \rightarrow K_c$  from below and studied the dynamics of a system with a symmetric distribution consisting of four frequencies which are equally probable. (We have checked that a humped distribution qualitatively gives the same results.) In the present case, we have simply taken the very same frequencies as in Fig. 1.5. Then  $\langle \omega \rangle = 1.75$  and  $K_c = 3.4748$ , as follows from the fixed-point equation (1.22). Figure 1.6a with  $K = 3.4747 < K_c = 3.4748$  shows that, for a discrete distribution, the absence of partial locking below  $K_c$  is quite universal. Here too we have verified that the two sublattices  $I(1.5)$  and  $I(2.0)$  associated with frequencies near the center of the distribution lock nowhere below  $K_c$ . One may wonder, though, how the system “feels” that  $K$  is approaching  $K_c$  from below. To this end we have studied the recurrence time  $T_{\text{rec}}$  of the asymptotic phase difference between  $I(1.5)$  and  $I(2.0)$  as a function of  $(K_c - K)$ ; cf. Fig. 1.6b. As  $K$  approaches  $K_c$  from below, the amplitude of the phase difference does not vary but the recurrence time  $T_{\text{rec}}$  does: it diverges to infinity. As in a *first-order* phase transition, we do not find a pure power law behavior. For  $K > K_c$  the phase difference is asymptotically fixed and  $T_{\text{rec}}$  is infinite.

There are, apparently, classes of oscillator models with different generic behavior. That is, there are different *universality classes*. The absolutely continuous distributions [33,35,37,85–88,123–125] belong to one class and the discrete distributions to another one. The former give rise to two transitions at  $K_c$  and  $K_{pc}$  whereas the latter appear to have only a single transition at  $K_c$ .

In summary: It is the very existence of a Lyapunov function  $\mathcal{H}$  that allows a physically transparent treatment of a phase-locked state of the Kuramoto model as a ground state. In fact, the argument holds for any equivalent model with finite-range interactions [34,100,112,113]. The Lyapunov function  $\mathcal{H}$  has two constituents, an  $XY$  ferromagnet with coupling strength  $K > 0$  and a random field  $(\omega_i - \langle \omega \rangle)$ ; cf. Eq. (1.10). In the case of the Kuramoto model, the  $XY$  ferromagnet dominates for  $K > K_c$ , the sublattices have a homogeneous phase, and their phases lock with respect to each other. The larger  $|\omega - \langle \omega \rangle|$ , the larger the phase shift. For  $K < K_c$ ,



**Fig. 1.6.** “Critical behavior” as  $K \rightarrow K_c$  from below. (A) Phase difference ( $\phi_1 - \phi_3$ ) between two oscillators near the center of a symmetric, discrete distribution as a function of time. They were taken out of a population of eight oscillators ( $N = 8$ ) with  $\omega_1 = \omega_2 = 2$ ,  $\omega_3 = \omega_4 = 1.5$ ,  $\omega_5 = \omega_6 = 4$ , and  $\omega_7 = \omega_8 = -0.5$ . The initial conditions are as in Fig. 1.5b. Furthermore,  $K = 3.4747 < K_c = 3.4748$ . The recurrence time  $T$  between two subsequent peaks is about 231. It is plain that the sublattices  $I(1.5)$  and  $I(2.0)$  do not lock, even though their frequencies are near the center  $\langle \omega \rangle = 1.75$ ,  $K$  is only slightly below  $K_c$ , and, between the peaks, the system does look “partially phase locked.” (B) To verify that dependence of  $T$  upon  $(K_c - K)$  has a first-order character, we have plotted  $\ln[1/\ln(T/T_0)]$  against  $\ln(K_c - K)$  where  $K_c = 3.474\ 828$  and  $T_0 = 2.2$ . The open circles represent numerically obtained data points. A pure power law behavior à la  $T = T_0 (K_c - K)^x$  for some  $x < 0$  does not occur. The first-order dependence is in agreement with the nature of the transition as  $K$  approaches  $K_c$  from above; cf. Fig. 1.3 and the discussion below (Eq. (1.22)). Taken from [62].

the sublattices still have a homogeneous phase as  $t \rightarrow \infty$  (sublattice phase locking) but  $\mathcal{H}$  has no stationary point and the “minimizing path” is determined by the distribution of the  $\omega$ ’s, as is brought out by Eq. (1.43). We now turn to models which are more specific to the observed synchronization phenomena in the primary visual cortex.

### 1.2.2 OSCILLATOR MODELS IN ACTION

Despite their limitations, oscillator models have been quite popular as theoretical vehicles to describe coherent oscillations as found in the primary visual cortex [8,28,51,52,118]. Here we analyze the essentials of proposals due to Schuster and Wagner [115,116] and Sompolinsky et al. [119,120]. We think that the work provided by these authors is more fundamental than that of König and Schillen [77,114], who simply make a phenomenological ansatz to describe a single cortical column [75,84] as a nonlinear oscillator; the interested reader is referred to them. It is quite surprising how well experimental facts can be explained qualitatively by such a global description.

Schuster and Wagner start with the neurons which constitute a single column, then model a combination of two columns, reduce its dynamical behavior to that of two *phases*, and build a network out of these simple constituents, the phases, so as to end up with a Kuramoto-type model. The neurons are taken to be analog neurons and described by a *rate* coding [70,131], which presupposes a time window much wider than the temporal coherence of neuronal spiking, viz., 2–3 ms *vs* 50–100 ms needed to define a rate (here about 50 Hz). In the present subsection we simply take this contradiction for granted and continue.

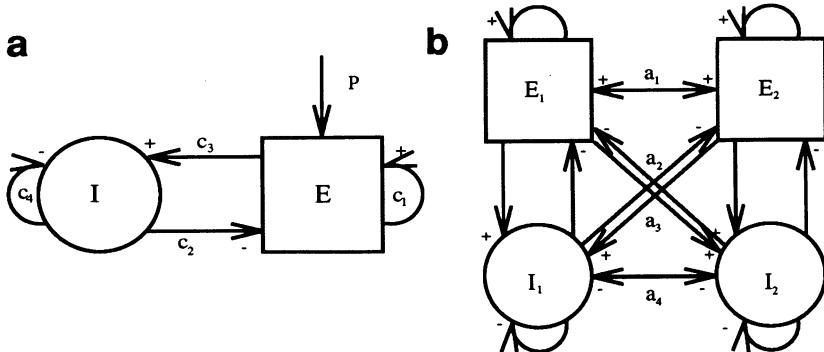
A column is defined to be a set of  $N_e$  excitatory and  $N_i$  inhibitory neurons which are globally coupled, that is, all to all; cf. Fig. 1.7a. Their firing rates  $e_k$  and  $i_k$  are governed by

$$\begin{aligned} \frac{de_k}{dt} &= -e_k + S \left( \beta_e \left[ c_1 \left( N_e^{-1} \sum_{l=1}^{N_e} e_l \right) - c_2 \left( N_i^{-1} \sum_{l=1}^{N_i} i_l \right) - \vartheta_k^e + p_k \right] \right) , \\ \frac{di_k}{dt} &= -i_k + S \left( \beta_i \left[ c_3 \left( N_e^{-1} \sum_{l=1}^{N_e} e_l \right) - c_4 \left( N_i^{-1} \sum_{l=1}^{N_i} i_l \right) - \vartheta_k^i \right] \right) . \end{aligned} \quad (1.45)$$

The  $\vartheta$ ’s are thresholds,  $p_k$  is the input of neuron  $k$ , and the  $\beta$ ’s determine the steepness of the sigmoid response function  $S(\beta x) = 1/[1 + \exp(-\beta x)]$ . The  $c$ ’s,  $\vartheta$ ’s, and  $\beta$ ’s are parameters which are at our disposal.

It is handy to introduce the collective variables

$$E = N_e^{-1} \sum_{l=1}^{N_e} e_l \quad \text{and} \quad I = N_i^{-1} \sum_{l=1}^{N_i} i_l \quad (1.46)$$



**Fig. 1.7.** Schematic representation of (a) the intracolumn and (b) the intercolumn interaction [115,116]. The excitatory and inhibitory subpopulations carry the labels  $E$  and  $I$ , respectively. Inside a single column, the couplings are denoted by  $c$ , between the columns by  $a$ . The + and - signs stand for excitatory and inhibitory couplings; cf. Eqs. (1.45) and (1.48).

and compute  $\dot{E} = dE/dt$  and  $\dot{I}$  by substituting Eq. (1.45) into Eq. (1.46) so as to arrive at

$$\begin{aligned}\dot{E} &= -E + S[\beta_e(c_1E - c_2I - \vartheta^e + P)] \quad , \\ \dot{I} &= -I + S[\beta_i(c_3E - c_4I - \vartheta^i)]\end{aligned}\quad (1.47)$$

for a homogeneous input  $p_k \equiv P$ . From a physical point of view, it is important to realize that  $E$  and  $I$  are collective variables that are to describe large populations of neurons. It turns out [115] that, for reasonable parameter values, there is an open interval  $(P_{cl}, P_{ch})$  such that for  $P$  outside this interval we have stable fixed points for  $E$  and  $I$  whereas for  $P \in (P_{cl}, P_{ch})$  we find oscillations. When  $P$  is moved through  $P_{cl}$  the system undergoes a Hopf bifurcation [67].

We now couple *two* of these columns (see Fig. 1.7b) in such a way that ( $k \neq l$ )

$$\begin{aligned}\dot{E}_k &= -E_k + S[\beta_e(A_k^e + \eta U_l)], \\ \dot{I}_k &= -I_k + S[\beta_i(A_k^i + \eta V_l)]\end{aligned}\quad (1.48)$$

replaces Eq. (1.47). Here  $A_k^e = c_1E_k - c_2I_k - \vartheta^e + P$  and  $A_k^i = c_3E_k - c_4I_k - \vartheta^i$  as before while

$$U_l = a_1E_l - a_2I_l, \quad V_l = a_3E_l - a_4I_l \quad (1.49)$$

takes into account the interaction between both columns. For the sake of convenience we put  $a_3 = a_4 = 0$  and thus  $V_l = 0$ . For  $P$  slightly above  $P_{cl}$ ,

where the Hopf bifurcation occurs with amplitudes  $E_k = E_{k0}$  and  $I_k = I_{k0}$ , we may write

$$E_k - E_{k0} \propto r_k \cos \phi_k, \quad I_k - I_{k0} \propto r_k \sin \phi_k \quad (1.50)$$

and develop everything in Eq. (1.48) in a power series with respect to these quantities up to third order. After a series of nontrivial transformations [87,115] and some wishful thinking we end up with

$$\begin{aligned} \dot{\phi}_1 &= \omega_1 - K_{12} \sin(\phi_1 - \phi_2) \quad , \\ \dot{\phi}_2 &= \omega_2 - K_{21} \sin(\phi_2 - \phi_1) \quad . \end{aligned} \quad (1.51)$$

What is important in Eq. (1.50) is  $\phi_k$  — and not  $r_k$  — as it allows  $E_k - E_{k0}$  and  $I_k - I_{k0}$  to change sign. That's physically why we concentrate on the  $\phi_k$ 's.

Equation (1.51) already exhibits phase locking. To see this, we assume  $\omega_2 > \omega_1$ , put  $\phi = \phi_2 - \phi_1$ , and get from Eq. (1.51),

$$\dot{\phi} = (\omega_2 - \omega_1) - (K_{12} + K_{21}) \sin \phi \quad . \quad (1.52)$$

Assuming phase locking, that is,  $\dot{\phi}_1 = \dot{\phi}_2$ , we find  $\dot{\phi} = 0$  and, thus,  $2K := (K_{12} + K_{21}) \geq (\omega_2 - \omega_1) =: 2K_c$  as a necessary and sufficient condition for this to occur. Throughout what follows we assume symmetry, viz.,  $K_{12} = K_{21} = K$ . In passing we note the alternative interpretation of  $\phi_k$  as a global variable describing a column. It stems from a normal form analysis of a Hopf bifurcation in a small neighborhood of the critical point  $(E_{k0}, I_{k0})$  [53,54]; its validity beyond that is unknown. A representation of Eq. (1.48) which is similar in spirit but more general than Eq. (1.52) has been derived by Ermentrout and Kopell [38].

Instead of just two columns we may also consider several ones. Each orientation and direction specific column is now modeled by an oscillator with phase  $\phi_k$  and eigenfrequency  $\omega_k$ . Generalizing Eq. (1.51) we obtain

$$\dot{\phi}_k = \omega(P_k) - \sum_l J_{kl} \sin(\phi_k - \phi_l) \quad , \quad (1.53)$$

where  $\omega(P_k) = \omega_k$ , if  $k$  is active. Moreover,  $J_{kl} = K$ , if both  $k$  and  $l$  are active. In all other cases,  $\omega(P_k)$  and  $J_{kl}$  vanish. That is the model.

For the moment we assume a globally connected network. We imagine columns in the primary visual cortex as vertices of a two-dimensional square lattice in some bounded domain; each column with index  $k$  is described by a phase  $\phi_k$ . We now want to mimick the experiments of Gray and Singer [51,52]. A clearly visible light bar induces an input  $P > P_{cl}$  in a connected domain  $\Delta$ . It has a specific orientation and thus activates only specific columns, namely, those in  $\Delta$ . By the very definition of the  $J_{kl}$  in Eq. (1.53) we then end up with a Kuramoto model based on  $\Delta$ . If the  $\omega_k$  do not

scatter too much, the system gets phase locked and all the phases rotate with a mean frequency  $\langle \omega \rangle$  defined by Eq. (1.3). Due to the phase shifts (1.17) some come a bit earlier, others a bit later. In experimental terms, one would say that we have a coherent oscillation with some “irregularities.”

Surprisingly, this does resemble the much more realistic *weak locking scenario* presented by Ritz et al. [108]: In addition to the neurons that do not join the club at all, the remaining ones, which lock, spike — to first approximation — with uniform frequency but some come a bit too early, others a bit too late. In this way they simulate a “phase shift.” There is a big difference, however, between the present and the weak locking scenario. There we have single neurons that spike. Here we have columns whose global behavior is described by fictitious phases.

Next we imagine a disconnected light bar  $\Delta = \Delta_1 \cup \Delta_2$  where  $\Delta_1$  and  $\Delta_2$  are disjoint. Let us also suppose that the  $\omega$ 's hardly scatter. Then we have another Kuramoto model on  $\Delta$  and there is still a globally coherent oscillation, though the coupling is not as strong as in the previous case due to the hole between  $\Delta_1$  and  $\Delta_2$ . This is consistent with experiment inasmuch as we still find coherence. It does not agree with experiment in that the temporal coherence between  $\Delta_1$  and  $\Delta_2$  is fixed whereas in experiment it is much weaker than in the previous case. More precisely, the cross-correlation is less pronounced.

Can we break this coherence but leave *each* of the  $\Delta_i$  phase locked? An evident and not so realistic solution consists in putting  $\omega_k = \omega_i$  for all  $k \in \Delta_i$  and  $i = 1, 2$  and taking  $K < K_c = |\omega_1 - \omega_2|/|\Delta|$ . This is a direct consequence of the proposition in the previous subsection and the fact that  $K$  in Eq. (1.53) has not been written  $K/N$  where  $N = |\Delta|$  is the size of the set  $\Delta$ . For fixed  $K$ , the difference between  $\omega_1$  and  $\omega_2$  is bound to be unrealistically large. Schuster and Wagner [116] therefore came up with a more intelligent alternative. Let  $J_{kl} := J_{ki}C_{kl}$  where  $C_{kl} = 1$  with probability  $\gamma \exp(|k - l|/\lambda)$  and  $C_{kl} = 0$  otherwise. Here  $\gamma$  is a suitable prefactor and  $|k - l|$  is the Euclidean distance between  $k$  and  $l$ . If the separation between  $\Delta_1$  and  $\Delta_2$  is large enough, then hardly any bond between the two sets has been left and the coherence between them is gone whereas the short-range couplings dominate in each of the sets separately and, thus, keep each of the  $\Delta_i$  coherently oscillating. The only proviso is a small *difference* between  $\omega_1$  and  $\omega_2$  — as is neatly demonstrated by Figs. 1a and 1b of [116].

Sompolinsky, Golomb, and Kleinfeld [119,120] have devised a model that has a richer structure but is based on the very same ideas as that of Schuster and Wagner. We now start with a collection of  $N$  neurons, a cluster, that all have a common receptive field. A cluster is a kind of *hypercolumn*, in which all the possible orientations are represented for a receptive field area [75,84]. The neurons are labeled by the spatial coordinate  $\mathbf{R}$  of the cluster and by their preferred orientation  $\theta$ , which is assumed to be uniformly distributed

between 0 and  $\pi$  within each cluster. Typically, only a few clusters respond to a stimulus.

The activity of a neuron  $(\mathbf{R}, \theta)$  is described through a *rate* coding:

$$P_{\mathbf{R}}(\theta, t) = V(\theta) [1 + \lambda \cos \phi_{\mathbf{R}}(\theta, t)] , \quad (1.54)$$

where  $V$  is the tuning curve that tells us how well the neuron responds to a stimulus and  $\lambda$  is a suitable prefactor. For example, given the neuron's orientation preference  $\theta$  and the stimulus' orientation  $\theta_0$  we may, and will, take a tentlike shape,

$$V(\theta) = V_0(1 - |\theta_0 - \theta|/2\sigma) \quad \text{for } |\theta_0 - \theta| < 2\sigma, \quad (1.55)$$

and  $V(\theta) = 0$  otherwise; that is to say, including the case of *no* stimulus. Comparing Fig. 1.8(a) with Fig. 1.1(a), we see that Eq. (1.55) is realistic, a typical value for  $\sigma$  being 0.4 rad ( $25^\circ$ ). Equation (1.54) incorporates the idea that  $V$  gives the firing rate averaged over the duration of the stimulus while  $P_{\mathbf{R}}$  takes into account some temporal structure through  $\phi_{\mathbf{R}}(\theta, t)$ . The dependence upon  $t$  is henceforth suppressed whenever there is no fear of confusion.

The dynamics of  $\phi_{\mathbf{R}}(\theta)$  is governed by

$$\dot{\phi}_{\mathbf{R}}(\theta) = \omega - \eta_{\mathbf{R}}(\theta) + \sum_{\mathbf{R}', \theta' (\neq \theta)} J_{\mathbf{R}\theta, \mathbf{R}'\theta'} \sin [\phi_{\mathbf{R}}(\theta) - \phi_{\mathbf{R}'}(\theta')] . \quad (1.56)$$

This is just an ansatz which is in the same category as Eq. (1.53). In Eq. (1.56) we have in addition the white-noise term  $\eta_{\mathbf{R}}(\theta, t)$ , a Gaussian process with mean zero and variance:

$$\langle \eta_{\mathbf{R}}(\theta, t) \eta_{\mathbf{R}'}(\theta', t') \rangle = 2T \delta_{\mathbf{R}, \mathbf{R}'} \delta_{\theta, \theta'} \delta(t - t') \quad (1.57)$$

that will drive the system to an equilibrium. Finally,  $\omega/2\pi$  is about 50 Hz, in agreement with the experimental data [28, 51, 52]. Throughout what follows we put  $\phi_{\mathbf{R}}(\theta, t) = \omega t + \varphi_{\mathbf{R}}(\theta, t)$  and thus eliminate the trivial dynamics — as in Eq. (1.4). It is implicitly understood that  $\omega$  vanishes for all those neurons which are not stimulated.

The interaction is such that only stimulated neurons exchange information,

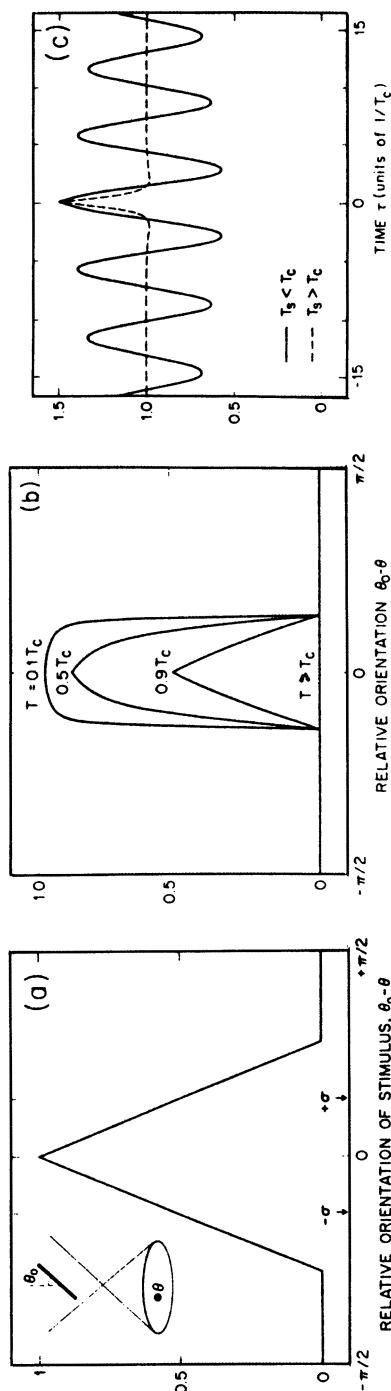
$$J_{\mathbf{R}\theta, \mathbf{R}'\theta'} = V(\theta) W_{\mathbf{R}\mathbf{R}'}(\theta, \theta') V(\theta') . \quad (1.58)$$

We specify  $W$ . Each neuron interacts with other neurons in the *same* cluster through

$$W_{\mathbf{R}\mathbf{R}}(\theta, \theta') = \frac{W_S}{N} \quad (1.59)$$

whatever the preferential direction. The *intercluster* interaction is direction dependent and taken to be

$$W_{\mathbf{R}\mathbf{R}'}(\theta, \theta') = \frac{W_L}{N} F_L(\theta - \theta') \quad (1.60)$$



**Fig. 1.8. Individual and collective neuronal response of the oscillator model (1.5).** (a) Tuning curve telling us how well a neuron with direction preference  $\theta$  responds to a stimulus with orientation  $\theta_0$ ; cf. the experimental response as given by Fig. 1.1a. (b) The dependence of the long-time limit of the average phase  $m_{\bar{\theta}}(\theta)$  as given by Eq. (1.68) upon temperature  $T$ , that is, noise, and the orientation  $\theta_0$  of the stimulus relative to the preferred neuronal orientation  $\theta$ ; here we consider the special case  $\mathbf{m} = (\mathbf{x}, 0)$ . (c) Normalized auto-correlograms including the oscillatory part  $\cos(\omega\tau)$  for neurons that are stimulated by a moving bar with their preferred direction ( $\theta = \theta_0$ ); cf. Eq. (1.74). The solid line is for  $T < T_c$ , the dashed one for  $T > T_c$ . For  $T < T_c$ , the cross-correlograms between optimally and suboptimally (e.g.,  $\theta - \theta_0 \approx 0.35$  rad ( $20^\circ$ )) stimulated neurons in a single cluster turn out to be nearly identical to the auto-correlograms. Taken from [120].

with  $W_L = \varepsilon W_S$  and  $0 \leq \varepsilon \ll 1$ . So the long-range intercluster interactions are much weaker than the short-range intracluster ones, which seems reasonable. Both  $W_L$  and  $W_S$  are positive.

There is a problem though [120, p. 6995]. A realistic  $F_L$  would have a Mexican hat form with a rather broad “negative rim” and, hence, induce a phase shift of  $\pi$  in those regions where  $F_L(\theta - \theta')$  is negative. This is not too surprising since for negative bonds  $XY$  spins prefer being antiparallel, which is equivalent to having a phase shift  $\pi$ . What we really want is that the nonactivated neurons spike much more slowly than the activated ones and *not* — as here — that they fire with the very same rate as their stimulated counterparts. In other words, they should have a smaller  $\omega$  and not a different phase. A convenient way out is the ansatz  $F_L(\theta - \theta') = N\delta_{\theta,\theta'}$ , which we adopt from now on.

Given a stimulus, we have to solve Eq. (1.56) for  $\varphi_{\mathbf{R}}(\theta, t)$ . Due to the white noise (1.57), we cannot specify the  $\varphi$ ’s themselves but we do know [130] their equilibrium distribution:

$$\varrho(\varphi) = C \exp(-\beta \mathcal{H}) , \quad (1.61)$$

where  $C$  is a normalization constant,  $\beta = 1/T$ , and  $\mathcal{H}$  is the Hamiltonian in a coordinate system that rotates with angular velocity  $\omega$ ,

$$\begin{aligned} \mathcal{H} = & \frac{1}{2} \sum_{\mathbf{R}} \sum_{\theta \neq \theta'} V(\theta) \frac{W_S}{N} V(\theta') \cos[\varphi_{\mathbf{R}}(\theta) - \varphi_{\mathbf{R}}(\theta')] \\ & - \frac{1}{2} \sum_{\mathbf{R} \neq \mathbf{R}'} \sum_{\theta, \theta'} V(\theta) \frac{W_L}{N} F_L(\theta - \theta') V(\theta') \cos[\varphi_{\mathbf{R}}(\theta) - \varphi_{\mathbf{R}'}(\theta')]. \end{aligned} \quad (1.62)$$

The eating of the pudding consists of the following procedure. We suppose that the equilibrium Eq. (1.61) is reached so fast that it can be assumed throughout. We then drop the second term from Eq. (1.62) which is proportional to  $\varepsilon \ll 1$ . The first term represents a ferromagnet and, as  $T \rightarrow 0$ , the activated “spins” in each cluster are to be parallel and, thus, the phases get locked. That is, the clusters are decoupled and each of them behaves like a big “spin” with the phases “glued together.” This more or less remains so as long as  $T < T_c$  where  $T_c$  is a critical temperature that will be computed shortly. On the other hand, for  $\varepsilon > 0$  but  $T = 0$  we have a perfect phase locking of the *whole* system since the intercluster term is ferromagnetic (excitatory) as well. Increasing the temperature  $T$  we decrease the cross-correlation between two disjoint regions,  $\Delta_1$  and  $\Delta_2$ , which have been equipped with parallel stimuli. We finish the argument by observing that in a real system we also have noise, and we are done. In passing we note that by filling the hole between  $\Delta_1$  and  $\Delta_2$  we increase the cross-correlation. Furthermore,  $\Delta_1$  and  $\Delta_2$  get decorrelated at a temperature  $T_l \approx \varepsilon T_c$ .

We turn to the details. Equation (1.61) tells us that we have to study the thermal equilibrium states, more precisely, the ergodic components [32],

associated with the Hamiltonian (1.62). If  $\varepsilon = 0$ , the clusters decouple and, dropping  $\mathbf{R}$  for the moment, it suffices to study a single cluster,

$$\mathcal{H} = -\frac{N}{2} W_S \left( N^{-1} \sum_{\theta} V(\theta) \mathbf{S}_{\theta} \right)^2, \quad (1.63)$$

which is the direct equivalent of Eq. (1.9) with  $K := W_S$  and an extra  $V(\theta)$  weighing the spins. We change labels  $\theta \rightarrow \theta_i, 1 \leq i \leq N$ , and rewrite Eq. (1.63):

$$\begin{aligned} -\beta \mathcal{H} &= \frac{\beta N}{2} W_S \left[ \left( N^{-1} \sum_{i=1}^N V(\theta_i) \cos(\varphi_i) \right)^2 + \left( N^{-1} \sum_{i=1}^N V(\theta_i) \sin(\varphi_i) \right)^2 \right] \\ &:= \frac{\beta N}{2} W_S [x_N^2 + y_N^2] \equiv NF(\mathbf{x}). \end{aligned} \quad (1.64)$$

The free energy  $f$  belonging to  $\mathcal{H}$  is

$$-\beta f(\beta) = \lim_{N \rightarrow \infty} N^{-1} \ln \left[ \int d\mathbf{S}^N \exp(-\beta \mathcal{H}) \right] \quad (1.65)$$

with  $dS_i = d\varphi_i / 2\pi$  and  $-\pi \leq \varphi_i \leq \pi$ . It can be evaluated straightforwardly through the large-deviations technique [61, p. 13]. In our context, the cumulant-generating function is defined by

$$\begin{aligned} c(t_1, t_2) &= \lim_{N \rightarrow \infty} N^{-1} \ln \left\{ \prod_{i=1}^N \int \frac{d\varphi_i}{2\pi} \exp[(t_1 \cos \varphi_i + t_2 \sin \varphi_i) V(\theta_i)] \right\} \\ &= \lim_{N \rightarrow \infty} N^{-1} \sum_{i=1}^N \ln \int \frac{d\varphi_i}{2\pi} \exp[(t_1 \cos \varphi_i + t_2 \sin \varphi_i) V(\theta_i)] \\ &= \int_0^\pi \frac{d\theta}{\pi} \ln \left\{ \int \frac{d\varphi}{2\pi} \exp[(t_1 \cos \varphi + t_2 \sin \varphi) V(\theta)] \right\}. \end{aligned} \quad (1.66)$$

In obtaining the final equality we have exploited the strong law of large numbers [89] and the fact that the  $\theta_i$  are uniformly distributed on  $[0, \pi]$ . The free energy  $f(\beta)$ , which provides us with global information on phase transitions and other macroscopic phenomena, is given by [61, p. 13]

$$-\beta f(\beta) = \max_{\mathbf{m}} [-\frac{\beta}{2} W_S(x^2 + y^2) + c(\beta W_S x, \beta W_S y)], \quad (1.67)$$

where  $\mathbf{m} = (x, y)$  satisfies the fixed-point equation  $\mathbf{m} = \nabla c[\nabla F(\mathbf{m})]$ ,

$$\begin{aligned} \begin{pmatrix} x \\ y \end{pmatrix} &= \int \frac{d\theta}{\pi} V(\theta) \frac{\int d\varphi \begin{pmatrix} \cos \varphi \\ \sin \varphi \end{pmatrix} \exp[\beta V(\theta) W_S(x \cos \varphi + y \sin \varphi)]}{\int d\varphi \exp[\beta V(\theta) W_S(x \cos \varphi + y \sin \varphi)]} \\ &\equiv \int \frac{d\theta}{\pi} V(\theta) \begin{pmatrix} m_x \\ m_y \end{pmatrix}. \end{aligned} \quad (1.68)$$

We note that both  $(x,0)$  and  $(0,y)$  are special solutions to Eq. (1.68). In the case  $(x,0)$ , the quantity  $m_x$  has been plotted as a function of  $\theta_0 - \theta$  in Fig. 1.8(b).

To see where  $\mathbf{m}$  comes from, we vary  $\beta$  and look for a bifurcation from  $(0,0)$ . Equation (1.68) is of the form  $G(\mathbf{x}, \beta) = 0$  and we know  $\mathbf{x} = \mathbf{0}$  is always a solution, whatever  $\beta$ . A bifurcation occurs at  $\beta_c$ , if the derivative  $\partial_{\mathbf{x}}G(\mathbf{0}, \beta_c)$  is not invertible. (If it is, we can *uniquely* solve for  $\mathbf{x}(\beta)$  in a suitably small  $\beta$  neighborhood due to the implicit function theorem and, hence, no bifurcation can occur.) A simple computation gives

$$\begin{aligned}\partial_{\mathbf{x}}G(\mathbf{0}, \beta) &= \left[ \beta W_S \int_0^\pi \frac{d\theta}{2\pi} V^2(\theta) - 1 \right] \mathbb{I} \Rightarrow T_c \\ &= W_S \int_0^\pi \frac{d\theta}{2\pi} V^2(\theta) = \frac{2W_S\sigma}{3\pi},\end{aligned}\quad (1.69)$$

that is, we obtain a bifurcation at a double eigenvalue [73]. At  $\beta_c$ , nonzero  $\mathbf{m}$  bifurcate from the trivial solution  $\mathbf{m} = \mathbf{0}$ , which is unstable for  $\beta > \beta_c$ .

Once we have understood the main idea for treating a single cluster we can turn to the case of, say, two clusters,  $\Delta_1$  and  $\Delta_2$ , at positions  $\mathbf{R}$  and  $\mathbf{R}'$ , respectively. For  $T \ll T_c$  ( $\beta \gg \beta_c$ ), the activated spins in each cluster are parallel and frozen — except for those which have  $V(\theta) \approx 0$  — and they remain so for  $\epsilon > 0$  but small. Fixing  $\epsilon$  we can choose the temperature  $T$  so high ( $T > T_l$ ) that the two clusters are decorrelated and still low enough ( $T_l \approx \epsilon T_c$ ) so as to guarantee that the clusters themselves remain frozen. We now spell out the formalities. The Hamiltonian reads in self-explanatory notation:

$$\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2 - \sum_{\theta, \theta'} \frac{W_L}{N} V(\theta) F_L(\theta - \theta') V(\theta') \cos[\varphi_1(\theta) - \varphi_2(\theta')]$$

which for the specific choice  $F_L(\theta - \theta') = N\delta_{\theta, \theta'}$  may be rewritten

$$\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2 - \sum_{i=1}^N W_S V(\theta_i) \cos[\varphi_1(\theta_i) - \varphi_2(\theta_i)]. \quad (1.70)$$

All three terms in Eq. (1.70) are proportional to  $N$ , as should be the case in view of Eq. (1.65).

Also in the case of two (or any finite number of) clusters the free energy can be evaluated exactly. The computation is straightforward and left to the reader. We only note that Eq. (1.70) attains a more physical appearance once it is reformulated in terms of  $XY$  spins,

$$\mathcal{H} = -NW_S \left[ \frac{1}{2} \sum_{\alpha=1}^2 \left( N^{-1} \sum_{i=1}^N V(\theta_i) \mathbf{S}_{i,\alpha} \right)^2 + \frac{\epsilon}{N} \left( \sum_{i=1}^N V^2(\theta_i) \mathbf{S}_{i,1} \cdot \mathbf{S}_{i,2} \right)^2 \right]. \quad (1.71)$$

It makes explicit that in the limit  $T \rightarrow 0$  all spins are parallel and, thus, all phases are perfectly locked as this configuration has minimal energy. Equation (1.71) also makes explicit that  $V(\theta_i)$  operates as a weight, that vanishes if the neuron is not stimulated, and attains its maximal value 1, if the stimulus orientation agrees with that of the neuron. For the sake of definiteness we take an ergodic component that has its magnetization parallel to the  $x$  axis; this corresponds to  $\mathbf{m} = (x, 0)$ .

The advantage of the above setup is that the computation of auto- and cross-correlation functions is clear, though cumbersome, sailing. Due to (1.54) the auto-correlogram  $\langle P_{\mathbf{R}}(\theta, t)P_{\mathbf{R}}(\theta, t + \tau) \rangle_t$ , an average over time, is  $V^2(\theta)$  times

$$(1 + \lambda \mathbf{S}_t \cdot \mathbf{e}_1 + \lambda \mathbf{S}_{t+\tau} \cdot \mathbf{e}_1 + \lambda^2 (\mathbf{S}_t \cdot \mathbf{e}_1)(\mathbf{e}_1 \cdot \mathbf{S}_{t+\tau}))_t. \quad (1.72)$$

Here  $\mathbf{e}_1 = (1, 0)$  and we have used Eq. (1.54). Since  $\phi_{\mathbf{R}}(\theta, t) = \omega t + \varphi_{\mathbf{R}}(\theta, t)$  we imagine ourselves in a coordinate system that rotates with angular velocity  $\omega$ . In this frame, we have the gradient dynamics  $\dot{\varphi} = -\nabla \mathcal{H} + \xi$  with  $\xi$  representing white noise with variance (1.57) and the system thus converges to the stationary distribution  $C \exp(-\beta \mathcal{H})$  — more precisely, to the distribution restricted to a specific ergodic component. Its expectation value is denoted by  $\langle \dots \rangle_\beta$ . In the original coordinate system we obtain a stochastic behavior in this component superimposed upon a uniform rotation with angular velocity  $\omega$ . The corresponding expectation value is written  $\langle \dots \rangle_{\beta, \omega}$ . Because of the uniform rotation the terms with prefactor  $\lambda$  in Eq. (1.72) vanish and we are left with

$$1 + \frac{1}{2} \lambda^2 \langle (\mathbf{S}_t \cdot \mathbf{e}_1)(\mathbf{e}_1 \cdot \mathbf{S}_{t+\tau}) + (\mathbf{S}_t \cdot \mathbf{e}_2)(\mathbf{e}_2 \cdot \mathbf{S}_{t+\tau}) \rangle_{\beta, \omega} = 1 + \frac{1}{2} \lambda^2 \langle \mathbf{S}_0 \cdot \mathbf{S}_\tau \rangle_{\beta, \omega}, \quad (1.73)$$

where  $\mathbf{S}_\tau$  is the state at time  $\tau$  that has evolved from  $\mathbf{S}_0$ . We write out what  $\langle \mathbf{S}_0 \cdot \mathbf{S}_\tau \rangle_{\beta, \omega}$  means,

$$\begin{aligned} \langle \mathbf{S}_0 \cdot \mathbf{S}_\tau \rangle_{\beta, \omega} &= \langle \cos [\varphi(\theta, 0) - \varphi(\theta, \tau) - \omega \tau] \rangle_\beta \\ &= \cos(\omega \tau) \langle \cos [\varphi(\theta, 0) - \varphi(\theta, \tau)] \rangle_\beta \\ &= \cos(\omega \tau) \langle \mathbf{S} \cdot \mathbf{S}_\tau \rangle_\beta. \end{aligned} \quad (1.74)$$

Here we have exploited the fact that the ergodic component is such that  $\mathbf{m} = (x, 0)$  so that  $\langle \sin [\varphi(\theta, 0) - \varphi(\theta, \tau)] \rangle_\beta$  vanishes. Furthermore, the stochastic time evolution transforms  $\mathbf{S}_0 = \mathbf{S}$  into  $\mathbf{S}_\tau$ . Since it is mixing [130], we can evaluate Eq. (1.74) in the limit  $\tau \rightarrow \infty$  so as to find

$$\lim_{\tau \rightarrow \infty} \langle \mathbf{S} \cdot \mathbf{S}_\tau \rangle_\beta = \langle \mathbf{S} \rangle_\beta^2. \quad (1.75)$$

For a plot of Eq. (1.74) as a function of  $\tau$  we refer to Fig. 1.8(c). The right-hand side of Eq. (1.75), viz.,  $\langle \mathbf{S} \rangle_\beta$ , is easy to obtain. For a single cluster,  $\langle \mathbf{S} \rangle_\beta = \mathbf{m} = (x, 0)$  where  $x$  follows from Eq. (1.68),

$$x = \int_0^\pi \frac{d\theta}{\pi} V(\theta) I_1 [\beta V(\theta) W_S x] / I_0 [\beta V(\theta) W_S x] \quad (1.76)$$

and  $I_n$  is the modified Bessel function of order  $n$  (9, Eq. (9.6.19)),

$$I_n(z) = \int_0^\pi \frac{d\vartheta}{\pi} e^{z \cos \vartheta} \cos(n\vartheta).$$

The computation of a cross-correlation function proceeds in a similar vein. For two clusters  $\mathbf{R}$  and  $\mathbf{R}'$  we obtain

$$\langle P_{\mathbf{R}}(\theta, t) P_{\mathbf{R}'}(\theta', t + \tau) \rangle_t = V(\theta)V(\theta')[1 + \frac{1}{2}\lambda^2 \langle \mathbf{S}(\mathbf{R}) \cdot \mathbf{S}_\tau(\mathbf{R}') \rangle_\beta]. \quad (1.77)$$

Here too the dynamics on the right is that of the  $\varphi$ 's. Furthermore,  $\mathbf{S}$  has preferential direction  $\theta$  and  $\mathbf{S}_\tau$  prefers  $\theta'$ . For an approximate calculation of Eqs. (1.74) and (1.77) with  $\tau$  finite, we refer to Sompolinsky et al. [120].

Stepping back for an overview, we have to face the question: How sensible is this type of model? First of all, it uses a *rate* coding to describe a phenomenon that has a temporal resolution much finer than the time window needed to determine the rate. This problem can be avoided, if we interpret the  $P_{\mathbf{R}}(\theta, t)$  in Eq. (1.54) as a spatially averaged activity comparable with a local field potential (cf. Singer [118, Fig. 1]) instead of a time-averaged firing rate of a single neuron. If we do so, however, we end up with pure phenomenology.

A second drawback turns up in pattern segmentation. Suppose we have a composite pattern, say a black and white picture, which is built up out of  $n$  parts, collections  $\Delta_\mu$  of black pixels (“active” columns) with  $1 \leq \mu \leq n$ ; the rest is white. For the sake of simplicity we assume for the moment that the sizes  $|\Delta_\mu|$  and the eigenfrequencies  $\omega_i$  are all equal. Formally, the fractions  $p_\mu = |\Delta_\mu|/N$  do not depend on  $\mu$ . Neither does  $\omega_i$  depend on  $i$ . Furthermore, the  $\Delta_\mu$  are taken to be disjoint. The question is: How can we separate them?

Since the “neurons” are described by phases and a phase lives in the interval  $[-\pi, \pi]$ , the best we can do is (i) producing a sublattice phase locking on each of the  $\Delta_\mu$ , which means that there all the  $XY$  spins are parallel and adding up to a sublattice magnetization  $\mathbf{m}_\mu = |\Delta_\mu|^{-1} \sum_{i \in \Delta_\mu} \mathbf{S}_i$  of unit length, and (ii) inducing a phase shift of  $2\pi/n$  between two subsequent  $\mathbf{m}_\mu$ . Realizing (i) is easy. Let  $J_{ij} = (K/N) \sum_\mu n_i^\mu n_j^\mu$  where  $n_i^\mu = 1$ , if  $i \in \Delta_\mu$ , and zero otherwise. This is an interaction which either vanishes or is ferromagnetic. The corresponding Hamiltonian, our Lyapunov function, is taken to be

$$\mathcal{H} = -\frac{1}{2} \sum_{i,j} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j = -\frac{KN}{2} \sum_\mu (N^{-1} \sum_{i \in \Delta_\mu} \mathbf{S}_i)^2 = -\frac{KN}{2} \sum_\mu p_\mu^2 \mathbf{m}_\mu^2.$$

The  $\Delta_\mu$  have been decoupled and each of them is phase locked. Adding a bit of temperature and, lo and behold, no correlation between the  $\mathbf{m}_\mu$  is left. That is no good since we wanted a starlike configuration of the vectors

$\mathbf{m}_\mu$  in the plane. To this end we aim at a Hamiltonian of the form

$$\mathcal{H} = -\frac{KN}{2} \sum_\mu p_\mu^2 \mathbf{m}_\mu^2 + \frac{\gamma}{2} \left( \sum_\mu \mathbf{m}_\mu \right)^2 \quad (1.78)$$

with  $\gamma > 0$  chosen suitably. The first term on the right is as before. The ground state of  $\mathcal{H}$  has all spins on each sublattice  $\Delta_\mu$  parallel and thus  $\mathbf{m}_\mu^2 = 1$ , and at the same time  $\sum_\mu \mathbf{m}_\mu = 0$ , that is, a starlike configuration, so as to minimize the second term. The latter is anti-ferromagnetic and corresponds to  $J_{ij}^{AF} = -(\gamma/p^2) \mathbb{1}_i \mathbb{1}_j$  where  $p = p_\mu$  and  $\mathbb{1}_i = 1$ , if  $i \in \cup_\mu \Delta_\mu$ , and zero otherwise.

How good is this description? It is of course translationally invariant but hinges on the  $\Delta_\mu$  being disjoint. For visual impressions this seems reasonable. We had also assumed that all  $\Delta_\mu$  had about equal size. If we drop this assumption and still want to end up with the same antiferromagnetic term in Eq. (1.78), we have to take

$$J_{ij}^{AF} = -\gamma \sum_{\mu,\nu} (n_i^\mu / p_\mu) (n_j^\nu / p_\nu).$$

Plainly, we can do so but it is a complete mystery — at least to us — how a synapse, that gathers local information from a pre- and postsynaptic neuron only, can collect *global* information à la  $p_\mu$  concerning pattern  $\mu$ .

Finally, does this setup really *separate* the  $\Delta_\mu$ ? The phase difference of the  $\varphi_i$  is at best  $2\pi/n$ . If the  $\omega_i$  scatter, we get phase shifts (cf. Eq. (1.17)) which soon bridge the gap  $2\pi/n$ ; the best deal, then, is for  $n = 2$ . A similar, additional, fading of the phase differences occurs, if  $\beta$  is not high enough or, equivalently, if the noise is not low enough.

Yet another drawback exists. It is more subtle, though. Nowhere does the amplitude of the stimulus appear explicitly. A pure  $XY$  model does not provide enough degrees of freedom to account for the stimulus strength — except for changing the eigenfrequencies  $\omega$  by hand. The trick therefore is to put  $V(\theta)$  in the interaction (cf. Eq. (1.58)). This is pretty weird since the hardware (wetware), for example, the connectivity, is given and should not depend on the state of the receiving neurons.

Moreover, the dynamics should be derived — not postulated. In particular, the ansatz of both proposals is that all active neurons “oscillate” coherently with eigenfrequency  $\omega$ . The  $\omega$  is not a consequence of the hard- and software (synaptic efficacies) but a priori given. Also an interaction of the form  $\sin(\phi_i - \phi_j)$  is hard to obtain from neuronal data. If  $\omega$  is the same for all neurons, phase locking for low enough temperatures (i.e., noise) is a trivial consequence of the model; cf. Eqs. (1.7)–(1.9). Though in the present models the frequencies either hardly scatter or do not scatter at all, it is a nice property of the Kuramoto model that phase locking holds as long as the frequencies do not scatter “too” much. In other words, locking is robust.

Because of all these serious drawbacks of phase oscillators it is time to study another coding, namely, that of *spiking* neurons.

## 1.3 Spiking Neurons

In the preceding sections we have shown how oscillator models can be used to describe neuronal activity. Real neurons are, of course, not mathematical oscillators coupled by some interaction term, but *spiking* units that generate action potentials. The spikes are transmitted along the axon and across the synapses to other neurons where they evoke some postsynaptic response. A network model with spiking neurons should therefore include three aspects. First, it should incorporate a mechanism that *generates spikes* with an appropriate interval distribution. Second, it should describe axonal and synaptic signal *transmission* and, finally, it should account for the *integration* process on the dendritic tree. We start by reviewing some common models of a *single* spiking neuron. If spiking neurons are to be connected in a large network, the computational complexity allowed for the description of a single neuron is limited. Nevertheless it is possible to construct a network of spiking neurons which takes into account the above requirements and yet is simple enough to allow analytical solution as well as large-scale computer simulation. The presentation of such a model ansatz forms the main part of this section.

### 1.3.1 HODGKIN–HUXLEY MODEL

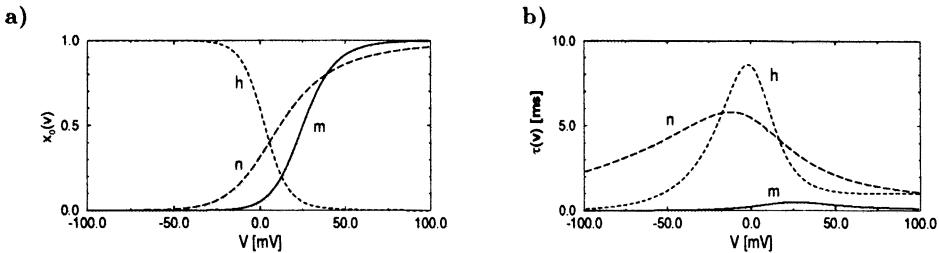
The classic description of neuronal spiking dates back to Hodgkin and Huxley [68] who have summarized their extensive experimental studies on the giant axon of the squid in four differential equations. The first describes the conservation of electric charge on a piece of membrane of capacitance  $C$  under the influence of a current and a voltage  $V$ :

$$C \frac{dV}{dt} = \sum_i I_i + I, \quad (1.79)$$

where  $I$  is some external driving current and  $\sum_i I_i$  is the sum of the ionic currents through the cell membrane,

$$\sum_i I_i = -g_{Na}m^3h(V - V_{Na}) - g_Kn^4(V - V_K) - g_L(V - V_L). \quad (1.80)$$

The constants  $V_{Na} = 115$  mV,  $V_K = -12$  mV, and  $V_L = 10.6$  mV are the equilibrium potentials of the three components sodium, potassium, and “leakage,” the  $g$ ’s are parameters of the respective ion conductance which



**Fig. 1.9.** Equilibrium function (a) and time constant (b) of the Hodgkin–Huxley model.

also depends on the variables  $m$ ,  $n$ , and  $h$  determined by the differential equations

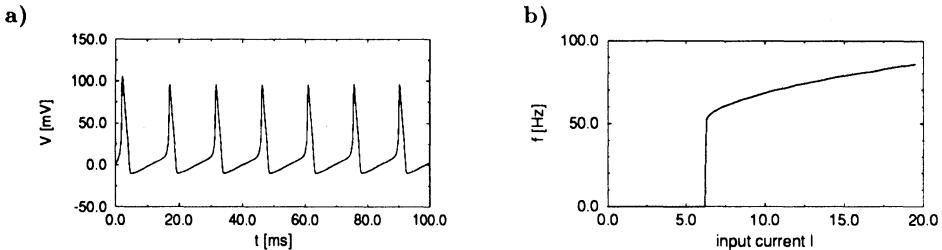
$$\begin{aligned}\dot{m} &= \alpha_m(V)(1-m) - \beta_m(V)m, \\ \dot{n} &= \alpha_n(V)(1-n) - \beta_n(V)n, \\ \dot{h} &= \alpha_h(V)(1-h) - \beta_h(V)h\end{aligned}\quad (1.81)$$

with  $\dot{m} = dm/dt$ , and so on. The  $\alpha$  and  $\beta$  are given functions of  $V$  that have been adjusted empirically to fit the data of the giant axon of the squid. In appropriate units the prefactors are  $g_{\text{Na}} = 120$ ,  $g_{\text{K}} = 36$ , and  $g_{\text{L}} = 0.3$ .

To get an idea of how the model works it is more convenient to write Eq. (1.81) in the form

$$\dot{x}(V) = -\tau^{-1}(V)[x(V) - x_0(V)], \quad (1.82)$$

where  $x$  stands for  $m$ ,  $n$ , and  $h$ . The equilibrium value  $x_0(V)$  and the time constant  $\tau(V)$  are given by the transformation  $x_0(V) = \alpha_x(V)/[\alpha_x(V) + \beta_x(V)]$  and  $\tau^{-1}(V) = [\alpha_x(V) + \beta_x(V)]$ . Using the parameters given by Hodgkin and Huxley [68], we have plotted the functions  $x_0(V)$  and  $\tau(V)$  in Fig. 1.9. The function  $x_0(V)$  has a sigmoidal shape with some “threshold”  $\theta_x$  and  $\tau$  is significant only in a limited range around  $\theta_x$ . Note that  $m$  and  $n$  increase with  $V$  whereas  $h$  decreases. Thus, if the membrane voltage is raised by some external input, the ion conductance of sodium (Na) increases due to increasing  $m$  and sodium can flow into the cell. This raises the membrane potential further and further. At high values of  $V$  the sodium conductance decreases due to the factor  $h$ . Also potassium (K) efflux sets in and lowers the potential. Due to the longer time constant  $\tau_n(V)$ , the potassium concentration slowly reaches its equilibrium potential. The overall effect is a short action potential as shown in Fig. 1.10(a). Numerical integration of Eqs. (1.79)–(1.81) shows that a constant input  $I$  larger than a critical value  $I_\theta$  results in a regular spike train. If the number of spikes during a large interval  $T$  is counted and divided by  $T$ , a firing rate results. The firing rate as a function of the input  $I$  is called the *gain function* of the Hodgkin–Huxley model. It is shown in Fig. 1.10(b).

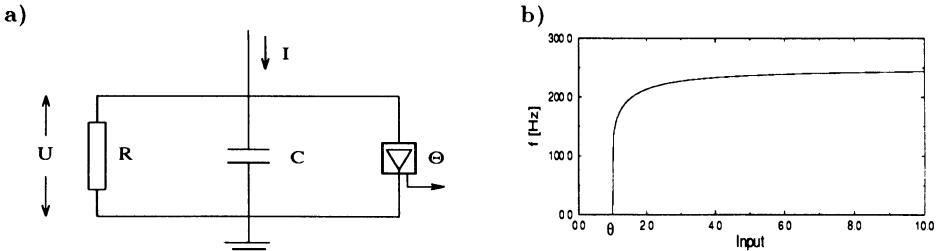


**Fig. 1.10.** Spike train (a) and gain function (b) of the Hodgkin–Huxley model.

Using the above equations and an appropriate set of parameters, Hodgkin and Huxley could describe an enormous number of data from experiments on the giant axon of the squid. Due to the success of the model in this special system there have subsequently been several attempts to generalize it in order to describe other experimental situations as well (for a review see, e.g., [74]). Whereas the model had originally been designed to describe the form and temporal change of an action potential during *axonal* transmission, a model ansatz completely analogous to Eqs. (1.79) to (1.81) has been used recently to describe spike generation at the *soma* of the neuron [15, 21, 30, 105, 126, 132, 134]. The main difference is that additional ion channels have to be included, in particular those that account for  $\text{Ca}^{++}$  and the slow components of the potassium current. For each type of ion channel  $i$  a current  $I_i = g_i m_i^n (V - V_i)$  is added where the conductance parameters  $g_i$  and  $n_i$  and the function  $x_i(V, t)$  are adjusted to fit experimental data.

Signal transmission and integration on the dendrites has traditionally been described by passive electric processes, most prominently in the dendritic cable theory of Rall; see, for example, [104]. Such an approach can account even for some nonlinear dendritic phenomena [2, 15, 105]. A different approach to dendritic integration is based again on Hodgkin–Huxley type equations. The only change with respect to the theory of *axonal* signal transmission is that a different set of parameters for the conductivity of various channels are to be used on the dendrites. If several spatial compartments of the neuron are put together, a set of some 20 coupled nonlinear differential equations results (see, e.g., [127]). Numeric solution of the system shows good agreement with experiments. This is an important indication that a Hodgkin–Huxley type analysis is a useful tool in understanding the properties of single neurons. It is however obvious that such an approach is too detailed, if the description of a large network of neurons is attempted.

There have been several attempts to reduce the complexity of the Hodgkin–Huxley equations [40, 98]. Abbott and Kepler [1] have shown that a separation of time scales in the model of Hodgkin and Huxley is possible. Their systematic reduction yields two different simplified models, viz., on the one hand a binary model with a continuous variable  $u$  and a two-state



**Fig. 1.11.** Schematic diagram (a) and gain function (b) of the integrate-and-fire model.

variable  $S$  and, on the other hand, an integrate-and-fire model with a single variable  $h$ . The latter will be considered in the following subsection.

### 1.3.2 INTEGRATE-AND-FIRE MODEL

The integrate-and-fire model is a simple mechanism of spike generation and dendritic integration. A description of the form of an action potential and the axonal transmission process is not attempted. In the model, a neuron is reduced to a capacity  $C$  that is parallel to a resistance  $R$ ; cf. Fig. 1.11(a). The capacity is charged by a current  $I$  which may be some external source or else the sum of all postsynaptic currents. The voltage  $U$  changes according to

$$\dot{U} = C^{-1}I - U/RC. \quad (1.83)$$

If  $U$  reaches a threshold  $\theta$ , an action potential is generated and the voltage is reset to 0. For constant driving current  $I$ , Eq. (1.83) can be integrated and yields

$$U(t) = IR[1 - \exp(-t/RC)]. \quad (1.84)$$

The condition  $V(t_\theta) = \theta$  provides us with the loading time  $t_\theta = RC \log [I/(I - I_\theta)]$  for  $I > I_\theta = \theta/R$ . The total interval between two subsequent spikes can be calculated from the loading time and the absolute refractory period  $\gamma^{\text{refr}}$ . The *firing rate* of the model neuron is *defined* as the inverse of the interspike interval. It is given by

$$f(I) = \left[ \gamma^{\text{refr}} + RC \log \frac{I}{I - I_\theta} \right]^{-1}. \quad (1.85)$$

The firing rate as a function of the driving current  $I$  is shown in Fig. 1.11(b).

Stein [122] and Jack et al. [74] give a thorough discussion of the properties of the integrate-and-fire model for different sets of parameters. If the constant driving current  $I$  is replaced by a noisy input (generated, e.g., by stochastically arriving synaptic potentials), the sharp threshold at  $I = I_\theta$  disappears and the mean firing rate depends smoothly on  $I$ .

To summarize, the integrate-and-fire model gives a simple picture of neuronal spike generation which is primarily based on the *loading time* of the cell. Indeed, Abbott and Kepler [1] have shown that the more complex Hodgkin–Huxley model can be reduced to an integrate-and-fire model in the limit that the membrane loading time is the dominant time scale of the neuronal dynamics. Due to its computational simplicity, several integrate-and-fire neurons can be connected so as to form a large neural network. The stationary states of the fully connected system can be described in a straightforward approach [13]. The dynamics of such a network is more complicated, but it has recently been analyzed by several groups, using a mean-field approach [3,14,127,128]. Before turning to larger networks, however, we will discuss another model of a single neuron which is at the same level of description as the integrate-and-fire neuron, but more suitable to network problems.

### 1.3.3 SPIKE RESPONSE MODEL

The integrate-and-fire model concentrates on the membrane loading time as the dominant factor during spike generation. The spike response model (SRM) takes a different point of view. Here the reduced excitability (or increased threshold) of a neuron after emission of a first spike is seen as the limiting factor during spike generation. The effects of dendritic integration, on the other hand, are included into the SRM by an explicit description of a typical postsynaptic potential.

The SRM [43,44,46–48] allows one to describe both spike reception and spike generation in a single formalism, viz., by two response functions. First, incoming spikes evoke a postsynaptic potential modeled by a response kernel  $\epsilon(s)$ . Second, spike emission induces refractoriness which is described as an internal response  $\eta^{\text{refr}}(s)$ . We postpone a detailed description of spike reception to later and start with the problem of spike generation.

#### Spike generation

The state of the model neuron is described by an internal variable  $h$  which can, but need not [109], be considered as the membrane voltage. If, at time  $t^f$  the membrane voltage  $h$  reaches the threshold  $\theta$ , an action potential is generated which is transmitted along the axon to other neurons. At the same time a negative contribution  $\eta^{\text{refr}}$  is added to the variable  $h$  which describes the reduced excitability of a neuron. After several spikes at times  $t^1, t^2, \dots$  the total voltage is given by

$$h(t) = h^{\text{syn}}(t) + h^{\text{refr}}(t) \quad (1.86)$$

with

$$h^{\text{refr}}(t) = \sum_{f=1}^F \eta^{\text{refr}}(t - t_i^f), \quad (1.87)$$

where  $t^f$  with  $1 \leq f \leq F$  are the most recent firing times of the model neuron;  $h^{\text{syn}}$  describes the input from synapses on the dendritic tree. The upper limit  $F$  is a useful parameter that allows one to switch between different types of model neuron. If  $F = 1$ , only the last spike is relevant and the mathematics of the model is particularly simple. If  $F > 1$ , then several spikes contribute to the refractory voltage  $h^{\text{refr}}$  and the model exhibits adaptation effects.

The *refractory function*  $\eta^{\text{refr}}$  is one of the two response functions in the SRM. It describes the after-potential which follows each spike of a neuron. From a different point of view, we can take  $-\eta^{\text{refr}}$  to describe the reduced excitability immediately after a spike or, equivalently, an increased dynamic threshold [71]. The excitability is determined by the time constants which govern the opening and closing of various ion channels in the cell membrane. In principle, it could be calculated from the microscopic dynamics of the  $\text{Na}^+$ ,  $\text{K}^+$ , and  $\text{Ca}^{++}$  channels. After a few ms the influence of the  $\text{Na}^+$  channels and some fast  $\text{K}^+$  channels disappears and the after-potential is dominated by the slow dynamics of the  $\text{Ca}^{++}$  and  $\text{Ca}^{++}$ -dependent potassium ( $\text{K}$ ) channels. This leads to a prolonged after-*hyperpolarizing* potential (AHP) which may last for a 100 ms and more [23,90]. In some neurons an additional period of *depolarization* (depolarizing after-potential, for short DAP) has been found which takes place on a medium time scale of 10–50 ms. In the following we do not intend to calculate the after-potential from some model of the microscopic channel dynamics. Instead we will guess a reasonable refractory field and require that the resulting spike trains look realistic. For example, we will consider the refractory function

$$\eta^{\text{refr}}(s) = \begin{cases} -\infty & \text{for } s \leq \gamma^{\text{refr}} \\ \eta_0/(s - \gamma^{\text{refr}}) & \text{for } s > \gamma^{\text{refr}} \end{cases} \quad (1.88)$$

with an absolute refractory time  $\gamma^{\text{refr}} = 4$  ms and a prefactor  $\eta_0 = 1$  ms. This is the refractory function shown in Fig. 1.12(a).

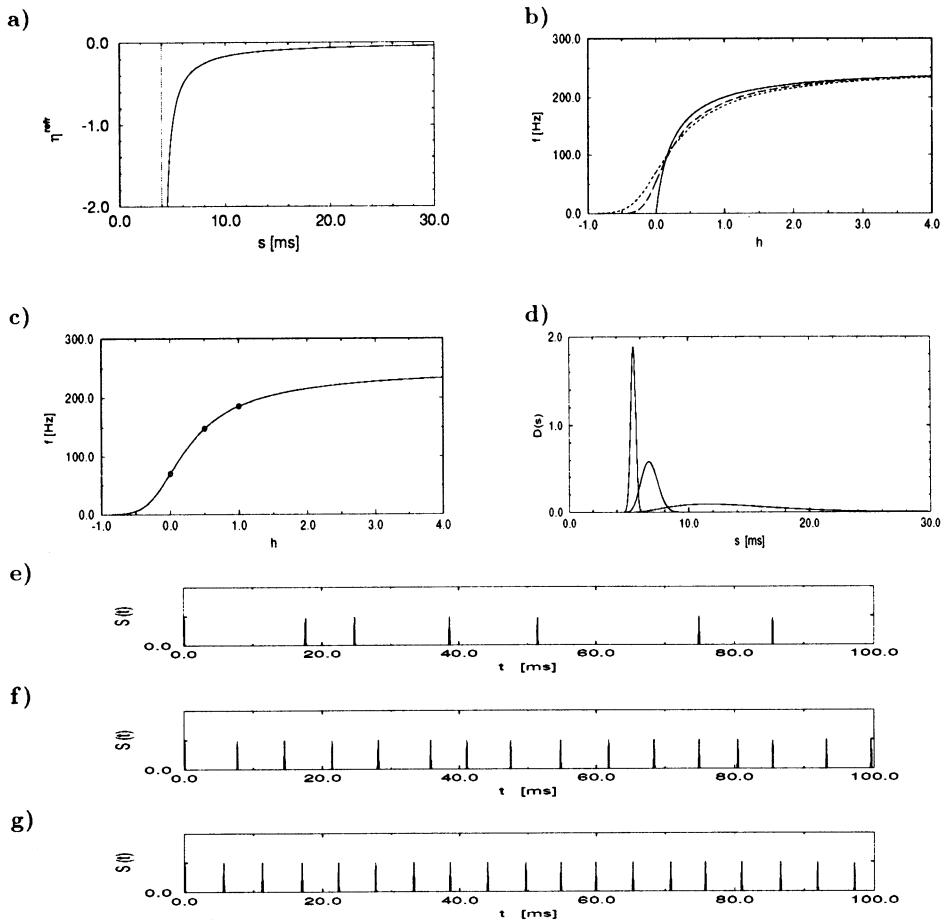
So far the model had no noise. To include noise into the system we introduce the probability  $P_F$  of firing during an infinitesimal time  $\delta t$ :

$$P_F(h; \delta t) = \tau^{-1}(h) \delta t. \quad (1.89)$$

In a realistic model, the response time  $\tau(h)$  should be long if  $h < \theta$  (to prevent firing) and should vanish if  $h \gg \theta$ . In analogy to a chemical reaction constant we make the ansatz

$$\tau(h) = \tau_0 \exp[-\beta(h - \theta)], \quad (1.90)$$

where  $\tau_0$  is the response time at threshold. The parameter  $\beta$  determines the amount of noise in the system. For  $\beta \rightarrow \infty$  we find a strict threshold condition. The neuron does not fire, if  $h < \theta$  ( $\tau \rightarrow \infty$ ), and fires immediately, if  $h > \theta$ . Typical spike trains of the model with finite noise ( $\beta < \infty$ ) are shown in Figs. 1.12(e)–1.12(g).



**Fig. 1.12.** Standard model neuron. (a) Refractory function  $\eta^{\text{refr}}(s)$ . It is taken to  $-\infty$  during the absolute refractory period ( $0 \leq s \leq 4$  ms). (b) Gain function with noise ( $\beta = 8$  dotted,  $\beta = 15$  dashed) and noiseless (solid). (c) Gain function for  $\beta = 8$ . The marked points (open circles) indicate the values at which the interval distribution (d) and typical spike trains (e)–(g) have been calculated. (d) Interval distribution with  $\beta = 8$  and constant stimulus ( $h^{\text{ext}} = 0$ , relatively flat;  $h^{\text{ext}} = 0.5$ , middle;  $h^{\text{ext}} = 1$ , narrow distribution). (e)–(g) Typical spike trains with  $\beta = 8$ : (e)  $h^{\text{ext}} = 0$ ; (f)  $h^{\text{ext}} = 0.5$ ; (g)  $h^{\text{ext}} = 1.0$ . All results are for a nonadaptive neuron ( $F = 1$ ).

Equations (1.86)–(1.90) define the process of spike generation in the SRM. We may now ask what the *spiking statistics* of a SRM neuron with an arbitrary refractory field  $\eta^{\text{refr}}$  and a given amount of noise  $\beta$  looks like.

### Spiking statistics

Let us consider a neuron driven by a synaptic potential  $h^{\text{syn}}(t)$ . A first spike may have occurred at time  $t_1^f$ . Thus the total field for  $t > t_1^f$  is

$$h(t) = h^{\text{syn}}(t) + \eta^{\text{refr}}(t - t_1^f). \quad (1.91)$$

For the moment we assume that the system evolves in short but discrete time steps of length  $\delta t$ . Given the first spike we would like to calculate the *conditional probability*  $\hat{P}_F^{(2)}(t|t_1^f)$  of finding a second spike after  $n$  time steps. The second spike occurring after  $n$  steps, the neuron must stay quiescent for  $n - 1$  time steps and fire thereafter. With the definition of the firing probability  $P_F(t; \delta t)$  we find

$$\hat{P}_F^{(2)}(t|t_1^f) = \prod_{k=1}^{n-1} [1 - P_F(h_k; \delta t)] P_F(h_n; \delta t), \quad (1.92)$$

where  $h_k = h(t_k)$  is the membrane potential at time  $t_k$ . We assume that the time step  $\delta t$  is short enough in order to guarantee  $P_F(h_k; \delta t) \ll 1$ . In this case the product on the right-hand side of Eq. (1.92) can be written as  $\prod_{k=1}^{n-1} [1 - P_F(h_k; \delta t)] = \exp\{-\sum_{k=1}^{n-1} P_F(h_k; \delta t)\}$ . Taking the limit  $\delta t \rightarrow 0$  while keeping  $t - t_1^f$  fixed we find

$$\hat{P}_F^{(2)}(t|t_1^f) = \tau^{-1}[h(t)] \exp\left\{-\int_{t_1^f}^t \tau^{-1}[h(s)] ds\right\} \delta t. \quad (1.93)$$

Equation (1.93) is the conditional probability of finding a spike in a small interval  $\delta t$  around  $t$  given a *first* spike at  $t_1^f$ .

For pedagogical reasons, Eq. (1.93) has been derived from a discrete time-step approach and the limit  $\delta t \rightarrow 0$  has been taken only at the end of the calculation. For analytical considerations it is, however, more convenient to study the model in continuous time. Before doing so we would like to calculate the probability that a neuron stays quiescent during a time  $s$  given a first spike at  $t_1^f$ . To simplify the argument we consider an ensemble of  $N(t_1^f)$  equivalent and independent neurons which have all fired at time  $t_1^f$  and are subject to the same synaptic potential  $h^{\text{syn}}(t)$ . The number  $N(t)$  of neurons which have “survived” a time  $t$  without firing decays according to

$$\dot{N} = -\tau^{-1}[h(t)] N \quad (1.94)$$

since neurons spike with the rate  $\tau^{-1}[h(t)]$ . Integration yields the generalized “survivor” function [102]

$$S(t|t_1^f) \equiv N(t)/N(t_1^f) = \exp\left\{-\int_{t_1^f}^t \tau^{-1}[h(s)] ds\right\}. \quad (1.95)$$

In order to get the conditional probability of firing between  $t$  and  $t + \delta t$  we have to multiply Eq. (1.95) with  $P_F(t; \delta t)$  which yields

$$\hat{P}_F^{(2)}(t|t_1^f) = \tau^{-1}[(h(t)]S(t|t_1^f) \quad (1.96)$$

which is identical to Eq. (1.93). In a continuous time analysis it is more convenient to consider the probability density  $P_F^{(2)}(t|t_1^f)$  defined by

$$P_F^{(2)}(t|t_1^f) = \lim_{\delta t \rightarrow 0} \frac{\hat{P}_F^{(2)}(t|t_1^f)}{\delta t}. \quad (1.97)$$

We note the identity

$$P_F^{(2)}(t|t_1^f) = -\frac{d}{dt}S(t|t_1^f) \quad (1.98)$$

which will be useful below.

For nonadaptive neurons ( $F = 1$ ) each spike leads to a reset of the refractory potential and only the last spike counts. This is called a renewal process [24,31,102]. In this case  $P_F^{(2)}(t|t_1^f)$  is sufficient to describe the spike statistics of the model neuron. For  $F > 1$  the probability of finding an action potential depends on  $F$  earlier spikes. Let us consider a neuron that has fired at  $t_1^f, t_2^f, \dots, t_F^f$ . The conditional probability  $P_F^{(F)}(t|t_1^f, t_2^f, \dots, t_F^f)$  of finding the next spike in an interval  $\delta t$  around  $t$  is given analogously to Eqs. (1.95) and (1.96) by

$$P_F^{(F)}(t|t_1^f, t_2^f, \dots, t_F^f) = \tau^{-1}[h(t)] \exp\left\{-\int_{t_1^f}^t \tau^{-1}[h(s)]ds\right\}, \quad (1.99)$$

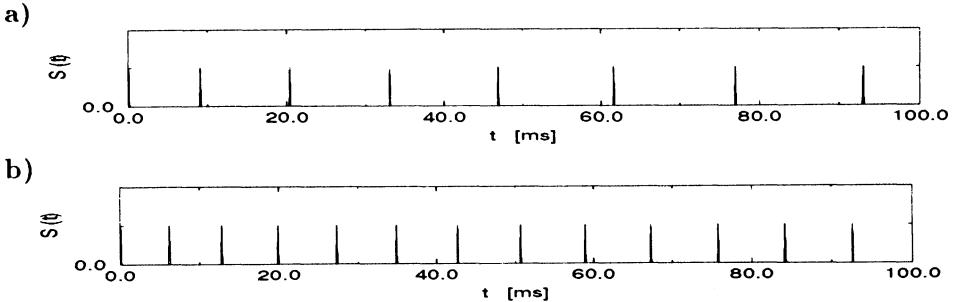
where  $h(t)$  is now given by

$$h(t) = h^{\text{syn}}(t) + \sum_{i=1}^F \eta^{\text{refr}}(t - t_i^f). \quad (1.100)$$

Thus the spiking statistics can be described similarly to Eq. (1.93) but is slightly more complex.

In the following we restrict our arguments to nonadaptive neurons ( $F = 1$ ). The spiking behavior of a neuron is traditionally classified by the *mean firing rate* and the *interval distribution* for a given input potential  $h^{\text{syn}}(t) \equiv h_0$ . The interval distribution  $D(s)$  is defined by the condition that  $D(s)\delta s$  is the probability of finding an interval of length  $s \dots s + \delta s$  in a given spike train. Comparison with the definition  $P_F^{(2)}(t|t_1^f)$  shows that  $D(s)$  is identical to  $P_F^{(2)}(s|0)$ , except that we have to consider the special case of constant external input  $h_i^{\text{ext}}(t) \equiv h_0$ . Using Eq. (1.93) with  $h(t) = h_0 + \eta^{\text{refr}}(t)$  we find

$$D(s) = \tau^{-1}[h_0 + \eta^{\text{refr}}(s)] \exp\left\{-\int_0^s \tau^{-1}[h_0 + \eta^{\text{refr}}(s')]ds'\right\}. \quad (1.101)$$



**Fig. 1.13.** *Spike trains of an adaptive model neuron.* Adaptation means that, for a given constant input, the interspike intervals get longer as time proceeds. The refractory function is the same as in the previous figure. In contrast to the nonadaptive case, *several* spikes contribute to the total refractory potential ( $F > 1$ ). Noise is slightly reduced to make adaptation more pronounced. (a)  $h^{\text{ext}} = 0.2$ ; (b)  $h^{\text{ext}} = 0.5$ .

The interval distribution is shown in Fig. 1.12 for various values of  $h_0$ , with  $\eta^{\text{refr}}$  given by the standard ansatz (1.88). We note that the interval distribution is sharply peaked at short intervals, if  $h > \theta$ , and shows a long exponential tail for  $h \approx \theta$ . This behavior corresponds well to experimental results.

The *mean interval*  $\bar{s}$  is given by

$$\bar{s} = \int_0^\infty s D(s) ds = \int_0^\infty S(s|0) ds. \quad (1.102)$$

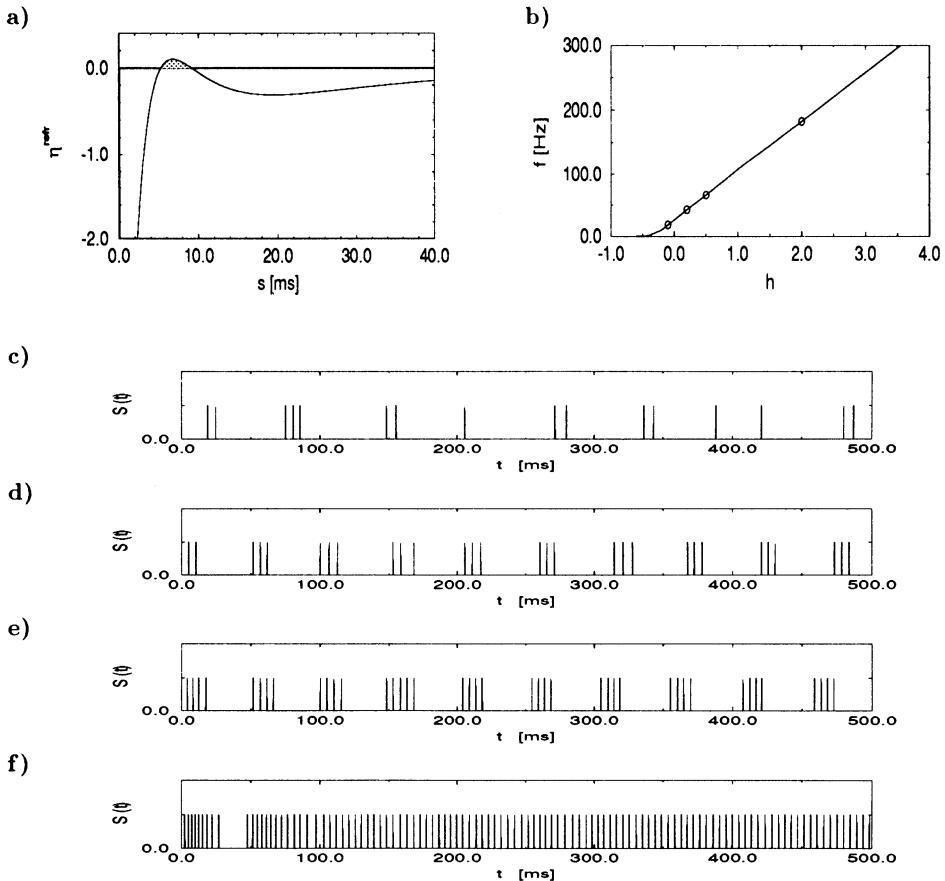
The first equality is the definition of  $\bar{s}$ , the second is found through an integration by parts using Eqs. (1.98) and (1.101). The function  $S(s|0) = \exp\{-\int_0^s \tau^{-1}[h_0 + \eta^{\text{refr}}(s')]ds'\}$  is the *survivor function* [102] of a neuron driven by  $h^{\text{syn}} \equiv h_0$ .

The *mean firing rate* can be defined as the inverse of the mean interspike interval  $\bar{s}$  [102]. The gain function is the mean firing rate  $f$  as a function of the input  $h_0$ . Due to expression (1.102), it can be written

$$f(h_0) = \left[ \int_0^\infty ds \exp\left\{-\int_0^s \tau^{-1}[h_0 + \eta^{\text{refr}}(s')]ds'\right\} \right]^{-1}. \quad (1.103)$$

The gain function of a standard model neuron as calculated from Eq. (1.103) is shown in Fig. 1.12(b). Typical spike trains of the same model are shown in Figs. 1.12(e)–(g).

So far we have concentrated on nonadaptive neurons ( $F = 1$ ). The spike train of a model neuron with the same refractory field as before, but  $F > 1$  is plotted in Fig. 1.13. It shows strong adaptation effects. Intervals are short at the beginning and get longer afterwards, even *though the input stays constant*:  $h^{\text{syn}}(t) \equiv h_0$ .



**Fig. 1.14.** *Bursting model neuron.* (a) shows the refractory function  $\eta^{\text{refr}}(s)$  which can be described by three exponentially decaying components. One of them has a depolarizing effect (shaded). This yields a periodic bursting as shown in the spike trains (c)–(f). In (b) the gain function is plotted ( $\beta = 15$ ). The values of  $h$  at which the spike trains (c)–(f) have been calculated have been marked by open circles. (c) Noise induced firing ( $h^{\text{ext}} = -0.1$ ). (d) Periodic activity bursts of three spikes each ( $h^{\text{ext}} = 0.2$ ). (e) Periodic bursts of 4–5 spikes ( $h^{\text{ext}} = 0.5$ ). (f) Oscillatory behavior ( $h^{\text{ext}} = 2.0$ ).

As an additional example we consider a neuron with an oscillatory refractory field that includes a period of depolarization (Fig. 1.14). This leads to a spike train with *burst* structure as seen in part (d) and (e) of the figure. At high input level the model exhibits fast regular spiking (Fig. 1.14(c)).

The above results are based on a continuous-time analysis of the model dynamics. In fact, real neurons also operate in continuous time so this is rather an advantage than a disadvantage. Simulations, however, rely on a discretization of time into finite intervals of some length  $\Delta t$ , say, 1 ms. For

a correct discretization, we need the probability of firing  $P_F(h; \Delta t)$  within a *finite* interval  $\Delta t$ . We cannot use Eq. (1.89) directly since we should have  $0 \leq P_F(h; \Delta t) \leq 1$ , but  $\tau^{-1}$  goes to  $\infty$  for  $h > \theta$ , if  $\beta \rightarrow \infty$ . The correct answer can be found using the survivor function (1.95) which is derived from Eq. (1.93) by integration over time. The probability of firing in an interval  $\Delta t$  is the same as the probability of not surviving  $\Delta t$  and thus

$$P_F(h; \Delta t) = 1 - S(t + \Delta t | t) = 1 - \exp\left\{-\int_t^{t+\Delta t} \tau^{-1}[h(s)] ds\right\}. \quad (1.104)$$

In a discrete-time model the integral can be replaced by a single term  $\tau^{-1}(h)\Delta t$  and we find

$$P_F(h; \Delta t) = 1 - \exp\{-\tau^{-1}(h)\Delta t\}. \quad (1.105)$$

It follows immediately that  $0 \leq P_F(h; \Delta t) \leq 1$ , as should be the case. Equation (1.105) can be used in simulations of noisy model neurons in discrete time and should replace the more common relation  $(1+\tanh \beta h)/2$  for a Glauber dynamics [61].

### Spike transmission and reception

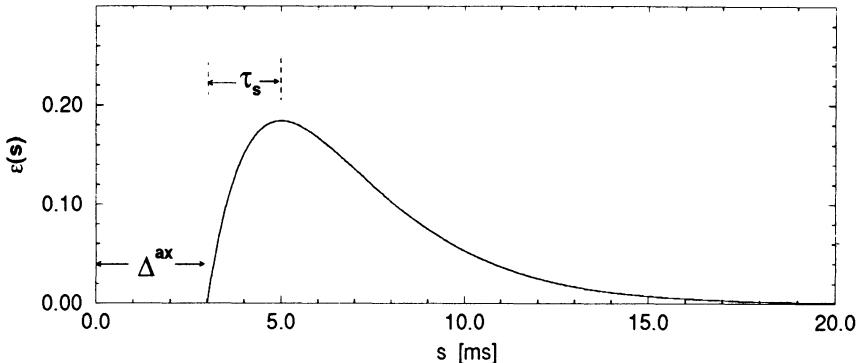
So far we have discussed the process of spike generation in the SRM neuron — but how can axonal transmission and dendritic integration be included into the neuron model? To answer this question let us consider signal transmission from a neuron  $j$  to a postsynaptic neuron  $i$ . The sender  $j$  generates a spike train:

$$S_j(t) = \sum_{f=1}^{n_f} \delta(t - t_j^f) \quad (1.106)$$

with firing times  $t_j^f$ . After a delay  $\Delta^{\text{ax}}$  the spikes arrive at a synapse connecting  $j$  and  $i$  and some signal is transmitted to the postsynaptic neuron. The detailed processes during synaptic transmission may be disregarded. Indeed, the only object that is relevant for further spike generation is the shape of the signal when it arrives at the *soma* of the postsynaptic neuron. Here it is no longer a sharp pulse but it has a much broader shape. This is partly due to the chemical transmission process at the synapse and partly due to passive electrical transmission along the dendritic tree. The form of the signal at the postsynaptic neuron may be described by an  $\alpha$  function [74] (see also [19])

$$\epsilon(s) = \begin{cases} 0 & \text{for } 0 \leq s \leq \Delta^{\text{ax}} \\ [(s - \Delta^{\text{ax}})/\tau_s]^2 \exp[-(s - \Delta^{\text{ax}})/\tau_s] & \text{for } s > \Delta^{\text{ax}} \end{cases}. \quad (1.107)$$

The parameter  $\Delta^{\text{ax}}$  is the axonal transmission delay and  $\tau_s$  the rise time of the response (see Fig. 1.15). Realistic parameter values are  $2 \leq \tau_s \leq 5$  ms and  $1 \leq \Delta^{\text{ax}} \leq 5$  ms for signal transmission within (one hemisphere of) the



**Fig. 1.15.** *Excitatory postsynaptic potential, EPSP.* If a presynaptic neuron fires a spike at  $s = 0$ , then an extended pulse arrives — after a delay  $\Delta^{ax}$  — at the soma or the postsynaptic neuron.  $\tau_s$  describes the rise time of the pulse.

cortex. Depending on the sign of the postsynaptic pulse, the response is called an excitatory or inhibitory postsynaptic potential (EPSP or IPSP).

Before we proceed to a larger network let us summarize our considerations. A presynaptic neuron  $j$  connected to neuron  $i$  evokes a postsynaptic potential:

$$h_i^{\text{syn}}(t) = J_{ij} \int_0^\infty \epsilon(s) S_j(t-s) ds, \quad (1.108)$$

where  $J_{ij}$  is the synaptic efficacy of the connection and  $\epsilon$  is the response kernel. The refractory potential (1.87) can be written in a very similar way, viz.,

$$h_i^{\text{refr}}(t) = \int_0^\infty \eta^{\text{refr}}(s) S_i(t-s) ds, \quad (1.109)$$

where  $\eta^{\text{refr}}$  is the refractory function. With this set of symmetric response formulas we close our discussion of a single SRM neuron.

## 1.4 A Network of Spiking Neurons

In the following we assume that we have a large network of SRM neurons at our disposal. We also assume that the shape of the postsynaptic potential is the same for all pairs of neurons and synapses. The *amplitude* of the response, however, depends on the efficacy of the synaptic coupling  $J_{ij}$  and may be different for each synapse. The simplest case is that of homogeneous coupling, that is,  $J_{ij} = J_0/N$ , where  $N$  is the total number of neurons in the net. This corresponds to a completely unstructured network. A convenient way to impose *some* structure is to divide the network into a finite number of ensembles denoted by  $x$ , each containing infinitely many and equivalent neurons. To be precise, the number of neurons  $N(x)$  in an ensemble is

proportional to the total number of neurons  $N(\mathbf{x}) = p(\mathbf{x})N$  as  $N \rightarrow \infty$ . The coupling of neurons within an ensemble  $\mathbf{x}$  is  $J(\mathbf{x}, \mathbf{x})$  and the coupling between a neuron of ensemble  $\mathbf{z}$  presynaptic to a neuron of ensemble  $\mathbf{x}$  is  $J(\mathbf{x}, \mathbf{z})$ . To give a biological interpretation,  $\mathbf{x} = (\mathbf{r}, \theta)$  could denote a column in the visual cortex containing neurons with receptive fields at location  $\mathbf{r}$  and with angle preference  $\theta$ .

In the following we mostly assume a coupling matrix as in the standard model of associative memory [69,70] (for a review, see [61])

$$J_{ij} = 2 \frac{J_0}{N} \sum_{\mu=1}^q \xi_i^\mu \xi_j^\mu. \quad (1.110)$$

The parameter  $J_0$  determines the overall coupling in the network. The factor  $2/N$  will turn out to be a useful normalization constant. The index  $\mu$  labels the patterns  $1 \leq \mu \leq q$  which are defined as sets of  $N$  random numbers  $\{\xi_i^\mu | 1 \leq i \leq N\}$  where the  $\xi_i^\mu = \pm 1$  are independent, identically distributed random variables with  $\text{prob}(\xi_i^\mu = \pm 1) = 1/2$ . The coupling matrix (1.110) which can be generated by a Hebbian learning rule [64,65] is known to yield good storage properties for random patterns ([11,12], for a review see [61]).

At first sight, a coupling matrix of the form (1.110) looks completely different from an ensemble coupling  $J(\mathbf{x}, \mathbf{z})$ . Nevertheless, we will show below that the Hebb matrix (1.110) can be rewritten in a form  $J(\mathbf{x}, \mathbf{z})$ , in other words, we can find ensembles  $\mathbf{x}$  and  $\mathbf{z}$  of equivalent neurons which all have the same synaptic efficacy pattern. In order to allow comparison to earlier models of neural networks, we use the coupling (1.110) as a standard ansatz. Most of our results, however, can easily be translated to arbitrary coupling  $J(\mathbf{x}, \mathbf{z})$ .

The above coupling (1.110) is restricted to random patterns with zero mean, that is,  $a \equiv \text{prob}(\xi_i^\mu = +1) - \text{prob}(\xi_i^\mu = -1) = 0$ . There is, however, a straightforward generalization to biased random patterns with  $a \neq 0$ . We simply set

$$J_{ij} = \frac{2J_0}{N(1-a^2)} \sum_{\mu=1}^q \xi_i^\mu (\xi_j^\mu - a). \quad (1.111)$$

This coupling yields good results for both sequences of patterns [59] and the problem of pattern separation [108]. Note that for low levels of the mean activity,  $a \rightarrow 0$ , the coupling is highly asymmetric. This implies that during learning a synapse is changed only if the presynaptic neuron is active. The state of the postsynaptic neuron determines whether the synaptic efficacy is enhanced or reduced. In a biological context it seems implausible to have a synapse change from excitatory to inhibitory influence. This problem can, however, be avoided if we allow for an appropriately chosen threshold and/or inhibitory interneurons.

To sum up the above considerations, there are several different choices for a synaptic matrix  $J_{ij}$ . To be specific, we assume in the following the standard form (1.110). Before proceeding further, let us put together the dynamic equations of the network. If a neuron  $j$  fires at time  $s = 0$ , a (delayed) postsynaptic signal  $J_{ij}\epsilon(s)$  is evoked at the soma of the postsynaptic neuron  $i$ . The total postsynaptic potential is the sum of all incoming signals:

$$h_i^{\text{syn}}(t) = \sum_{j, j \neq i}^N J_{ij} \int_0^\infty \epsilon(s) S_j(t-s) ds. \quad (1.112)$$

Here we have assumed that all dendritic contributions add up linearly.

The *internal* contributions can be written similarly as

$$h_i^{\text{refr}}(t) = \int_0^\infty \eta^{\text{refr}}(s) S_i^{(F)}(t-s) ds. \quad (1.113)$$

Here  $\eta^{\text{refr}}(s)$  is the refractory function as explained in the preceding subsection. The upper index  $(F)$  is a reminder that only the  $F$  most recent spikes of neuron  $i$  contribute to its refractory field. If  $F = 1$ , then the simple relation  $h^{\text{refr}}(t) = \eta^{\text{refr}}(t - t_1^f)$  holds. This is the case which we will study below.

The firing probability in an infinitesimal interval is given by Eq. (1.89) and is repeated here for completeness:

$$P_F(h; \delta t) = \tau^{-1}(h) \delta t. \quad (1.114)$$

Equations (1.112)–(1.114) describe the dynamics of a network of SRM neurons. For the sake of simplicity we assume in the following that the response function  $\epsilon(s)$  and the refractory function  $\eta^{\text{refr}}(s)$  are identical for all neurons, but the model can easily be generalized so as to include different types of neuron and synapse in the same network [45,46].

### Solution of the network dynamics: Shortcut

In the preceding subsection we have discussed the dynamics of a single neuron that may or may not be part of a larger network. Now we would like to analyze the state of a network of SRM neurons as a whole. Can we make any predictions concerning the *global dynamics* of the system? Let us consider a large network of  $N$  neurons with  $N \rightarrow \infty$ . In a first step, we classify the neurons  $1 \leq i \leq N$  according to the *sublattices* [57,58,61]

$$L(\mathbf{x}) = \{i | \xi_i = \mathbf{x}\}. \quad (1.115)$$

All neurons that are equivalent in the sense that they have learned the same information  $(\xi_i^1, \dots, \xi_i^q) = (x^1, \dots, x^q) = \mathbf{x}$  (see Fig. 1.16) belong to the same sublattice. Now let us turn to the neurons in a specific sublattice  $L(\mathbf{z})$ . The total number of those neurons will be denoted by  $|L(\mathbf{z})| = Np(\mathbf{z})$ .

	i	j	k	l	m	
pattern 1	1	1	1	-1	1	...
pattern 2 ...	-1	-1	-1	-1	-1	...
pattern 3	1	-1	1	1	-1	...

**Fig. 1.16.** Sublattices. A neural network has learned three different patterns  $\{\xi_i^1\}, \{\xi_i^2\}, \{\xi_i^3\}$ . Each pattern consists of a sequence of numbers  $\{\xi_i^\mu = \pm 1; 1 \leq i \leq N\}$ , out of which five are shown. Each row contains the values of one pattern and each column the values for one neuron. Neurons  $i$  and  $k$  have learned the same information and thus belong to the same sublattice. Neuron  $j$  and  $m$  belong to a different sublattice.

The *activity*  $A(\mathbf{z}, t)$  is a measure of the number  $N_{\text{sp}}$  of spikes fired by neurons of this sublattice in a short time  $\Delta t$ . To be precise,

$$A(\mathbf{z}, t) = |L(\mathbf{z})|^{-1} \lim_{\Delta t \rightarrow 0} \frac{N_{\text{sp}}(\mathbf{z}, t)}{\Delta t}. \quad (1.116)$$

With these abbreviations the synaptic potential of a neuron  $i \in L(\mathbf{x})$  is

$$\begin{aligned} h_i^{\text{syn}}(t) &= 2J_0 \sum_{\mu=1}^q \xi_i^\mu \sum_{j=1}^N \xi_j^\mu \int_0^\infty \epsilon(s') S_j(t-s') ds' \\ &= 2J_0 \sum_{\mu=1}^q x^\mu \sum_{\mathbf{z}} z^\mu \int_0^\infty \epsilon(s') \frac{|L(\mathbf{z})|}{N} |L(\mathbf{z})|^{-1} \sum_{j \in L(\mathbf{z})} S_j(t-s') ds' \\ &= 2J_0 \sum_{\mu=1}^q x^\mu \sum_{\mathbf{z}} z^\mu \int_0^\infty \epsilon(s') p(\mathbf{z}) A(\mathbf{z}, t-s') ds' \\ &= \sum_{\mathbf{z}} J(\mathbf{x}, \mathbf{z}) \int_0^\infty \epsilon(s') p(\mathbf{z}) A(\mathbf{z}, t-s') ds', \end{aligned} \quad (1.117)$$

where we have set  $J(\mathbf{x}, \mathbf{z}) = 2J_0 \sum_{\mu=1}^q x^\mu z^\mu$ . We note that the index  $i$  on the right-hand side of Eq. (1.117) has disappeared and only a dependence upon the sublattice vector  $\mathbf{x}$  remains. That is, all neurons  $i \in L(\mathbf{x})$  are subject to the *same* synaptic potential  $h_i^{\text{syn}}(t) = h(\mathbf{x}, t)$  which depends on the activity of all other sublattices  $A(\mathbf{z}, t-s')$  convoluted with the synaptic kernel  $\epsilon(s')$ .

Do we know anything about the activity  $A(\mathbf{x}, t-s)$ ? A term of the form  $A(\mathbf{x}, t-s)$  picks out all neurons that have fired at time  $(t-s)$  and belong to the same sublattice  $L(\mathbf{x})$ . It follows that they are subject to the *same* synaptic potential  $h^{\text{syn}}(\mathbf{x}, t)$ . Now let us return to the conditional probability  $P_F^{(2)}(t|t_1^f)$ . Remembering the definition of this conditional probability

we note that  $P_F^{(2)}(t|t-s)A(\mathbf{z}, t-s)$  is the portion of neurons that have fired at  $t-s$  and fire again at time  $t$ . Integration over all past firing times yields

$$A(\mathbf{z}, t) = \int_0^\infty P_F^{(2)}(t|t-s)A(\mathbf{z}, t-s)ds. \quad (1.118)$$

Equations (1.117) and (1.118) give a closed solution to the global dynamics of the network. They are the main result of this subsection.

### Explicit solution

The way we arrived at Eq. (1.118) was rather handwaving and the reader might feel uneasy about it. In the following we present a rigorous derivation of the above results which makes all necessary assumptions explicit. To make the arguments as clear as possible we consider the model in discrete time. For example, we may think of a time step  $\Delta t$  of about 1 ms. The spike train of a neuron  $i$  can be described by

$$S_i(t) = \sum_{f=1}^{n_f} \delta[t, \hat{t}_i^f]. \quad (1.119)$$

The term  $\delta[n, m]$  denotes the Kronecker  $\delta$  and  $\hat{t}_i^f$  is the discrete version of the firing time. The membrane potential can be rewritten

$$h_i(t) = 2 \frac{J_0}{N} \sum_{\mu=1}^q \sum_{j=1, j \neq i}^N \xi_i^\mu \xi_j^\mu \sum_{s=0}^{\infty} \hat{\epsilon}(s) S_j(t-s) + \sum_{s=0}^{s_m} \hat{\eta}^{\text{refr}}(s) \sum_{f=1}^F \delta[t-s, \hat{t}_i^f]. \quad (1.120)$$

We have placed a hat over  $\epsilon(s)$  and  $\eta^{\text{refr}}(s)$  to remind the reader that here we work with a discrete model.

We would like to gather all neurons with the same membrane potential into a common class. To this end we have to subdivide the sublattices further and collect all neurons with identical synaptic potential *and* identical refractory field into a common subclass:

$$L(\mathbf{x}, s, t) = \{i | \xi_i = \mathbf{x}, h_i^{\text{refr}}(t) = \eta^{\text{refr}}(s)\}. \quad (1.121)$$

We assume that  $\eta^{\text{refr}}(s) = 0$  for all  $s > s_m$ . Then  $s$  is limited to the range  $0 \leq s \leq s_m$ . We also require the total number of patterns  $q$  to be finite. Then the total number of subclasses is finite. The number of neurons in each subclass  $L(\mathbf{x}, s, t)$ , to be denoted by  $|L(\mathbf{x}, s, t)|$ , may change from each time step to the next. In the limit of  $N \rightarrow \infty$ , however, each subclass contains infinitely many elements (or none at all). The neurons in each subclass are equivalent and independent. Thus the activity

$$\hat{A}(\mathbf{x}, s, t + \Delta t) = |L(\mathbf{x}, s, t)|^{-1} \sum_{i \in L(\mathbf{x}, s, t)} S_i(t + \Delta t) \quad (1.122)$$

is given by the probability that one of the neurons fires in one time step:

$$\lim_{N \rightarrow \infty} \hat{A}(\mathbf{x}, s, t + \Delta t) = P_F[h(\mathbf{x}, s, t); \Delta t]. \quad (1.123)$$

This is an immediate result of the law of large numbers [89].

If the activity of all subclasses is known, the total activity of a sublattice can be calculated by a summation over all possible values of  $s$ , that is,

$$\begin{aligned} \hat{A}(\mathbf{x}, t + \Delta t) &= |L(\mathbf{x})|^{-1} \sum_{i \in L(\mathbf{x})} S_i(t + \Delta t) \\ &= \sum_{s=0}^{s_m} \frac{|L(\mathbf{x}, s, t)|}{|L(\mathbf{x})|} \hat{A}(\mathbf{x}, s, t + \Delta t). \end{aligned} \quad (1.124)$$

Introducing  $\hat{p}(\mathbf{x}, s, t) = |L(\mathbf{x}, s, t)|/|L(\mathbf{x})|$  we have

$$\begin{aligned} \hat{A}(\mathbf{x}, t + \Delta t) &= \sum_{s=0}^{s_m} \hat{p}(\mathbf{x}, s, t) |\hat{A}(\mathbf{x}, s, t + \Delta t)| \\ &= \sum_{s=0}^{s_m} \hat{p}(\mathbf{x}, s, t) P_F[h(\mathbf{x}, s, t); \Delta t]. \end{aligned} \quad (1.125)$$

The function  $P_F[h; \Delta t]$ , Eq. (1.89), and the membrane potential

$$h(\mathbf{x}, s, t) = 2J_0 \sum_{\mu=1}^q x^\mu \sum_{\mathbf{z}} z^\mu p(\mathbf{z}) \sum_{s'=0}^{\infty} \hat{\epsilon}(s') A(\mathbf{z}, t - s') + \hat{\eta}^{\text{refr}}(s) \quad (1.126)$$

are known. The only — so far — unknown variable is  $\hat{p}(\mathbf{x}, s, t)$ . It denotes the portion of those neurons in sublattice  $L(\mathbf{x})$  that have spent a time  $s$  without firing. Can we say anything about  $\hat{p}(\mathbf{x}, s, t)$ ? A neuron that belongs at time  $t$  to  $L(\mathbf{x}, s, t)$  will be at time  $t + \Delta t$  in  $L(\mathbf{x}, s - \Delta t, t + \Delta t)$ , unless it has fired — which it does with probability  $P_F[h(\mathbf{x}, s, t); \Delta t]$ . Formalizing this argument we arrive at [44]

$$\begin{aligned} \hat{p}(\mathbf{x}, 0, t + \Delta t) &= \sum_{s=0}^{s_m} \hat{p}(\mathbf{x}, s, t) P_F[h(\mathbf{x}, s, t); \Delta t], \\ \hat{p}(\mathbf{x}, s, t + \Delta t) &= \hat{p}(\mathbf{x}, s - \Delta t, t) \{1 - P_F[h(\mathbf{x}, s - \Delta t, t); \Delta t]\}, \\ \hat{p}(\mathbf{x}, s_m, t + \Delta t) &= \hat{p}(\mathbf{x}, s_m, t) \{1 - P_F[h(\mathbf{x}, s_m, t); \Delta t]\} \\ &\quad + \hat{p}(\mathbf{x}, s_m - \Delta t, t) \{1 - P_F[h(\mathbf{x}, s_m - \Delta t, t); \Delta t]\}. \end{aligned} \quad (1.127)$$

The last equation looks slightly different since all neurons with  $s > s_m$  have the same refractory field  $\eta^{\text{refr}}(s) = 0$  and we gather them into the same subclass  $L(\mathbf{x}, s_m, t)$ . This equation can be dropped, if all values  $s > s_m$  are included in Eq. (1.127).

The set of Eqs. (1.125)–(1.127) describes the macroscopic dynamics of the network in discrete time. As the final step of our derivation we now take the limit  $\Delta t \rightarrow 0$ . Before doing so let us look at the scaling properties of the various quantities. We have

$$\begin{aligned} P_F[h; \Delta t] &= \tau^{-1}[h]\Delta t, \\ \hat{p}(\mathbf{x}, s, t) &= p(\mathbf{x}, s, t)\Delta t, \\ \hat{A}(\mathbf{x}, t) &= A(\mathbf{x}, t)\Delta t. \end{aligned} \quad (1.128)$$

Using Eq. (1.128) we find from Eqs. (1.125) and (1.126),

$$A(\mathbf{x}, t) = \int_0^\infty ds p(\mathbf{x}, s, t)\tau^{-1}[h(\mathbf{x}, s, t)] \quad (1.129)$$

with

$$h(\mathbf{x}, s, t) = 2J_0 \sum_{\mu=1}^q x^\mu \int_0^\infty \epsilon(s') \sum_{\mathbf{z}} z^\mu p(\mathbf{z}) A(\mathbf{z}, t-s') ds' + \eta^{\text{refr}}(s). \quad (1.130)$$

If we introduce  $J(\mathbf{x}, \mathbf{z}) = 2J_0 \sum_{\mu=1}^q x^\mu z^\mu$  this can also be written in the form

$$h(\mathbf{x}, s, t) = \sum_{\mathbf{z}} J(\mathbf{x}, \mathbf{z}) \int_0^\infty \epsilon(s') p(\mathbf{z}) A(\mathbf{z}, t-s') ds' + \eta^{\text{refr}}(s). \quad (1.131)$$

The evolution of  $\hat{p}$ , viz., Eq. (1.127), yields

$$p(\mathbf{x}, 0, t) = \int_0^\infty p(\mathbf{x}, s, t) \tau^{-1}[h(\mathbf{x}, s, t)] ds \quad (1.132)$$

and

$$\frac{\partial}{\partial t} p(\mathbf{x}, s, t) = -\{\tau^{-1}[h(\mathbf{x}, s, t)] + \frac{\partial}{\partial s}\} p(\mathbf{x}, s, t). \quad (1.133)$$

Here we have assumed that  $s$  can take all values:  $s > 0$ . This assumption allows us to disregard the third of Eqs. (1.127). Normalization of  $p(\mathbf{x}, s, t)$  requires

$$\lim_{s \rightarrow \infty} p(\mathbf{x}, s, t) = 0 \quad (1.134)$$

and

$$\int_0^\infty p(\mathbf{x}, s, t) ds = 1. \quad (1.135)$$

If we compare Eq. (1.132) with Eq. (1.129) we find  $p(\mathbf{x}, 0, t) = A(\mathbf{x}, t)$ . This is what we should expect since  $p(\mathbf{x}, 0, t)$  is the portion of neurons that have been quiescent for a time  $s = 0$ . In other words, it describes those neurons which are firing *now*, at time  $t$ .

Next, we would like to discuss the time evolution of  $p(\mathbf{x}, s, t)$ ; cf. Eq. (1.133). The portion of neurons that have stayed quiescent for a time  $s$

changes due to firing (with rate  $\tau^{-1}$ , first term), and also since other neurons join or leave this group (drift term  $d/ds$ ). Equation (1.133) can be integrated and yields

$$p(\mathbf{x}, s, t) = p(\mathbf{x}, 0, t-s) \exp\left\{-\int_0^s \tau^{-1}[h(\mathbf{x}, s', t-s+s')]ds'\right\}. \quad (1.136)$$

This allows the following interpretation. A neuron which has fired at time  $t-s$  belongs to the subclass  $L(\mathbf{x}, 0, t-s)$  (first factor on the right-hand side). If it has not fired since then (survival probability  $S(t|t-s)$ , second factor), then it now belongs to the subclass  $L(\mathbf{x}, s, t)$  (left-hand side). Combining Eq. (1.136) with Eq. (1.129) we obtain

$$A(\mathbf{x}, t) = \int_0^\infty ds A(\mathbf{x}, t-s) \tau^{-1}[h(\mathbf{x}, s, t)] \exp\left\{-\int_0^s \tau^{-1}[h(\mathbf{x}, s', t-s+s')]ds'\right\}. \quad (1.137)$$

This is the solution which we have “guessed” before; cf. Eq. (1.118).

### Discussion of the solution

Before we proceed further let us look back and discuss the results. The essential step in the derivation of Eqs. (1.129) and (1.136) is the formation of appropriate (sub)classes which contain independent and equivalent neurons. This allows us to replace the spiking activity  $A$  of the subclass by the spiking probability  $P_F$  of a typical neuron in this class. For this to be exact we have to take the limit  $N \rightarrow \infty$ . Afterwards the limit  $\Delta t \rightarrow 0$  can be taken which yields the above results.

As we have indicated before, the integration kernel in Eq. (1.137) is the conditional probability  $P_F^{(2)}(t|t-s)$ . The exponent of  $P_F^{(2)}(t|t-s)$  contains another integration over time and still another integration has to be performed to calculate the membrane potential  $h(t)$ ; cf. Eq. (1.130). This intricate dependence of the system upon its own history reflects the fact that it is a globally connected network of *noisy* pulse-emitting elements subject to their own refractory field  $\eta^{\text{refr}}(s)$  and coupled through a response kernel  $\epsilon(s)$ . In order to get an idea of the effect of the integration kernel  $P_F^{(2)}(t|t-s)$  we consider the limit  $\beta \rightarrow \infty$ . In this case  $P_F^{(2)}(t|t-s)$  reduces to

$$\lim_{\beta \rightarrow \infty} P_F^{(2)}(t+s|t) = \delta[s - s_F(t)], \quad (1.138)$$

where  $s_F(t)$  fulfills the threshold condition

$$s_F(t) = \inf\{s | h(\mathbf{x}, s, t+s) = \theta\}. \quad (1.139)$$

The interpretation of this result is obvious. Noiseless neurons that have fired at  $t$  will fire again, if the threshold is crossed the next time. For finite  $\beta$ , the integration kernel is no longer a  $\delta$  function, but has finite width. The structure of the results, however, remains the same.

Before we turn to some specific solutions let us conclude this subsection with a couple of more general considerations. Though we have started with a specific model, that is, the SRM neuron, the results (1.118) or, equivalently, (1.137) are valid for any network of nonadaptive spiking neurons with noise. The specific way refractoriness and noise are introduced — see Eqs. (1.89) and (1.90) — is just a convenient manner to calculate the spiking statistics  $P_F^{(2)}(t|t_1^f)$ . This is the only entity that enters the further analysis. Mathematically, firing in a nonadaptive model ( $F = 1$ ) can be considered as a renewal process [24, 31, 102]. Other “renewal” models of a spiking neuron, for example, the integrate-and-fire neuron with noise, would lead to a similar spiking behavior. If the spiking statistics  $P_F^{(2)}(t|t_1^f)$  can be calculated, then it is also possible to find a closed solution of the type (1.137). Similarly, the detailed model of synaptic transmission and dendritic integration does not matter for the global dynamics. The only relevant quantity is the net effect of all incoming signals at the soma of the postsynaptic neuron. As long as dendritic summation is linear the result always has the structure of Eq. (1.130). Thus, our results are of a more general nature and allow some insight into the global dynamics in networks of spiking neurons.

Finally, we conclude this subsection with a remark regarding the problem of coding by ensembles *vs* coding by single (spiking) neurons. Our network equations (1.130) and (1.137) show that in a large and fully connected network an exact transition from single neurons to “ensembles” — the sublattices  $L(\mathbf{x})$  — is possible. In our mathematical model the classification of neurons into appropriate sublattices of equivalent neurons is straightforward; cf. Eq. (1.115). In a real biological network, however, this can be much more difficult. If we describe an area of the visual cortex which has a columnar structure [72], we can assume that equivalent neurons are situated next to each other in a single column. Then a description by ensembles is equivalent to a spatial average over one column. In this case, we can give Eq. (1.131) a different interpretation. The sublattices  $L(\mathbf{x})$  correspond to columns where  $\mathbf{x} = (\mathbf{r}, \theta)$  denotes the location  $\mathbf{r} \in \mathbb{R}^2$  and preferred angle  $\theta = [0, 2\pi]$  of the receptive fields of neurons in a given column, and  $A(\mathbf{x}, t)$  is the mean activity of a column. The interaction may be chosen to depend on the distance and the angle difference  $J(\mathbf{x}, \mathbf{z}) = J_1(|\mathbf{r} - \mathbf{r}'|)J_2(\theta - \theta')$ . This leads to dynamic equations between local ensembles with finite-range interactions. In general, however, equivalent neurons are not neighbors, but distributed all over the network. In this case local averaging spoils information which is hidden in the spatio-temporal firing pattern of single neurons. This will be explained in more detail in Sec. 1.5.

#### 1.4.1 STATIONARY SOLUTIONS — INCOHERENT FIRING

Our results (1.130) and (1.137) describe the macroscopic dynamics in a network of nonadaptive SRM neurons. In the following we will analyze

two special solutions in more detail, viz., coherent and incoherent firing.

Incoherent firing can be defined by the condition of *stationary* activity, that is,

$$A(\mathbf{x}, t) \equiv A(\mathbf{x}). \quad (1.140)$$

By definition  $A(\mathbf{x}, t)$  is calculated as an average over a large number of neurons; cf. Eq. (1.116). If all neurons fire incoherently, the averaged activity may be constant in time even though every neuron emits a spike train with a specific time structure.

In case of constant  $A(\mathbf{x}, t)$ , Eq. (1.137) is trivially fulfilled since we have  $\int_0^\infty P_F^{(2)}(t|t-s)ds = 1$ . The membrane potential (1.130) simplifies to

$$h(\mathbf{x}, s, t) = h^{\text{syn}}(\mathbf{x}) + \eta^{\text{refr}}(s) \quad (1.141)$$

with

$$h^{\text{syn}}(\mathbf{x}) = \sum_{\mathbf{z}} J(\mathbf{x}, \mathbf{z}) p(\mathbf{z}) A(\mathbf{z}), \quad (1.142)$$

where  $J(\mathbf{x}, \mathbf{z}) = 2J_0 \sum_{\mu=1}^q x^\mu z^\mu$ . Equation (1.136) together with the normalization (1.135) yields

$$A(\mathbf{x}) = \left\{ \int_0^\infty ds \exp \left\{ - \int_0^s \tau^{-1} [h^{\text{syn}}(\mathbf{x}) + \eta^{\text{refr}}(s')] ds' \right\} \right\}^{-1}, \quad (1.143)$$

where we have used  $A(\mathbf{x}) \equiv A(\mathbf{x}, t) = p(\mathbf{x}, 0, t)$ . If we compare this result with the definition of the mean firing rate (1.103) we find that both expressions are identical. We conclude that the activity  $A(\mathbf{x})$  — defined as the spiking activity averaged over a large number of neurons — is equivalent to the mean firing rate as defined by the inverse of the mean interspike interval in the spike train of a single neuron. This result which we have derived here for a specific model of spiking neurons is much more general. It is a mean-field result which is valid in the limit of  $N \rightarrow \infty$  in every globally connected network — *independent of the neuron model*. The only condition on the network is that it stores only a finite number of patterns. Under this condition we can find infinitely large groups of equivalent neurons, the so-called sublattices  $L(\mathbf{x})$ . Whenever the system is in a stationary state, averaging over time (mean firing rate  $f$ ) is the same as averaging over the group of identical neurons (activity  $A$ ). In other words, we have proven *universality* [45]: a network in a stationary state is described fully by the *gain function* (mean firing rate *vs* input) and the synaptic efficacies of the neurons — whatever the neuron model. Thus, universality explains why associative properties have been found in networks with a variety of different model neurons [13, 20, 22, 70, 71, 92].

If the gain function of the model neurons is known, an analytical expression for a stationary network state can be derived. The activity on a sublattice  $L(\mathbf{x})$  is given by the fixed-point equation

$$A(\mathbf{x}) = f \left[ \sum_{\mathbf{z}} J(\mathbf{x}, \mathbf{z}) p(\mathbf{z}) A(\mathbf{z}) \right], \quad (1.144)$$

where  $f(h^{\text{syn}})$  is the gain function (1.103) of the model neuron and we have set  $J(\mathbf{x}, \mathbf{z}) = 2J_0 \sum_{\mu=1}^q x^\mu z^\mu$ . Thus we have to solve a set of simple fixed-point equation (1.144) instead of the coupled dynamic equations (1.137).

Before presenting some illustrative examples let us perform a change of variables and introduce the *overlap*  $m_\mu$ . In the theory of neural networks,  $m_\mu$  is used as a measure of the correlation between the momentary state of the network and a specific pattern  $\mu$ ; see, for example, [61]. In a discrete-time model with time intervals  $\Delta t$ , the overlap is defined by

$$m_\mu(t) = 2 \sum_{i=1}^N S_i(t) \xi_i^\mu \quad (1.145)$$

with  $S_i(t) = \delta[t, \hat{t}_i^f] \in \{0, 1\}$ . In passing we note that if  $S_i(t) \in \{-1, 1\}$ , then the prefactor 2 must be dropped. Taking the limit  $\Delta t \rightarrow 0$  and using the definition of the sublattice activity (1.116) we find

$$m_\mu(t) = 2 \sum_{\mathbf{x}} x^\mu p(\mathbf{x}) A(\mathbf{x}, t). \quad (1.146)$$

Let us now assume that there are macroscopic correlations with a single pattern  $\nu$ , that is,  $m_\mu = m \delta_{\mu,\nu}$ . A stationary state with  $m > 0$  is called a *retrieval state*. The value of  $m$  is given by the nontrivial solutions of the fixed-point equation

$$\begin{aligned} m &= 2 \sum_{\mathbf{x}} x^\nu p(\mathbf{x}) f[J_0 \sum_{\mu=1}^q x^\mu m_\mu] \\ &= f[J_0 m] - f[-J_0 m], \end{aligned} \quad (1.147)$$

where we have assumed random patterns with  $p(x^\nu = +1) = p(x^\nu = -1) = 1/2$ . We now turn to some specific examples so as to illustrate the above results.

### Examples of stationary solutions

As a simple, but instructive example, let us consider a model with an absolute refractory period only, that is,

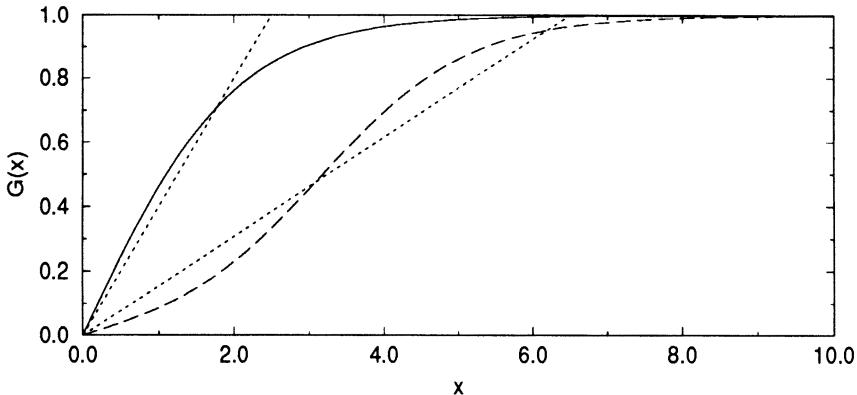
$$\eta^{\text{refr}}(s) = \begin{cases} -\infty & \text{for } 0 \leq s \leq \gamma^{\text{refr}} \\ 0 & \text{for } s > \gamma^{\text{refr}} \end{cases}. \quad (1.148)$$

Using Eq. (1.103) of Sec. 3.3 we find the gain function

$$f(h) = \{\gamma^{\text{refr}} + \tau_0 \exp[-\beta(h - \theta)]\}^{-1}. \quad (1.149)$$

Assuming that there is overlap with a single pattern only, that is,  $m_\mu = m \delta_{\mu,\nu}$ , we obtain

$$\gamma^{\text{refr}} m = \frac{\sinh(J_0 \beta m)}{r + \cosh(J_0 \beta m)}, \quad (1.150)$$

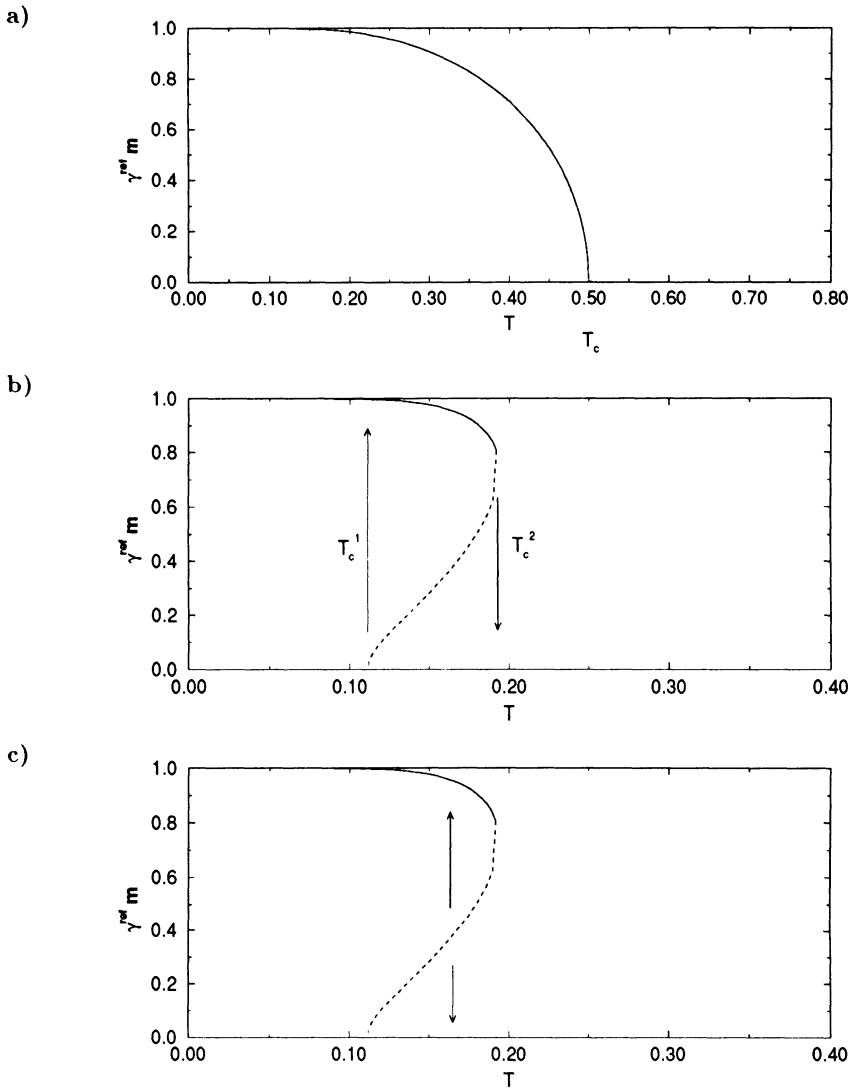


**Fig. 1.17.** Model with absolute refractory period only. The function  $G(x) = \sinh x / (r + \cosh x)$  is plotted in dependence upon  $x = J_0 \beta m$ . For  $r = 1$  it yields the solid curve,  $r = 12$  gives the dashed one. From the intersection with a straight line of slope  $\gamma^{\text{refr}} / J_0 \beta$  (dotted) we find the overlap  $\gamma^{\text{refr}} |m|$  in a stationary state. For large  $r$ , the critical temperature  $T_c$  is lower than for  $r = 1$  and an additional unstable solution appears. (From [44].)

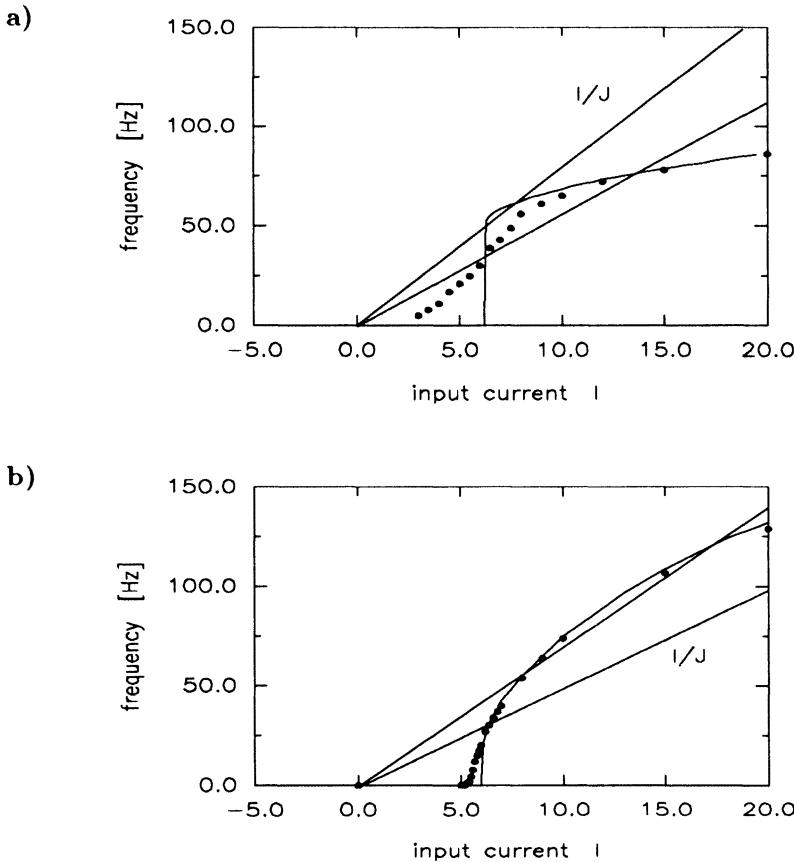
where we have introduced the parameter  $r = (1/2)[(\gamma^{\text{refr}}/\tau_0) + (\tau_0/\gamma^{\text{refr}})]$ . In Fig. 1.17 we have plotted the function  $G(x) = \sinh x / [r + \cosh x]$  for different values of the parameter  $r$ . The intersection of  $G(x)$  with a straight line of slope  $1/\beta$  yields the stationary overlap  $m$ . In Fig. 1.18 we show  $m$  as a function of  $T = \beta^{-1}$ . For parameters  $r < 2$  (Fig. 1.18a), the overlap decreases continuously with temperature and reaches 0 at a critical noise level  $T_c = (1+r)^{-1}$ . The trivial state  $m = 0$  is unstable for  $T < T_c$  and stable for  $T > T_c$ . We thus have a continuous phase transition similar to the one that has been found for the Hopfield network [11,69] to which the system formally reduces in the limit  $r \rightarrow 0$ .

The behavior, however, is completely different for parameter values  $r > 2$ . In this case (Fig. 1.18(b), (c)), a regime of bistability occurs in which both the trivial state and the retrieval state are stable *stationary* states of the system. (This does not exclude that these states might be unstable with respect to oscillatory solutions; see below.) The bistability has two effects on the behavior of the system. To discuss the first one, we keep the noise level  $\beta$  fixed and assume that the system starts with some initial overlap. Depending on whether the overlap is larger or smaller than the overlap  $m'$  of the unstable fixed point, the system either reconstructs the pattern and goes to the retrieval state, or the overlap is reduced and it goes to the trivial state. Thus, the unstable fixed point acts as a threshold value for the retrieval of the pattern (Fig. 1.18(c)).

The second effect is a hysteresis behavior as shown in Fig. 1.18(b)). If we start at a high noise temperature and ‘cool’ the system down, the overlap changes discontinuously from  $m = 0$  to  $m \approx 1$  at the lower critical temper-



**Fig. 1.18.** Overlap  $m$  as a function of temperature  $T = 1/\beta$  for a model with absolute refractory period only. In (a) we have  $r = 1$  and a continuous phase transition. In (b) and (c) we have set  $r = 8$  which yields a regime of bistability where hysteresis (b) can be found. The unstable solution is the retrieval threshold as shown in (c). (From [44].)



**Fig. 1.19.** Retrieval behavior for (a) Hodgkin-Huxley model and (b) integrate-and-fire model with parameters  $\gamma^{\text{refr}} = 4 \text{ ms}$ ,  $\tau_s = 10 \text{ ms}$ ,  $I_\theta = 6 \mu\text{A}$ . We have plotted the function  $g(I) = f(I) - f(-I)$  which is essentially identical to the gain function  $f(I)$  since  $f(-I) \equiv 0$  (noiseless neuron — solid line; with noise — dotted). The intersection of  $g(I)$  with a straight line of slope  $1/J_0$  yields the stationary overlap in a retrieval state. (Adapted from [45].)

ature  $T_c^{(1)} = 1/(1+r)$ . On the other hand, if we ‘heat’ the network starting at zero temperature, the network stays in the retrieval state until the upper critical temperature  $T_c^{(2)}$  is reached where it jumps discontinuously to the trivial state. In terms of thermodynamics this behavior indicates a first-order phase transition. Thus, a network of neurons with large  $r$ , that is, very long or extremely short refractory periods compared to the intrinsic time constant  $\tau_0$ , belongs to a different universality class than the Hopfield network.

The above example of a network of refractory neurons is a very simple case and we may ask what happens if a more realistic model of a spiking

neuron is considered. In order to answer this question, let us have another look at the Hodgkin–Huxley model and the integrate-and-fire model. Equation (1.147) tells us that we should calculate the gain function and evaluate it at  $\pm h^{\text{syn}}$  where  $h^{\text{syn}}$  is to be found self-consistently by solving

$$\begin{aligned} m &= J_0^{-1} h^{\text{syn}}, \\ m &= f(h^{\text{syn}}) - f(-h^{\text{syn}}). \end{aligned} \quad (1.151)$$

The graphical solution is shown in Fig. 1.19 where  $g(h) = f(h) - f(-h)$  is plotted for both types of model neuron. The intersection of  $g(h)$  with a straight line of slope  $1/J_0$  yields the overlap  $m = h^{\text{syn}}/J_0$ . We immediately see that a retrieval solution exists, if the global synaptic efficacy  $J_0$  is strong enough.

Equations (1.151) are often called a mean-field solution of the network [2,13] — in analogy to the mean-field approximation used in the Ising model of ferromagnetism. Whereas a mean-field ansatz is a bad approximation in all models with (short-ranged) nearest-neighbor interaction, it is a good approximation if there is sufficiently strong long-ranged interaction. Indeed, a model-independent derivation of Eqs. (1.151) shows that this is the *exact* solution in a *fully* connected network with a finite number of patterns [45]. Full connectivity is equivalent to infinitely ranged interaction, and in this case the mean-field result is *exact* in the limit of  $N \rightarrow \infty$ .

### Stability analysis

The above examples show that there exist stationary solutions which have a macroscopic overlap with one of the patterns — but are these solutions stable? In order to analyze the stability we have to consider the response of the system to a small time-dependent perturbation  $h^1$  around the stationary state. To make the argument as clear as possible we first consider the noiseless limit ( $\beta \rightarrow \infty$ ).

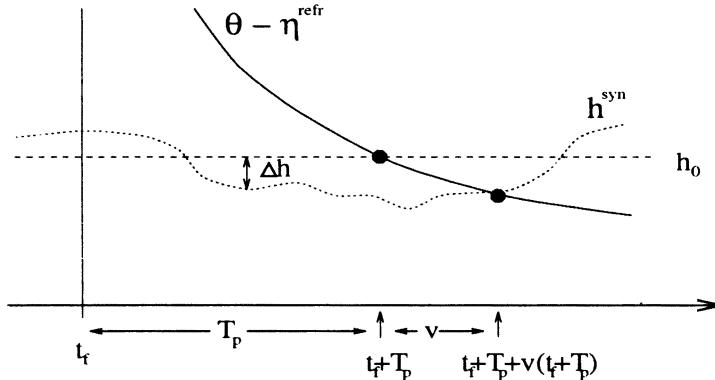
We start with the macroscopic dynamics of the activity; cf. Eq. (1.137). In the noise-free limit, the kernel in Eq. (1.137) reduces to a  $\delta$  function, and with the rules of  $\delta$  function algebra we find

$$\begin{aligned} A(\mathbf{x}, t) &= \int ds A(\mathbf{x}, t-s) \delta[s - s_F(\mathbf{x}, t-s)] \\ &= A(\mathbf{x}, t - s_F) \left[ 1 + \frac{d}{dt} s_F|_{t-s_F} \right]^{-1}. \end{aligned} \quad (1.152)$$

The interval  $s_F(\mathbf{x}, t)$  can be found from the threshold condition  $s_F(\mathbf{x}, t) = \inf \{s | \eta^{\text{refr}}(s) + h^{\text{syn}}(\mathbf{x}, t+s) = \theta\}$  which yields

$$s_F(\mathbf{x}, t) = T_p + v(t + T_p). \quad (1.153)$$

Here  $T_p$  is the interval in the unperturbed case, that is,  $h^{\text{syn}}(\mathbf{x}, t) \equiv h_0(\mathbf{x})$ ,



**Fig. 1.20.** Effect of a perturbation. The synaptic potential  $h^{\text{syn}}$  and the effective threshold  $\theta - \eta^{\text{refr}}$ . Neuron  $i$  has fired at time  $t_f$ . It will fire again if the threshold condition  $h^{\text{syn}} = \theta - h^{\text{refr}}$  is fulfilled. In the stationary state  $h^{\text{syn}} = h_0$  is constant (dashed) and firing occurs after a time  $T_p$ . With a perturbation  $\Delta h$  (dotted) the firing is late by a time  $v$ .

and  $v$  is the first-order correction due to the perturbation  $h^1(t)$ , viz.,

$$v(t + T_p) = -\frac{h^1(t + T_p)}{\frac{d}{dt}\eta^{\text{refr}}|_{T_p}}. \quad (1.154)$$

An increase in the synaptic potential leads to a shorter interspike interval and *vice versa*. This is shown graphically in Fig. 1.20.

We now use Eq. (1.153) and expand to first order in  $v$ . The result is a *continuity equation* of the activity [48,49]

$$A(t) - A(t - T_p) = -\frac{d}{dt}[A(t - T_p)v(t)]. \quad (1.155)$$

Here we have dropped the sublattice index  $x$  since we consider in the following only neurons with  $x^\nu = +1$ . Neurons with  $x^\nu = -1$  are assumed to stay quiescent. In other words, we assume that there is a macroscopic overlap with a single pattern only, that is,  $m_\mu = m\delta_{\mu,\nu}$ .

In order to study the stability of the stationary state we now focus on *periodic* perturbations with a frequency  $\omega$ , viz.,

$$A(t) = A_0 + A_1 \exp i\omega t. \quad (1.156)$$

Due to the perturbation  $A_1$  the synaptic potential oscillates according to

$$\begin{aligned} h^{\text{syn}}(t) &= J_0 \int_0^\infty \epsilon(s) A(t - s) ds \\ &= h_0 + |\Delta h| \exp i(\omega t - \alpha) \end{aligned} \quad (1.157)$$

with  $|\Delta h| = J_0 A_1 |\tilde{\epsilon}(\omega)|$  and  $\int_0^\infty \epsilon(s) \exp(-i\omega s) ds = |\tilde{\epsilon}(\omega)| \exp(-i\alpha)$ . The term  $h_0 = J_0 A_0$  is the synaptic potential in the unperturbed case. The

change in  $h^{\text{syn}}$  influences the duration of the interspike interval; cf. Eqs. (1.153) and (1.154). If we substitute the ansatz (1.156) into the continuity equation (1.155), we find

$$[1 - \exp(-i\omega T_p)] = A_0 \frac{J_0|\tilde{\epsilon}(\omega)|}{\frac{d}{dt}\eta^{\text{refr}}|_{T_p}} i\omega \exp(-i\alpha). \quad (1.158)$$

This is the first important result of our stability analysis. The terms on both sides of the equation can be expressed by amplitude and phase of a complex variable which yields [48,49]

$$2|\sin(\omega T_p/2)| = \frac{J_0|\tilde{\epsilon}(\omega)|}{\frac{d}{dt}\eta^{\text{refr}}|_{T_p}} A_0 \omega \quad (1.159)$$

and

$$\begin{aligned} \omega T_p/2 &= \alpha & \text{for } 2k\pi \leq \omega T_p/2 \leq (2k+1)\pi, \\ \omega T_p/2 &= \alpha - \pi & \text{for } (2k+1)\pi \leq \omega T_p/2 \leq (2k+2)\pi. \end{aligned} \quad (1.160)$$

In the second equation we have included a phase shift of  $\pi$  at the value  $\omega T_p/2 = \pi$ .

The above results are correct in the low-noise limit ( $\beta \rightarrow \infty$ ), but what happens for finite  $\beta$ ? In this case the kernel  $P_F^{(2)}(t|t_1^f)$  in Eq. (1.137) is no longer a  $\delta$  function, but has finite width. An ansatz similar to Eqs. (1.156) and (1.157) yields to first order in  $A_1$ :

$$[1 - I_1(\omega)] = A_0 J_0 |\tilde{\epsilon}(\omega)| \exp(-i\alpha) I_2(\omega). \quad (1.161)$$

Here,  $I_1(\omega)$  is the Fourier transform of the *interval distribution* in the stationary state, that is,

$$\begin{aligned} I_1(\omega) &= \int_0^\infty ds \exp(-i\omega s) \tau^{-1}[h_0 + \eta^{\text{refr}}(s)] \\ &\times \exp \left\{ - \int_0^s \tau^{-1}[h_0 + \eta^{\text{refr}}(s')] ds' \right\} \end{aligned} \quad (1.162)$$

which reduces to  $\exp(-i\omega T_p)$  in the limit of  $\beta \rightarrow \infty$ . The term  $I_2(\omega)$  is another integral,

$$I_2(\omega) = \int_0^\infty ds \exp(-i\omega s) \left[ -\frac{d}{ds} S_1(s) \right], \quad (1.163)$$

where

$$\begin{aligned} S_1(s) &= - \left\{ \int_0^s \exp(i\omega s') \frac{d}{dh} \tau^{-1}[h_0 + \eta^{\text{refr}}(s')] ds' \right\} \\ &\times \exp \left\{ - \int_0^s \tau^{-1}[h_0 + \eta^{\text{refr}}(s')] ds' \right\} \end{aligned} \quad (1.164)$$

is the leading term in an expansion of the survivor function to first order in  $\Delta h$ . This is equivalent to saying that  $-\frac{d}{ds}S_1(s)$  is the first-order correction in an expansion of the interval distribution; cf. Eq. (1.98).

In order to get a better understanding of the term  $I_2(\omega)$ , we assume  $\beta \gg 1$  but still finite. In this case the interval distribution has a narrow peak in the neighborhood of  $T_p$  and we can linearize the refractory function so as to get  $\eta^{\text{refr}}(s) = \eta^{\text{refr}}(T_p) + \frac{d}{ds}\eta^{\text{refr}}|_{T_p}(s - T_p)$ . Using  $\tau^{-1}[h_0 + \eta^{\text{refr}}(s')] \approx \tau_0^{-1} \exp\{\beta[\frac{d}{ds}\eta^{\text{refr}}|_{T_p}(s' - T_p)]\}$  we can perform the integration in the first term of Eq. (1.164) and find

$$S_1(s) \approx -\frac{\beta}{\beta \frac{d}{ds}\eta^{\text{refr}}|_{T_p} + i\omega} \exp(i\omega s) P_F^{(2)}(s|0). \quad (1.165)$$

If we use  $\int_0^\infty P_F^{(2)}(s|0)ds = 1$ , we find from Eq. (1.164) through integration by parts:

$$I_2(\omega) = \frac{i\omega}{\frac{d}{ds}\eta^{\text{refr}}|_{T_p} + (i\omega/\beta)} \quad (1.166)$$

which reduces in the limit of  $\beta \rightarrow \infty$  to  $i\omega/\frac{d}{ds}\eta^{\text{refr}}|_{T_p}$ . This is the noise-free result which we have found before; cf. Eq. (1.158).

We now return to the noise-free case and continue with an analysis of Eqs. (1.159) and (1.160). The most important factor on the right-hand side of Eqs. (1.159) and (1.160) is the amplitude and, respectively, phase of the EPSP. If  $\epsilon(s)$  is a delayed  $\alpha$  function, Eq. (1.107), the Fourier integral can be done and yields the amplitude

$$|\tilde{\epsilon}(\omega)| = \frac{1}{(1 + \omega^2\tau_s^2)} \quad (1.167)$$

and phase

$$\alpha = \omega\Delta^{\text{ax}} + 2\arctan\omega\tau_s. \quad (1.168)$$

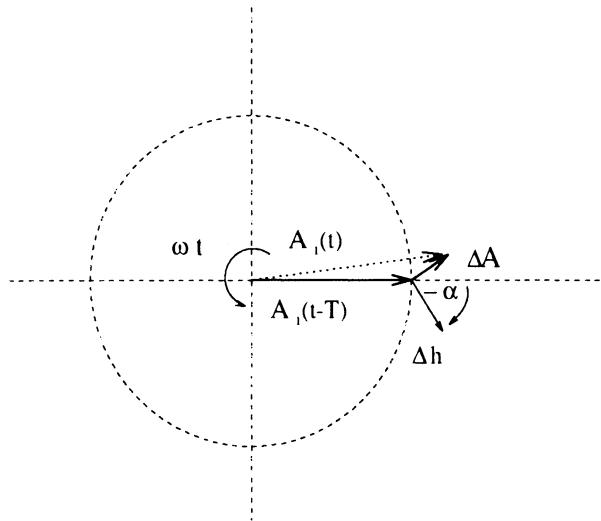
If  $\omega\tau_s$  is large (or small), the phase equation (1.168) reduces to

$$\alpha = \begin{cases} \omega\Delta^{\text{ax}} - 2/(\omega\tau_s) + \pi & \text{for } \omega\tau_s \gg 1 \\ \omega\Delta^{\text{ax}} + 2\omega\tau_s & \text{for } \omega\tau_s \ll 1 \end{cases}. \quad (1.169)$$

For a realistic set of parameters we have  $A_0J_0/\frac{d}{dt}\eta^{\text{refr}}|_{T_p} \approx 1 \text{ ms}$  and  $\omega\tau_s \gg 1$ . The term  $\omega|\tilde{\epsilon}(\omega)|$  vanishes due to Eq. (1.167). Thus the right-hand side of Eq. (1.159) is small. It follows that oscillatory solutions will have a frequency

$$\omega_n = n(2\pi/T_p)(1 + \kappa_n), \quad (1.170)$$

where  $\kappa_n \ll 1$ . The principal mode ( $n = 1$ ) has a period  $T_p$  equivalent to the interspike intervals in the stationary state. The other modes ( $n \geq 2$ ) are higher harmonics.



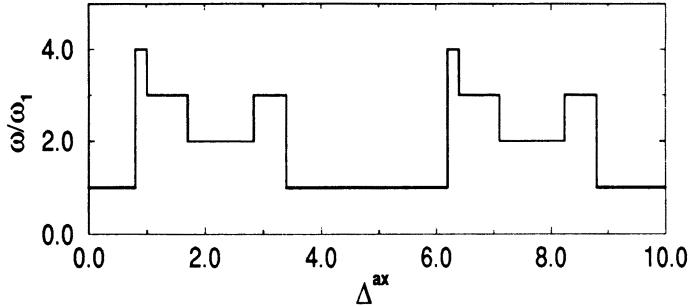
**Fig. 1.21. Phase relations.** At time  $t - T_p$ , the oscillatory part of the activity  $A_1 \exp(i\omega t)$  has phase  $\phi = 0$ . After one period the phase has approximately been shifted by  $2\pi$ . The postsynaptic potential  $\Delta h$  now has a phase  $-\alpha$  and causes a change in activity  $\Delta A$  with phase  $\pi/2 - \alpha$ . The new activity is now  $A_1(t) = A_1(t - T_p) + \Delta A$ . The oscillation is maximally reinforced if  $\alpha = \pi/2$ .

For a given mode  $\omega_n$  the delay  $\Delta^{ax}$  can now be used to match the phase condition (1.160). In the case  $\omega\tau_s \gg 1$ , we find

$$\Delta^{ax}/T_p = (1/2)(1 - 1/n) + (\kappa_n/2n) + (T_p/\tau_s)(1/2\pi^2 n^2)(1 - 2\kappa_n). \quad (1.171)$$

In order to visualize the phase relations we have plotted the various terms of Eq. (1.158) in the complex plane. At time  $t - T_p$  the phase of the oscillatory activity is taken to be  $\phi = 0$ . The perturbation of the activity evokes a synaptic potential with phase  $-\alpha$ ; cf. Eq. (1.157). The change in  $h^{syn}$  causes a shift  $v$  in the interspike interval, Eq. (1.154), and with respect to Eq. (1.155) a change of the activity,  $\Delta A(t) = -A_0 \frac{d}{dt} v(t)$  with phase  $\pi/2 - \alpha$ . It is obvious from Fig. 1.21 that the oscillation will be reinforced maximally, if  $\alpha = \pi/2$ . It decreases if  $\alpha = -\pi/2$ . The solutions of Eq. (1.158) are those values of  $\alpha$  where an oscillation neither increases nor decreases in time.

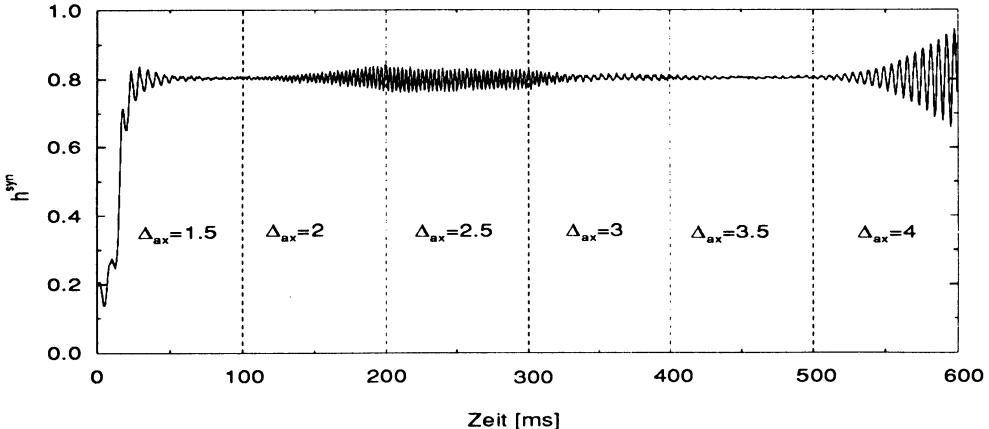
For a given delay  $\Delta^{ax}$ , some of the modes  $\omega_n$  are increasing, others decreasing. As time proceeds, the modes develop with different growth factors. The mode  $\omega_k$  with the largest growth factor will be called the *dominant mode*. In Fig. 1.22 the frequency of the dominant mode is plotted as a function of the delay  $\Delta^{ax}$ . The parameters of the specific model system are  $\gamma^{refr} = 4$  ms,  $J_0 = 4.5$  ms, and  $\eta_0 = 1$  ms. This set of parameters yields an interspike interval  $T_p = 5.5$  ms in the stationary state which accounts for the periodicity in the stability diagram.



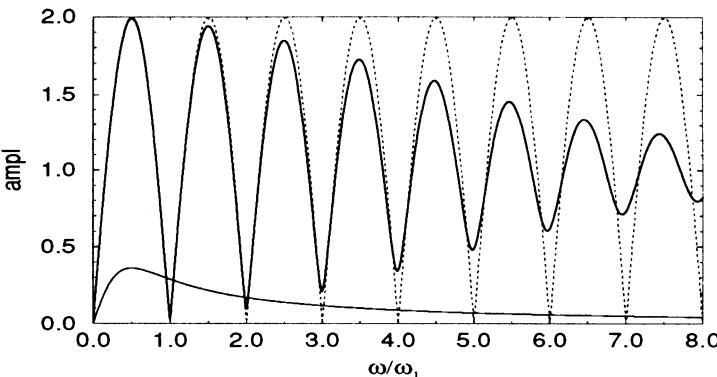
**Fig. 1.22.** *Stability diagram in the noise-free limit.* The frequency of the dominant oscillatory mode is plotted as a function of the delay  $\Delta^{0x}$ . Whatever the delay, the stationary state is unstable. For some delays, oscillations start with the principal mode  $\omega_1$  whereas for others with higher harmonics ( $\omega_n = n\omega_1$  with  $n \geq 2$ ). In real systems with noise higher harmonics are suppressed. Here the parameter values are  $\gamma^{\text{refr}} = 4$  ms,  $J_0 = 4.5$  ms, and  $\eta_0 = 1$  ms. (Adapted from [49].)

To compare our noise-free analysis with a real system, we have simulated a network of  $N = 1000$  neurons with the above parameters at a finite noise level ( $\beta = 20$ ). The result is shown in Fig. 1.23. For  $\Delta^{0x}$  in a range between 2 and 2.5 ms, a fast  $n = 2$  oscillation is seen, for  $\Delta^{0x} > 3.5$  the  $n = 1$  mode is dominant. Higher harmonics ( $n \geq 3$ ) are suppressed by noise. The suppression of fast oscillations is explained by the stability analysis of the noisy system; cf. Eqs. (1.161)–(1.166). Figure 1.24 shows the amplitude condition as derived from Eqs. (1.158) and (1.161) for a noise-free and, respectively, noisy system. For the latter graph, we have assumed a noise level of  $\beta = 15 \gg 1$  and used the limit  $|I_2(\omega)| = \omega / \frac{d}{ds} \eta^{\text{refr}}|_{T_p}$ , as derived from Eq. (1.166). As before, we find for our set of model parameters  $A_0 J_0 / \frac{d}{ds} \eta^{\text{refr}}|_{T_p} \approx 1$  leading to the amplitude condition  $|1 - I_1(\omega)| = \omega / (1 + \omega^2 \tau_s^2)$ ; cf. Eq. (1.161). This is the relation which has been used in Fig. 1.25. The results show that higher harmonics ( $n > 2$ ) cannot occur for  $\beta = 15$  (solid) whereas they are possible for  $\beta \rightarrow \infty$  (dotted).

As a summary of this subsection the following important results should be kept in mind. First, stationary solutions exist and can be found by a straightforward mean-field approach. Second, these solutions need not be stable. Indeed, a stability analysis shows that in the noise-free case the stationary state is always *unstable*, whatever the delay  $\Delta^{0x}$ . Thus *noise plays an extremely important role in stabilizing the system*. For a given noise level, the system switches from unstable to stable and back to unstable behavior, if the delay  $\Delta^{0x}$  is increased. For a given delay, the system can be tuned from incoherent to coherent firing, if the noise level is reduced. Similar results hold, if not the delay, but the width of the EPSP is changed.



**Fig. 1.23.** *Simulation results.* The postsynaptic potential  $h^{\text{syn}}$  is plotted as a function of time. Every 100 ms the delay  $\Delta^{\text{ax}}$  has been increased by 0.5 ms. In the stationary state, active neurons fire regularly with rate  $T_p^{-1} = 1/5.5 \text{ ms}$ . For a delay  $\Delta^{\text{ax}} > 3.5 \text{ ms}$ , oscillations with period  $\omega_1 = 2\pi/T_p$  build up rapidly. For intermediate delays  $2 \leq \Delta^{\text{ax}} \leq 2.5$  small-amplitude oscillations with twice the frequency occur. Higher harmonics are suppressed by noise ( $\beta = 20$ ). Simulation results should be compared with Fig. 14 and the theoretical prediction given in the text. The parameters are as in Fig. 14.



**Fig. 1.24.** *Amplitude condition.* The Fourier transform of the EPSP multiplied by  $\omega$ , viz.,  $\omega|\tilde{\epsilon}(\omega)| = \omega/(1 + \omega^2 \tau_s^2)$  (thin solid line), and the Fourier transform of the interval distribution  $I_1(\omega)$  (shifted by 1 along the  $y$  axis, viz.,  $1 - I_1(\omega)$ ) are plotted as a function of  $\omega/2\pi T_p^{-1}$  (fat line,  $\beta = 15$ ; dashed, noise-free system). The intersection points yield the frequency of possible instabilities; cf. Eqs. (1.161) and (1.166). For  $\beta = 15$  only first- and second-order instabilities may occur whereas in the noise-free case all higher harmonics are possible as well.

### 1.4.2 COHERENT FIRING — OSCILLATORY SOLUTIONS

So far we have restricted our arguments to incoherent firing. In this case, the sublattice activity  $A(\mathbf{x}, t)$  is constant in time and both activity and firing rate are equivalent ways of describing the network state. The above stability analysis of the stationary state, however, shows that a state of incoherent firing is not always stable and that oscillations may occur. If these oscillations remain small, firing is still mainly incoherent and has only a small coherent part. It can therefore be called a partially coherent state.

In this subsection we will consider a completely different state which can be characterized as a perfectly coherent or locked oscillation. We will present an analytical expression for the period of coherent oscillations and analyze its stability. Comparison of the two approaches, viz., instability of the stationary state and stability of the coherent state, reveals that both types of solution are completely independent of each other. Indeed, simulations show that partially and perfectly coherent solutions can coexist in the very same network [49]. Different initial conditions allow one to activate either one or the other solution.

#### Existence and stability of collective oscillations

We analyze coherent firing in the limit of  $\beta \rightarrow \infty$ . In this case the kernel  $P_F^{(2)}(t|t_1^f)$  in Eq. (1.137) reduces to

$$\lim_{\beta \rightarrow \infty} P_F^{(2)}(t|t_1^f) = \delta[s - s_F(\mathbf{x})], \quad (1.172)$$

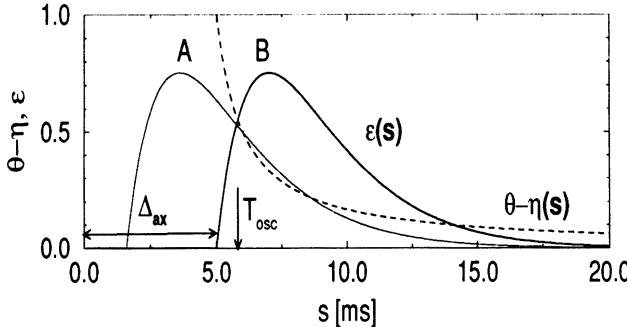
where  $s_F$  is given by the threshold condition

$$s_F(\mathbf{x}) = \inf \{s | \eta^{\text{refr}}(s) + h^{\text{syn}}(\mathbf{x}) = \theta\}. \quad (1.173)$$

This condition can be used to derive the period  $T_{\text{osc}}$  of a coherent oscillation. In order to do so, we assume that all neurons which are active during retrieval of pattern  $\nu$  ( $\xi_i^\nu = +1$ ) have fired regularly and coherently in the past. To be specific, we assume that firing has occurred at times  $t = 0, -T_{\text{osc}}, -2T_{\text{osc}}, \dots$ . We also assume that neurons with  $\xi_i^\nu = -1$  have been quiescent in the past. Thus we require that a collective oscillation with period  $T_{\text{osc}}$  has existed in the past. In the following the period  $T_{\text{osc}}$  will be determined self-consistently.

With the above assumptions the activity on a sublattice with  $x^\nu = +1$  is

$$A(x^\nu = +1, t) = \sum_{n=0}^{\infty} \delta(t + nT_{\text{osc}}). \quad (1.174)$$



**Fig. 1.25.** Determination of the period of a coherent oscillation. The intersection of the scaled EPSP  $\epsilon(s)$  (solid line) with the increased threshold (dashed line)  $\theta - \eta^{\text{refr}}(s)$  yields the oscillation period  $T_{\text{osc}}$ . It is stable if the slope of  $\epsilon(s)$  is positive at the intersection, as in case B.

Due to Eq. (1.117) or (1.130) the coherent activity will evoke a postsynaptic potential:

$$h^{\text{syn}}(t) = \begin{cases} J_0 \sum_{n=0}^{\infty} \epsilon(t + nT_{\text{osc}}) & \text{for } x^\nu = +1 \\ -J_0 \sum_{n=0}^{\infty} \epsilon(t + nT_{\text{osc}}) & \text{for } x^\nu = -1 \end{cases}. \quad (1.175)$$

For  $t > 0$  the refractory potential is ( $F = 1$ )

$$h^{\text{refr}}(t) = \begin{cases} \eta^{\text{refr}}(s)\delta(t - s) = \eta^{\text{refr}}(t) & \text{for } x^\nu = +1 \\ 0 & \text{for } x^\nu = -1 \end{cases}. \quad (1.176)$$

We now must check whether the oscillation will continue for  $t > 0$ . For this to be true the neurons with  $x^\nu = +1$  must fire again at  $t = T_{\text{osc}}$ . In other words, the threshold condition (1.173) must be fulfilled at  $t = T_{\text{osc}}$ . This implies

$$J_0 \sum_{n=1}^{\infty} \epsilon(nT_{\text{osc}}) + \eta^{\text{refr}}(T_{\text{osc}}) = \theta. \quad (1.177)$$

Neurons with  $x^\nu = -1$  should stay quiescent and thus

$$- J_0 \sum_{n=0}^{\infty} \epsilon(t + nT_{\text{osc}}) < \theta \quad \text{for } t \geq 0. \quad (1.178)$$

In case  $\theta > 0$ ,  $J_0 > 0$ , and  $\epsilon(s) \geq 0$  for all  $s > 0$ , Eq. (1.178) is always true. Equation (1.177) can be used to determine  $T_{\text{osc}}$ .

For a simple graphical interpretation of Eq. (1.177) we assume that the EPSP is shorter than or comparable to the period  $T_{\text{osc}}$ . To be specific, we require  $\sum_{n=2}^{\infty} \epsilon(nT_{\text{osc}}) \ll \epsilon(T_{\text{osc}})$ . In this case, Eq. (1.177) can be reduced to

$$J_0 \epsilon(T_{\text{osc}}) + \eta^{\text{refr}}(T_{\text{osc}}) = \theta. \quad (1.179)$$

The graphical solution is indicated in Fig. 1.25. The intersection of  $J_0\epsilon(s)$  with the effective threshold  $\theta - \eta^{\text{refr}}(s)$  yields the period  $T_{\text{osc}}$ . In case of several intersections,  $T_{\text{osc}}$  is given by the *first* crossing point (shortest interval  $s$ ) since neurons fire the first time that threshold is reached; cf. the threshold condition (1.173).

A condition similar to Eq. (1.177) can be derived for adaptive neurons ( $F > 1$ ). The generalized result is

$$J_0 \sum_{n=1}^{\infty} \epsilon(nT_{\text{osc}}) + \sum_{n=1}^F \eta^{\text{refr}}(nT_{\text{osc}}) = \theta. \quad (1.180)$$

From Eq. (1.180) we can find the period  $T_{\text{osc}}$  of collective oscillations in a network of adaptive neurons ( $F > 1$ ) connected by synapses with an arbitrary response function  $\epsilon(s)$ .

Once we know that a locked oscillation exists, we have to check its stability. In order to do so, we assume that all neurons with  $x^\nu = +1$  have fired coherently at  $t = 0$  and  $t = -nT_{\text{osc}}$  with  $n \in \mathbf{N}$ , — except neuron  $j$  which has been slightly delayed and has fired at  $t = \Delta t$  instead of  $t = 0$ . Stability requires that the delay  $|\Delta t|$  has decreased after one period. In other words, the neuron should be “pulled back” into the coherent oscillation.

Neuron  $j$  which has fired at  $t = \Delta t$  fires again, if the threshold condition

$$h_j^{\text{syn}}(t') + h_j^{\text{refr}}(t') = \theta \quad (1.181)$$

is fulfilled the next time. The contribution  $h^{\text{syn}}$  is the postsynaptic potential evoked by all the other neurons and is identical to the unperturbed case. The term  $h_j^{\text{refr}}$ , however, is different due to the delay of neuron  $j$  and equals

$$h_j^{\text{refr}}(t) = \eta^{\text{refr}}(t - \Delta t) \quad (1.182)$$

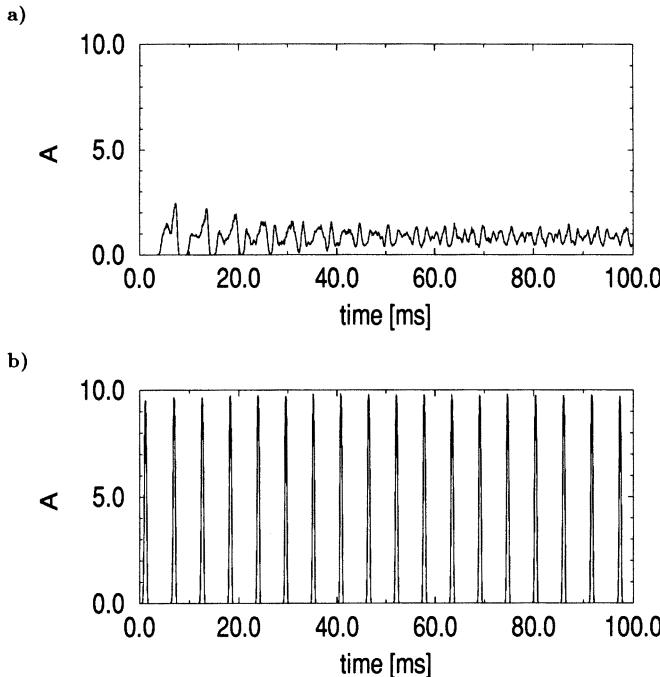
if  $F = 1$ . Expansion of Eqs. (1.181) and (1.182) to first order in  $\Delta t$  yields

$$\frac{t' - T_{\text{osc}}}{\Delta t} = \frac{d}{dt} \eta^{\text{refr}}|_{T_{\text{osc}}} \left[ \frac{d}{dt} h^{\text{syn}}|_{T_{\text{osc}}} + \frac{d}{dt} \eta^{\text{refr}}|_{T_{\text{osc}}} \right]^{-1}. \quad (1.183)$$

For a standard refractory function, for example, Eq. (1.88), we have  $\frac{d}{ds} \eta^{\text{refr}}(s) > 0$  for all  $s > \gamma^{\text{refr}}$ . In this case, we arrive at the simple locking condition

$$\left\{ \frac{d}{dt} h^{\text{syn}}|_{T_{\text{osc}}} > 0 \right\} \iff \{(t' - T) < \Delta t\}. \quad (1.184)$$

A graphical interpretation of this result is given in Fig. 1.25. An oscillation is stable if the first intersection of  $J_0\epsilon(s)$  with  $\theta - \eta^{\text{refr}}(s)$  occurs at a point where  $\epsilon(s)$  has positive slope. If  $\epsilon(s)$  is decreasing, a collective oscillation is unstable. These results have been tested in a simulation of a network of 1000 neurons at a finite noise level ( $\beta = 20$ ). The result is shown in Fig. 1.26. In passing we note that a similar, though slightly more involved, argument applies to the case where the neurons have a purely *inhibitory* interaction and, in addition, an excitatory input.



**Fig. 1.26. Stability of the coherent oscillation.** We start the system in a perfectly coherent state. (a) For short axonal delay,  $\Delta^{ax} = 1.6$  ms (case A of previous figure), the oscillation decays, and a stationary state is assumed. (b) For  $\Delta^{ax} = 5$  ms locking is possible and a stable coherent oscillation emerges. (Activity A normalized to a maximum of 10/ms.)

### Information processing by means of oscillations

In the introduction of this chapter, we pointed out that coherent firing may be important in the context of feature linking and figure-ground segregation. Experimental data show that the period of coherent activity bursts is in the range of 20–30 ms. Our simple model network, however, predicts a much shorter oscillation period of about 5–6 ms — so what is wrong?

So far we have considered a network of *identical* neurons, all connected by Hebbian synapses. In biological systems it is known that several types of neuron exist, for example, pyramidal and stellate cells. Stellate cells are found to have local connections that mostly end at inhibitory synapses. If we take a theorist's point of view, we can mimic the local inhibitory interaction by connecting each neuron of our network with an inhibitory partner neuron. If one of the neurons fires, its partner neuron is excited and sends an inhibitory signal back to the first neuron. The inhibitory feedback can be included in our ansatz by replacing the refractory function  $\eta^{\text{refr}}$  by

a more generalized inhibitory function:

$$\eta^{\text{refr}}(s) \implies \eta^{\text{inh}}(s) \quad (1.185)$$

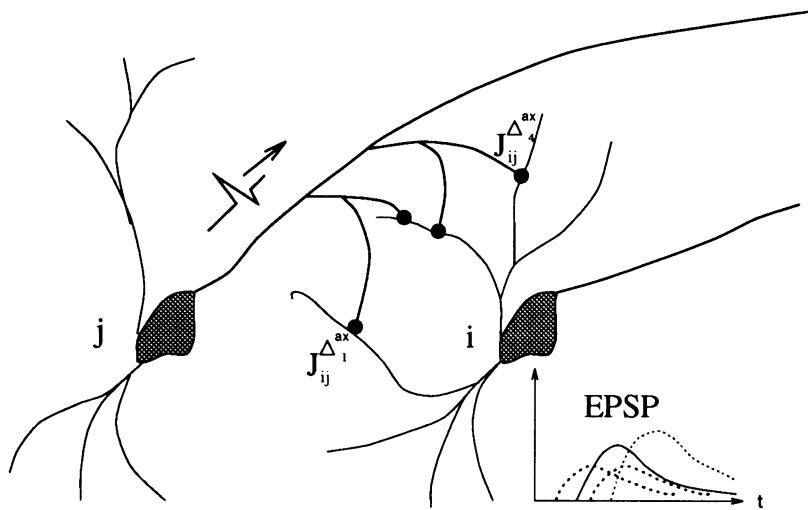
that includes both refractoriness and inhibition. Due to inhibition which may last over 10–20 ms the new  $\eta$  function has a broad regime of highly negative values. The net effect is a longer oscillation period of 20–25 ms which is within the experimental range. All results concerning existence and stability of locked solutions can be transferred into the new system. The above arguments can also be generalized so as to include neurons with different time constants of refractoriness, axonal transmission, and inhibition [46]. The application of the generalized model to the problem of binding and pattern separation is presented in the paper by Ritz et al. in this volume [108].

## 1.5 Hebbian Learning of Spatio-Temporal Spike Patterns

In the preceding sections we have considered incoherent and coherent firing. Incoherent states can be described by the mean activity of an appropriately chosen ensemble or else by the mean firing rate of a typical neuron in the ensemble. Coherent states are often called collective oscillations if all active neurons fire periodically and *in synchrony*; cf. Sec. 1.2 and 1.4.2 of this chapter. Locked oscillations can be considered as a simple case of a time-dependent network state — all neurons fire with the same temporal time structure.

In this section we will study a more general class of time-dependent states where different neurons fire at different times [47]. Thus, if we plot the action potentials of different neurons as a function of time, we find a complex spatio-temporal spike pattern. If these states exist, two questions can be asked: First, is the spatio-temporal excitation pattern predetermined by the network structure or is it possible to *learn* fairly arbitrary spatio-temporal states by a Hebbian learning rule? Second, if an arbitrary time-dependent excitation is indeed possible — what is the maximum time resolution in such a system state?

With respect to the fairly long postsynaptic integration time of 5–50 ms it has often been argued that the time resolution is limited by the duration of a single EPSP. Here we show that a resolution down to 1 ms can be realized and used to distinguish different spike patterns which cannot be classified by means of an evaluation scheme based on mean firing rates or averaged activity.



**Fig. 1.27.** *Neuronal connectivity and signal transmission.* Each pair of neurons is connected by several synapses with different synaptic delay  $J_{ij}^{\Delta ax}$ . The four excitatory postsynaptic potentials received at the soma of  $i$  due to a spike of neuron  $j$  are shown on the right. Note the different delays. (From [48].)

### 1.5.1 SYNAPTIC ORGANIZATION OF A MODEL NETWORK

In the preceding sections we have considered a homogeneous network where each neuron pair  $(i, j)$  is connected by a single synapse for signal transmission from  $j$  to  $i$  and another one from  $i$  to  $j$ . The synaptic efficacy as calculated by the “Hebbian” rule (1.110) is independent of the axonal delay  $\Delta^{ax}$ . In the following, the delay will be considered not only during retrieval, but also during learning. As a result, the efficacy of a connection from  $j$  to  $i$  depends on the delay  $\Delta^{ax}$ , that is,  $J_{ij} \rightarrow J_{ij}^{\Delta ax}$ . The exact formula will be explained in the next subsection. In the following we consider  $J_{ij}^{\Delta ax}$  as given.

For the sake of simplicity, we assume that each neuron pair is connected by several synapses; cf. Fig. 1.27. (How this condition can be dropped will be indicated below.) They are located at different positions along the axon of the presynaptic neuron. If one of the neurons — say neuron  $j$  — fires, a spike is transmitted along the axon and induces — after a delay  $\Delta^{ax}$  — a postsynaptic signal at the receiving neuron  $i$ . The delay  $\Delta^{ax}$  depends on the position of the specific synapse on the axon and varies in our simulation between  $\Delta_{\min} = 1$  ms and  $\Delta_{\max} = 4$  ms. To phrase it differently: Given neuron  $j$  fires, neuron  $i$  is affected after 1, 2, 3, and 4 ms but the synaptic efficacies may, and in general will, differ for the four delays. Such a range of delays may seem a rather implausible assumption since synaptic contacts of a single neuron to another neuron are unlikely to span time differences

as large as 4 ms. The model can, however, be changed easily so as to allow for dilution such that three of the four connections are cut at random. The remaining connection has only one delay with  $1 \leq \Delta^{\text{ax}} \leq 4$  ms which seems to be a realistic range. The important point is that the synaptic efficacy of a connection from  $j$  to  $i$  depends not only on  $j$  and  $i$  but also on the *delay*. This will be explained below.

The synaptic signals received by neuron  $i$  can be added and give the total postsynaptic potential

$$h^{\text{syn}} = \sum_{j=1}^N \sum_{\Delta^{\text{ax}}=\Delta_{\min}}^{\Delta_{\max}} \sum_{f=1}^{n_f} J_{ij}^{\Delta^{\text{ax}}} \int_0^\infty ds \epsilon(s) \delta(t - s - \Delta^{\text{ax}} - t_j^f). \quad (1.186)$$

Here  $\epsilon(s) = (s/\tau_s) \exp(-s/\tau_s)$  with  $\tau_s = 3$  ms describes the shape of a single EPSP and — in contrast to Eq. (1.107) — the axonal delay shows up now in the argument of the  $\delta$  function. Note that, for the sake of simplicity, we have assumed in Eq. (1.186) that all contributions add up linearly but it is straightforward to allow for more complicated interactions.

In addition to the excitatory synaptic contribution we include a global inhibitory signal which is identical for all neurons,

$$h_i^{\text{inh}}(t) = - \sum_{j=1}^N \sum_{\Delta^{\text{ax}}=\Delta_{\min}}^{\Delta_{\max}} \sum_{f=1}^{n_f} J^{\text{inh}} \int_0^\infty ds \epsilon(s) \delta(t - s - \Delta^{\text{ax}} - t_j^f). \quad (1.187)$$

In contrast to Eq. (1.186), the inhibitory synaptic efficacy  $J^{\text{inh}}$  does not depend on local properties of neurons and synapses. It is adjusted according to the firing probability  $p$  in the spike patterns the system has been trained on. During the simulations in the following subsections it is set to  $J^{\text{inh}} = p = 0.025$ .

Equations (1.186) and (1.187) show that the postsynaptic neuron sums over the signals from all other neurons (spatial average) and integrates over the past (temporal average). Based on this observation, it has often been argued that quantities that average over space (ensemble activity) or time (mean firing rate) should be sufficient to describe the states of the network. The results of this section will, however, show that the network is capable of storing and retrieving patterns with a high resolution in space and time — despite the postsynaptic averaging. This clearly underlines the potential relevance of nonaveraged or local information in neural networks.

In order to store those time-resolved excitation patterns, the efficacies have to be adjusted according to a Hebbian learning rule. How this can be done will be explained now.

### 1.5.2 HEBBIAN LEARNING

As a basic principle of synaptic learning, Hebb postulated in 1949 that synaptic efficacies are enhanced during periods of simultaneous activity

of the presynaptic and postsynaptic neurons [55]. Long-term potentiation (LTP) which has been found in hippocampal brain slices seems to support this postulate [19,76]. In most experiments on LTP and in many models of Hebbian learning the terminology of “simultaneous activity” is understood as a time-averaged quantity, for example, mean firing rate or mean membrane voltage. In our approach we shift emphasis to single spikes and EPSPs and consider the momentary state of the pre- and postsynaptic membrane directly at the synapse. The main idea is that only those quantities which are locally available *at the synapse* can lead to a change of synaptic efficacy [64,65].

In order to make these ideas more precise, we consider a synapse  $J_{ij}^{\Delta ax}$  from neuron  $j$  to neuron  $i$ ; cf. Fig. 1.28. If neuron  $j$  fires at time  $t_j^f$ , a spike is transmitted along the axon and arrives after a time  $\Delta^{ax}$  at the synapse  $J_{ij}^{\Delta ax}$ . Here it induces the release of neurotransmitter substance which finally evokes an EPSP at the postsynaptic neuron  $i$ . Let us now assume that neuron  $i$  has also fired, but at a different time  $t_i^f$ . When and how does information about the postsynaptic spiking arrive at the synapse? As a solution to this question we tentatively suggest that firing causes not only emission of an action potential along the axon, but also some kind of *dendritic* pulse that spreads by (active or passive) transport over the dendritic tree. While active transport is known in the hippocampus [134], its existence in the neocortex is somewhat still unclear. The relevance of these pulses to synaptic coding has not been discussed so far. The important point in this context is that *active* processes would allow the pulse to remain a pointlike event, *local in space and time*. A synapse could thus receive precise time resolved information on postsynaptic spiking.

As to synaptic learning we propose that a strengthening of a synapse occurs only if the dendritic pulse arrives at the synapse concurrently with (or slightly after) the neurotransmitter (or some other messenger substance) released from the presynaptic terminal. At this point this assumption is purely speculative. Most experiments have not studied the detailed time structure of LTP, but rather mean activity. Let us assume that the arrival of transmitter substance initiates some chemical activation process necessary for the induction of LTP. If we suppose that the activation reaches its maximum after a delay of  $\Delta^{chem} \approx 1$  ms and declines sharply afterwards, then this process defines a narrow time window  $\nu(t)$  for a match with the dendritic pulse

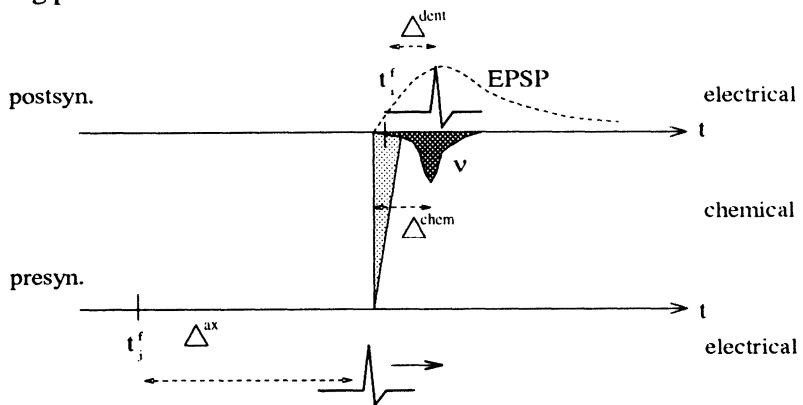
$$\nu(t) = \exp\left[-\frac{1}{2}\left(\frac{t - \Delta^{chem}}{\tau_{chem}}\right)^2\right]. \quad (1.188)$$

See also Fig. 1.28. Synaptic efficacy is enhanced only if the signal of the dendritic pulse arrives during the time of transmitter release  $\nu(t)$  due to a presynaptic spike. This is our precise interpretation of the phrase simultaneous activity.

Adaptation of synaptic efficacies is done during a separate learning ses-

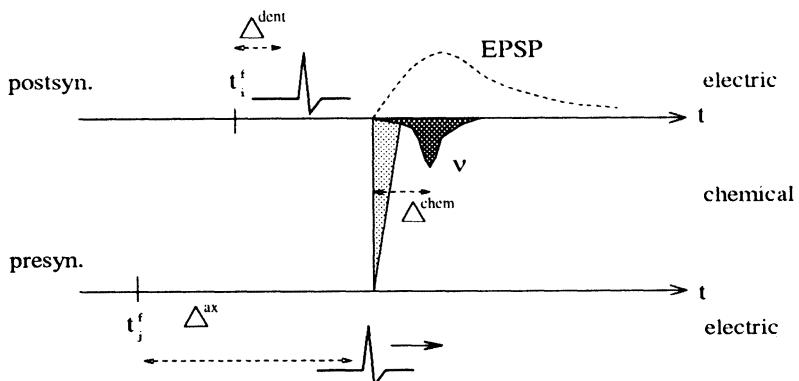
a)

strong potentiation



b)

no potentiation



**Fig. 1.28. Hebbian learning at a synapse.** A synapse is strengthened only if the presynaptic signal and a postsynaptic “dendritic” pulse arrive simultaneously at the postsynaptic side of the synapse. In both (a) and (b) the pre- and postsynaptic neurons fire at  $t_j^f$  and  $t_i^f$ , respectively. When an action potential of  $j$  arrives at a synapse to  $i$  we get a neurotransmitter release (dotted region) which initiates a narrow time window (shaded) during which the dendritic pulse due to  $i$ 's firing has to arrive in order to strengthen the synapse. In (a) dendritic pulse and the time window coincide and potentiation of the synapse occurs whereas in (b) neuron  $i$  fires too early and there is no effect. (From [47].)

sion of total length  $T = \sum_{\mu=1}^q T^\mu$ . During the training period  $T^\mu$  a spatio-temporal excitation pattern  $\mu$  with  $1 \leq \mu \leq q$  is impressed on the network. Let  $t_i^f(\mu)$  and  $t_j^g(\mu)$  denote the firing times of the post- and presynaptic neuron. The presynaptic spike needs a time  $\Delta^{ax}$  to travel to the synapse  $J_{ij}^{\Delta^{ax}}$ , the postsynaptic pulse arrives there after a time  $\Delta^{dent}$ . To keep arguments simple we assume that the dendritic delay  $\Delta^{dent}$  is independent from the position of the synapse on the dendritic tree and relatively short, say  $\Delta^{dent} = 1$  ms. This assumption is supported by experimental findings of Stuart and Sakmann [126]. In view of the above considerations, a change of efficacy for a synapse with delay  $\Delta^{ax}$  is — after the presentation of pattern  $\mu$  — given by

$$\begin{aligned} \delta J_{ij}^{\Delta^{ax}}(\mu) \sim & \sum_{f=1}^{F_i^\mu} \int_0^{T^\mu} dt \delta[t - \Delta^{dent} - t_i^f(\mu)] \\ & \times \sum_{g=1}^{F_j^\mu} \int_0^\infty ds \nu(s) \delta[t - \Delta^{ax} - t_j^g(\mu) - s]. \end{aligned} \quad (1.189)$$

The total number of spikes neuron  $i$  has to fire for generating the pattern  $\mu$  is denoted by  $F_i^\mu$ . The first  $\delta$  function on the right-hand side of Eq. (1.189) describes the *postsynaptic* activity at the location of the synapse, the second  $\delta$  function the *presynaptic* activity. A strengthening of the synapse occurs only if both activities coincide within a time window defined by  $\nu(s)$ . The integrals over time can be carried out yielding

$$\delta J_{ij}^{\Delta^{ax}}(\mu) = \alpha^\mu \sum_{f=1}^{F_i^\mu} \sum_{g=1}^{F_j^\mu} \nu[t_i^f(\mu) - t_j^g(\mu) + \Delta^{dent} - \Delta^{ax}], \quad (1.190)$$

where  $\alpha^\mu$  is some proportionality factor. It is a free parameter and can be used to normalize the right-hand side of Eq. (1.190). To keep things simple we assume that all neurons have the same mean firing rate and fire exact  $F_i^\mu = F_j^\mu \equiv F^\mu$  spikes during the training session  $T^\mu$ . A useful normalization is then  $\alpha^\mu = 1/F^\mu$ .

Several spatio-temporal excitation patterns can be subsequently learned by the same procedure. We assume that the final strength of the synapse is given by a *linear superposition* of all contributions:

$$J_{ij}^{\Delta^{ax}} = \sum_{\mu=1}^q \delta J_{ij}^{\Delta^{ax}}(\mu) = \sum_{\mu=1}^q \sum_{f,g=1}^{F^\mu} \frac{1}{F^\mu} \nu[t_i^f(\mu) - t_j^g(\mu) + \Delta^{dent} - \Delta^{ax}]. \quad (1.191)$$

After the training session the efficacies are kept at a fixed value. Retrieval of the excitation patterns can be induced by a short external signal as is shown in the next subsection.

Before proceeding to simulations of a special system we would like to explain how the system works in general. A neuron which has fired at  $t = 0$  is inhibited thereafter by its own refractory field. As a consequence it has a reduced excitability (higher threshold). Only if the postsynaptic potential  $h^{\text{syn}}$  is strong enough to exceed the effective threshold  $h^{\text{syn}} > \theta - h^{\text{refr}}$ , it can fire again. The postsynaptic potential,  $h^{\text{syn}}$ , is the sum of all incoming EPSP's. Let us now assume that the network has to learn a spike pattern where neuron  $i$  should fire again at  $t_i^f$ . The Hebb rule (1.191) enhances *only those* connections where the EPSP has the correct timing so as to cause  $i$  to fire at  $t_i^f$ . This argument shows that the above learning rule should allow to store fairly arbitrary spike patterns. For the sake of simplicity we have restricted our simulations to the special case of cyclic patterns. The main results, however, are much more general.

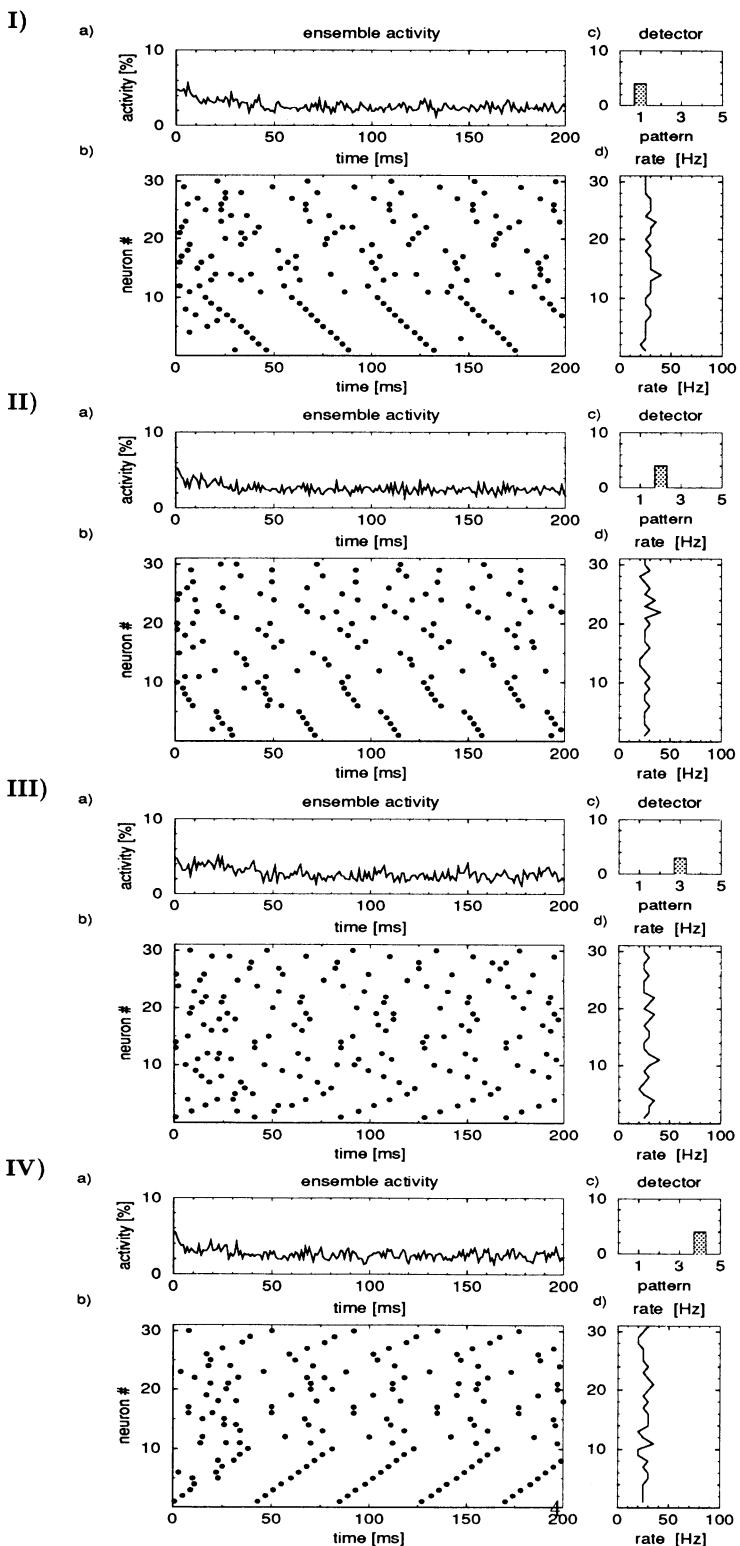
### 1.5.3 LOW-ACTIVITY RANDOM PATTERNS

A network of  $N = 1000$  adaptive neurons ( $F = 10$ ) has been trained on  $q = 4$  patterns. The procedure easily takes care of many more patterns but  $q = 4$  is just convenient for simulation. Each pattern consists of a sequence of spikes from different neurons during a time window of  $T = 40$  time steps, that is, 40 ms. The sequence is then repeated. A spike pattern  $\mu$  ( $1 \leq \mu \leq q$ ) is defined here by exactly one firing time  $t_i^f(\mu)$  for each single neuron:  $1 \leq i \leq N$ . The firing time is drawn from a random distribution with equal probability  $p = 0.025$  for all times  $1 \leq t_i^f \leq 40$ . Thus, in an ideal and noiseless pattern all neurons fire regularly with a rate of 25 Hz, but the firing of different neurons is randomly correlated.

During the training session all spike patterns  $1 \leq \mu \leq q$  are impressed on the neurons and the synaptic efficacies are adjusted according to Eq. (1.191). In order to check whether the patterns are now stable attractors of the neuronal dynamics, retrieval of the patterns has to be studied. A retrieval session is started by a short external stimulus of duration  $t_{\text{init}} = 5$  ms. It consists of a spatio-temporal sequence of short pulses  $h_i^{\text{ext}}(t) = \delta[t, t_i^f - T]$  for  $-5 \leq t \leq 0$  ms. In other words, the network is initialized during 5 ms in a state consistent with one of the learned patterns. The pattern  $\mu$  that is matched should be completed and cyclically retrieved afterwards.

The results of four retrieval sessions with different stimuli are shown in Fig. 1.29. For the four patterns (I–IV), the ensemble activity (spatial average) during retrieval is plotted in (a), the spike raster of a few selected neurons during the retrieval session is shown in (b). In (c), the spike pattern is classified and, in (d), the mean firing rate of the neurons of (b) is calculated.

Let us first turn to the *spike raster*, Fig. 1.29, Ib–IVb. In a way similar to experimental multielectrode recordings, we have selected 30 neurons whose label has been plotted along the  $y$  axis. Time is plotted along the  $x$  axis.



**Fig. 1.29.** Retrieval of spatio-temporal spike patterns. See text for details. (From [47].)

The origin  $t = 0$  marks the *end* of the stimulus, thus for all  $t \geq 0$  we have  $h_i^{\text{ext}}(t) = 0$ . All spikes of neuron # $i$  appear as black dots along a line parallel to the  $x$ -axis. If we count the number of spikes along this line and divide by the total recording time ( $T = 200$  ms), we find the *mean firing rate* of the neuron. This is the quantity which is plotted in (d), to the right of the spike raster. We see that all neurons have approximately the same firing rate, 25 Hz. Thus, if we consider the mean firing rate only, we can not detect any significant structure in the firing behavior of the neurons.

Instead of averaging over time we can also average over space to define the ensemble activity of the network. If we count the number of spikes in every ms (along a vertical line in the spike raster) and divide by the total number of neurons, we find the *ensemble activity* plotted in (a). We see that immediately after the stimulus the ensemble activity is high ( $\approx 5\%$ ), but due to adaptation it decreases. After 50 ms it has settled to an average of 2.5% and no significant structure is left. Nevertheless, if we look at the spike raster (b), we see that the network remains in a regular firing state. The specific spike pattern has been induced by the stimulus and is different for Figs. 1.29, I–IV.

For ease of visualization of this fact, we have used a little trick. Neurons with index #1 to #10 did not learn random patterns but “meaningful” objects like diagonal stripes; cf. the spike raster Ib–IVb. Ten neurons hardly disturb a network of  $N = 1000$  neurons, but they can help the human reader to find regularities and to recognize different patterns.

The *pattern detector* (c) distinguishes the patterns through a mechanism based on the correlations between the spikes of all neurons in the net. A pattern  $\mu$  is detected, if the firing times during the last 40 ms match the time shifted spike raster of pattern  $\mu$ . To be more specific, we define a correlation function

$$\text{corr}^\mu(t) = \frac{1}{N} \sum_{i=1}^N \sum_{\tau=0}^{39} \delta[t - t^f, \tau] \delta[40 - t_i^f(\mu), \tau], \quad (1.192)$$

where  $\delta[x, y]$  denotes the Kronecker  $\delta$  with  $\delta[x, y] = 1$  for  $x = y$  and 0 otherwise. A detector signal  $D^\mu = +1$  is given, if  $\text{corr}(t) \geq 0.5$  during the simulation run ( $0 < t < 200$ ). Thus we require that 50% of the spikes must be correlated exactly with spike pattern  $\mu$ .

If we analyze the series of Figs. 1.29 I–IV, a number of conclusions can be drawn. First of all, it is indeed possible to store and retrieve spatio-temporal spike patterns with a time resolution of 1 ms in a neural network with biological constraints. This is quite remarkable in view of the typical duration of an EPSP (approximately 5–15 ms) which is much longer. Second, several patterns can be stored in the same network. In analogy to the storage capacity calculations of Gardner [42] we expect that the number of spatio-temporal spike patterns of duration  $T$  that can be stored in an *optimally* designed network of  $N$  neurons is proportional to  $NC/(Tp \ln|p|)$

where  $C$  is the connectivity per neuron pair and  $p \ll 1$  is the globally averaged activity of a typical spatio-temporal pattern. From the static case it is known that optimal storage requires a long iterative learning procedure [80] whereas a standard single-step learning rule yields a reduced capacity  $\alpha_c^H = 0.14$  [12]. The same result,  $\alpha_c = 0.14$ , also holds for cyclic patterns in a model network with a given distribution of delays [66]. These arguments lead to the estimate that our model network of  $N = 1000$  neurons with  $p = 1/T = 0.025$  should be capable of storing about 100 patterns. Exact answers to the question of storage capacity are not known yet.

Comparing Fig. 1.29, Id–IVd and Ia–IVa, we see that neither the mean firing rate nor the ensemble activity contain significant structures that would suffice to distinguish the four patterns. Nevertheless, the time-resolved spike raster (Fig. 1.29, Ib–IVb) shows that the four patterns can be retrieved and distinguished. This is confirmed by the pattern detector (Fig. 1.29, Ic–IVc) which measures the correlations in the spiking of different neurons, but of course, this could also be done by a neuron or cluster of neurons which exactly learned this classification in the same way as described before. Methods that are based on mean firing rates or ensemble activities would miss the information contained in the time resolved spike raster. Indeed, the above examples clearly show that single spikes can carry important information.

#### 1.5.4 DISCUSSION

We have presented a rather crude model of a neural network with time-resolved learning and retrieval capabilities. Spike emission and reception is governed by two response functions, viz., the refractory function  $\eta(s)$  and the EPSP  $\epsilon(s)$  (spike response model, SRM). In our present network, there is only a single type of model neuron, whereas several classes of neurons with a variety of intrinsic characteristics are found in real neural structures. Each pair of neurons in the model of this section is connected by exactly four synapses whereas the connectivity in real neural systems is much more complex. Inhibition has been modeled in a coarse way by a single activity-dependent neuron, whereas in reality inhibition is local. We have described a *homogeneous* network whereas a real brain is organized in several distinct areas and layers. In all these respects, the present model network can only be a rough approximation of real neural systems.

On the other hand, the spike response model (SRM) is carefully crafted with regard to the time constants of neural systems. We have assumed realistic axonal delays. We have also attempted to describe the EPSP realistically. Our model EPSP has a rise time of 3 ms and a slowly decaying tail. The overall duration (width of half-maximum) is approximately 6 ms, in good agreement with experimental data [18,19]. The model of a single neuron in our network includes an absolute refractory time of 3 ms and an after-hyperpolarizing potential with long time constants. The exact nu-

merical values are somewhat arbitrary, but they lie in a reasonable regime. As a result, both spike train and gain function have a realistic shape; cf. the section on the SRM neuron.

Based on our careful modeling of biological time constants, we have tried to answer the central question of this section: What is the physiologically relevant time scale of neuronal signal processing? Regarding the long post-synaptic integration time and spatial summation performed on the dendritic tree, it has been argued that relevant changes should occur on a time scale of 100 ms and more [13]. The results of this section, however, show that *it is possible* to store and retrieve information with a time resolution of 1 ms.

Our approach which emphasizes spatio-temporal spike patterns is related to the concept of *synfire chains* introduced by Abeles some years ago [6]. The central idea is that a neuron is excited much more efficiently if several EPSP's arrive simultaneously instead of arriving at random. If the activity of a number of presynaptic neurons converges synchronously onto a set of postsynaptic neurons, the latter will fire and can drive another set of neurons. Thus, an "activity wave" propagates through a network of neurons connected by appropriately diverging and converging links. Abeles calls this "synfire activity" and the set of neurons participating in such an activity wave a "synfire chain." This concept allows one to explain the experimental finding that accurate timing of the order of 1 ms can be seen over intervals as long as 100 ms and more; cf. [7] for a review of the experimental evidence. Simulations presented in the same paper [7] indicate that synfire chains can in principle be embedded in a neural network with appropriately designed connectivity.

In the context of our model each synfire chain corresponds to a specific spatio-temporal pattern. Every neuron can participate in several different synfire chains. A given chain (or spatio-temporal pattern) is activated by a short stimulus on a subset of neurons and propagates subsequently through the whole net. For the sake of convenience every chain in our model is a closed loop of 40 ms but longer, open chains (noncyclic patterns) would also be possible. Our model shows how a synfire chain with a specific spatio-temporal order can be formed using a Hebbian learning rule. We note that fixed delays are not required. In fact, using a range of different delays helps to stabilize the spatio-temporal ordering of the patterns. During retrieval, a temporal precision of 1–2 ms can be achieved over very long intervals.

The results of the present section are based mainly on computer simulations. They do not imply that all biological systems actually operate on such a fast time scale. On the other hand, we may ask why nature should not make use of this possibility. Experience tells us that evolution grasps most opportunities of improvement. In all cases where a "fast" system is somehow "better" than a slow one, we would therefore expect that single spikes and their cross-correlations become important. A fly rushing around at high speed through an unknown environment is an example of an animal

with the need of a fast reaction time. Indeed, it has been shown experimentally that single spikes are of eminent importance for signal processing in the neural system of the fly [16,111].

If it is true that spikes are relevant, some traditional methods of data analysis and several assumptions of model approaches need to be reconsidered carefully. Future experiments should be checked carefully for the appropriate time resolution. This can be done by analyzing the information content of the experimental data with variable resolution. A time resolution of down to 1 ms could be necessary [6–8,16,111], but a resolution of 10 or 30 ms seems to suffice in most cases where such an analysis has been carried through so far [39,83,101].

What are the conclusions with respect to theoretical approaches to network modeling? As we have seen in the preceding sections, a network of analog neurons should be used only in those cases where the relevant states can be defined in terms of mean firing rates. This condition is fulfilled, if we have reason to believe that the biological system which is to be described is in a state of incoherent stationary firing. In all time-dependent problems, however, spiking is essential and the type of model neuron that is used as the basic unit of a neural network has to be chosen carefully. In particular, if activity patterns that change on a time scale of 50 ms or less are considered (as in the case of collective oscillations in the cortex), the time constants of spiking and signal transmission become important. In these cases, a *spiking model neuron* should be preferred.

## 1.6 Summary and Conclusions

We are now at the end of our considerations and should look back on the whole chapter in order to recall what has been done. We have started with the problem of neural coding and asked the question of what aspect of the neuronal signal conveys the information on the environment. To be specific, is the information contained in single spikes, in time-averaged firing rates, or in ensemble averaged activities? There is, of course, no definite answer yet and this basic question is still open.

In principle, single spikes or spike patterns can be of great importance as we have seen in the last section. This does not mean, however, that a spike or pulse code is used in every biological system. Biology could still use a rate code or ensemble code, if circumstances allow one to do so. From general arguments we would expect that in all situations where fast reactions are required (e.g., during flight) a rate code may be too slow. Similarly, if the number of units is limited due to scarce space there may not be enough neurons to allow an ensemble averaged activity code. If both limitations hold we would expect a spike code — as it was found in the H1 neuron of the fly which is sensitive to fast horizontal movements [16,111].

The question of neural coding leads to the problem of the correct de-

scription of neuronal activity. Since, in general, we do not know what kind of code is used in a given neuronal system it is, to say the least, dangerous to construct a model that is based on mean firing rates or ensemble averaged activities for the basic variables. Such an approach is equivalent to an *a priori* assumption with respect to the neural code — and the code is, and remains, an open problem.

In certain cases, for example, area 17 of the visual cortex of cat and monkey, it is known that neighboring neurons have similar properties which leads to the idea of columnar structure [72]. In this case, it seems likely that an ensemble code is used and a network model of these areas could, in principle, be based on ensemble averaged activities. If we try to construct such a model we should realize that it is difficult to guess the correct interaction between different ensembles or columns. Similarly, in the case of stationary memory patterns, a rate code is, in general, sufficient. Within a pure rate description, however, it is impossible to check whether the stationary states are stable with respect to collective oscillations. This can only be done in a more detailed model. For these reasons, it is almost always preferable to start with a model of *spiking* neurons. In this case the question of the neural code need not be solved on an *a priori* basis but can be treated *a posteriori* within the context of possible model solutions.

The above approach has been adopted in Secs. 1.3 and 1.4 of this chapter. We have first studied the models of single spiking neurons and have then turned to a large and fully connected network of those neurons (spike response model, in short SRM) which has to store a finite number of static patterns. Based on a dynamic mean-field solution which is exact in the limit of  $N \rightarrow \infty$  we could derive the *correct* interaction term that allows a description by ensemble averaged activities. Each ensemble or assembly comprises a group of equivalent neurons which can be distributed all over the network. If we require a *stationary* activity, a rate description is appropriate. Stationary states, however, are stable only if the noise level is sufficiently high and if the delays are adjusted appropriately. Otherwise the system has a tendency to form collective oscillations. Thus oscillatory solutions appear quite naturally in a network of spiking neurons. In the SRM, both the period and the stability of coherent oscillations can be determined analytically.

The basic structure of locking and collective behavior has been analyzed in Sec. 1.2 in the context of a Lyapunov function approach. It is based on a rather abstract phase-variable model and has no direct relation to neurobiology, but it allows one to study the mechanism of synchronization within a simple framework. The “meaning” of coherence is still an open question, but there are theoretical concepts as well as preliminary experimental results which suggest that synchronization may be important in the context of feature linking and pattern segmentation. Other chapters of the present volume will return to this question and present different approaches. From our point of view, the effect of synchronization or coherent spiking which

has now been seen in many experiments is a strong indication that the time structure of neuronal signals is relevant — beyond the mean firing rate.

*Acknowledgments.* The authors are most grateful to Raphael Ritz for stimulating discussions and some of the plots. They also gratefully acknowledge several illuminating discussions with and helpful advice from Steve Strogatz. JLvH thanks Bard Ermentrout (Pittsburgh, PA) for an enlightening discussion on ground state degeneracy in  $XY$  ferromagnets and more general models. He also thanks Reinhard Lang (Heidelberg) for a hint to the work of Varadhan. This work has been supported by the Deutsche Forschungsgemeinschaft under Grant No. HE 1729/2-1.

## REFERENCES

- [1] Abbott LF, Kepler TB (1990) Model neurons: From Hodgkin Huxley to Hopfield. In: *Statistical Mechanics of Neural Networks*, L. Garrido (Ed.), Lecture Notes in Physics **368** (Springer, Berlin) pp. 5–18
- [2] Abbott LF (1991) Realistic synaptic inputs for model neural networks. *Network* **2**:245–258
- [3] Abbott LF, van Vreeswijk C (1993) Asynchronous states in a network of pulse-coupled oscillators. *Phys. Rev. E* **48**:1483–1490
- [4] Abeles M, Lass Y (1975) Transmission of information by the axon. *Biol. Cybern.* **19**:121–125
- [5] Abeles M (1982) *Local Cortical Circuits* (Springer, Berlin)
- [6] Abeles M (1991) *Corticonics: Neural Circuits of the Cerebral Cortex* (Cambridge University Press, Cambridge)
- [7] Abeles M, Prut Y, Bergman H, Vaadia E, Aertsen A (1993) Intergration, synchronicity, and periodicity. In: *Brain Theory*, A. Aertsen (Ed.) (Elsevier, Amsterdam)
- [8] Abeles M (1994) Firing rates and well-timed events in the cerebral cortex. This volume, Ch. 3
- [9] Abramowitz M, Stegun IA (1965) *Handbook of Mathematical Functions* (Dover, New York)
- [10] Adrian ED (1926) The impulses produced by sensory nerve endings. *J. Physiol. (London)* **61**:49–72
- [11] Amit DJ, Gutfreund H, Sompolinsky H (1985) Spin-glass models of neural networks. *Phys. Rev. A* **32**:1007–1032
- [12] Amit DJ, Gutfreund H, Sompolinsky H (1987) Statistical mechanics of neural networks near saturation. *Ann. Phys. (NY)* **173**:30–67

- [13] Amit DJ, Tsodyks MV (1991) Quantitative study of attractor neural networks retrieving at low spike rates. I. Substrate-spike rates and neuronal gain. *Network* **3**:259–274
- [14] Bauer HU, Pawelzik K (1993) Alternating oscillatory and stochastic dynamics in a model for a neuronal assembly. To appear in *Physica D* **69**
- [15] Bernander Ö, Douglas RJ, Martin KAC, Koch C (1991) Synaptic background activity influences spatio-temporal integration in pyramidal cells. *Proc. Natl. Acad. Sci. U.S.A.* **88**:11,569–11,573
- [16] Bialek W, Rieke F, de Ruyter van Stevenick RR, Warland D (1991) Reading a neural code. *Science* **252**:1854–1857
- [17] Bindman L, Christofi G, Murphy K, Nowicky A (1991) Long-term potentiation (LTP) and depression (LTD) in the neocortex and hippocampus: An overview. In: *Aspects of Synaptic Transmission*, T.W. Stone (Ed.) (Taylor and Francis, London), Vol. 1
- [18] Brown TH, Johnston D (1983) Voltage-clamp analysis of mossy fiber synaptic input to hippocampal neurons. *J. Neurophysiol.* **50**:487–507
- [19] Brown TH, Ganong AH, Kairiss EW, Keenan CL, Kelso SR (1989) Long-term potentiation in two synaptic systems of the hippocampal brain slice. In: *Neural Models of Plasticity*, J.H. Byrne and W.O. Berry (Eds.) (Academic Press, San Diego, CA), pp. 266–306
- [20] Buhmann J, Schulten K (1986) Associative recognition and storage in a model network with physiological neurons. *Biol. Cybern.* **54**:319–335
- [21] Bush P, Douglas RJ (1991) Synchronization of bursting action potential discharge. *Neural Comput.* **3**:19–30
- [22] Choi MY (1988) Dynamic model of neural networks. *Phys. Rev. Lett.* **61**:2809–2812
- [23] Connors BW, Gutnick MJ, Prince DA (1982) Electrophysiological properties of neocortical neurons in vitro. *J. Neurophysiol.* **48**:1302–1320
- [24] Cox DR (1962) *Renewal Theory* (Methuen, London)
- [25] Davis JL, Eichenbaum H (Eds.) (1991) *Olfaction. A Model System for Computational Neuroscience* (MIT Press, Cambridge, Mass.)
- [26] Dinse HRO, Krüger K, Best J (1991) Temporal structure of cortical information processing: Cortical architecture, oscillations, and nonseparability of spatio-temporal receptive field organization. In: *Neuronal Cooperativity*, J. Krüger (Ed.) (Springer, Berlin) pp. 68–104
- [27] Eckhorn R, Grüsser OJ, Kröller J, Pellnitz K, Pöpel B (1976) Efficiency of different neural codes: Information transfer calculations for three different neural systems. *Biol. Cybern.* **22**:49–60
- [28] Eckhorn R, Bauer R, Jordan W, Brosch M, Kruse W, Munk M, Reitboeck HJ (1988) Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biol. Cybern.* **60**:121–130

- [29] Eckhorn R, Krause F, Nelson JI (1993) The RF cinematogram: A cross-correlation technique for mapping several visual fields at once. *Biol. Cybern.* **69**:37–55
- [30] Ekeberg Ö, Wallen P, Lansner A, Traven H, Brodin L, Grillner S (1991) A computer based model for realistic simulations of neural networks. *Biol. Cybern.* **65**:81–90
- [31] Eggermont JJ (1990) *The Correlative Brain* (Springer, Berlin)
- [32] van Enter ACD and van Hemmen JL (1984) Statistical-mechanical formalism for spin glasses. *Phys. Rev. A* **29**:355–365
- [33] Ermentrout GB (1985) Synchronization in a pool of mutually coupled oscillators with random frequencies. *J. Math. Biol.* **22**:1–9
- [34] Ermentrout GB (1985) The behavior of rings of coupled oscillators. *J. Math. Biol.* **23**:55–74
- [35] Ermentrout GB (1990) Oscillator death in populations of “all to all” coupled nonlinear oscillators. *Physica D* **41**:219–231
- [36] Ermentrout GB (1992) Stable periodic solutions to discrete and continuum arrays of weakly coupled nonlinear oscillators. *SIAM J. Appl. Math.* **52**:1665–1687
- [37] Ermentrout GB, Kopell N (1990) Oscillator death in systems of coupled neural oscillators. *SIAM J. Appl. Math.* **50**:125–146, and references quoted therein
- [38] Ermentrout GB, Kopell N (1991) Multiple pulse interactions and averaging in systems of coupled neural oscillators. *J. Math. Biol.* **29**:195–217
- [39] Eskandar EN, Richmond BJ, Hertz JA, Optican LM, Troels K (1992) Decoding of neuronal signals in visual pattern recognition. In: *Advances in Neural Information Processing 4*, J.E. Moody et al. (Eds.) (Morgan Kaufmann, San Mateo, CA), pp. 356–363
- [40] FitzHugh R (1961) Impulses and physiological states in theoretical models of nerve membranes. *Biophys. J.* **1**:445–66
- [41] Freeman WJ (1975) *Mass Action in the Nervous System* (Academic Press, New York)
- [42] Gardner E (1988) The space of interactions in neural network models. *J. Phys. A: Math. Gen.* **21**:257–270
- [43] Gerstner W (1990) Associative memory in a network of “biological” neurons. In: *Advances in Neural Information Processing Systems 3*, R.P. Lippmann, J.E. Moody, D.S. Touretzky (Eds.) (Morgan Kaufmann, San Mateo, CA), pp. 84–90
- [44] Gerstner W, van Hemmen JL (1992a) Associative memory in a network of spiking neurons. *Network* **3**:139–164
- [45] Gerstner W, van Hemmen JL (1992b) Universality in neural networks: The importance of the mean firing rate. *Biol. Cybern.* **67**:195–205

- [46] Gerstner W, Ritz R, van Hemmen JL (1993a) A biologically motivated and analytically soluble model of collective oscillations in the cortex. I. Theory of weak locking. *Biol. Cybern.* **68**:363–374
- [47] Gerstner W, Ritz R, van Hemmen JL (1993b) Why spikes? Hebbian learning and retrieval of time-resolved excitation patterns. *Biol. Cybern.* **69**:503–515
- [48] Gerstner W (1993) *Kodierung und Signalübertragung in neuronalen Systemen — Assoziative Netzwerke mit stochastisch feuernenden Neuronen* (Verlag Harri Deutsch, Frankfurt), Reihe Physik, Bd. **15**
- [49] Gerstner W, van Hemmen JL (1993) Coherence and incoherence in a globally coupled ensemble of pulse-emitting units. *Phys. Rev. Lett.* **71**:312–315
- [50] Grensing D, Kühn R (1986) Random site spin glass models *J. Phys. A: Math. Gen.* **19**:L1153–L1157
- [51] Gray CM, Singer W (1989) Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* **86**:1698–1702
- [52] Gray CM, König P, Engel AK, Singer W (1989) Oscillatory responses in cat visual cortex exhibit intercolumnar synchronization which reflects global stimulus properties. *Nature* **338**:334–337
- [53] Guckenheimer J, Holmes P (1983) *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer, Berlin)
- [54] Hassard B, Wan BY (1978) Bifurcation formulae derived from center manifold theory. *J. Math. Anal. Appl.* **63**:297–312
- [55] Hebb DO (1949) *The Organization of Behavior* (Wiley, New York)
- [56] van Hemmen JL, Kühn R (1986) Nonlinear neural networks. *Phys. Rev. Lett.* **57**:913–916
- [57] van Hemmen JL, Grensing D, Huber A, Kühn R (1986) Elementary solution of classical spin glass models. *Z. Phys. B* **65**:53–63
- [58] van Hemmen JL, Grensing D, Huber A, Kühn R (1988) Nonlinear neural networks I and II. *J. Stat. Phys.* **50**:231–257 and 259–293
- [59] van Hemmen JL, Gerstner W, Herz AVM, Kühn R, Sulzer B, Vaas M (1990) Encoding and decoding of patterns which are correlated in space and time. In: *Konnektionismus in Artificial Intelligence und Kognitionsforschung*, G. Dorffner (Ed.) (Springer-Verlag, Berlin), pp. 153–162
- [60] van Hemmen JL, Ioffe LB, Kühn R, Vaas M (1990) Increasing the efficiency of a neural network through unlearning. *Physica A* **163**:386–392
- [61] van Hemmen JL, Kühn R (1991) Collective phenomena in neural networks. In: *Models of Neural Networks*, E. Domany, J.L. van Hemmen, K. Schulten (Eds.) (Springer-Verlag, Berlin), pp. 1–105
- [62] van Hemmen JL, Wreszinski WF (1993) Lyapunov function for the Kuramoto model of nonlinearly coupled oscillators *J. Stat. Phys.* **72**:145–166

- [63] Hertz J, Krogh A, Palmer RG (1991) *Introduction to the Theory of Neural Computation* (Addison-Wesley, Redwood City, CA)
- [64] Herz AVM, Sulzer B, Kühn R, van Hemmen JL (1988) The Hebb rule: Storing static and dynamic objects in an associative neural network. *Europhys. Lett.* **7**:663–669.
- [65] Herz AVM, Sulzer B, Kühn R, van Hemmen JL (1989) Hebbian learning reconsidered: Representation of static and dynamic objects in associative neural nets. *Biol. Cybern.* **60**:457–467
- [66] Herz AVM, Li Z, van Hemmen JL (1991) Statistical mechanics of temporal association in neural networks with transmission delays. *Phys. Rev. Lett.* **66**:1370–1373
- [67] Hirsch MW, Smale S (1974) *Differential Equations, Dynamical Systems, and Linear Algebra* (Academic Press, New York), Chap. 9
- [68] Hodgkin AL, Huxley AF (1952) A quantitative description of ion currents and its applications to conduction and excitation in nerve membranes. *J. Physiol. (London)* **117**:500–544
- [69] Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* **79**:2554–2558
- [70] Hopfield JJ (1984) Neurons with graded response have computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. U.S.A.* **81**:3088–3092
- [71] Horn D, Usher M (1989) Neural networks with dynamical thresholds. *Phys. Rev. A* **40**:1036–1044
- [72] Hubel DH, Wiesel TN (1977) Functional architecture of macaque monkey visual cortex. *Proc. R. Soc. London Ser. B* **198**:1–59
- [73] Iooss G, Joseph DD (1980) *Elementary Stability and Bifurcation Theory* (Springer, Berlin), Chap. V; to be fair, this book is explicit but not “elementary.”
- [74] Jack JJB, Noble D, Tsien RW (1975) *Electric Current Flow in Excitable Cells* (Clarendon Press, Oxford)
- [75] Kandel ER, Schwartz JH (1985) *Principles of Neural Science*, 2nd Ed. (Elsevier, Amsterdam)
- [76] Kelso SR, Ganong AH, Brown TH (1986) Hebbian synapses in hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* **83**:5326–5330
- [77] König P, Schillen TB (1991) Stimulus-dependent assembly formation of oscillatory responses: I. Synchronization. *Neural Comput.* **3**:155–166
- [78] Konishi M (1986) Centrally synthesized maps of sensory space. *Trends in Neurosci.* **9**:163–168
- [79] Kopell N (1986) Phase methods for coupled oscillators and related topics: An annotated bibliography. *J. Stat. Phys.* **44**:1035–1042

- [80] Krauth W, Mézard M (1987) Learning algorithms with optimal stability in neural networks. *J. Phys. A: Math. Gen.* **20**:L745–L752
- [81] Krüger J (1983) Simultaneous individual recordings from many cerebral neurons: Techniques and results. *Rev. Physiol. Biochem. Pharmacol.* **98**:177–233
- [82] Krüger J, Aiple F (1988) Multielectrode investigation of monkey striate cortex: Spike train correlations in the infragranular layers. *J. Neurophysiol.* **60**:798–828
- [83] Krüger J, Becker JD (1991) Recognizing the visual stimulus from neuronal discharges. *Trends in Neurosci.* **14**:282–286
- [84] Kuffler SW, Nicholls JG, Martin AR (1984) *From Neuron to Brain*, 2nd Ed. (Sinauer, Sunderland, Mass.)
- [85] Kuramoto Y (1975) Self-entrainment of a population of coupled nonlinear oscillators. In: *International Symposium on Mathematical Problems in Theoretical Physics*, H. Araki (Ed.) (Springer, Berlin), pp. 420–422
- [86] Kuramoto Y (1984) Cooperative dynamics of oscillator community. *Progr. Theor. Phys. Suppl.* **79**:223–240
- [87] Kuramoto Y (1984) *Chemical Oscillations, Waves, and Turbulence* (Springer, Berlin), pp. 68–77
- [88] Kuramoto Y, Nishikawa I (1987) Statistical macrodynamics of large dynamical systems. Case of a phase transition in oscillator communities. *J. Stat. Phys.* **49**:569–605
- [89] Lamperti J (1966) *Probability* (Benjamin, New York), Chap. 7
- [90] Lancaster B, Adams PR (1986) Calcium-dependent current generating the afterhyperpolarization of hippocampal neurons. *J. Neurophysiol.* **55**:1268–1282
- [91] Larson J, Lynch G (1986) Induction of synaptic potentiation in hippocampus by patterned stimulation involves two events. *Science* **232**:985–988
- [92] Little WA, Shaw GL (1978) Analytical study of the memory storage capacity of a neural network. *Math. Biosci.* **39**:281–290
- [93] MacKay DM, McCulloch WS (1952) The limiting information capacity of a neuronal link. *Bull. of Math. Biophys.* **14**:127–135
- [94] von der Malsburg C, Buhmann J (1992) Sensory segmentation with coupled neural oscillators. *Biol. Cybern.* **67** 233–242
- [95] Matthews PC, Strogatz SH (1990) Phase diagram for the collective behavior of limit-cycle oscillators. *Phys. Rev. Lett.* **65**:1701–1704
- [96] Mirollo RE, Strogatz SH (1990) Jump bifurcation and hysteresis in an infinite-dimensional dynamical system of coupled spins. *SIAM J. Appl. Math.* **50**:108–124
- [97] Mirollo RE, Strogatz SH (1990) Synchronization of pulse coupled biological oscillators. *SIAM J. Appl. Math.* **50**:1645–1662

- [98] Nagumo J, Arimoto S, Yoshizawa S (1962) An active pulse transmission line simulating nerve axon. Proc. IRE **50**:2061–2070
- [99] Neven H, Aertsen A (1992) Rate coherence and event coherence in the visual cortex: A neuronal model of object recognition. Biol. Cybern. **67**:309–322
- [100] Niebur E, Kammen DM, Koch C, Rudermann D, Schuster HG (1991) Phase-coupling in two-dimensional networks of interacting oscillators. In: *Advances in Neural Information Processing Systems 3*, R.P. Lippmann, J.E. Moody, D.S. Touretzky (Eds.) (Morgan Kaufmann, San Mateo, CA), pp. 123–127
- [101] Optican LM, Richmond BJ (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior cortex: III. Information theoretic analysis. J. Neurophysiol. **57**:162–178
- [102] Perkel DH, Gerstein GL, Moore GP (1967) Neuronal spike trains and stochastic point processes. I. The single spike train. Biophys. J. **7**:391–418
- [103] van der Pol B (1927) Forced oscillations in a circuit with nonlinear resistance (reception with reactive triode). The London, Edinburgh, and Dublin Philos. Mag. and J. Sci. **3**:65–80
- [104] Rall W (1964) Theoretical significance of dendritic trees for neuronal input–output relations. In: *Neural Theory and Modeling*, R.F. Reiss (Ed.) (Stanford University Press), pp. 73–97
- [105] Rapp M, Yarom Y, Segev I (1992) The impact of parallel fiber background activity on the cable properties of cerebellar Purkinje cells. Neural Comput. **4**:518–533
- [106] Reitboeck HJA (1983) A multielectrode matrix for studies of temporal signal correlations within neural assemblies. In: *Synergetics of the Brain*, E. Basar et al. (Eds.) (Springer, Berlin), pp. 174–182
- [107] Richmond BJ, Optican LM, Podell M, Spitzé H (1987) Temporal encoding of two-dimensional patterns by single units in primate inferior cortex: I. Response characteristics. J. Neurophysiol. **57**:132–146
- [108] Ritz R, Gerstner W, van Hemmen JL (1994) Associative binding and segregation in a network of spiking neurons. This volume, Ch. 5
- [109] Rotter S, Heck D, Aertsen A, Vaadia E (1993) A stochastic model for networks of spiking cortical neurons: Time-dependent description on the basis of membrane currents. In: *Gene, Brain, Behavior*, H. Elsner and M. Heisenberg (Eds.) (Thieme, Stuttgart), p. 491
- [110] Rudin W (1974) *Real and Complex Analysis* (McGraw-Hill, New York), p. 63
- [111] de Ruyter van Steveninck RR, Bialek W (1988) Real-time performance of a movement-sensitive neuron in the blowfly visual system: Coding and information transfer in short spike sequences. Proc. R. Soc. London Ser. B **234**:379–414

- [112] Sakaguchi H, Shinomoto S, Kuramoto Y (1987) Local and global self-entrainments in oscillator lattices. *Progr. Theor. Phys.* **77**:1005–1010
- [113] Sakaguchi H, Shinomoto S, Kuramoto Y (1988) Mutual entrainment in oscillator lattices with nonvariational-type interactions. *Progr. Theor. Phys.* **79**:1069–1079
- [114] Schillen TB, König P (1991) Stimulus-dependent assembly formation of oscillatory responses. II. Desynchronization. *Neural Comput.* **3**:167–177
- [115] Schuster HG and Wagner P (1990a) A model for neuronal oscillations in the visual cortex: 1. Mean-field theory and derivation of the phase equations. *Biol. Cybern.* **64**:77–82
- [116] Schuster HG and Wagner P (1990b) A model for neuronal oscillations in the visual cortex: 2. Phase description and feature dependent synchronization. *Biol. Cybern.* **64**:83–85
- [117] Singer W (1991) The formation of cooperative cell assemblies in the visual cortex. In: *Neural Cooperativity*, J. Krüger (Ed.) (Springer, Berlin), pp. 165–183
- [118] Singer W (1994) The role of synchrony in neocortical processing and synaptic plasticity. This volume, Ch. 4
- [119] Sompolinsky H, Golomb D, and Kleinfeld D (1990) Global processing of visual stimuli in a neural network of coupled oscillators. *Proc. Natl. Acad. Sci. U.S.A.* **87**:7200–7204
- [120] Sompolinsky H, Golomb D, Kleinfeld D (1991) Cooperative dynamics in visual processing. *Phys. Rev. A* **43**:6990–7011
- [121] Stein RB (1967) The information capacity of nerve cells using a frequency code. *Biophys. J.* **7**:797–826
- [122] Stein RB (1967) The frequency of nerve action potentials generated by applied currents. *Proc. R. Soc. London Ser. B* **167**:64–86
- [123] Strogatz SH, Mirollo RE (1988) Collective synchronization in lattices of nonlinear oscillators with randomness. *J. Phys. A: Math. Gen.* **21**:L699–L705
- [124] Strogatz SH, Mirollo RE (1991) Stability of incoherence in a population of coupled oscillators. *J. Stat. Phys.* **63**:613–635
- [125] Strogatz SH, Mirollo RE, Matthews PC (1992) Coupled nonlinear oscillators below the synchronization threshold: Relaxation by generalized Landau damping. *Phys. Rev. Lett.* **68**:2730–2733
- [126] Stuart GJ, Sakmann B (1994) Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. *Nature* **367**:69–72
- [127] Traub RD, Wong RKS, Miles R, Michelson H (1991) A model of a CA3 hippocampal pyramidal neuron incorporating voltage-clamp data on intrinsic conductances. *J. Neurophysiol.* **66**:635–650

- [128] Tsodyks M, Mitkov I, Sompolinsky H (1993) Patterns of synchrony in inhomogeneous networks of oscillators with pulse interaction. *Phys. Rev. Lett.* **71**:1281–1283
- [129] Usher M, Schuster HG, Niebur E (1993) Dynamics of populations of integrate-and-fire neurons, partial synchronization and memory. *Neural Comput.* **5**:570–586
- [130] Varadhan SRS (1980) *Diffusion Problems and Partial Differential Equations* (Springer, Berlin), p. 266 et seq.
- [131] Wang D, Buhmann J, von der Malsburg C (1990) Pattern segmentation in associative memory. *Neural Comput.* **2**:94–106
- [132] Wilson HR, Cowan JD (1972) Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* **12**:1–24
- [133] Wilson MA, Bhalla US, Uhley JD, Bower JM (1989) GENESIS: A system for simulating neural networks. In: *Advances in Neural Information Processing Systems*, D. Touretzky (Ed.) (Morgan Kaufmann, San Mateo, CA), pp. 485–492
- [134] Wong RKS, Prince DA, Basbaum AI (1979) Intradendritic recordings from hippocampal neurons. *Proc. Natl. Acad. Sci. U.S.A.* **76**:986–990
- [135] Yamada WM, Koch C, Adams PR (1989) Multiple channels and calcium dynamics. In: *Methods in Neuronal Modeling, from Synapses to Networks*, C. Koch and I. Segev (Eds.) (MIT Press, Cambridge, Mass.)

## 2

# The Correlation Theory of Brain Function

Christoph von der Malsburg<sup>1</sup>

with 1 figure

**Synopsis.** A summary of brain theory is given insofar as it is contained within the framework of *localization theory*. Difficulties of this “conventional theory” are traced back to a specific deficiency: there is no way to express relations between active cells (as for instance their representing parts of the same object). A new theory is proposed to cure this deficiency. It introduces a new kind of dynamical control, termed synaptic modulation, according to which synapses switch between a conducting and a nonconducting state. The dynamics of this variable is controlled on a fast time scale by correlations in the temporal fine structure of cellular signals. Furthermore, conventional synaptic plasticity is replaced by a refined version. Synaptic modulation and plasticity form the basis for short-term and long-term memory, respectively. Signal correlations, shaped by the variable network, express structure and relationships within objects. In particular, the figure-ground problem may be solved in this way. Synaptic modulation introduces flexibility into cerebral networks which is necessary to solve the invariance problem. Since momentarily useless connections are deactivated, interference between different memory traces can be reduced, and memory capacity increased, in comparison with conventional associative memory.

## Foreword

This is a slightly revised, fourth edition of a paper which appeared for the first time in 1981 as Internal Report 81–2 of the Department of Neurobiology of the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany.

---

<sup>1</sup>Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany.

## 2.1 Introduction

The purpose of this paper is to point out a specific deficiency in existing brain theory and to propose a way in which this gap could be filled. Although it leaves open a number of technical questions and presents more of a program than an accomplished theory, at least a language is developed to describe processes from the cellular to the cognitive level.

Searching for the function of the brain in all generality may be regarded as a venture bound to fail in view of the diversity of function of even a single brain. It is clear that any answer to the question can only be of a very general kind, much as a "general theory of painting" can only be a description of the process by which pigments are prepared and applied to canvas with a brush, and could say nothing about art, subject, style, and interpretation.

There is every reason to believe in the existence of general principles governing the function of the brain. Cerebral anatomy is surprisingly homogeneous in spite of the diversity of functional modalities represented in its different parts. The rapid cultural development of man has created fields of mental activity for which the brain cannot have been prepared by phylogeny in any detailed way. Both arguments seem to force the conclusion that the brain is governed by general principles of organization.

## 2.2 Conventional Brain Theory

The literature on brain theory is vast and cannot be summarized here. This section concentrates on a set of ideas, which is fairly consistent in itself, and with experiments. My account passes over most of the rich and sometimes ingenious detail to which these ideas have been worked out in the literature. However, I try to bring out points where the ideas fail.

### 2.2.1 LOCALIZATION THEORY

#### The Macroscopic Level

Observation of behavioral defects caused by localized lesions of the brain has firmly established that different parts of the brain are preoccupied with different *modalities*, topics of mental activity [12]. Examples are vision, audition, motor control, basic emotions, and drives (e.g., aggression, pleasure, and hunger), various aspects of speech, and long-term planning of action. The ability to lay down long-term memory can be destroyed by a specific local lesion; however, already existing long-term memory is not affected thereby. Memory traces and the ability to recall seem to be localized together with the modalities to which they refer.

Several kinds of hierarchy can be construed on the basis of the modalities.

For instance, sleep-waking regulation, drives, emotions, and planning all exert global control over the rest of the mind. However, I will treat all localized topics as on the same level. The term hierarchy will be reserved to forms of cooperation of localized objects.

There are aspects of the mind's function which cannot be localized in parts of the brain. Among these are consciousness and reason.

### The Microscopic Level

In recent decades *localization theory* has been refined down to the microscopic level. The information carrying units are thought to be nerve cells. These produce signals in the form of short (1 ms) electrochemical pulses, the so-called spikes, which can be recorded with the help of fine electrodes.

How are these signals to be interpreted? For the more central stages of the brain neurophysiology has answered this question in an operational way. In a vast majority of experiments, signals are evaluated in terms of a peri-event or poststimulus time histogram (PSTH). An event is constituted by a stimulus presented to the brain or a response evoked from the brain. The time shortly before and after the event is divided into a raster of small intervals (typically 1 to 10 ms). The event is repeated and the mean number of spikes falling into each interval is recorded. The mean frequency reacts in return on the event, if at all, by a short increase or decrease. The art of the experimenter consists in finding an event which influences the activity of the cell he is recording from. In this way a topic can be assigned to the cell. A typical topic is "an edge of light with a particular spectral component moving in a particular direction over a point of the retina."

The success of neurophysiology with this type of experiment has been tremendous. It is true, not all of the brain has been mapped in this way, and in fact it may not be practical to do so because some of the events may be difficult to find. Nevertheless, many scientists are ready to extrapolate the microscopic version of localization theory to all of the central nervous system. In this ultimate form, localization theory can be summarized as "one cell—one topic": cells are the atoms of meaning.

### The Brain as a Projection Screen

The topology of the periphery is preserved in the central representations, for example, neighboring points of a sensory surface project to neighboring points on a central sheet of cells. To each point of a sense organ (retina, cochlea, skin) there corresponds a small region centrally, often called *hyper-column*. Single cells in that region are specialized to particular combinations of quality values describing the point of the sensory surface (e.g., spectral distribution, direction of movement, stereo depth).

A single peripheral surface is represented by several central fields, which may vary in their emphasis on different qualities, and which usually are connected by topologically ordered fiber projections.

Again, this picture has to be extrapolated to cover all of the brain. It presents an experimental challenge to find the precise terms of this extrapolation. Let us suppose it will turn out to be possible. Then the physiological-anatomical picture is that of a screen on which patterns appear corresponding to input, output, and internal processing (e.g., emotional, planning), similar to the moving pictures on a color television screen.

### 2.2.2 THE PROBLEM OF NERVOUS INTEGRATION

#### The General Question

The picture of the brain as a projection screen is very suggestive, and in its principal traits it is well-founded in experimental observation. However, the picture poses a problem, that of nervous integration: In what way do the different patterns of activity interact? To be sure, the machinery for cellular interactions is conspicuously there in that each cell is contacted on its dendritic and somatic membrane by many synapses through which its membrane potential can be raised or lowered upon arrival of nerve impulses. The axonal fibers and branches for the transport of impulses from cells to synapses fill the larger part of the brain's volume. More precisely, the nervous integration question asks how this machinery is organized. The problem calls for new concepts, and at present it cannot be attacked experimentally.

One can train or ask a subject to react to the presentation of an apple by pressing a button. On command the subject can organize its brain so that upon the appearance of a certain class of patterns in the visual modality another pattern of a certain structure is created in the motor modality. This simple everyday example alone combines in it several complex organization processes which will now be named.

#### Representation of Structured Objects

In our cultural world we form symbols of a higher order by the juxtaposition (in time or space) of symbols of a lower order; for example, words out of letters or phonemes. According to localization theory, neurons are the basic symbols in the brain. Their position is fixed and cannot be used to form groupings. Another code is required to represent association of cells into patterns forming symbols on a higher level.

When we analyze a complex visual scene it is important to break it down into patterns which are simple enough so that we can hope to recognize them, for example, identify them with objects we saw before. A single pattern in turn has to be broken down into subpatterns, possibly through several stages; for example, man-arm-hand-finger-joint (cf. [15]). It should be possible to group neurons into such a hierarchy of patterns in a flexible way, without the introduction of new hardware for new patterns.

## Invariance

It is an everyday experience that there are objects of a relatively fixed structure, which affect our senses in an enormously variable way. For example, the picture of an apple can vary in perspective, and in position and size on the retina, depending on the relative coordinates of eye and apple. It is important to reduce this variability to an *invariant* representation of the intrinsic structure of the object, in order to be able to generalize, that is, draw the same conclusions from the perception of an object, irrespective of variations in its appearance; cf. [15]. An analogous discussion applies to motor patterns.

## Memory

There must be a physical basis for the gradual acquisition of information. We usually discuss it taking two aspects into account. According to the first, the brain changes to acquire new abilities. This will be the subject of the subsequent paragraph. The second is the ability of the brain to recall complex patterns which were active at an earlier time. With this memory in the narrow sense, recall should be evoked by an activity pattern or an input which is sufficiently close to the original pattern.

## Self-Organization

Self-organization refers to the ability of the brain to organize structures and activity patterns. The term “organize” implies that the process is directed toward some useful goal, which still has to be defined. A goal we already mentioned is the retention of earlier states of activity. In this way the brain can become independent of external sources of information and can build models for phenomena. Other goals will be defined in later sections.

The ability to organize itself sets the brain in sharp contrast to the computer, which relies entirely on a human programmer. It is also the basis of the reliability of the brain, being able to “repair” deviations from an optimal configuration. Self-organization puts heavy constraints on possible functional schemes for the brain.

## Control of Action

The metaphor of the brain as a projection screen assigns a passive role to it. In reality we know that the brain is spontaneously active. The “projector” is an integral part of it — to stay with the metaphor. Accordingly, a solution to the nervous integration problem has to include a scheme for the control of processes and the global integration of action.

### 2.2.3 PROPOSED SOLUTIONS

Localization theory — cf. Sec. 2.2.1 — proposes a basic frame into which any functional scheme for the brain has to be fitted. It poses the nervous integration problem, some aspects of which have been presented in Sec. 2.2.2. In this subsection we discuss some potential solutions which have been proposed in the literature, and point out some problems which they do not solve.

#### Synaptic Plasticity

A synaptic connection can be characterized by the size of the postsynaptic conductance transient or postsynaptic potential (PSP) which is produced in the postsynaptic cell upon arrival of a nerve impulse in the presynaptic fiber. The PSP size may slowly change under the control of the neural signals on the presynaptic and the postsynaptic side. This leads to a feedback situation: PSP size (together with the presynaptic signal) has an influence on the postsynaptic signal, which in turn controls the change of the PSP. If this feedback is positive and if the changes impressed on the PSP are permanent (nondecaying) one speaks of synaptic plasticity. The formation of new synapses may be included in the definition of synaptic plasticity [9]. In the case of an excitatory synapse, the EPSP (excitatory PSP) is increased (or a synapse established) after coincidence of neural activity on the presynaptic and postsynaptic side. In the framework of localization theory this is a straightforward implementation of the idea of an association and of Pavlov's conditioned reflex. It is usually assumed that plastic synaptic changes need seconds to become established and hours to consolidate (show full effect and stabilize). Synaptic plasticity has been shown experimentally to exist [1,4], although it is, in the presence of controlling signals, intrinsically difficult to demonstrate its nondecaying nature.

The instability which is caused by positive feedback has to be controlled somehow. Several schemes have been proposed: An upper limit to the synaptic weight (PSP size for a single spike), limitation of the sum of all synaptic weights converging on a cell or diverging from a cell, and stabilization of the mean level of activity in the postsynaptic cell. The third proposal means that, if the time average (over, say, several hours) of cell activity exceeds a certain value, either all excitatory synapses converging onto the cell are reduced in weight by a certain factor (and if this average is too low, the synapses are increased in weight), or the inhibitory synapses are increased in weight.

Synaptic plasticity is thought to be the basis of memory. The positive feedback involved in it leads to the kind of instability that is required for pattern generation and self-organization. In this sense, synaptic plasticity is analogous to self-reproduction in biological evolution.

## Feature Detectors

In the context of theoretical discussions within the frame of localization theory a cell in the sensory part of the brain is termed a feature detector, feature being the term for the event by which the cell is triggered; see, for example, [15]. Feature detectors may differ in level. On the lowest level, they respond to the signal of a peripheral sensory cell. On the highest conceivable level feature detectors respond to the appearance of entire objects [2]. They are then referred to as *cardinal* or grandmother cells. Feature detectors of intermediate level are found experimentally (a typical feature has been described when we discussed “the microscopic level” of localization theory).

Fairly specific feature detectors are proposed in many models of perception as a basis for the discrimination between complex patterns. The postulated level of feature detectors is regulated by a tradeoff. The higher the level (the more specific the cells) the smaller the overlap of the sets of cells responding to different patterns and easier the task of discriminating between them. High-level features, on the other hand, mean large numbers of cells, less flexibility (because specific trigger features must be adapted to particular pattern universes) and a low-duty cycle for each particular cell.

Many models employ cardinal cells because they seem to solve the problem which we have indicated when discussing the “representation of structured objects.” In reality that problem is not solved by cardinal cells. Either a cardinal cell is able to represent a whole class of objects. Then the individual object cannot be represented in detail, because the signal of a single cardinal cell is too crude. Or there has to be a cardinal cell for each pattern (a person with a new facial expression constituting a new pattern!). The number of cardinal cells required would then be forbidding (even if the invariance problem had been solved somehow), and it would be impossible to recognize new patterns which differed from familiar ones merely in detail. In addition, a cardinal cell would have to be silent (possibly for decades) until its pattern appeared again, but there is every reason to believe that a cell which is forced to stay silent for a day (e.g., by deafferentation) will change its synaptic makeup to become active again.

From this discussion it follows that high-level feature detectors do not solve any of the nervous integration problems. Low-level feature detectors, on the other hand, are an experimental fact and have to be the building blocks of any theory under the roof of localization theory.

## Cell Assemblies

Stimulation of some part of the brain will switch on many cells simultaneously. In the context of localization theory, it therefore appears natural to regard sets of simultaneously activated neurons as the basic internal objects. The nervous integration problem requires that such sets should not just be passively activated by stimuli but that they should rather be dy-

namental units, integrated by interactions. The *cell assembly* [8] is a model idea describing a certain system of such interactions.

A cell assembly is a set of neurons cross-connected such that the whole set is brought to become *simultaneously* active upon activation of appropriate subsets, which have to be sufficiently similar to the assembly to single it out from overlapping other assemblies. In view of the fluctuating nature of cellular signals, activation of cells in an assembly is simultaneous only on a coarse time scale, longer than, say, 50 ms.

Assembly reconstitution, or its completion from parts, has been proposed as the fundamental process of brain function. Important special cases would be the attachment of abstract symbolic representations to sensory patterns (recognition), the reactivation of supplementary information stored by past experience, and the generation of a response pattern which has previously been associated with a stimulus. According to this view, the action of the brain is controlled by a succession of such completion processes, alternating with the (partial) decay of assemblies (due to some delayed disinhibition) leaving residues which, together with new stimuli, form the germs for other assemblies.

Analysis of long periods of brain activity would reveal a hierarchy of subpatterns which appear as part of many assemblies. The dynamics of assembly completion could possibly be interpreted as interaction among subassemblies, analogous to excitation and inhibition exchanged between single cells. Subassembly interactions would have to be realized with the help of the synaptic interactions of the constituent cells.

It is an unsolved problem whether assembly interactions with these specifications are possible [11]. However, the assembly concept has a more fundamental flaw. When a particular assembly is active, there is no means to analyze it in terms of subassemblies: *It just consists of a bunch of simultaneously active cells.* (The above analysis in terms of subassemblies was only possible as a Gedankenexperiment.) This must lead to serious ambiguities. For instance, when we see a visual pattern, it is not only necessary to know which collection of features apply to it, but also in which way they are grouped. Even if the feature set is so complete that it can only be combined in one way into an image, it is important to know this combination. (When we see two people in the street we usually do not confuse which jacket is worn together with which trouser by one of them.) In particular, it must be possible to represent the result of a successful figure-ground discrimination.

### Associative Memory

Assemblies are supposed to be formed by synaptic plasticity. A pair of simultaneously stimulated cells establishes or strengthens its synaptic connection (in case there is a fiber bridging their distance). If this happens to many pairs of cells in a repeatedly activated pattern, an assembly can

be formed. Several detailed schemes for this process have been proposed and analyzed under the name of *associative memory*. Analysis has been made possible by simplifying assumptions; for example, linearity, a single parallel update, small overlap. It has been shown that many overlapping assemblies can be stored and retrieved in the same network without too much interference between them.

The lack of internal structure in assemblies leads to a serious difficulty of associative memories: Each memory trace recalls a fixed subset of cells without possible variation — apart from noise. However, cognitive psychology makes it obvious that realistic memory traces often correspond to a network of subpatterns connected in a flexible way to fit a particular situation.

### Visual Perception, Perceptrons

Visual perception presents two outstanding problems, viz., *figure-ground discrimination* and *extraction of invariants*. The perceptron approach [14] to perception, which makes use of most of the ideas reviewed so far, demonstrates quite explicitly the inadequacies of those ideas to solve the two problems mentioned.

Perceptrons are meant to be models for sensory subsystems of the brain. A typical perceptron consists of threshold units (neurons) of three kinds,  $S$ ,  $A$ , and  $R$ , that is, sensory, associative, and response units. These are arranged in layers which are sequentially connected:  $S \rightarrow A \rightarrow R$ . Cross-connections within a layer, or even backward connections, may also exist. The  $A$  units play the role of feature detectors. The  $A \rightarrow R$  connections are modifiable by synaptic plasticity.

The prominent feature of a perceptron is its ability to reorganize itself in response to the repeated activation of a certain subset  $s$  of  $S$  units such that subsequently a specific  $R$  unit fires precisely when  $s$  is presented.

The invariance problem calls for a single  $R$  unit to respond to the presentation of a pattern  $p$  in *any* position in  $S$ . Rosenblatt [14] proposed to solve the problem by the introduction of a second layer  $A'$  of feature detectors, sandwiched between  $A$  and  $R$ :  $S \rightarrow A \rightarrow A' \rightarrow R$ . A unit  $a' \in A'$  responds to the presentation of a certain feature in arbitrary position within  $S$ . Unit  $a'$  is enabled to do so by the presence of units  $a_i \in A$  with  $1 \leq i \leq N_{a'}$ , each of which responds to the particular feature in a position  $x_i$  in  $S$ . All units  $a_i$  have a connection to  $a'$ , which fires, if at least one cell of the set  $\{a_i\}$  is activated. Many different feature detectors analogous to  $a'$  are present in  $A'$ . The pattern  $p$  will activate the same subset of  $A'$ , independent of its position in  $S$ . A specific  $R$  unit can now be trained to react to this subset. Activity in a unit of  $A'$  makes a statement about the presence of a particular subpattern (or feature) of  $p$ . In order to generalize with respect to the position in  $S$ , information about the position of the subpattern is discarded. If the features are sufficiently complex, it may be possible — at

least in principle — to recover the relationships of overlap and reconstruct the full pattern  $p$  in a way analogous to solving a jigsaw puzzle. This reconstruction, however, is done nowhere in a perceptron, and the recognition of  $p$  has to be done on the basis of the uncorrelated feature set represented by the active units in  $A'$ . This is possible only if the features represented in  $A'$  are of sufficiently high level, which means that they are very numerous, or specialized to a particular universe of patterns in  $S$ , or both. The machinery needed, particularly in  $A$ , is gigantic (as is demonstrated by a simulated version of a perceptron [5]). It is evident that an enormous improvement over the perceptron could be made with the help of a scheme by which the overlap conditions of subpatterns would be made explicit.

An  $R$  unit can suppress all  $a'$  units not belonging to its own trigger set, if appropriate inhibitory back-couplings  $R \rightarrow A'$  are present. Rosenblatt [14] suggested to solve the selective-attention problem in this way. He recognized, however, that this is not a solution to the general figure-ground problem, since learning and recognition of a figure have to precede the suppression of its background. He admitted that new concepts were needed for the problem of figural unity [14, p. 555]. Again, this calls for a scheme by which cells in  $A'$  could express their relationship in terms of separate figures.

## 2.3 The Correlation Theory of Brain Function

### 2.3.1 MODIFICATIONS OF CONVENTIONAL THEORY

This subsection introduces (i) a scheme for the interpretation of cellular signals which is a refinement of the one given in Sec. 2.2.1 (the microscopic level), and (ii) a short-term analog of synaptic plasticity.

#### Correlations between Cellular Signals

In Sec. 2.2.1 I have discussed the experimental procedure by which the correlation between a cellular signal and an event is detected. The averaging in the poststimulus time histogram (PSTH) method is important to get rid of an apparently random time structure within the cellular response. This time structure will now become important.

Consider the spike trains emitted by two cells in the central nervous system. These signals may be evaluated in terms of a *correlation*.<sup>2</sup> It is supposed to measure the similarity between the two signals and should at least discriminate between synchrony and asynchrony in their temporal fine structure (with a resolution of 2 to 5 ms). It has to be assumed that

---

<sup>2</sup>The term “correlation” is not meant to imply a specific mathematical formulation.

the trivial state in which all cells are globally correlated with each other is suppressed by a system of inhibitory connections which permits only a small fraction of all cells to be active at any time.

### Synaptic Modulation

The synaptic connection between brain cells  $i$  and  $j$  is characterized by a strength  $w_{ij}$ . It is a measure for the size of the postsynaptic potential evoked in cell  $i$  upon arrival of a spike from cell  $j$ . Here I postulate that the weight  $w_{ij}$  of an excitatory synapse depends on two variables with different time scales of behavior,  $a_{ij}$  and  $s_{ij}$ . The set  $\{s_{ij}\}$  constitutes the permanent network structure. Its modification (synaptic plasticity) is slow and constitutes the basis for long-term memory. The new dynamic variable  $a_{ij}$ , termed *state of activation* of synapse  $ij$ , changes on a fast time scale (fractions of a second) in response to the correlation between the signals of cells  $i$  and  $j$ . With no signals in  $i$  and  $j$ ,  $a_{ij}$  decays towards a resting state  $a_0$ , within times typical for short-term memory. With strong correlation between the signals the value  $a_{ij}$  changes such that  $w_{ij}$  increases (*activation*). With uncorrelated signals,  $a_{ij}$  changes such that  $w_{ij}$  decreases to zero (*inactivation*). This behavior of the variable  $a_{ij}$  will be referred to as *synaptic modulation*. It can change the value of  $a_{ij}$  significantly within a fraction of a second. Not all synapses from a given cell to other cells can grow at the same time, since the inhibitory system referred to above (“correlations between cellular signals”) prevents those target cells from firing all at the same time. Also the synapses received by a cell compete with each other — for the same reason. The physical basis for synaptic modulation is not clear. It might correspond to the accumulation or depletion of some chemical substance at a strategic location in or near the synapse. Here we take the relevant postsynaptic signal to be the cell’s output spike train, but it may also be a more local dendritic signal. As a simple example one could assume  $w_{ij} = a_{ij}s_{ij}$  with  $0 \leq a_{ij} \leq 1$  and a resting state  $a_0$  within the interval  $(0,1)$ .

The variables  $\{s_{ij}\}$  are controlled by what I shall call *refined synaptic plasticity*: strong correlation between the temporal fine structure in the signals of cells  $i$  and  $j$  causes  $s_{ij}$  to grow. This growth may be thought to be limited in the usual way; for example, by sum rules. Absence of correlation does not directly reduce  $s_{ij}$ .

The analogy between synaptic modulation and refined plasticity is apparent. Both are controlled by correlations in the signals of cell pairs in a positive feedback fashion. They differ in time scale of decay (seconds for  $a_{ij}$ , decades to permanent for  $s_{ij}$ ) and of buildup. Moreover, they differ in the way they are controlled. The  $a_{ij}$  react only to the two locally available signals and are both increased and decreased by correlations and their absence. The  $s_{ij}$  are only increased by local signals and are decreased in response to the growth of other synapses.

### 2.3.2 ELEMENTARY DISCUSSION

#### Sources of Correlations

Correlations between the signals of cells can be caused by time structure in sensory signals exciting the cells. However, there is a more important source of correlations. Time structure in cellular signals can be created spontaneously, for example, by a tendency of cells to form bursts of spikes. Correlations arise when this time structure is transmitted to other cells by excitation or inhibition.

#### Effects of Correlations

One effect of the correlation between signals in cells  $i$  and  $j$  has already been mentioned: activation of the synaptic weight  $w_{ij}$ . Specific connection patterns (e.g., reflex arcs) can be created in this way, and a pluripotent network can be turned into a specialized machine.

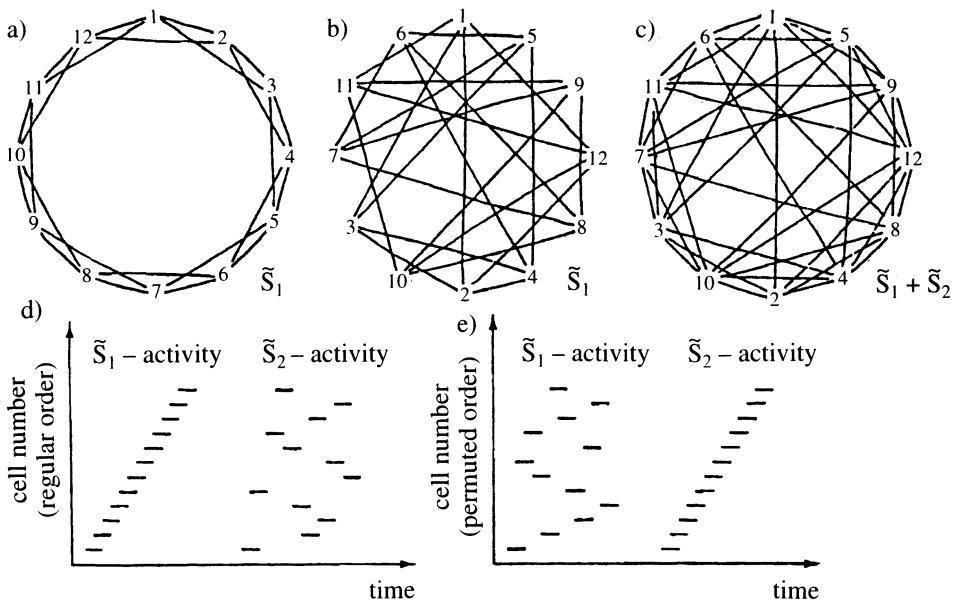
Second, a correlation between the signals of cells  $i$  and  $j$  enables them to positively interact so as to excite a third cell  $k$  (if  $w_{ki}, w_{kj} \neq 0$ ). The individual signals may not be able to transcend the threshold of cell  $k$ , whereas simultaneously arriving signals may. Two subnetworks with uncorrelated activity patterns may coexist without interfering.

Third, correlations control (refined) synaptic plasticity. The absence of correlations between two activity objects, even if they sometimes coexist within the same span of a second, keeps them from developing mutual synapses.

#### Correlation Dynamics

The dynamical system introduced so far for the cellular signals and temporary synaptic strengths forms a basis for organization. The correlation between the signals in cells  $i$  and  $j$  and the states of activation  $a_{ij}$  and  $a_{ji}$  of their common synapses form a positive feedback loop, driving  $a_{ij}$  and  $a_{ji}$  away from the resting state  $a_0$  and shifting the signal pair away from a structureless uncorrelated state. In this way correlations can stabilize their own existence and cease to be transitory and shaky statistical epiphenomena. Different synapses on one cell compete with each other, as was pointed out above, and certain sets of synapses cooperate with each other. For instance, the direct pathway from cell  $i$  to cell  $j$  cooperates with the indirect pathway from  $i$  via cell  $k$  to cell  $j$ . These dynamical interactions between synapses and corresponding signals tend to stabilize certain optimal connectivity patterns (together with their corresponding signal patterns). These can be characterized locally as having sparse connections (to avoid competition), which are arranged so as to have optimal local cooperation between them.

The slow component  $s_{ij}$  of synaptic strength is plastically modified only by strong correlations, that is, mainly when connectivity forms an optimal



**Fig. 2.1.** Two topological networks on the same set  $S$  of 12 cells.  $E$  is the two-dimensional plane,  $p = 4$ , and the mapping  $m$  assigns the cells in natural order to equidistant points on a circle. (a) shows the resulting topological network  $\tilde{S}_1 = (S, m, E, p)$ . (b) shows  $\tilde{S}_1$ , however with a permutation  $P$  applied to cell positions. In (c) the second topological network  $\tilde{S}_2 = (S, Pm, E, p)$  is added. (d) and (e) show cellular activity (correlates) in the form of traveling waves as shaped by  $\tilde{S}_1$  (left-hand side of the graphs) or  $\tilde{S}_2$  (right-hand side). Line segments symbolize short bursts of activity. In (d) the order of cells on the ordinate is as in (a), in (e) the same activity is plotted with the order of cells on the ordinate as in (b) and (c). Each correlate activates its own network and deactivates the other. Simulations of cell activity and synaptic modulations have shown that  $\tilde{S}_1$  or  $\tilde{S}_2$  can be dynamically recovered from the superposition (c).

pattern. Therefore the structure of the permanent network tends to be a superposition of optimal connectivity patterns. When input to the network activates certain cells (possibly in a correlated fashion), a dynamical process of organization sets in, as a result of which synapses forming an optimal network are fully activated and all other synapses are deactivated.

### 2.3.3 NETWORK STRUCTURES

It is not clear how optimal connectivity patterns can be characterized globally. This subsection proceeds on the basis of the conjecture that they have a topological structure, the neighborhoods of which may correspond to overlapping sets of coactive cells. (Mathematical rigor is not attempted in the notation.)

### The Topological Network

Let  $S$  be a set of  $n$  cells,  $E$  an appropriate space of low dimensionality,  $m$  a map assigning to each cell of  $S$  a point in  $E$ , and  $p$  a natural number with  $p \ll n$ . The topologically structured network  $\tilde{S} = (S, m, E, p)$  is constructed by connecting each cell of  $S$  with its  $p$  nearest neighbors in  $E$  by excitatory synapses. I will refer to  $\tilde{S}$  simply as a *topological network*.

A topological network embedded in ordinary space is a very common idea; cf., for example, [4]. The point made here is that there is no need for a network to stay with this natural embedding. This has the important consequence that there is a huge number of topological networks on the same set  $S$ , even if  $E$ ,  $p$ , and the set of assigned points in  $E$  are kept fixed. Namely, instead of  $m$  one can consider  $Pm$ , with any permutation  $P$  of the assigned points in  $E$ . See Fig. 2.1 for an example.

### The Correlate of a Topological Network

Before one can deduce the dynamical behavior of the network activity much more detail has to be specified. However, here we are only interested in certain aspects of the dynamics and, thus, a few assumptions suffice.

Let the synaptic weights be constant for the moment. Since an activated excitatory synapse between two cells creates correlations in their signals, “neighbors” in a topological network are correlated. The topological aspect is important since on the one hand the topological network is cooperative: each synapse is helped by others in establishing a correlation; for example, 1–3 and 1–2–3 in Fig. 2.1(a) cooperate. On the other hand, the network can still be decomposed into many weakly coupled subnetworks, in contrast to globally cross-connected networks.

Two kinds of signal pattern can exist in a topological network. In one there are waves running through the network (consult, e.g., [4]; for an example see Fig. 2.1). The diffuse network of inhibition keeps the cells from firing in global simultaneity. The other kind of pattern stresses the analogy to the spin lattice of a ferromagnet [11]: A cell randomly switches between an active and a silent state. In so doing it is influenced by its neighbors. If a majority of neighbors is active, it is more likely that the cell will fire. If a majority of neighbors is silent, it is more likely that the cell will be silent. The strength of coupling between the behavior of the cell and its environment can be characterized by a parameter  $T$ , analogous to the temperature of the spin lattice. For  $T = 0$  the coupling is rigid: all cells in the network switch up and down simultaneously. For infinite  $T$  there is no effective coupling and the cells are independent. Here we are interested in intermediate values of  $T$ , for which a cell is correlated with its neighbors and this correlation decreases over a characteristic “distance” which is a fraction of the “diameter” of the network. With either kind of cellular activity the structure of the network is expressed in the signals by correlations. Such a signal structure will be called a *correlate*.

Now allow the synaptic activities  $a_{ij}$  to vary. Consider a set  $C$  of cells which are excited by input activity. Suppose  $C$  is part of several sets of cells  $S', S'', \dots$  which are internally connected by topological networks  $\tilde{S}', \tilde{S}'', \dots$ . If these topologies are independent and all synapses are in the resting state,  $C$  is globally coupled in a nontopological fashion. The connectivity in  $C$  is then probably unstable. A stable state can be reached after one of the topological networks, say  $\tilde{S}'$ , has been activated and the others have been inactivated. In order for this to happen, the complement of  $C$  in  $S'$  has to be invaded so as to fill the holes left by  $C$  in the topology of  $\tilde{S}'$ . After the network with the topology of  $\tilde{S}'$  has been activated, activity can no longer invade the rest of the other sets  $S'', \dots$ , because the  $p$  environments of their cells — even if they are active in  $S'$  — never fire synchronously.

Correlate reconstruction is the fundamental process of correlation theory. It must take place on the fast time scale of thought processes. Its synergetics is a complicated matter and needs further detailed work. We now turn to an important special case.

### Projection between Topological Networks

Consider two structurally identical networks,  $\tilde{S}_1$  and  $\tilde{S}_2$ , on disjoint sets  $S_1$  and  $S_2$  of  $n$  cells each. The two sets are connected to each other by a one-to-one projection  $R$  of activated synapses connecting cells in corresponding position, so that  $R$  corresponds to an isomorphism. This defines on  $S_1 \cup S_2$  again a topological structure which can carry a correlate, with correlations at short distance in  $\tilde{S}_1$  and  $\tilde{S}_2$ , and between cells in  $S_1$  and  $S_2$  which correspond to each other according to  $R$ . This special kind of topological correlate can be approached from different starting configurations, as will now be discussed.

$\diamondsuit$ 1. Consider first the case with  $R$  in the resting state and correlates corresponding to  $\tilde{S}_1$  and  $\tilde{S}_2$  active in  $S_1$  and  $S_2$  but not mutually correlated. The projection  $R$  will have a weakly synchronizing influence on pairs of corresponding cells in  $\tilde{S}_1$  and  $\tilde{S}_2$ . The correlations thereby induced will activate the synapses of  $R$  and strengthen the  $S_1$ - $S_2$  correlations, until the stationary state is reached with fully activated  $R$  and the activity strongly correlated between  $S_1$  and  $S_2$ . On the other hand, if on  $S_2$  a network with a considerably different topological structure were activated,  $R$  would be deactivated.

The case is very reminiscent of the basic two-cells-one-synapse situation: correlation in (internal structure of) the correlates on  $S_1$  and  $S_2$  leads to  $R$  activation whereas lack of correlation leads to deactivation. In this sense  $S_1$  and  $S_2$  can be regarded as composite analogs to single cells.

$\diamondsuit$ 2. Let  $\bar{R}$  be a system of synapses connecting each cell of  $S_1$  with each cell of  $S_2$ . Let  $\tilde{S}_1$  and  $\tilde{S}_2$  be isomorphic topological networks on  $S_1$  and  $S_2$ . The synapses of  $\bar{R}$  are initially in their resting state. A very similar sys-

tem, referring to an ontogenetic problem, has been simulated by Willshaw and von der Malsburg [17] with two-dimensional  $E$ , and has been treated analytically for one-dimensional  $E$  by Häussler and von der Malsburg [8]. There it has been shown that  $\bar{R}$  can dynamically reduce to a one-to-one projection between the isomorphically corresponding cells in  $\tilde{S}_1$  and  $\tilde{S}_2$ . The system is able to spontaneously break the symmetry between several possible projections.

$\diamond 3$ . Several topological networks  $\tilde{S}_2, \tilde{S}'_2, \tilde{S}''_2, \dots$  may exist in  $S_2$  (in addition to  $\tilde{S}_1$  and  $\bar{R}$ ). Before a topological correlate can be established on  $S_1 \cup S_2$ , several decisions have to be made, between  $\tilde{S}_2, \tilde{S}'_2, \tilde{S}''_2, \dots$  and between possible one-to-one mappings corresponding to one of the structures on  $S_2$ . These decisions have to be made simultaneously. This is likely to cause chaos instead of a specific correlate. However, if symmetries between the various structures are already slightly broken in the initial state, an ordered stationary state may be reached, as is made plausible by extrapolation from experience with a case similar to the one treated in the previous paragraph.

### Composite Elements

In Sec. 2.3.1 I have introduced the basic machinery of correlation theory — in terms of cells: correlations of their signals, synapses, and their modulation. The above discussion “Projection between topological networks” has prepared the way to the use of a very similar language on a higher level. The idea consists in considering sets of topologically connected cells instead of single cells as network elements. The sets may then be termed *composite elements*. Likewise, the ensemble of cellular signals of a set may be regarded as a *composite signal*, and the ensemble of fibers connecting two composite elements as a *composite connection*. The correlation between two cellular signals was defined in terms of synchrony and asynchrony between spike trains. Correlation between the signals of two composite elements has to be defined as a structural correspondence between the composite signals in terms of the composite connection between the elements. Each single synapse between two composite elements should be modulated by a globally evaluated correlation between the composite signals. This is made possible by the fact that a temporal correlation in the signals locally available to the synapse can only be established in the context of a global correlation between the elements, as we have just seen under  $\diamond 1$ .

Composite elements can again form networks  $S_1, S_2, \dots$ , with composite connections  $R^{ij}$ . In order to form a correlate between the composite elements it is necessary that the different composite connections be locally consistent with each other. Let us introduce an arbitrary but fixed numbering of cells in each element. A one-to-one composite connection  $R^{ij}$  between composite elements  $i$  and  $j$  is then equivalent to a permutation matrix  $P^{ij}$  in which each nonzero element corresponds to a synapse. In a triplet of ele-

ments  $S_i, S_j, S_k$  the permutation  $P^{ik}$  must be the same as  $P^{ij}P^{jk}$  in order to be consistent. Stated differently, the composite permutation matrix corresponding to a closed chain of connections must be unity:  $P^{ij}P^{jk}P^{ki} = \mathbb{I}$ . (The condition can be relaxed for chains of elements which are longer than the correlation length. This opens the door to the whole complexity and richness of nontrivial fiber bundle or gauge field structures.) Also on this new level the dynamical interactions between signals and synapses stabilize certain preferred connectivity patterns and correlations, and again it may be conjectured that they have a topological structure.

In applications it may be necessary to introduce supercomposite elements. Section 2.3.4 ("Interpretation of an Image") will give an example. The elaboration of particular structures is, however, a complex dynamical and mathematical problem.

### The Synergetic Control of Action

How can the dynamical behavior of the brain's network structure be characterized globally? Suppose the state of the brain at time  $t$  could be understood as a superposition of structures, termed modes, with the following properties. A mode is a subnetwork of active cells and synapses which, if left to itself, would reproduce its form, possibly change its amplitude. (For a neuronal system, decomposition into modes has been carried out rigorously by Häussler and von der Malsburg [8].) To predict the state of the brain at time  $t + \Delta t$ , we decompose its state at time  $t$  into modes, let each of them grow or decay for the interval  $\Delta t$ , and superpose the results again. With the help of a global control parameter it often can be achieved that only one or a few modes grow and all others decay. It is conceivable that such global control exists in the brain. If only one mode grows it soon dominates the state of the system. If several modes are related by a symmetry they grow or decay with the same speed. This is the reason why symmetry breaking, that is, the selection of one of the related modes, is difficult to achieve.

The distinguishing feature which allows a mode to grow fast is maximal local self-amplification and optimal cooperation of locally converging dynamical influences, for example, correlation between signals converging onto one cell.

If growth of a mode is sufficiently slow there is time for the exchange of signals between all parts of the network. All locally available information is then integrated into the one global decision — growth or decay. After a mode has grown and established itself for some time, conditions may cease to be favorable to it, either because the mode has prepared the way for a successor mode which starts to compete successfully, or because the environment has changed, or simply because of some kind of fatigue. Thus the brain is ruled by a succession of modes. This view emphasizes the analogy to many other self-organizing systems [7], and would put the brain

into sharp contrast to the computer and other man-made systems with detailed central control.

Memory may be thought of as the ability of the brain to change its network so as to improve the success of modes which were once active. In the extreme case an entire global mode which once dominated the brain's state for a short moment can be reestablished. A physical basis for this ability is synaptic plasticity, which reinforces those networks which are strongly activated.

### 2.3.4 APPLICATIONS OF CORRELATION THEORY

#### Visual Elements

Light stimulation of one retinal point can directly affect several thousand neurons in visual cortex. Together they form a composite element of lowest level, a *primary visual element*. Each neuron is specifically sensitive to a particular combination of quality values characterizing the stimulus: level of brightness or darkness, spectral distribution, spatial distribution of light within a receptive field, stereo depth, direction, and speed of movement. The visual cortex contains multiple representations of the retina. These are interconnected by diffusely retinotopic fiber projections. Primary visual elements may be composed of cells in several visual areas and even in the thalamus. The part of the brain formed by primary visual elements will be termed *V*.

Consider a particular visual element while the eyes are slowly scanning a scene. When a light edge crosses the receptive field of the element, a subset of cells is activated simultaneously. The subset describes the quality of the edge of light. This simultaneous excitation triggers activation of synapses and formation of a correlate within the active subset of the element under consideration. A subnetwork results which now represents a composite feature detector. Its signal expresses a composite quality which can be recognized even from mixtures of signals stemming from different visual elements. Confusion is excluded by signal correlations within a set of fibers coming from one primary visual element.

Visual elements have been introduced here as those collections of cells which are affected from one retinal point. One could possibly also consider somewhat larger patches of cortex (and thalamus) as elements. Those larger elements would then be capable of forming correlates corresponding to patches of visual texture. There is no need for the brain's "hardware" to contain complex feature detector cells. Only cells responding to rather simple stimuli are required, from which complex composite feature detectors can be "manufactured on the spot" by activation of synaptic networks.

### Figure-Ground Discrimination

Suppose all visual elements in the primary region  $V$  are integrated by a fiber system which connects feature sensitive cells in one element with cells specific for the same local quality in many other elements. Suppose two elements interconnected this way are stimulated by a similar composite quality and correlates have formed in both of them, so that the situation described in Sec. 2.3.3 ( $\diamond 1$ ) is given. In due course the connection between the elements will be activated and the composite signals of the elements will correlate with each other. On the other hand, if the two elements were stimulated by radically different composite qualities, mutual synapses would be deactivated and the signals would decouple.

Suppose a visual scene contains a region  $F$  characterized by local qualities which change continuously from point to point inside  $F$  and which change discontinuously across the boundary of  $F$ . (A prominent role among these qualities will be played by differential velocity caused by perspective movement or object displacement.) The mechanism just described will set up a network of activated synapses and a correlate which fills the region of primary visual cortex excited by the image of  $F$ . All elements taking part in it will signal this fact by mutual local correlations. There will be no correlations across the boundary of the network.

In this way the scene is decomposed into a patchwork of figures. Moreover, a figure is decomposed into a hierarchy of parts, the strongest correlations signaling affiliation to one part of the figure, weaker ones affiliation to adjacent parts, and so on. This decomposition of the visual scene into a hierarchy of correlates starts already prior to recognition of patterns, a stage of the process which was termed “preattentive vision” by B. Julesz [10].

### Invariant Image Representation

The correlation structure described in the last paragraph has to be built up anew for each image fixation. Another part of the visual system, to which I will refer as  $I$ , can accumulate visual information over longer periods of time and independently of retinal image location. A physical prerequisite for this is a fiber system which connects each element of  $V$  with each element of  $I$ . (This strong assumption can later be relaxed considerably.) If all these fibers were activated at the same time, a great jumble of signals would converge on the elements in  $I$ . It is, however, possible to deactivate most connections and activate only topologically ordered one-to-one projections between  $V$  and  $I$ .

I assume that the elements in  $V$  and in  $I$  are tied together by topological networks  $N_V$  and  $N_I$ , respectively. (This is a statement about permanent weights.) The topology is the natural one of two-dimensional visual space. Consider for simplicity a situation in  $V$  with just two active correlates,  $F$  and  $G$ . Here  $F$  refers to figure and  $G$  to ground. Correlations in both

$F$  and  $G$  are topologically structured by activated subnetworks of  $N_V$ . The components of  $N_V$  connecting  $F$  with  $G$  are deactivated. Initially there may only be spontaneous activity in  $I$ , the correlations of which are topologically structured by  $N_I$ . Connections from  $F$  and  $G$  which converge on one element of  $I$  carry noncorrelated signals and cannot cooperate to cause excitation, correlation, or synaptic activation. If one considers  $I$  and just the  $F$  part of  $V$  as two superelements, the situation is that of Sec. 2.3.3 ( $\diamond 2$ ). As was pointed out there, a stationary state will be reached in which a one-to-one projection is activated that connects neighboring elements in  $F$  to neighboring elements in  $I$ . If symmetries are not broken by other influences, the scale and orientation of the  $F \rightarrow I$  projection will be such that the image of  $F$  fits best into  $I$ . At the same time the correlate structure of the intra- and interelement networks in  $F$  is transferred to the corresponding elements in  $I$ . (This is analogous to the transfer of retinal markers to the tectum in [14].)

The simulations of Willshaw and von der Malsburg [18] have shown that the simultaneous presence of two independent correlates, like  $F$  and  $G$ , can lead to a final state with two independent mappings of the kind described, one for  $F$  and one for  $G$ . The network  $I$  can then tune its correlate to  $F$  or  $G$ .

New mappings between  $V$  and  $I$  have to be set up for each new image fixation. This is enormously facilitated by relevant structures in  $I$ , built up during previous fixations. Relative image distortions between fixations are absorbed by distortions in the projections which are established. Over the course of many fixations more and more information about a figure, although arriving from different parts of the retinas, can be gradually accumulated in  $I$ .

After a mapping between  $V$  and  $I$  has been activated, information can be transferred from  $I$  back to  $V$ . The afferent information can thus be scrutinized by the retrograde activation of composite feature correlates.

In distinction to the perceptron approach to the invariance problem, the geometrical structure of the figure is explicitly represented in  $I$ . There is no need to recover it from the distribution of active feature detectors; cf. Sec. 2.2.3 ("Visual Perception, Perceptrons").

### Interpretation of an Image

Before an image can be recognized it must be brought into contact with ideas of shapes and objects stored in memory. Let us invoke a part  $M$  of the brain. To a neurophysiologist  $M$  would appear similar to  $I$ . However, it would be dominated by specific connection patterns which have been laid down previously by synaptic plasticity and which correspond to abstract schemata of objects. These can be revived by resonance with structures in  $I$  to carry correlates. Recognition between structures in  $I$  and in  $M$  is possible on the basis of a correspondence of detailed network structure,

which in turn is expressed in terms of correlations between signals. The situation was discussed in Sec. 2.3.3 ( $\diamond 3$ ). Several relevant memory traces may be activated simultaneously or consecutively.

The representation of an object in  $I$  has to be fairly insensitive to image size, position, orientation, and (slight) distortion. It therefore lacks information about these parameters and it is necessary for structures such as  $M$  to have access to the primary image in  $V$ . This is possible with the help of full direct fiber systems connecting  $M$  with  $V$ . These can be easily structured during a fixation because all elements of an image in  $V$  are functionally labeled by correlations with the corresponding elements in  $I$ . The original image can be scrutinized by the selective setup of part-networks referring to parts of it. A full interpretation of an image is constituted by a correlate in a supernetwork composed of many superelements such as  $V$ ,  $I$ , and  $M$ , partly belonging to other modalities of the brain.

Memory in its direct form precisely reestablishes correlates which were previously active. The observed great flexibility of memory traces could be explained, if memory in this extreme form were restricted to certain substructures of the brain, like the  $M$  mentioned above. For instance, we know that the memory trace corresponding to a human face leaves unspecified all accidental aspects, for example, perspective, illumination, and expression. The trace has to be complemented by particular correlates in other areas, like  $V$  and  $I$ , before it can be compared with a real image. This flexibility cannot be accounted for by cell assemblies which cannot be analyzed in terms of their parts.

## 2.4 Discussion

### 2.4.1 THE TEXT ANALOGY

The relationship between correlation theory and conventional brain theory may be clarified through the analogy to the way our cultural world makes use of symbols. When we write text we employ a set of basic symbols, letters, or ideograms. Out of these we form higher symbols, words, sentences, and paragraphs. We do so by producing many copies of the basic symbols and arranging these in spatial patterns. According to localization theory there is a fixed set of basic symbols with fixed locations in the brain. According to conventional theory, higher level symbols are formed by the simultaneous activity of many of these. This is analogous to summarizing a paragraph with the help of letter statistics. According to conventional synaptic plasticity, experience is accumulated in a way analogous to a measurement of the probability with which letters appear simultaneously in large pieces of text (large enough to represent the amount of information which we hold momentarily in our mind). Conventional theory tries

to mend the obvious deficiency by introducing more and more letters and ideograms. This, however, creates more problems than it solves.

In correlation theory higher symbols are constructed out of lower ones with the help of correlations and modulating synapses. The full hierarchy of patterns is thereby represented. The refined plasticity of correlation theory is analogous to measurement of the probabilities of letters or words to be adjacent. Dynamical selection of synergetic networks is analogous to a grammar which allows only strings of letters with a certain structure to be formed. Correlation theory thus allows for much more complete information to be stored, and used for the reconstruction of previously active patterns, than does conventional theory.

Another point may be clarified through the text analogy. Microscopic localization theory (Sec. 2.2.1) attributes to each cell its own meaning. Also a letter or ideogram has its own meaning. However, if a symbol is placed in a certain context, its meaning is enormously refined or even completely changed. The crude meaning attributed to a cell by the poststimulus time histogram of neurophysiology can, according to correlation theory, become refined. For instance, a primary visual cell may, at a particular moment, signify the red component in the image of the planet Mars in the night sky — not as interpreted by the proverbial little green man sitting in our brain, but due to the dynamical integration of the cell into a complex network.

#### 2.4.2 THE BANDWIDTH PROBLEM

Important elementary thought processes take place within 200 to 400 ms. This sets an upper limit to the interval in which correlations can be measured. A lower limit to the time resolution with which signals can be observed is in the range of several milliseconds due to differences in transmission delays.<sup>3</sup> Since in addition noise must be tolerated, a bandwidth of 100 Hz for the cellular signal seems to be an optimistic estimate. With this bandwidth not too many signals can be distinguished in an observation time shorter than 200 ms. On the other hand, the discussion in Sec. 2.3.4 has demonstrated how important it is that the activity patterns express highly differentiated structures in terms of signal correlations. It is, however, not necessary that all this structure be expressed in parallel within short time intervals (of, say, 200 ms) by stationary correlations. For instance, two cells may have significantly anticorrelated signals for a short period of time. All synaptic pathways between them and to common target cells will then be inactivated. If the inactivated state of the synapses is rather insensitive to uncorrelated signals, then there is no need for continued expression of the anticorrelation in the signals of the two cells. In short, the effects of earlier correlations are stored in short-term memory. From an undifferentiated

---

<sup>3</sup>Meanwhile we know that life is slightly more complicated — and interesting.

initial state in which every part is cross-coupled with every other a highly differentiated state can thus be reached in a sequential manner. In each step some parts of the network are decoupled from each other so that later on no confusion between their signals can arise. The price to be paid is the time necessary for this sequential process and the path dependence of the final state.

Synaptic plasticity can be regarded as a means to store past signal correlations on a more permanent basis, so that the signals are freed from the necessity to make certain distinctions. The time required to reach certain final states is thereby reduced. This increased efficiency is partly paid for by a restriction in generality. Composite elements have more complicated signals than single cells and can express correlation structures much more efficiently. They are enabled to do so by their special connectivity patterns. In an extreme case a nervous system (or subsystem) is completely specialized by its permanent network structure and no further synaptic modulation is necessary for the completion of its tasks. This is the case studied by what was called here conventional brain theory.

### 2.4.3 FACING EXPERIMENT

There is an enormous body of empirical knowledge about the brain, gathered in such diverse fields as clinical observation, anatomy, neurophysiology, psychophysics, and cognitive psychology. Correlation theory must be judged mainly as an attempt at integrating the heterogeneous perspectives offered by the various methods into one coherent picture.

Localization theory, although strongly supported by experiment, created unrest because it posed the nervous integration problem, which has not been solved by conventional theory. In this situation many scientists have resigned from inquiring about the brain's function or have contented themselves with philosophical answers. Correlation theory tries to make sense out of localization theory and proposes a solution to the nervous integration problem.

Neurophysiology tells us that cellular signals fluctuate. Since this temporal fine structure is not locked to the stimulus it is often taken as a sign of unreliability. According to correlation theory such a rapid modulation of the signal is essential to brain function.

According to cognitive psychology and common observation, there is a short-term memory with a lifetime of many seconds. Cellular signals have a temporal correlation length considerably shorter than 100 ms. Synaptic plasticity on the other hand is too slow. Synaptic modulation introduces a dynamical variable as a basis for short-term memory which could have the correct characteristic time constant.

The most convincing support for the theory will have to be found on the microscopic level: demonstration of organized correlations, synaptic modulation, and synaptic plasticity controlled by precise correlations.

The theory requires validation of its claims also on another, the functional, level. Elementary processes can be simulated on computers and analyzed mathematically. More complex processes will have to be demonstrated with the help of specialized machines.

#### 2.4.4 CONCLUSIONS

The brain and its function present us with a tangle of interrelated problems, none of which can be solved, or even be precisely formulated, on its own. In this situation a concept of global brain organization is needed which partitions the integral problem into defined technical questions. Correlation theory may be regarded as such a global scheme.

### References

- [1] Ariëns Kappers CU, Huber GC, Crosby EC (1936) *The Comparative Anatomy of the Nervous System of Vertebrates, Including Man* (MacMillan, New York); a reprint appeared in 1960 (Hafner, New York)
- [2] Baranyi A, and Feher O (1981) Synaptic facilitation requires paired activation of convergent pathways in the neocortex. *Nature* **290**:413–415
- [3] Barlow HB (1972) Single units and sensation: A neuron doctrine for perceptual psychology? *Perception* **1**:371–394
- [4] Beurle RL (1956) Properties of a mass of cells capable of regenerating pulses. *Philos. Trans. R. Soc. London Ser. B* **240**:55
- [5] Bliss TVP, Gardner-Medwin AR (1973) Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetized rabbit following stimulation of the perforant path. *J. Physiol.* **232**:357–374
- [6] Fukushima K (1980) Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybern.* **36**:193–202
- [7] Haken H (1978): *Synergetics, An Introduction* (Springer, Berlin)
- [8] Häussler AF, von der Malsburg C (1983) Development of retinotopic projections, an analytical treatment. *J. Theor. Neurobiol.* **2**:47–73
- [9] Hebb DO (1949) *The organization of behavior* (Wiley, New York). See also Hayek FA (1952) *The sensory order* (Univ. of Chicago Press, Chicago)
- [10] Julesz B (1981) Textons: The elements of texture perception, and their interactions. *Nature* **290**:91–97
- [11] Little WA (1974) The existence of persistent states in the brain. *Math. Biosci.* **19**:101–120
- [12] Legendy CR (1967) On the scheme by which the human brain stores information. *Math. Biosci.* **1**:555–597

- [13] Luria AR (1973) *The Working Brain, An Introduction to Neuropsychology* (Penguin, New York)
- [14] von der Malsburg C, Willshaw DJ (1977) How to label nerve cells so that they can interconnect in an ordered fashion. Proc. Natl. Acad. Sci. U.S.A. **74**:5176–5178
- [15] Rosenblatt F (1962) *Principles of Neurodynamics. Perceptrons and the Theory of Brain Mechanisms* (Spartan, Washington, DC)
- [16] Sutherland NS (1968) Outlines of a theory of visual pattern recognition in animals and man. Proc. R. Soc. London Ser. B **171**:297–317
- [17] Willshaw DJ, von der Malsburg C (1976) How patterned neural connections can be set up by self-organization. Proc. R. Soc. London Ser. B **194**:431–445
- [18] Willshaw DJ, von der Malsburg C (1979) A marker induction mechanism for the establishment of ordered neural mappings: Its application to the retinotectal problem. Philos. Trans. R. Soc. London Ser. B **287**:203–243

# 3

## Firing Rates and Well-Timed Events in the Cerebral Cortex

M. Abeles<sup>1</sup>

with 7 figures

**Synopsis.** The prevailing view is that the firing times of a neuron are the realizations of a stochastic point process with variable underlying rate functions. The immediate outcome of this view are a host of procedures that estimate this putative underlying rate and its dependence on other observable events. In this contribution some of the most frequently used estimation procedures will be described. There are several pieces of experimental evidence that do not conform to the view that all that there is to neuronal firing is the underlying firing rate. The implication of these observations to neural network modeling is discussed.

### 3.1 Measuring the Activity of Nerve Cells

The accepted dogma is that the nerve cell in the central nervous system of vertebrates operates by summing up currents that flow from the dendrites into the cell body. These currents determine the voltage at the cell body. Somewhere near the cell body there is a region which has the lowest threshold. This region is often called the *spike trigger zone*. When the voltage there reaches threshold, an action potential is generated. The action potential propagates along the axon releasing transmitter as it traverses the various synapses. In many cases the spike trigger zone is located at the initial segment of the axon. Once started, the action potential also invades the cell body and possibly also some distance into the dendritic trunks.

The two most popular techniques for measuring the activity of nerve cells in the vertebrate's brain are the intracellular and the extracellular techniques. In the intracellular technique, the voltage difference across the cell membrane is measured. In the extracellular technique, the voltage dif-

---

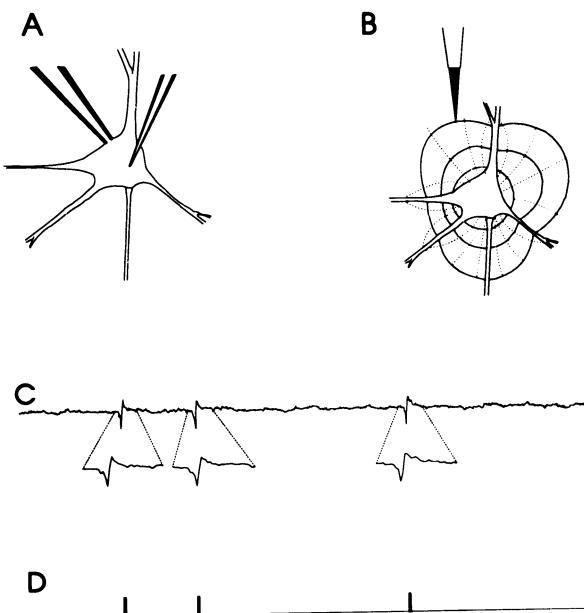
<sup>1</sup>Dept. of Physiology and The Center of Neural Computations, The Hebrew University of Jerusalem, P.O. Box 1172, Jerusalem 91-010, Israel.

ference between a point near an active neuron and another, remote (reference) point is measured. The rest of this section describes each of these techniques, and summarizes their advantages and shortcomings.

To obtain an intracellular recording, one must establish electrical contact with the interior of the cell. This is achieved mostly by using either a micropipette, or a patch electrode. Both types of electrodes are prepared by pulling a heated glass tubing into a very fine pipette, which is then filled with a salt solution. For the micropipette the tip diameter is very small (say  $0.2\text{ }\mu\text{m}$ ) and very sharp. The micropipette is used by pushing it into the cell, cutting through the cell's membrane on its way in. The patch electrode has a somewhat larger tip (say  $1\text{ }\mu\text{m}$ ) and the tip is dull. It is used by bringing it into contact with the cell membrane, applying a mild suction to seal the rim of the electrode's mouth to the membrane, and then breaking the membrane in the mouth by stronger suction (this situation is called, in the neurophysiological jargon, the *whole cell recording*). Fig. 3.1(A) illustrates these two methods of establishing electric contact with the interior of the cell. The voltage in the fluid filled pipette is measured by means of an electrode that can exchange ions with the fluid in a reversible chemical reaction. Most often a silver wire coated with a silver-chloride precipitate is used.

Intracellular recording can provide information about membrane potentials, conductivities, and synaptic currents as well as the time of the action potential. It also allows for injecting dyes to mark the cell and reconstructing its morphology from histological sections later on. Recordings through pipettes damp the high frequency component of the record due to the very high series resistance introduced by the fluid in the pipette's end and the capacitive shunt provided by the thin wall of the pipette. They also affect the composition of the intracellular fluid by diffusion to and fro the pipette's fluid. Despite these shortcomings, the intracellular method is a highly desirable way of recording the activity of nerve cells. However, in the living brain, it is very difficult to obtain prolonged recordings from a single neuron with such electrodes because they require great mechanical stability. A few  $\mu\text{m}$  movement would disrupt the contact between the electrode and the cell and would often result in an irreversible damage to the cell. The brain is a very soft (jelly-like) tissue, it may move around mechanically, expand, or shrink when the blood pressure changes. Even slight variations of venous pressure, following each breathing cycle can generate appreciable movement on the brains' surface. Thus, while there is abundant information which was obtained by intracellular recordings from brain slices in a dish, there is much less for the intact brain of anesthetized animals, and even less for behaving animals.

Most of the available data on activity of neurons in the brains of behaving animals was obtained by the extracellular recording technique. In this technique a very fine needle electrode is placed near the nerve cell body. Whenever the cell body is invaded by an action potential large currents are



**Fig. 3.1.** Techniques for recording the activity of a neuron. (A) A micropipette (right), or a patch electrode (left) are used to gain access into the interior of the cell. (B) A metal microelectrode is used to record the electric field around the cell's body. Some hypothetical currents (dotted lines) and isopotential surfaces (solid lines), which might be generated during an action potential, are drawn around the cell. (C) An example of extracellular activity, as recorded by a metal microelectrode. 70 ms of activity is shown. Activity was amplified, filtered in the range of 100–5000 Hz, and sampled at 12,000 samples per second. Below, magnified versions of the three spikes are plotted. (D) An abstract point process is derived from the activity recorded in (C).

flowing in the extracellular fluid between the cell body and the proximal dendrites (dashed lines in Fig. 3.1(B)). The current establishes a potential field according to its density, the geometry, and resistance of the extracellular space. An electrode in the vicinity of the cell body records the action potential as a brief (0.5 to 2 ms for cortical neurons) wave riding on a slowly varying background activity (Fig. 3.1(C)). The shape of these brief waves gave them the name *spikes*.

The exact size and shape of the spikes depend on the shape and properties of the neuron as well as on the electrode's position relative to the neuron. As long as these conditions are stable one expects to record the same shape. When a repeated appearance of the same spike shape is observed, it is assumed that it is generated by the same neuron. To be on the safe side, one refers to a series of spikes with a similar shape as a *single unit activity* (rather than a single neuron activity). The information in the single unit

activity is embedded in the time of occurrence of the spikes. Thus, if the activity was composed of spikes in times:

$$\{t_0, t_1, \dots, t_n\}$$

it is represented by a series of Dirac delta functions at these times:

$$f(t) = \sum_{i=0}^n \delta(t - t_i).$$

The single unit activity is often presented graphically as a sequence of dots or short bars as in Fig. 3.1(D).

## 3.2 Rate Functions and Stationary Point Processes

The simplest type of stochastic point process is the homogenous Poisson process, in which the probability of observing a spike at every small time window  $(t, t + \Delta t)$  is constant and independent of the history of firing:

$$\Pr\{\text{a spike in } (t, t + \Delta t)\} = \lambda \cdot \Delta t. \quad (3.1)$$

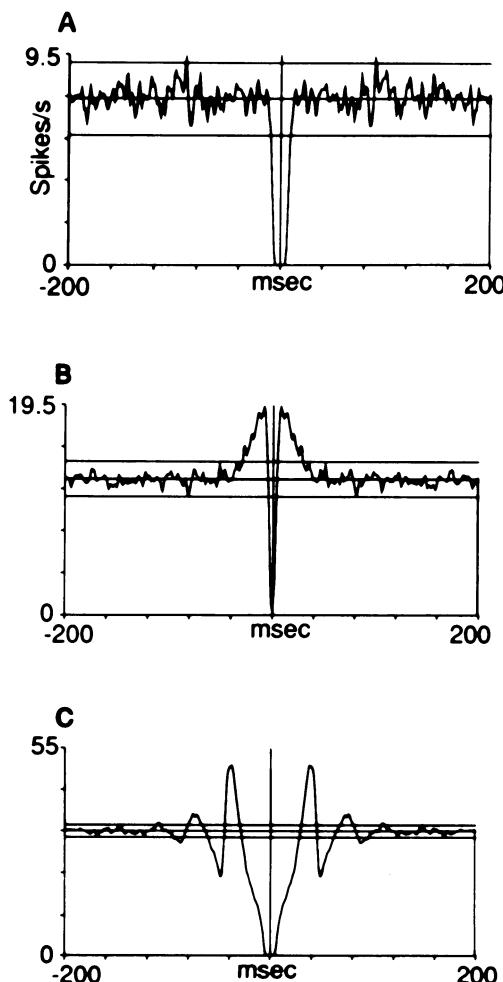
The parameter  $\lambda$  is the rate of the process. Equation (3.1) is valid when  $\Delta t$  is small enough such that the probability of seeing two spikes in  $(t, t + \Delta t)$  is zero. In neurophysiological reality, neuronal firing never achieve such Poissonian properties. Foremost because of the absolute (and relative) refractory period which prevents (or reduces the probability of) firing at short delays after the cell fired. At best firing statistics may behave like a renewal process, in which the firing probability depends only on the time elapsed after the occurrence of a spike (see [14] for discussion of such processes).

It is possible to evaluate the effect of a spike on the probability of observing subsequent spikes in a renewal process by

$$\Pr\{\text{a spike in } (t + \tau, t + \tau + \Delta t) | \text{a spike at } t\} = \lambda_{i,i}(\tau) \cdot \Delta \tau. \quad (3.2)$$

The rate function  $\lambda_{i,i}$  is the *renewal density* function. However, in neurophysiology it is usually called the *auto-correlation* function, and in the following text I shall adhere to this acceptable name (see [31] and [2] for further elaboration on the use of auto-correlations in neurophysiology).

A firing pattern which departs from a Poisson process only due to refractoriness would look like Fig. 3.2(A). The auto-correlation seen there was computed from data collected during 1000 s. The finite amount of data collected resulted in an estimated auto-correlation which is not flat. How is it possible to tell apart fluctuations which represent underlying changes in the auto-correlation from random fluctuations due to the finite amount



**Fig. 3.2.** Auto-correlations of spike trains. The auto-correlations were computed from 1000 s of simulated spiking activity. The average firing rate is plotted as a horizontal line. The range that should contain 99% of the data points (had all the fluctuations occurred by chance) is marked by the lines below and above the average rate. Note that as the firing rate increases the relative size of random fluctuations decreases. (A) Poisson-like spiking with refractoriness. (B) Refractoriness followed by an increased tendency to fire. (C) Noisy oscillator.

of data? In a Poisson process, the probability of observing  $n$  events when  $x$  are expected is given by

$$\Pr\{n|x\} = \frac{\exp(-x) \cdot x^n}{n!}. \quad (3.3)$$

The expected number of observations ( $x$ ) in a time window  $(\tau, \tau + \Delta\tau)$  can be calculated from

$$x = N \cdot \lambda \cdot \Delta\tau,$$

where  $N$  is the total number of observations made in that time window. For the auto-correlation function after each spike we look whether another spike occurred at that time window. Therefore the number of observations is equal to the total number of spikes:

$$N = \lambda \cdot T,$$

yielding

$$x = \lambda^2 \cdot T \cdot \Delta\tau. \quad (3.4)$$

It is possible to evaluate which of the fluctuations might be attributed to chance by using Eqs. (3.4) and (3.3). In Fig. 3.2 there are three horizontal lines. The line at the center is set at the average firing rate of the unit. The top and bottom lines set the region which should include 99% of the data if the process was Poissonian (see [3] for more details).

The shape of autocorrelation which is found most frequently in the cortex of awake animals is illustrated in Fig. 3.2(B). There, after a brief refractoriness the firing rate is elevated above the average and then slowly decays back to the average rate. The decay time may vary over a wide range (10–1000 ms). When such an autocorrelation is observed experimentally, one often says that the unit shows a tendency to *burst* (i.e., fire a group of spikes followed by a long silence). Had the unit fired in bursts of  $n$  spikes, the area under the hump should have been

$$\int_0^T \lambda_{i,i}(t) dt - \lambda T = (n^2 - n)/2, \quad (3.5)$$

where  $T$  is the duration of the hump and  $\lambda$  is the average firing rate of the cell. However, the area under the hump is usually less than 1, indicating that there are no real bursts, but shortly after recovery from the refractory period the cell is more likely to fire again.

The source of this hump in the auto-correlogram is a neglected topic in physiology. To date it is not known when deviations from flatness represent internal processes of the neuron, and when they result from a globally organized activity in the network.

In recent years, periodic oscillations in neurons of the cerebral cortex aroused much attention [12,20,27]. The autocorrelation function is well

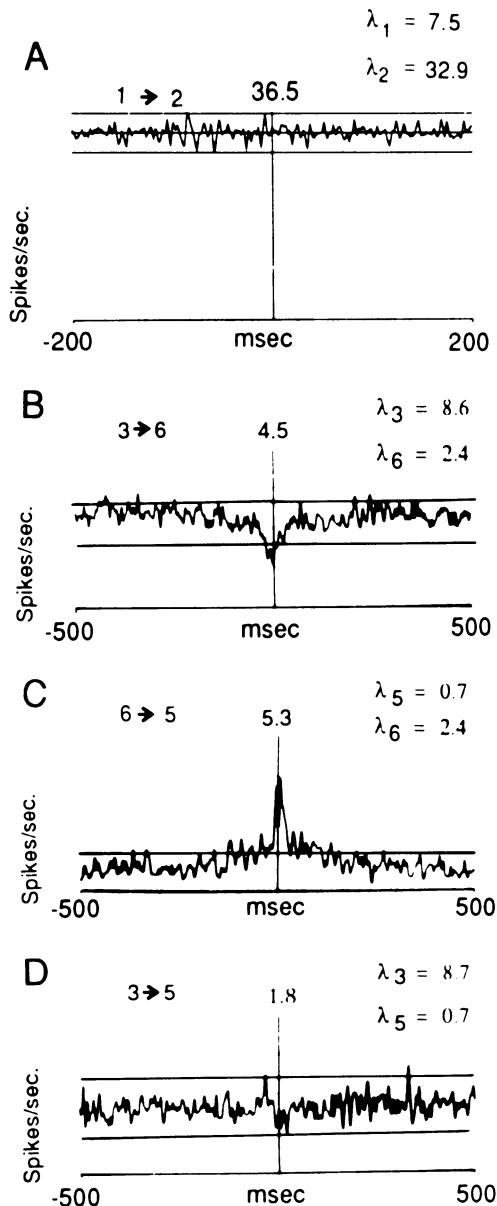
suited for revealing periodic oscillations. Figure 3.2(C) shows the auto-correlation of a simulated oscillating neuron when subject to noisy synaptic input. In the cortex periodic oscillations at low frequencies (2–12 Hz) are widely distributed. Those are not associated directly with information processing as evident from their prevalence during periods of drowsiness and sleep. It was experimentally shown that higher frequencies (20–60 Hz) can be associated with mental processes under the appropriate conditions. Despite the wonderful modeling possibilities that oscillations provide, it must be stressed that in the behaving animal, most of the cells in most cortical regions do not show oscillations most of the time.

The relations between the firing times of one neuron to the other are usually studied by means of cross-correlations. This function describes the variations in firing rate of one single unit around the time that another single unit fired. That is, the probability of seeing a spike from single unit  $j$  within  $(\tau, \tau + \Delta\tau)$  after unit  $i$  fired is given by

$$\Pr\{\text{a spike of } j \text{ at } (\tau+\tau, \tau+\tau+\Delta\tau) | \text{ a spike of } i \text{ at } \tau\} = \lambda_{i,j}(\tau) \cdot \Delta\tau, \quad (3.6)$$

where  $\lambda_{i,j}$  is the cross-correlation function. Again, as for the autocorrelation, what is computed is an estimator for an intensity function [15], and not a correlation function, but the name cross-correlation is widely used in the neurophysiological jargon, and I shall adhere here to this name (see [3, 7, 32] for more details on how these functions are computed and their relationships to true correlation functions).

Figure 3.3 shows several cross-correlations. Figure 3.3(A) shows the cross-correlation between the two simulated spike trains whose auto-correlations were shown in Figs. 3.2(A) and 3.2(C). Although within each of these trains the firing times had a well-defined structure (as seen in the nonflat autocorrelation), there is no structure to be seen in the cross-correlation. The small fluctuations seen in Fig. 3.3(A) are due to the finite size of the sample used to estimate the cross-correlation. Figures 3.3(B), 3.3(C), and 3.3(D) show estimates of cross-correlations between three next-neighbors in the auditory cortex of unanesthetized cats (designated units #3, #5, and #6). The correlation between units #3 and #6 (Fig. 3.3(B)) shows a trough around zero delay, meaning that the activity of the two units had negative correlation — when one unit fires the other tends to stay quiet. The correlation between units #6 and #5 (Fig. 3.3(C)) shows a hump around zero delay, meaning that the activity of the two units had positive correlation — when one unit fires the other also tends to fire. However, units #3 and #5 show no correlation (Fig. 3.3(D)). This seems peculiar. If (A) is correlated with (B) and (B) is anticorrelated with (C), then (A) should be anticorrelated with (C). However, since the correlations between units #3 and #6, and #6 and #5 are weak, the amount of correlation carried over from #3 to #5 (through #6) might be too weak to be detected with the amount of available data.



**Fig. 3.3. Cross-correlations of spike trains.** (A) Between two simulated spike trains. The auto-correlations of these spike trains were shown in Figs. 3.2(A) and 3.2(C). In the simulations there were no cross-relations between the firing times of the two cells. (B), (C), and (D) For three single units in the cortex. All three units were recorded through the same microelectrode, and therefore, are next neighbors.

In the behaving monkey most of the cross-correlations are flat (as in Fig. 3(D)) or show a small hump (as in Fig. 3.3(C)) around zero delay. Few show a trough (as in Fig. 3.3(B)) and still fewer show a narrow peak limited to positive (or negative) delays only. This last feature is taken as evidence for a direct excitatory interaction between the two units. While the humps (or troughs) that straddle the origin are taken as proof of some common source that affects both units (see [21] for a comprehensive review of cross-correlations among cortical cells and their interpretations).

In awake and behaving monkeys, the form of the cross-correlation function is not stable, it often changes from no correlations into positive correlation and back, or even from positive to negative correlation, depending on the arousal level, external stimuli, or the “psychological set” of the monkey. Such changes can take place within a fraction of a second (e.g., [35]). Even correlations that indicate direct interactions among the single units often show large changes with time. Thus, it is very likely that what is observed in cross-correlations is not simply the expression of the stationary anatomical connectivity, but more the outcome of the dynamics of the network within which the units are embedded.

Third-order correlations ( $\lambda_{i,j,k}(\tau_1, \tau_2)$ ) may be computed too. They estimate the probability of seeing a spike of single unit  $i$  at a delay  $(\tau_1, \tau_1 + \Delta\tau)$  and a spike of single unit  $j$  at a delay  $(\tau_2, \tau_2 + \Delta\tau)$  after the single unit  $k$  fired [4,33]. Such correlations and higher order correlation functions are seldom used although a lot of information about the interactions among neurons can be extracted from these correlations (e.g., [2], Chap. 6).

### 3.3 Rate Functions for Nonstationary Point Processes

All the rate functions described in the previous chapter ( $\lambda, \lambda_{i,i}, \lambda_{i,j}, \lambda_{i,j,k}$ ) describe the various rate dependencies accurately for stationary processes. When the process is not stationary an extra argument  $t$  must be added to all the functions. Thus, the constant firing rate  $\lambda$  becomes a variable rate  $\lambda(t)$ , and the cross-correlation  $\lambda_{i,j}(t)$ , which depends only on the delay between the two spikes becomes  $\lambda_{i,j}(t_1, t_2)$ , etc... While these functions are well defined for nonstationary stochastic point processes, the method for experimentally estimating them is not obvious.

Suppose there exists a population,  $\Omega$ , of single units with nonstationary firing rates and we wanted to estimate their firing rates around time  $t$ . We would sample a large number,  $N$ , of these neurons simultaneously and count how many of these neurons,  $n$ , fired during  $(t, t + \Delta t)$ . The firing rate  $\lambda(t)$  is estimated by

$$\lambda(t) \approx \frac{n}{N \cdot \Delta t}. \quad (3.7)$$

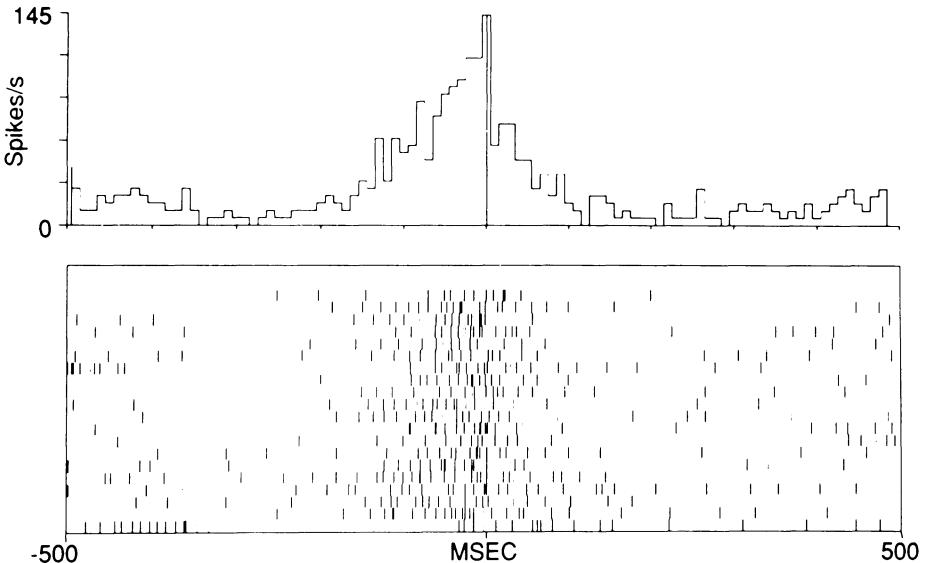
This estimation method cannot be carried experimentally, but can be approximated under the following conditions:

Suppose the variable firing rate  $\lambda(t)$  is linked to an external event (such as a stimulus or a movement), in such a way that it varies only over a limited time  $(-T, T)$  around this event. The activity of a single unit may be recorded around many,  $N$ , occurrences of the event (provided they are spaced more than  $2T$  apart), and then the record may be broken into  $N$  pieces, each lasting  $(-T, T)$  seconds around one event. These  $N$  pieces are treated as if they have been  $N$  independent realizations of the stochastic process, with time reset to 0 at the instance of each event. Equation (3.7) is used to estimate the variations of the firing rate around the event. When so doing,  $t$  is no more the absolute time, but the time around the external event. Estimates of the firing rates obtained in this way are called PSTH (poststimulus time histogram), or PETH (peri event time histogram). Note that this procedure estimates the firing rates of a hypothetical population of neurons having the same response properties as the sampled single unit. This is by no means the same as a population of neurons which is defined by other criteria (such as all the neurons within a certain brain territory). (See [6], Chap. 3, or [7] for a more rigorous description of this procedure.)

Figure 3.4 illustrates this procedure for estimating  $\lambda(t)$ . The data is taken from recording of the activity of a single unit in the frontal cortex of a behaving monkey. The monkey was trained to localize sound and light sources and when so instructed, to touch the target from which either sound or light was delivered 1–2 s earlier (see [7] for more details). The lower part of Fig. 3.4 shows the activity of single unit #4 around the time that the monkey touched target number 3. The action potentials of this single unit are displayed as short bars. All action potentials that happened during 500 ms around the first trial in which the monkey touched this target are displayed in the bottom line of Fig. 3.4. All the spikes around the second trial in which this event occurred are displayed above those of the first trial, and so on for all 20 repetitions of this event. This kind of display is usually called a *raster dot display*. The 20 lines of dots represent 20 realizations of a stochastic point process. By setting up 100 time bins the firing rate of the process is estimated, as shown in Fig. 3.4. It is possible to obtain better estimates than the one shown in Fig. 3.4, by using time bins with finer resolution and applying various smoothing functions [3]. Time varying auto- and cross-correlations can be computed in a similar way. A cross-correlation of this sort  $\lambda_{i,j}(t_1, t_2)$  gives the probability of observing a spike of single unit  $i$  at  $(t_1, t_1 + \Delta t)$  and also a spike of single unit  $j$  at  $(t_2, t_2 + \Delta t)$  by

$$\Pr\{\dots\} = \lambda_{i,j}(t_1, t_2) \cdot \Delta t^2. \quad (3.8)$$

This cross-correlation function is estimated by a similar procedure to the one used for the rate function. Activity of two single units is measured



**Fig. 3.4.** *Peri event time histogram (PETH).* The plots are derived from a recording of a single unit in the frontal cortex of a behaving monkey. *Lower panel:* A raster dot display showing all the spikes of a single unit around the time that the monkey touched a given target. Activity in successive occurrences are plotted above each other. *Upper panel:* The estimated firing rate of the unit, as a function of time around the instant that the monkey touched the target.

around some external event. The sections around  $N$  repeated instantiations of the event are treated as  $N$ -independent realizations of a two-dimensional point process. The plane spanned by the two time variables  $(t_1, t_2)$  is parcelled into square bins and then one counts the number of times  $n(t_1, t_2)$  in which both a spike from spike train  $i$  and a spike from spike train  $j$  occurred in  $(t_1, t_1 + \Delta t)$  and  $(t_2, t_2 + \Delta t)$ , respectively. The cross-correlation is estimated by

$$\lambda_{i,j}(t_1, t_2) \approx \frac{n(t_1, t_2)}{N \cdot \Delta t^2}. \quad (3.9)$$

This estimated cross-correlation is known in the neurophysiological jargon as joint peri stimulus time histogram (JPSTH). It has been introduced by Aertsen et al. (1989) who also developed methods that allow to test whether the two single units ( $i$  and  $j$ ) show internal interactions beyond what is generated by the co-modulations of their firing rates by the stimulus, whether such internal interactions are stationary, and devised ways to estimate the probability of seeing fluctuations in the JPSTH by chance [29].

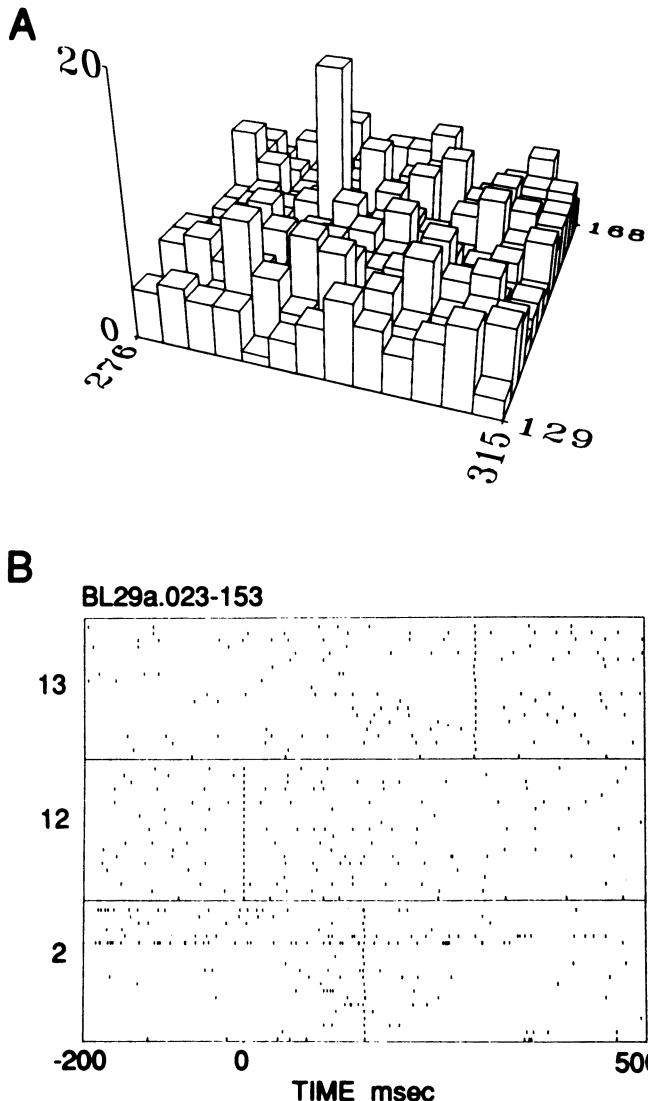
### 3.4 Rate Functions and Singular Events

The idea that firing rates are the way by which nerve cells communicate with each other became a well accepted dogma. Yet, there are many reports of fine structure of firing times that are inconsistent with the firing rate dogma [22,3 (Chap. 6), 4,16,25,24,19,8,36,13,37,7]. An example of such an event is shown in Fig. 3.5. It shows the interactions among three single units recorded in the frontal cortex of a behaving monkey. When the three-fold correlation  $\lambda_{12,13,2}(\tau_1, \tau_2)$  was estimated, one bin accumulated much more counts than what would be expected by chance. The correlogram was estimated for delays of 0 to 450 ms using 3 ms wide bins. The bin at  $(289 \pm 1, 151 \pm 1)$  accumulated 19 counts. That means that the event: a spike from single unit 12, at time  $t$ , and a spike from single unit 13 at  $(t + 289 \pm 1)$ , and also a spike from single unit 2 at  $(t + 151 \pm 1)$  repeated 19 times. The probability of seeing such an event by chance was 1 in 10 million. The actual events are shown in Fig. 3.5(B), where the activity of each of the involved single units is displayed in a different panel. Zero time is the time of the first spike (of unit #12) which was associated with this event. All the spikes within  $(-200, 500)$  ms around this spike are shown.

Two points jump to the eye when studying Fig. 3.5. First the spikes which are involved in the event are isolated. That is, they are not embedded in periods of strongly elevated firing rates. Second, despite the long delays (151 and 289 ms) the firing times stay accurate to within 3 ms. These properties do not conform the standard views of cortical function, as the multitude connections among cortical neurons, the weakness of each individual connection, and the prolonged integration time provided by synaptic potentials suggest that activity is generated by integrating multiple excitatory inputs which are dispersed in time, to generate a slowly varying firing rate. Yet, a systematic search for such tightly timed patterns revealed that they are widely spread [9].

There is no mathematical reason why a rate function  $\lambda_{i,j}(t_1, t_2)$  would not have a sharp, needle-like peaks, but intuitively, it seems that neuronal integration would not support such features of rate functions. It is argued that even if a pool of neurons receives a tightly synchronized volley of excitatory inputs, it would produce a dispersed output volley. Such dispersion is expected due to the gradual buildup and decay of the membrane depolarization and due to the noisiness of the synaptic transmission, the membrane potential, and the spike generating mechanism. Long delays are associated with transmission through several pools of neurons in tandem. Such arrangement is expected to cause further dispersion of the firing times. However, careful examination of neuronal properties reveals that this intuition is misleading.

The cortico-cortical excitatory synapses are weak [2, Chap. 4, 6,21,26] and the firing rates are low. In this situation the neurons are insensitive



**Fig. 3.5.** Relations among the spiking times of three single units. Data from recordings of activity in the frontal cortex of a behaving monkey. (A) A section from the threefold correlation among units #12, #2, and #13. The ordinate gives counts per bin —  $n(t_1, t_2)$ . The abscissae are the delays (in ms) from the firing of unit #12 to #13, and from #12 to #2. The expected counts per bin were 4.5, but one bin stood out with 20 counts. (B) A raster dot display of activity around the events that contributed counts to the outstanding bin in (A).

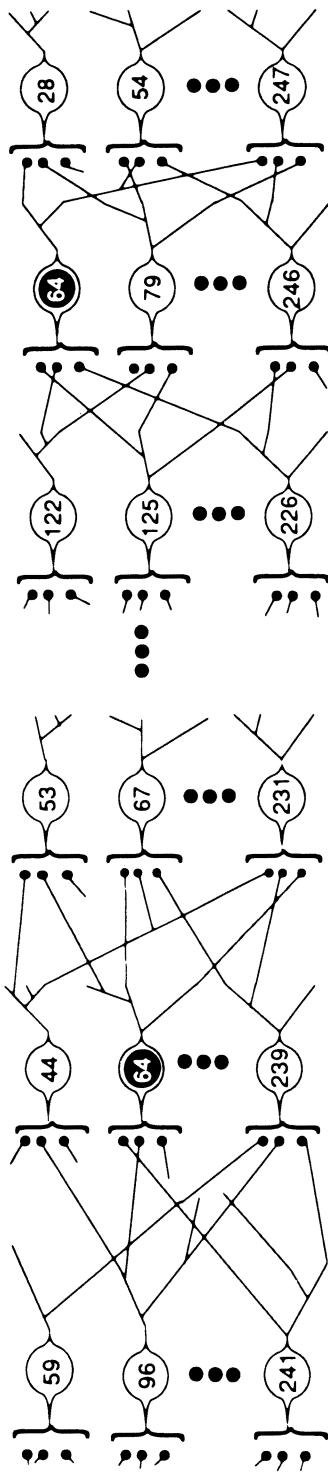
to dispersed excitatory inputs. They act like coincidence detectors not like integrators [1]. This selective sensitivity of cortical neurons to synchronous activation, was restressed by several investigators in recent years [23,28,34]. Thus, if a pool of neurons is activated by an asynchronous volley of spikes its output is likely to be more synchronized than its input. The weakness of cortico-cortical synapses also suggest that intracortical processing is carried by pools of neurons activating each other through multiple diverging and converging connections.

Suppose we are confronted with a chain of pools of neurons connected by diverging/converging connections as shown in Fig. 3.6. Suppose also that the first pool is excited by a continuous depolarizing current which produces an unsynchronized elevation of firing rate in the pool. If this activity is to propagate at all through the network, it would do so in a synchronized mode due to the following process. The second pool would preferentially pick up instances at which several of the neurons in the first pool fired synchronously by chance. Thus the activity of the neurons in the second pool would show a larger degree of internal synchrony. This would amplify itself further in the third pool, etc... [30,6 (Chap. 7), 7]. Put in other words a chain of neuronal pools linked by diverging-converging connections acts as a spatio-temporal filter for synchronism. Such a chain was called a *syn-fire chain*.

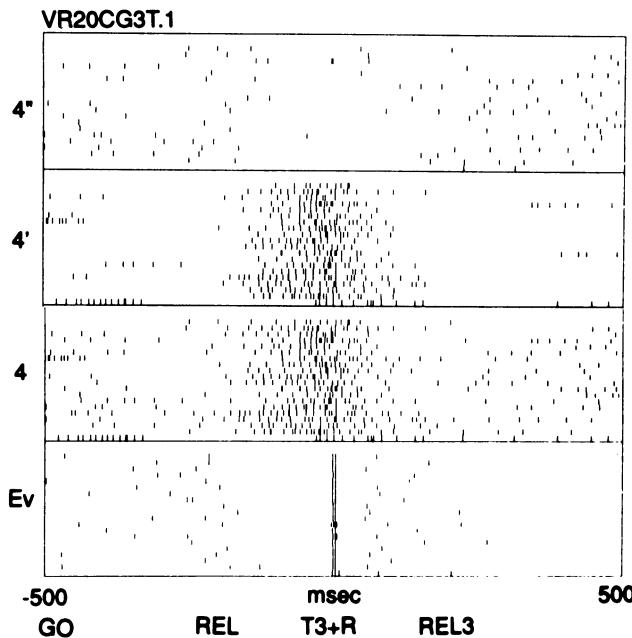
It is obvious that if the first link in a syn-fire chain is activated with absolute synchronism, its output would be dispersed in time, relative to the input. The next link may further increase this dispersion. However, the sensitivity of the neurons to synchronous activation would prevent the dispersion to build up too much. After a few links activity would reach a stable level of dispersion. From there on, any increment of dispersion, caused by chance, would be reduced back in the following links [30,6 (Chap. 7), 7]. This stable level of dispersion is expected to be of the same order of magnitude as the rise time of the excitatory synaptic potential (a few ms).

If during the experiment we happen to record from three neurons in the syn-fire chain, we expect to observe the same firing sequence whenever the chain is activated. Thus, the firing pattern seen in Fig. 3.5(B), is not a code of any sort. Rather, it is an indication (for the scientist) of the type of process taking place in the brain. The abundance of such patterns and their frequent association with external events (stimuli or movements) suggest that much of the intracortical processing is carried by syn-fire chains. This notion received strong support by the finding that the responses of single units in the cortex are often completely composed of accurately repeating firing patterns [7].

An example of this phenomenon is shown in Fig. 3.7 in which the results of the following analysis are presented. During session VR20 the activity of eight single units was recorded in parallel through six microelectrodes. The activity was divided into short (1 to 10 s) sections around various



**Fig. 3.6. The structure of a syn-fire chain.** The figure illustrates the idea of diverging-converging connections between sets of neurons. The structure is similar to a multilayered network, but with random connections between layers. The position of the cells in the drawing represents their sequence of activation and not their anatomical positions.



**Fig. 3.7.** *Random and well timed spikes.* The figure shows event-related activity of a single unit in the frontal cortex of a behaving monkey (the same data as in Fig. 3.4). The activity of single unit #4 was decomposed into well timed spikes (4'), that is, spikes associated with excessively repeating patterns, and random spikes (4''), that is, all the spikes of #4 which were not classified as 4'. The panels marked as 4, 4', and 4'' show the raster dot displays of these spikes around the time that the monkey touched target # 3. The panel marked Ev shows the times of the external events: GO — the instances at which the monkey was instructed to move, REL — the instance at which the movement started, T3+R — the instances at which the monkey touched the target and obtained the reinforcement (a drop of juice), REL3 — the instance at which the monkey left the target. T3+R was used to align all the rasters.

events. Among others, sections lasting from 0.5 s before the monkey was instructed to touch target # 3, until 0.5 s after he released the target were collated together into one data file. On this file a comprehensive search for all repeating firing patterns lasting up to 100 ms was carried. The search for repeating patterns used an algorithm that detects every repeating pattern regardless of the number of spikes in the pattern or its single unit composition [8]. The algorithm detected 19 distinct triplets each of which repeated at least 14 times. Had the firing times in the data been random only 0.57 such patterns would be expected. Note that the probability of observing 19 or more when only 0.57 are expected is given by

$$\begin{aligned}\Pr\{n \geq 19 | 0.57 \text{ is expected}\} &= \exp(-0.57) \times \sum_{i=19}^{\infty} \frac{0.57^i}{i!} \\ &< 0.000,000,000,000,000,000,000,2.\end{aligned}$$

Such extremely low probability often raises doubts about the assumptions used to calculate it. Yet, the computing algorithm was found to be very robust when tested against simulated data with variable firing rates and pairwise correlations [8]. Furthermore, although the expected value of 0.57 was obtained by pulling together triplets of all possible compositions (such as 1,1,1 ; 1,1,2 ; ...), all the 19 repeating patterns were composed of triplets of spikes from the single unit #4.

All these patterns occurred preferentially around the time that the monkey touched the target. Figure 3.4, which was used to illustrate the PETH of this unit, is based on the same data. Thus, the 19 well structured repeating patterns and the elevated firing rate of this unit are somehow associated. To clarify this relation all the spikes of unit #4 which composed one of the 19 patterns were marked as spikes of a pseudo single unit 4'. The results of this classification are shown in Fig. 3.7. Clearly all the activity which was associated with the movement is composed of the repeating patterns! Thus, it seems safe to conclude that the neural mechanism which elicited the activity of single unit #4 is generating strictly timed spikes even after long delays. It was proposed here that well timed spikes can be generated by syn-fire chains.

Yet, observations as described above seem at par with the simple syn-fire chain model. For the simple model the patterns are expected to be composed of spikes from different single units (as shown in Fig. 3.5), while here they are composed of spikes from the same single unit. For the simple syn-fire chain, it is expected to observe an association between a single pattern and a given process. Yet, here 19 different patterns were associated with touching the target. For the simple syn-fire chain, it is expected that a certain pattern of firing will appear every time the same process is repeated. Here, the same movement was repeated 20 times, yet most of the patterns repeated only 14–16 times.

The repeated appearance of spikes from the same single unit in the patterns suggests that the same single unit may participate more than once in the same syn-fire chain. This idea was illustrated in Fig. 3.6 where unit #64 appeared twice, once as part of the pool composed of {44, 64, ..., 239} and once in the pool {64, 79, ..., 246}. It was considered a part of the first pool because together with the other members of this pool it was excited by neurons {59, 96, ..., 241} and was exciting neurons {53, 67, ..., 246}. It was also a part of the other pool, because together with the other members of that pool it was excited by neurons {122, 125, ..., 226} and was exciting neurons {28, 54, ..., 247}.

Once such repeated participation of the same neuron in several pools takes place, the syn-fire chain ceases to be a simple feed-forward structure.

The activity that percolates through the chain may not stop after reaching the last pool, but may restart again at various places along the chain. When this happens we speak about *reverberations*. Simulations of syn-fire chains with repeated participation confirmed this conjecture [7]. Furthermore they showed that the reverberations need not be periodic. As there are several possible regions for reinitiating activity in the syn-fire chain, the spatio-temporal pattern of reverberating activity may be very complex, supporting the experimental observations of several preferred firing patterns.

The current neural network models have two major flavors: the multi-layered, feed-forward network and the recurrent, feed-back network. The first one is best suited for perceptron-like classifications, while the second is best suited for generating associative memory [17]. The reverberating syn-fire chain is somewhere in between. On a short time scale (ten ms) it acts as a strict feed-forward network, while on the longer time scale (hundred ms) it is similar to the reverberating attractor. It is possible that with this type of syn-fire chain, the strength of both network types can be exploited. In addition to these standard features of neural networks, E. Bienenstock [personal communication] suggested that the well timed spatio-temporal firing constellations which are generated by the syn-fire chains may be used by the brain to dynamically bind several syn-fire chains into a more complex structure. Such dynamic binding and unbinding ability is a prerequisite for any network that can sustain cognitive processes, particularly language.

*Acknowledgments.* The experimental work reported here was carried together with E. Vaadia, H. Bergman, I. Haalman, E. Margalit, Y. Prut, and H. Slovin. The author is indebted to V. Sharkansky for help in the artwork. This research was supported in part by a grant from the United States-Israeli Binational Science Foundation (BSF).

## REFERENCES

- [1] Abeles M. (1982a) Role of the cortical neuron: Integrator or coincidence detector? *Isr. J. Med. Sci.* **18**:83–92.
- [2] Abeles M. (1982b) Local Cortical Circuits: An Electrophysiological Study (Springer, Berlin)
- [3] Abeles M. (1982c) Quantification, smoothing, and confidence limits for single-units' histograms. *J. Neurosci. Meth.* **5**:317–325.
- [4] Abeles M. (1983) The quantification and graphic display of correlations among three spike trains. *IEEE Trans. BME* **30**:235–239
- [5] Abeles M. (1988) Neural codes for higher brain functions. In: *Information Processing by the Brain*, H.J. Markowitsch (Ed.) (Hans Huber Pub. Toronto), pp. 225–238

- [6] Abeles M. (1991) *Corticonics: Neural Circuits of the Cerebral Cortex* (Cambridge University Press, New York)
- [7] Abeles M. (1993) Analysis of single-unit activity in the cerebral cortex. In: *Neural Modeling and Neural Networks*, F. Ventriglia (Ed.) (Pergamon Press); Abeles M., Bergman H., Margalit E., Vaadia E. (1993a) Spatio-temporal firing patterns in the frontal cortex of behaving monkeys. *J. Neurophysiol.*, submitted
- [8] Abeles M., Gerstein G.L. (1988) Detecting spatio-temporal firing patterns among simultaneously recorded single neurons. *J. Neurophysiol.* **60**:909–924
- [9] Abeles M., Prut Y., Bergman H., Vaadia E., Aertsen A. (1993b) Integration, synchronicity, and periodicity. In: *Spatio-temporal Aspects of Brain Function*, A. Aertsen, W. von Seelen (Eds.) (Elsevier, Amsterdam)
- [10] Abeles M., de Ribaupierre F., de Ribaupierre Y. (1983) Detection of single unit responses which are loosely time-locked to a stimulus. *IEEE Trans. SMC* **13**:683–691
- [11] Aertsen A.M.H.J., Gerstein G.L., Habib M.K., Palm G. (1989) Dynamics of neuronal firing correlation: Modulation of “effective connectivity.” *J. Neurophysiol.* **61**:900–917
- [12] Ahissar E., Vaadia E. (1990) Single cell cortical oscillators in a somatosensory cortex of awake monkey. *Proc. Natl. Acad. Sci. U.S.A.* **87**:8935–8939
- [13] Bartanyan G.A., Pyrogov A.A. (1991) *Neurobiological Foundation of Higher Brain Functions* (Nauka, Leningrad) (in Russian)
- [14] Cox D.R. (1962) *Renewal Theory* (Methuen, London)
- [15] Cox D.R., Isham V. (1980) *Point processes* (Chapman and Hall, London)
- [16] Dayhoff J., Gerstein, G.L. (1983) Favored patterns in spike trains. II. Application. *J. Neurophysiol.* **49**:1349–1363
- [17] Domany E., Kinzel W., Meir R. (1989) Layered neural networks. *J. Phys. A: Math. Gen.* **22**:2081–2102
- [18] Douglas, R.J., Martin, K.A.C., Whitteridge D. (1991) An intracellular analysis of the visual responses of neurones in cat visual cortex. *J. Physiol. (London)* **440**:659–696
- [19] Frostig R., Frostig Z., Frysinger R., Schechtman V. (1985) Multineuron analysis reveals complex patterns of interaction among neurons. *Soc. Neurosci. Abstr.* **11**:1020
- [20] Gray C.M., Singer W. (1989) Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* **86**:1698–1702; Eckhorn R., Bauer R., Jorden W., Brosch M., Kruse W., Munk M., Reitboeck H.J. (1988) Coherent oscillations: A mechanism for feature linking in the visual cortex? *Biol. Cybern.* **60**:121–130
- [21] Fetz E., Toyama K., Smith W. (1991) Synaptic interactions between cortical neurons. In: *Cerebral Cortex*, A. Peters (Ed.) (Plenum Publ.), Vol. 9

- [22] Klemm, W.R., Sherry, C.J. (1981) Serial ordering in spike trains: What's it "trying to tell us"? *Int. J. Neurosci.* **14**:15–23
- [23] Koch, C., Douglas, R.J., Wehmeier, U. (1990) Visibility of synaptically induced conductance changes: Theory and simulations of anatomically characterized cortical pyramidal cells. *J. Neurosci.* **10**:1728–1744
- [24] Landolt J., Reinis S., Weiss D. (1985) Identification of local neuronal circuits in the visual cortex of the cat. *Soc. Neurosci. Abstr.* **11**:1010
- [25] Legendy C., Salcman M. (1985) Bursts and recurrences of bursts in the spike trains of spontaneously active striate cortex neurons. *J. Neurophysiol.* **53**:926–939
- [26] Mason A., Nicoll A., Straford K. (1991) Synaptic transmission between individual pyramidal neurons of the rat visual cortex in vitro. *J. Neurosci.* **11**:72–84
- [27] Murthy, V.N., Fetz E.E. (1992) Coherent 25–35 Hz oscillations in the sensorimotor cortex of awake behaving monkeys. *Proc. Natl. Acad. Sci. U.S.A.* **89**:5670
- [28] Nelken, I. (1988) Analysis of the activity of single neurons in stochastic settings. *Biol. Cybern.* **59**:201–215
- [29] Palm G., Aertsen A.M.H.J., Gerstein G.L. (1988) On the significance of correlations among neuronal spike trains. *Biol. Cybern.* **59**:1–11
- [30] Pantiliat S. (1986) Theoretical model of nervous system. Ph. D. thesis, the Hebrew University of Jerusalem
- [31] Perkel D.H., Gerstein G.L., Moore G.P. (1967a) Neuronal spike trains and stochastic point processes. I. The single spike train. *Biophys. J.* **7**:391–418
- [32] Perkel D.H., Gerstein G.L., Moore G.P. (1967b) Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. *Biophys. J.* **7**:419–440
- [33] Perkel D.H., Gerstein G.L., Smith M.S., Taton W.G. (1975) Nerve impulse patterns: A quantitative technique for three neurons. *Brain Res.* **100**:271–296
- [34] Reyes A.D., Fetz E.E. (1993) Effects of transient depolarizing potentials on the firing rate of neocortical neurons. *J. Neurophysiol.*, in press
- [35] Vaadia E., Aertsen A. (1992) Coding and computation in the cortex: Single-neuron activity and cooperative phenomena. In: *Information Processing in the Cortex*, A. Aertsen and V. Braatenberg (Eds.) (Springer, Berlin), pp 81–121
- [36] Villa A.E.P., Abeles M. (1990) Evidence for spatio-temporal firing patterns within the auditory thalamus of the cat. *Brain Res.* **509**:325–327
- [37] Villa A.E.P., Fuster J.M. (1992) Temporal correlates of information processing during visual short-term memory. *NeuroReport* **3**:113–116

# 4

## The Role of Synchrony in Neocortical Processing and Synaptic Plasticity

Wolf Singer<sup>1</sup>

with 3 figures

**Synopsis.** One of the basic functions in sensory processing consists of the organization of distributed neuronal responses into representations of perceptual objects. This process requires selecting responses on the basis of certain Gestalt criteria and establishing unambiguous relations among the subset of selected responses. This article proposes that response selection and the subsequent binding of selected responses on the basis of certain Gestalt criteria and establishing unambiguous relations among the subset of selected responses. This article proposes that response selection and the subsequent binding of selected responses is achieved by the temporal synchronization of neuronal discharges. The predictions derived from this hypothesis are examined by combining multielectrode recordings from the mammalian visual cortex with neuroanatomical and behavioral analyses. The results indicate that spatially distributed neurons can synchronize their discharges on a millisecond time scale. The probability of synchronization depends both on the functional architecture of cortical connections and on the configuration of stimuli and reflects some of the Gestalt criteria used for perceptual grouping, correlating well with the animals' perceptual abilities.

### 4.1 Introduction

The analysis of the activation patterns of individual nerve cells in the brain has revealed numerous and fascinating correlations between the responses of single neurons and complex behavioral patterns. This evidence led to the suggestion that the activation of individual neurons can represent a code for highly complex functions, a notion that is commonly addressed as

---

<sup>1</sup>Max Planck Institute for Brain Research, Deutschordenstrasse 46, D-60528 Frankfurt/Main, Germany.

the “single neuron doctrine.” However, there have always also been speculations that additional coding principles might be realized. Most of these begin with Donald Hebb’s proposal that representations of sensory or motor patterns should consist of assemblies of cooperatively interacting neurons rather than of individual cells. This coding principle implies that information is contained not only in the activation level of individual neurons but also, and actually to a crucial extent, in the relations between the activities of distributed neurons. If true, the description of a particular neuronal state would have to take into account not only the rate and the specificity of individual neuronal responses but also the relations between discharges of distributed neurons. Over the last decade, these speculations have received some support both from experimental results and theoretical considerations. Search for individual neurons responding with the required selectivity to individual objects was only partly successful and has so far revealed specificity only for faces and for a limited set of objects with which the animal had been familiarized extensively before (see below). Recordings from motor centers such as the deep layers of the tectum and areas of the motor cortex provided no evidence for command neurons such as exist in simple nervous systems and code for specific motor patterns. Rather, these studies provided strong support for a population code as the trajectory of a particular movement could be predicted correctly only if the relative contributions of a large number of neurons were considered [1–3]. Arguments favoring the possibility of relational codes have also been derived from the growing evidence that cortical processes are highly distributed [see 4–6]. Further indications for the putative significance of relational codes are provided by theoretical studies which attempted to simulate certain aspects of pattern recognition and motor control in artificial neuronal networks. Single cell codes were found appropriate for the representation of a limited set of well defined patterns but the number of required representational elements scaled very unfavorably with the number of representable patterns. Moreover, severe difficulties were encountered with functions such as figure-ground distinction because single cell codes turned out to be too rigid and inflexible, again leading to a combinatorial explosion of the required representational units. By implementing population or relational codes some of these problems could be solved or at least alleviated. Exploiting relational codes also opens up the possibility to use time as an additional coding space. By defining a narrow temporal window for the evaluation of coincident firing and by temporal patterning of individual neuronal responses, relations between the activities of spatially distributed neurons can be defined very selectively [7,8]. If such temporal coding is added to the principle of population coding the number of different patterns or representations that can be generated by a given set of neurons increases substantially. Moreover, it has been demonstrated that perceptual functions like scene segmentation and figure-ground distinction which require flexible association of features can in principle be solved if one relies on relational codes in which the related-

ness of distributed neurons is expressed by the temporary synchronization of their respective discharges [7–10].

Arguments emphasizing the importance of temporal relations between the discharges of cortical neurons have also been derived from recent data on connectivity and synaptic efficacy. Cortical cells receive many thousand synaptic inputs but on the average a particular cell contacts any of its target cells only with one synapse [11]. In vitro studies from cortical slices indicate that the efficacy of individual synapses is low and that not every presynaptic action potential triggers the release of transmitter [12]. Thus, many presynaptic afferents need to be activated simultaneously in order to drive a particular cell above threshold and to assure reliable transmission. Even more cooperativity is required for the induction of synaptic modifications such as long-term potentiation and long-term depression. These modifications have high thresholds and require substantial and prolonged postsynaptic activation [13,14]. Temporal coordination of cortical responses thus appears necessary both for successful transmission across successive processing stages and for the induction of use-dependent synaptic modifications.

Despite these numerous arguments supporting the putative importance of temporal relations among distributed neuronal responses in the neocortex systematic search for temporal relations among the activities of simultaneously recorded cortical neurons is still at an early stage. It is only recently that cross-correlation analysis has been applied to responses evoked by selected stimulus configurations in order to test whether the responses of spatially distributed cortical neurons exhibit temporal relations that are sufficiently consistent to serve a functional role in cortical processing. Before reviewing the results of these cross-correlation studies the conceptual background common to many of these experiments will be briefly reviewed, taking visual processes as an example.

## 4.2 Pattern Processing and the Binding Problem

Perceptual objects can usually be decomposed into components and in general the features of these components are not unique for a particular object. The individuality of objects results from the specific composition of elementary features and their relations rather than from the specificity of the component features. Hence, for a versatile representation of sensory patterns in the nervous system three basic functions have to be accomplished. (1) Elementary features need to be represented by neuronal responses, (2) responses to features constituting a particular object have to be distinguished from features of other objects and need to be bound together, and (3) the specific relations between these features have to be encoded and preserved.

One way to achieve such grouping of features and to establish an un-

ambiguous code for their specific relations is to connect the set of neurons which respond to the component features of a particular object to a higher order neuron which will represent the object. If the thresholds of these higher order neurons are adjusted so that each cell responds only to one particular combination of feature detectors, the responses of these higher order neurons would provide an unambiguous description of the relations between the component features and hence would be equivalent to the representation of the pattern. In this scheme the features of the object are bound together by convergence of fixed connections which link neurons representing component features with neurons representing the whole pattern.

This hypothesis led to the systematic search for pattern specific cells [15] at higher levels of processing which culminated in the discovery of face-selective neurons in the infero-temporal cortex of monkeys [16–20] and of cells responding selectively to objects which the animal had previously been familiarized with [21,22].

At about the same time indications accumulated that sensory processes in the neocortex are highly distributed. This has been demonstrated particularly clearly for the mammalian visual system. There are more than 20 cortical areas, distinguished either by cytoarchitectonical features or by their connectivity to which specific visual functions have been assigned recently [for review see 5,19,23,24]. Although there is a serial progression of signal transformation from occipital to more frontal cortical territories many of the newly described visual areas are organized in parallel as suggested by the symmetry of their interconnections. In addition, all areas receive massive feed-back connections from the areas to which they project but occasionally also from areas to which they are feeding signals only indirectly.

In the framework of the “binding by convergence” hypothesis most of these distributed activities would have to become reintegrated by convergence at some higher levels of processing in order to generate the postulated sets of object specific cells. However, not all of the predictions following from this latter assumption are supported by experimental evidence. First, while cells occupying higher levels in the processing hierarchy tend to be selective for more complex constellations of features than cells at lower levels, many continue to respond to rather simple patterns such as edges, gratings, and simple geometrical shapes [25,26]. Second, apart from cells responding preferentially to aspects of faces and hands [16–18,20,27] it has been notoriously difficult to find other object-specific cells except in cases where animals had been familiarized with a limited set of objects during extensive training [22]. Third, no single area in the visual processing stream has been identified so far which could serve as the ultimate site of convergence and which would be large enough to accommodate the vast number of neurons that are required if all distinguishable objects including their many different views were represented by individual neurons. Finally, the point has

been made that "binding by convergence" may not be flexible enough to account for the rapid formation of representations of new patterns.

Because these difficulties cannot be overcome easily in architectures noindent which solve the binding problem by serial recombination of converging feed-forward connections alternative proposals have been developed. These are all based on the assumption that representations consist of assemblies of a large number of simultaneously active neurons which may be contained in a single cortical area but which may also be distributed over many cortical areas [8,28-39]. The essential feature of assembly coding is that individual cells can participate at different times in the representation of different objects. The assumption is that just as a particular feature can be present in many different patterns, a neuron coding for this feature can be shared by many different representations. This reduces substantially the number of cells required for the representation of different objects and allows for considerably more flexibility in the generation of new representations.

Basic requirements for representing objects by such assemblies are: First, the responses of the cells responding to a visual scene need to be compared with one another and examined for possible, "meaningful" relations. Second, cells coding for features which can be related need to become organized into an assembly. This should be the case for the cells which are, for example, activated by the constituent features of a particular object. Third, if patterns change neurons should be able to rapidly change partners and to form new assemblies. Fourth, neurons which have joined a particular assembly must become identifiable as members of this very assembly. Their responses need to be tagged so that they can be recognized as being related, that is, the distributed responses of the assembly must be recognizable as representing a "whole." It is commonly assumed that these organizing steps, the probing of possible relations, the formation of an assembly, and the labeling of responses are achieved in a single self-organizing process by selective reciprocal connections between the distributed neuronal elements. The idea is that the probabilities with which neurons get organized into particular assemblies are determined, first, by the respective constellation of features in the pattern and, second, by the functional architecture of the assembly forming coupling connections. Several proposals have been made concerning the mechanisms by which these connections could serve to "label" the responses of neurons which have joined into the same assembly. Most of them assume that the assembly-generating connections are excitatory and reciprocal and serve to enhance and to prolong the responses of neurons which got organized in an assembly [34-36,38,40].

Another proposal is that assemblies should be distinguished in addition by a temporal code [8,9]. A similar suggestion, although formulated less explicitly, had been made previously by Milner [7]. Both hypotheses assume that the assembly forming connections should establish temporal coherence on a millisecond time scale between the responses of the coupled cells. Thus,

neurons having joined into an assembly would be identifiable as members of the assembly because of the synchronization of their discharges. Expressing relations between members of an assembly by the temporal coherence rather than the amplitude of their responses has several advantages: First, it reduces the ambiguities which result from the fact that discharge rates depend strongly on variables such as stimulus intensity and quality of fit between stimulus features and receptive field properties. Relying on temporal relations also preserves the important option to use discharge rates as a code for stimulus parameters. In systems exploiting coarse codes this is essential because the information about the presence of a particular feature and its precise location is contained in the graded responses of populations of cells. Second, by exploiting temporal relations the number of assemblies that can be active simultaneously without becoming confounded can be increased. In most cases simultaneously active assemblies will be distinguishable because of the spatial segregation introduced by topological maps and functional compartmentalization of cortical areas. But in order to allow for a flexible association of simultaneously active assemblies additional distinctions are required to avoid fusion of assemblies that should be kept apart [see, e.g., 8]. Responses of neurons could overlap on a coarse time scale but still remain distinguishable as coming from a particular assembly if they are correlated at a fine time scale. Third, cells which succeeded to synchronize their discharges have a stronger impact on target cells (see above). Thus, formation of coherently active assemblies can serve to enhance the saliency of responses to features that can be associated in a "meaningful" way. This may contribute to the segregation of object related features from unrelated features of the background. This concept of "binding by synchrony" has also been applied to intermodal integration [41] and even to high level processes underlying phenomena such as attention [30] and consciousness [42].

If assemblies are indeed distinguished by the temporal coherence of the responses of the constituting neurons a set of predictions can be derived.

1. Spatially segregated neurons should synchronize their responses if activated by features that can be grouped together. This should be the case, for example, for features constituting a single object.
2. Synchronization should be frequent among neurons within a particular cortical area but it should also occur across cortical areas.
3. The probability that neurons synchronize their responses both within a particular area and across areas should reflect some of the Gestalt criteria used for perceptual grouping.
4. Individual cells should be able to rapidly change the partners with which they synchronize their responses if stimulus configurations change and require new associations.

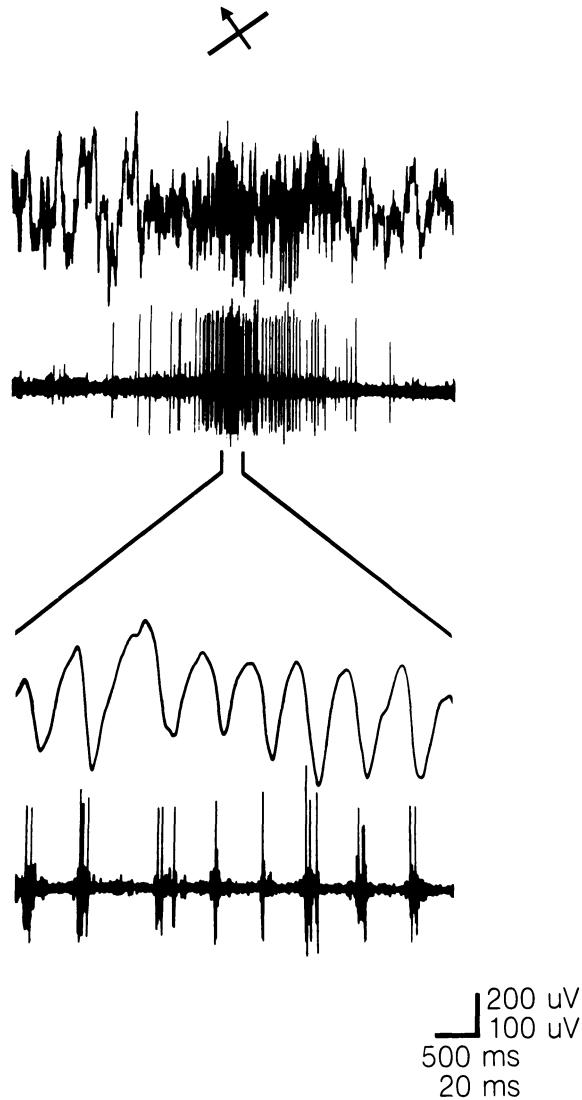
5. If more than one object is present in a scene several assemblies should form. Cells belonging to the same assembly should synchronize their responses while no consistent temporal relations should exist between the discharges of neurons belonging to different assemblies.
6. Synchronization probability should at least in part depend on the functional architecture of reciprocal cortico-cortical connections and should change if this architecture is modified.

### 4.3 Evidence for Dynamic Interactions Between Spatially Distributed Neurons

Systematic testing of the above listed predictions had been initiated by the observation that adjacent neurons in the cat visual cortex can transiently engage in highly synchronous discharges when presented with their preferred stimulus [43]. Groups of neurons recorded simultaneously with a single electrode were found to discharge in synchronous bursts which followed one another at intervals of 15 to 30 ms. These sequences of synchronous rhythmic firing appear preferentially when cells are activated with slowly moving light stimuli. They last no more than a few hundred milliseconds and may occur several times during a single passage of the moving stimulus (Fig. 4.1). Accordingly, autocorrelograms computed from such response epochs often exhibit a periodic modulation [43–48]. During such episodes of synchronous firing a large oscillatory field potential is recorded by the same electrode, the negative deflections being coincident with the cells' discharges. The occurrence of such a large field response indicates that many more cells in the vicinity of the electrode than those actually picked up by the electrode must have synchronized their discharges [44]. However, neither the time of occurrence of these synchronized response episodes nor the phase of the oscillations are related to the position of the stimulus within the neuron's receptive field. When cross-correlation functions are computed between responses to subsequently presented identical stimuli, these "shift predictors" reveal no relation between the temporal patterning of successive responses [44,46]. The rhythmic firing and the synchronization of the bursts are thus not related to some fine spatial structure in the receptive fields of cortical neurons but result from neuronal interactions.

This phenomenon of local response synchronization has been observed with multiunit and field potential recordings in several independent studies in different areas of the visual cortex of anesthetized cats (areas 17, 18, 19, and PMLS) [44–46,49,50], in area 17 of awake cats [51,52] in the optic tectum of awake pigeons [53], and in the visual cortex of anesthetized [48] and awake behaving monkeys [54].

Subsequently, it has been shown with multielectrode recordings in anesthetized and awake cats [51,54–57] and anesthetized and awake monkeys



**Fig. 4.1.** Multiunit activity (MUA) and local field potential (LFP) responses recorded from area 17 in an adult cat to the presentation of an optimally oriented light bar moving across the receptive field. Oscilloscope records of a single trial showing the response to the preferred direction of movement. In the upper two traces, at a slow time scale, the onset of the neuronal response is associated with an increase in high frequency activity in the LFP. The lower two traces display the activity at the peak of the response at an expanded time scale. Note the presence of rhythmic oscillations in the LFP and MUA (35–45 Hz) that are correlated in phase with the peak negativity of the LFP. Upper and lower voltage scales are for the LFP and MUA, respectively (adapted from [44]).

[54,57] that similar response synchronization can also occur between *spatially segregated* cell groups within the same visual area. Interestingly, the synchronization of responses over larger distances also occurs with zero phase lag. Hence, if the cross-correlograms show any interaction at all they typically have a peak centered around zero delay. The half-width at half-height of this peak is in the order of 2–3 ms indicating that most of the action potentials which showed some consistent temporal relation had occurred nearly simultaneously. This peak is often flanked on either side by troughs which result from pauses between the synchronous bursts. When the duration of these pauses is sufficiently constant throughout the episode of synchronization, the cross-correlograms show in addition a periodic modulation with further side peaks and troughs. But such regularity is not a necessary requirement for synchronization to occur. There are numerous examples from anesthetized cats [see, e.g., 58,59] and especially from awake monkeys [54] that responses of spatially distributed neurons can become synchronized and lead to cross-correlograms with significant center peaks without engaging in rhythmic activity that is sufficiently regular to produce a periodical modulation of averaged auto- and cross-correlograms. However, there are relations between oscillatory discharge patterns and response synchronization which will be discussed in detail in a later chapter.

#### 4.4 Stimulus-Dependent Changes of Synchronization Probability

As outlined above the probabilities with which distributed cells synchronize their responses should reflect some of the Gestaltcriteria applied in perceptual grouping. Another and related prediction is that individual cells must be able to change the partners with which they synchronize whereby the selection of partners should occur as a function of the patterns used to activate the cells. In this paragraph experiments are reviewed which were designed to address these predictions. Detailed studies in anesthetized cats and recently also anesthetized and awake monkeys have revealed that synchronization probability for remote groups of cells is determined both by factors within the brain as well as by the configuration of the stimuli [50,55–58,60,61]. In general, synchronization probability within a particular cortical area decreases with increasing distance between the cells. If cells are so closely spaced that their receptive fields overlap, the probability is high that their responses will exhibit synchronous epochs if evoked with a single stimulus. The latter condition requires that the orientation and direction preferences of the cell pairs are sufficiently similar or that their tuning is sufficiently broad to allow for coactivation by a single stimulus. As recording distance increases synchronization probability becomes more

and more dependent on the similarity between the orientation preferences of the neurons [55,62].

Concerning the dependence of synchronization probability on stimulus configuration single linearly moving contours have so far been found to be most efficient. Gray et al. [56] recorded multiunit activity from two locations in cat area 17 separated by 7 mm. The receptive fields of the cells were nonoverlapping, had nearly identical orientation preferences, and were spatially displaced along the axis of preferred orientation. This enabled stimulation of the cells with bars of the same orientation under three different conditions: two bars moving in opposite directions, two bars moving in the same direction, and one long bar moving across both fields coherently. No significant correlation was found when the cells were stimulated by oppositely moving bars. A weak correlation was present for the coherently moving bars. But the long bar stimulus resulted in a robust synchronization of the activity at the two sites. This effect occurred in spite of the fact that the overall number of spikes produced by the two cells and the oscillatory patterning of the responses were similar in the three conditions.

These results indicate clearly, that synchronization probability depends not only on the spatial segregation of cells and on their feature preferences, the latter being related to the cells' position within the columnar architecture of the cortex, but also and to a crucial extent on the configuration of the stimuli. So far, synchronization probability appears to reflect rather well some of the Gestaltcriteria for perceptual grouping. The high synchronization probability of nearby cells corresponds to the binding criterion of "vicinity," the dependence on receptive field similarities agrees with the criterion of "similarity," the strong synchronization observed in response to continuous stimuli obeys the criterion of "continuity" and the lack of synchrony in responses to stimuli moving in opposite directions relates to the criterion of "common fate."

Experiments have also been performed in order to test the prediction that simultaneously presented but different contours should lead to the organization of two independently synchronized assemblies of cells [57,58]. If groups of cells with overlapping receptive fields but different orientation preferences are activated with a single moving light bar they synchronize their responses even if some of these groups are suboptimally activated [55,58]. However, if such a set of groups is stimulated with two independent spatially overlapping stimuli which move in different directions, they split into two independently synchronized assemblies, those groups joining the same synchronously active assembly that have a preference for the same stimulus (Fig. 4.2). Thus, the two stimuli evoke simultaneous responses in a large array of spatially interleaved neurons but these neurons get organized in two assemblies which can be distinguished because of the temporal coherence of responses within and the lack of coherence between assemblies. Cells representing the same stimulus exhibit synchronized response

epochs while no consistent correlations occur between the responses of cells that are evoked by different stimuli. The parameters of the individual responses such as their amplitude or oscillatory patterning were not affected by changes in the global configuration of the stimuli. Thus, it is not possible to tell from the responses of individual cells whether they were activated by a single contour or by two different stimuli. Even if one evaluated the extent of coactivation of the simultaneously recorded cells on a coarse time scale as would be sufficient for the analysis of rate coded populations one would not be able to decide whether the cells had been activated by one composite figure whose features satisfy the preferences of the active cells or by two independent figures which excite the same set of cells. The only cue for this distinction was provided by the evaluation of synchronicity at a millisecond time scale.

The results of these experiments also prove that individual groups can change the partners with which they synchronize when stimulus configurations change. Cell groups that engaged in synchronous response episodes when activated with a single stimulus no longer did so when activated with two stimuli but then synchronized with other groups. One methodological caveat following from this is that cross-correlation analysis does not always reliably reflect anatomical connectivity [see also 63]. In agreement with the predictions from the assembly hypothesis interactions between distributed cell groups were found to be highly dynamic, variable, and strongly influenced by the constellation of features in the visual stimulus.

## 4.5 Synchronization Between Areas

Experiments have also been designed to test the prediction that cells distributed across different cortical areas should be able to synchronize their responses if they respond to the same contour. In the cat, interareal synchronization of unit responses has been observed between cells in areas 17 and 18 [45,49,59], between cells in areas 17 and 19 and 18 and 19 [49], between cells in area 17 and area PLMS, an area specialized for motion processing [50,64], and even between neurons in A 17 of the two hemispheres [49,60,65]. In the macaque monkey synchronous firing has been observed between neurons in areas V1 and V2 [66,67]. In all of these cases, whenever tested, synchronization depended on receptive field constellations and stimulus configurations, similar to the intra-areal synchronization [50,60].

## 4.6 The Synchronizing Connections

It is commonly assumed in interpretations of cross-correlation data that synchronization of neuronal responses with zero-phase lag is indicative of

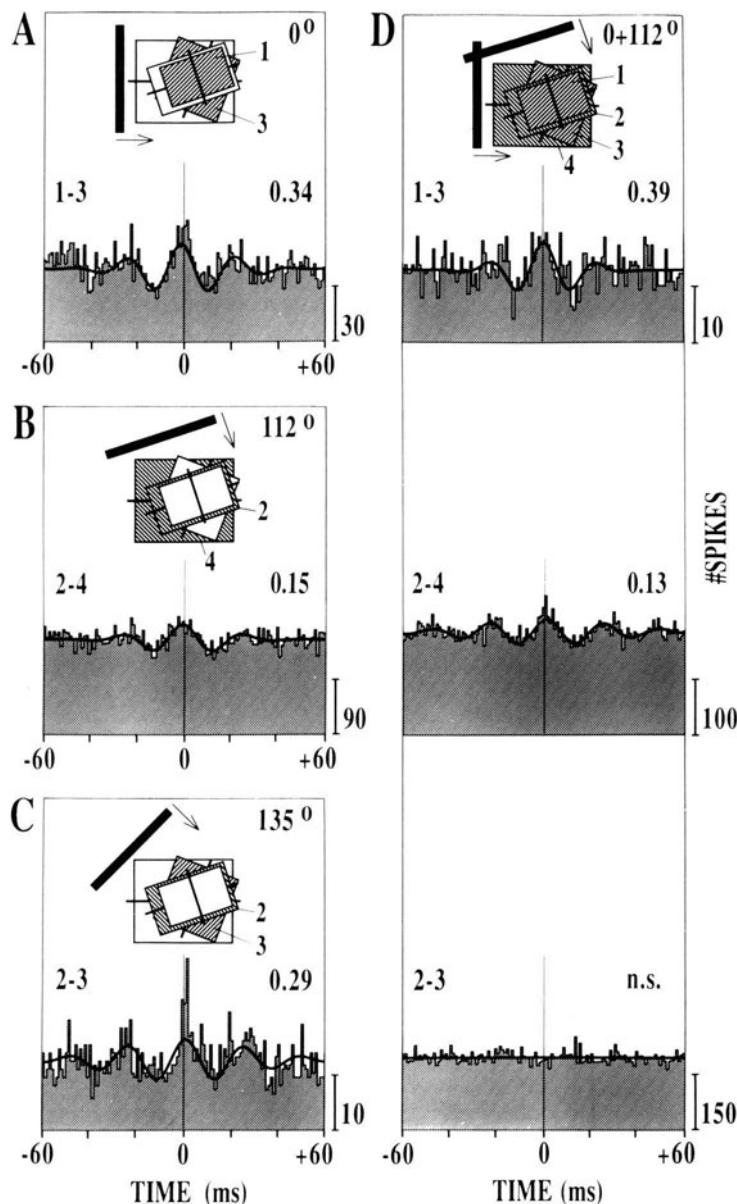


Fig. 4.2.

common input [68]. Because response synchronization occurred often in association with oscillatory activity in the  $\beta$  and  $\gamma$  range it has been proposed that the observed synchronization phenomena in the visual cortex are due to common oscillatory input from subcortical centers [69]. Oscillatory activity in the 30–60 Hz range has been described both for retinal ganglion cells and thalamic neurons [70–81]. In both structures oscillatory activities have been observed in about 20% of the cells. They occurred during spontaneous activity and were often uninfluenced by visual stimulation or even suppressed [77]. These oscillatory patterns in afferent activity are likely to contribute to oscillatory responses in the visual cortex but the possibility must also be considered that part of the thalamic oscillations are back-propagated from cortex by the cortico-thalamic projections.

If the synchronization phenomena observed at the cortical level were solely a reflection of common subcortical input this would be incompatible with the postulated role of synchronization in perceptual grouping. The

---

**Fig. 4.2. Stimulus dependence of short-range interactions.** Multiunit activity was recorded from four different orientation columns of area 17 of cat visual cortex separated by 0.4 mm. The four cell groups had overlapping receptive fields and orientation preferences of 22° (group 1), 112° (group 2), 157° (group 3), and 90° (group 4), as indicated by the thick line drawn across each receptive field in (A)–(D). The figure shows a comparison of responses to stimulation with single moving light bars of varying orientation (left) with responses to the combined presentation of two superimposed light bars (right). For each stimulus condition, the shading of the receptive fields indicates the responding cell groups. Stimulation with a single light bar yielded a synchronization between all cells activated by the respective orientation. Thus, groups 1 and 3 responded synchronously to a vertically orientated (0°) light bar (A), groups 2 and 4 to a light bar at an orientation of 112° (B), and cell groups 2 and 3 to a light bar of intermediate orientation. (C) Simultaneous presentation of two stimuli with orientations of 0° and 112°, respectively, activated all four groups (D). However, in this case the groups segregated into two distinct assemblies, depending on which stimulus was closer to the preferred orientation of each group. Thus, responses were synchronized between groups 1 and 3, which preferred the vertical stimulus, and between 2 and 4, which preferred the stimulus oriented at 112°. The two assemblies were desynchronized with respect to each other, and so there was no significant synchronization between groups 2 and 3. The cross-correlograms between groups 1 and 2, 1 and 4, and 3 and 4 were also flat (not shown). Note that the segregation cannot be explained by preferential anatomical wiring of cells with similar orientation preference (T'so et al., 1986) because cell groups can readily be synchronized in all possible pair combinations in response to a single light bar. The correlograms are shown superimposed with their Gabor function. The number to the upper right of each correlogram indicates the relative modulation amplitude. Abbreviation: ns, not significant. Scale bars indicate the number of spikes (taken from [58]).

hypothesis requires that synchronization probability depends to a substantial extent on interactions between the neurons whose responses actually represent the features that need to be bound together. As thalamic cells possess only very limited feature selectivity one is led to postulate that cortico-cortical connections should also contribute to the synchronization process. This postulate is supported by the finding that synchronization between cells located in different hemispheres is abolished when the corpus callosum is cut [60,64]. This is direct proof (i) that cortico-cortical connections contribute to response synchronization and (ii) that synchronization with zero-phase lag can be brought about by reciprocal interactions between spatially distributed neurons despite considerable conduction delays in the coupling connections. Thus, synchrony is not necessarily an indication of common input but may also be the result of a dynamic organization process which establishes coherent firing by reciprocal interactions.

Simulation studies are now available which confirm that synchrony can be established without phase lag by reciprocal connections even if they have slow and variable conduction velocities [82–85]. Use-dependent developmental selection of cortico-cortical connections could further contribute to the generation of architectures which favor synchrony. During early postnatal development cortico-cortical connections are susceptible to use-dependent modifications and are selected according to a correlation rule [86; see below]. This favors consolidation of connections whose activity is often in synchrony with the activity of their respective target cells. Hence, it is to be expected that connections are selected not only according to their feature specific responses but also as a function of conduction velocities which allow for a maximum of synchrony.

However, the possibility to achieve synchrony through reciprocal cortical connections does not exclude a contribution of common input to the establishment of cortical synchronization. Especially if temporal patterns of responses need to be coordinated across distant cortical areas, bifurcating cortico-cortical projections or divergent cortico-petal projections from subcortical structures such as the “nonspecific” thalamic nuclei, the basal ganglia and the nuclei of the basal forebrain could play an important role. By modulating in synchrony the excitability of selected cortical areas they could influence very effectively the probability with which neurons distributed across these selected areas engage in synchronous firing. A contribution of diverging cortical back-projections to long-range synchronization is suggested by the observation that unilateral focal inactivation of a pre-striate cortical area reduces intra-areal and interhemispheric synchrony in area 17 [65]. A contribution of thalamic mechanisms to the establishment of cortical synchrony has yet to be demonstrated.

## 4.7 Experience-Dependent Modifications of Synchronizing Connections and Synchronization Probability

The theory of assembly coding implies that the criteria according to which particular features are grouped together rather than others reside in the functional architecture of the assembly forming coupling connections. It is of particular interest, therefore, to study the development of the synchronizing connections, to identify the rules according to which they are selected, to establish correlations between their architecture and synchronization probabilities, and if possible, to relate these neuronal properties to perceptual functions.

In mammals cortico-cortical connections develop mainly postnatally [87–90] and attain their final specificity through an activity dependent selection process [91–94]. Recent results from strabismic kittens indicate that this selection is based on a correlation rule and leads to disruption of connections between cells which often exhibit decorrelated activity [86]. Raising kittens with artificially induced strabismus leads to changes in the connections between the two eyes and cortical cells so that individual cortical neurons become connected to only one eye [95]. Cortical neurons split into two subpopulations of about equal size, each responding rather selectively to stimulation of one eye only. Because of the misalignment of the two eyes it is also to be expected that there are no consistent correlations between the activation patterns of neurons driven by the two eyes. Recently, it has been found that strabismus, when induced in three week old kittens leads to a profound rearrangement of cortico-cortical connections. Normally, these connections link cortical territories irrespective of whether these are dominated by the same or by different eyes. In the strabismics, by contrast, the tangential intracortical connections come to link with high selectivity only territories served by the same eye. These anatomical changes in the architecture of cortico-cortical connections are reflected by altered synchronization probabilities. In strabismics response synchronization no longer occur between cell groups connected to different eyes while it is normal between cell groups connected to the same eye [61,96].

These results have several important implications. First, they are compatible with the notion that tangential intracortical connections contribute to response synchronization (see above). Second these results agree with the postulates of the assembly hypothesis that the assembly forming connections should be susceptible to use-dependent modifications and be selected according to a correlation rule.

These results are, for the least, compatible with the view that the architecture of cortico-cortical connections, by determining the probability of response synchronization, could set the criteria for perceptual grouping. Since this architecture is shaped by experience this opens up the possibility

that some of the binding and segmentation criteria are acquired or modified by experience.

## 4.8 Correlation Between Perceptual Deficits and Response Synchronization in Strabismic Amblyopia

Further indications for a relation between experience-dependent modifications of synchronization probabilities and functional deficits come from a recent study of strabismic cats who had developed amblyopia. Strabismus, when induced early in life, does not only lead to a loss of binocular fusion and stereopsis but may also lead to amblyopia of one eye [97]. This condition develops when the subjects solve the problem of double vision not by alternating use of the two eyes but by constantly suppressing the signals coming from the deviated eye. The amblyopic deficit usually consists of reduced spatial resolution and distorted and blurred perception of patterns. A particularly characteristic phenomenon in amblyopia is crowding, the drastic impairment of the ability to discriminate and recognize figures if these are surrounded with other contours. The identification of neuronal correlates of these deficits in animal models of amblyopia has remained inconclusive because the contrast sensitivity and the spatial resolution capacity of neurons in the retina and the lateral geniculate nucleus were found normal. In the visual cortex identification of neurons with reduced spatial resolution or otherwise abnormal receptive field properties remained controversial [for a discussion see 98,99]. However, multielectrode recordings from striate cortex of cats exhibiting behaviorally verified amblyopia have revealed highly significant differences in the synchronization behavior of cells driven by the normal and the amblyopic eye, respectively. The responses to single moving bars that were recorded simultaneously from spatially segregated neurons connected to the amblyopic eye were much less well synchronized with one another than the responses recorded from neuron pairs driven through the normal eye [100]. This difference was even more pronounced for responses elicited by gratings of different spatial frequency. For responses of cell pairs activated through the normal eye the strength of synchronization tended to increase with increasing spatial frequency while it tended to decrease further for cell pairs activated through the amblyopic eye. Apart from these highly significant differences between the synchronization behavior of cells driven through the normal and the amblyopic eye no other differences were found in the commonly determined response properties of these cells. Thus, cells connected to the amblyopic eye continued to respond vigorously to gratings whose spatial frequency had been too high to be discriminated with the amblyopic eye in the preceding behavioral tests. These results suggest that disturbed tem-

poral coordination of responses such as reduced synchrony may be one of the neuronal correlates of the amblyopic deficit. Indeed, if synchronization of responses at a millisecond time scale is used by the system for feature-binding and perceptual grouping, disturbance of this temporal patterning could be the cause for the crowding phenomenon as this can be regarded as a consequence of impaired perceptual grouping.

## 4.9 The Relation Between Synchrony and Oscillations

Because the synchronization of neuronal responses in the visual cortex is often associated with oscillatory discharge patterns many network models designed to explore functional implications of synchrony have used oscillatory activity as a carrier signal for the establishment of synchrony. However, experimental evidence indicates that oscillatory activity and response synchronization can be quite independent phenomena. It is necessary, therefore, to examine in some depth the relation between response synchronization on the one hand and oscillatory responses on the other. If cells exhibit oscillatory responses this does of course not imply that different cells discharge in synchrony. Likewise, the nonoccurrence of oscillations does not exclude synchrony. Thus, no inferences can be drawn from single cell recordings as to whether the responses of the recorded cell are synchronized with others irrespective of whether the recorded cell is found to discharge in an oscillatory manner or not. The situation is different when multiunit recordings are obtained with a single electrode. In this case periodically modulated auto-correlograms are always indicative not only of oscillatory firing patterns but also of response synchronization, at least among the local group of simultaneously recorded neurons. The reason is that such periodic modulations can only build up if a sufficient number of the simultaneously recorded cells are oscillating synchronously and at a sufficiently regular rhythm. However, not observing periodically modulated auto-correlograms of multiunit recordings does neither exclude that the recorded units oscillate, because nonsynchronized oscillations of individual cells would not be observable, nor does it exclude that the recorded cells actually fire in synchrony, because they could do so in a nonperiodic way. The same arguments are applicable to field potential and even more so to EEG recordings. If these recordings of mass activity exhibit an oscillatory pattern this always implies that a large number of neurons must have engaged in synchronized rhythmic activity because otherwise the weak fields generated by activation of individual synapses and neurons would not sum to potentials recordable with macroelectrodes. But again, the reverse is not true: Neither oscillatory discharge patterns nor response synchro-

nization can be excluded if macroelectrode recordings of mass activity fail to reveal oscillatory fluctuations.

Furthermore, it needs to be considered that single cell recordings may not be particularly well suited for the diagnosis of oscillatory activity. This is suggested by results from the visual cortex [44] and in particular from the olfactory bulb [101]. Individual discharges of single units may be precisely time locked with the oscillating field potential, which proves that these discharges participated in an oscillatory process and occurred in synchrony with those of many other cells, without, however, showing any sign of oscillatory activity in their auto-correlation function. The reasons for this apparent paradox are sampling problems and nonstationarity of the time series. If the single cell does not discharge at every cycle and if the oscillation frequency is not perfectly constant over a period of time sufficiently long to sample enough discharges for an interpretable auto-correlation function, the oscillatory rhythm to which the cell is actually locked will not be disclosable. Thus, the less active a cell and the higher and more variable the oscillation frequency, the less is it legitimate to infer from nonperiodically modulated auto-correlograms that a cell is not oscillating. Finally, some ambiguities associated with the term "oscillations" deserve discussion. Most commonly, oscillations are associated with periodic time series such as are produced by a pendulum or an harmonic oscillator. But there are also more irregular or even aperiodic time series which are still addressed as oscillatory. Such irregular oscillations typically occur in noisy linear or in nonlinear systems and cover a large spectrum of phenotypes from slightly distorted, periodic oscillations over chaotic oscillations to nearly stochastic time series. Oscillatory phenomena in the brain are rarely of the harmonic type and if so only over very short time intervals. Most often, oscillatory activity in the brain is so irregular that autocorrelation functions computed over prolonged periods of time frequently fail to reveal the oscillatory nature of the underlying time series.

The evidence that in most brain structures investigated so far phases of response synchronization tend to be associated with episodes of oscillatory activity raises the question as to whether oscillations and synchrony are causally related. One possibility is that oscillatory activity favors the establishment of synchrony and hence is instrumental for response synchronization. In oscillatory responses the occurrence of a burst predicts with some probability the occurrence of the next. It has been argued that this predictability is a necessary prerequisite to synchronize remote cell groups with zero phase lag, despite considerable conduction delays in the coupling connections [for review see 102]. This view is supported by simulation studies that have shown that zero phase lag synchronization can be achieved despite considerable conduction delays and variation of conduction times in the synchronizing connections if the coupled cell groups have a tendency to oscillate [83–85,103,104]. Another feature of networks with oscillatory properties is that network elements that are not linked directly can be

synchronized via intermediate oscillators [80]. This may be important, for instance, to establish relationships between remote cell groups within the same cortical area, or for cells distributed across cortical areas that process different sensory modalities. In both cases, linkages either via intermediate cortical relays or even via subcortical centers must be considered. These considerations suggest that oscillations, while not conveying any stimulus-specific information per se, may be instrumental for the establishment of synchrony over large distances.

However, it is also conceivable that oscillations occur as a consequence of synchrony. Simulation studies indicate that networks with excitatory and inhibitory feedback have the tendency to converge towards states where discharges of local cell clusters become synchronous [105–107]. Once such a synchronous voley has been generated, the network is likely to engage in oscillatory activity. Because of recurrent inhibition and because of  $\text{Ca}^{2+}$ -activated  $\text{K}^+$  conductances [108,109] the cells which had emitted a synchronous discharge will also become simultaneously silent. Upon fading of these inhibitory events, firing probability will increase simultaneously for all cells and this, together with maintained excitatory input and nonlinear voltage-gated membrane conductances such as the low threshold  $\text{Ca}^{2+}$  channels [109] will favor the occurrence of the next synchronous burst, and so on. Thus, oscillations are a likely consequence of synchrony and it actually becomes an important issue to understand how cortical networks can be prevented from entering states of global oscillations and, if they do, how these can be terminated. These issues have recently been addressed in a number of simulation studies [9,103,104,107,110–112].

## 4.10 Rhythm Generating Mechanisms

One possible substrate for the generation of  $\beta$  and  $\gamma$  band oscillations could be inhibitory feedback in much the same manner as that described for the olfactory bulb and piriform cortex [113]. Although direct evidence that recurrent inhibition underlies high frequency cortical oscillations is very limited [114] such a mechanism appears plausible since local circuit inhibitory interneurons exist in abundance in the neocortex [115,116], and artificial neuronal networks containing recurrent inhibitory connections readily exhibit oscillatory activity [82,113,117–119]. It is likely, however, that many forms of inhibition may coexist in the cortical network and the connectivity required to generate 30–60 Hz oscillatory responses may occur in only a subpopulation of neurons.

Alternatively or in addition oscillatory activity in cortex may arise as a consequence of intrinsic pacemaker properties of cortical neurons [108] much as it does in thalamocortical relay neurons [120] and cells of the thalamic reticular nucleus [121,122]. Support for this conjecture has recently been obtained from in vitro intracellular recordings in rat frontal [123]

and sensorimotor cortex [124]. In the former study intrinsic 10–50 Hz oscillations in membrane potential were observed in inhibitory interneurons of layer 4. These fluctuations were produced by membrane depolarization and were dependent on both sodium and potassium conductances [123]. In the latter study intrinsic low frequency 5–50 Hz oscillations produced by membrane depolarization were found in a subpopulation of layer 5 cells [124].

Recent intracellular recordings from cat striate cortex *in vivo* have attempted to distinguish among these possibilities. High frequency 30–60 Hz oscillations of membrane potential have been observed in response to visual stimuli [125]. These signals show a remarkable similarity to oscillatory activity recorded extracellularly [44]. The oscillations in membrane potential are stimulus-dependent and orientation specific and occur less often in simple cells receiving monosynaptic input from the LGN [125]. Action potentials occur on the depolarizing phase of the oscillation mirroring the negativity observed extracellularly [44]. Two further lines of evidence indicate that the oscillations result from extrinsic as opposed to intrinsic mechanisms [125]. Oscillations in membrane potential increase in amplitude in response to visual input when the cells are hyperpolarized suggesting that they arise from excitatory synaptic input rather than voltage-dependent membrane conductances. These data therefore suggest that oscillatory responses in visual cortex reflect a property of the cortical network.

## 4.11 The Duration of Coherent States

It has been argued that synchronous oscillatory activity is unlikely to serve a function in visual processing because the time required to establish and to evaluate synchrony would be incompatible with the short recognition times common in visual perception [126]. The following considerations suggest that such time constraints are probably not critical. In a study by Gray et al. [127] recordings of field potential and unit activity were performed at two sites in cat visual cortex having a separation of at least 4 mm. Field potential responses were chosen for analysis which displayed a particularly close correlation to the simultaneously recorded unit activity. The following variables were determined: (1) the onset latency of the synchronous activity; (2) the time-dependent changes in phase, frequency, and duration of the synchronous episodes within individual trials; and (3) the intertrial variation in each of these parameters.

The results, combined with previous observations [55], demonstrated that correlated responses in cat visual cortex exhibit a high degree of dynamic variability. The amplitude, frequency, and phase of the synchronous events vary over time. The onset of synchrony is variable and bears no fixed relation to the stimulus. Multiple epochs of synchrony can occur on individual trials and the duration of these events also fluctuates from one stimulus

presentation to the next. Most importantly, the results demonstrated that response synchronization can be established within 50–100 ms, a time scale consistent with behavioral performance on visual discrimination tasks [127].

Similar, rapid fluctuations between synchronous and asynchronous states have been observed in other systems, and recent methodological developments have made a quantitative assessment of these rapid changes possible. Using the joint PSTH and gravitational clustering algorithms [128,129] it has been possible to examine the time course of correlated firing among pairs and larger groups of neurons, respectively [130]. These findings clearly indicate that the formation of coherently active cell assemblies is a dynamic process. Patterns of synchronous firing can emerge from seemingly non-organized activity within tens of milliseconds, and can change as a function of stimulus and task conditions within similarly short time intervals. These findings suggest that the temporal constraint imposed by perceptual performance can be met by the dynamic processes which underly the organization of synchronously active cell assemblies.

Theoretical considerations point in the same direction. Assemblies defined by synchronous discharges need not oscillate at a constant frequency over prolonged periods of time. Rather, it is likely that neuronal networks which have been shaped extensively by prior learning processes can settle very rapidly into a coherent state when the patterns of afferent sensory activity match with the architecture of the weighted connections in the network. Such a good match can be expected to occur for familiar patterns which during previous learning processes had the opportunity to mould the architecture of connections and to optimize the fit. If, what matters for the nervous system is the simultaneity of discharges in large arrays of neurons, a single synchronous burst in thousands of distributed neurons may actually be sufficient for recognition. Obviously, the nervous system can evaluate and attribute significance to coherent activity even if synchrony is confined to a single burst because its parallel organization allows for simultaneous assessment of highly distributed activity.

Especially if no further ambiguities have to be resolved, or if no further modifications of synaptic connectivity are required it would actually be advantageous if the system would not enter into prolonged cycles of reverberation after having converged towards an organized state of synchrony. Rather, established assemblies should be erased by active desynchronization as soon as possible in order to allow for the build up of new representations. Thus, when processing highly familiar patterns or executing well trained motor acts which raise no combinatorial problem the system would function nearly as fast as a simple feed-forward network. The differential and flexible routing of activity which is required in order to organize the appropriate assemblies could be achieved by the weighted association fibers within only a few reentrant cycles if there are no ambiguities. The duration of such reentrant cycles would be of about the same order of magnitude as the delays for simple feed-forward processing, as the parallel organi-

zation of the system allows for simultaneous exchange of signals between distributed processes. It is even conceivable that the organization of a pattern characterized by simultaneous discharges of distributed neurons can be achieved faster than expected from the addition of conduction and integration times in single feed-forward architectures. The reason is that the cortical network is already active and continuously exchanging signals between neurons when visual signals become available. In order to organize a synchronous state it is thus not necessary to collect sufficient excitatory drive to reach in succession the thresholds of serially connected neurons. Rather, it appears to be sufficient to only shift the time of occurrence of action potentials to establish synchrony and this might be less time consuming.

## 4.12 Synchronization and Attention

The hypothesis that information about feature constellations is contained in the temporal relation between the discharges of distributed neurons, and in particular in their synchrony, has also some bearing on the organization of attentional mechanisms. It is obvious that synchronous activity will be more effective in driving cells at higher levels than nonorganized asynchronous discharges. Thus, those assemblies would appear as particularly salient and hence effective in attracting attention which succeed to make their discharges coherent with shorter latency and higher temporal precision than others. Conversely, responses of neurons reacting to features which cannot be grouped or bound successfully, and hence, cannot be synchronized with the responses of other neurons, would have only a low chance to be relayed further and to influence shifts of selective attention. It is thus conceivable that out of the many responses that occur at peripheral stages of visual processing only a few are actually passed on towards higher levels. These would either be responses to particularly salient stimuli causing strong and simultaneous discharges in a sufficient number of neurons or responses of cells which succeeded to get organized in sufficiently coherent assemblies. Thus, responses to changes in stimulus configuration or to moving targets have a good chance to be passed on even without getting organized internally because they would be synchronized by the external event. But responses to stationary patterns will require organization through internal synchronization mechanisms in order to be propagated. This interpretation implies that neuronal responses which attract attention and gain control over behavior should differ from nonattended responses not so much because they are stronger but because they are better synchronized among each other. Accordingly, shifting attention by top-down processes would be equivalent with biasing synchronization probability of neurons at lower levels by feedback connections from higher levels. These top-down influences could favor the emergence of coherent states in selected subpopulations of

neurons — the neurons which respond to contours of an “attended” object or pattern. Thus, the mechanism that allows for grouping and scene segmentation — the organization of synchrony — could also serve the management of attention. The advantage would be that nonattended signals do not have to be suppressed, which would hitherto eliminate them from competition for attention. Rather, cells could remain active and thus be rapidly recruitable into an assembly if changes of afferent activity or of feedback signals modify the balance among neurons competing for the formation of synchronous assemblies.

In a similar way shifts of attention across different modalities could be achieved by selectively enhancing synchronization probability, in particular sensory areas and not in others. This could be achieved, for example, by modulatory input from the basal forebrain or nonspecific thalamic nuclei. If these projection systems were able to modulate in synchrony the excitability of cortical neurons distributed in different areas would greatly enhance the probability that these neurons link selectively with each other and join into coherent activity. Such linking would be equivalent with the binding of the features represented in the respective cortical areas. Again, this view equates grouping or binding mechanisms with attentional mechanisms. The “attention” directing systems would simply have to provide a temporal frame within which distributed responses can then self-organize towards coherent states through the network of selective cortico-cortical connections. In doing so the attentional systems need not themselves produce responses in cortical neurons. It would be sufficient that they cause a synchronous modulation of their excitability. It is conceivable that the synchronous field potential oscillations that have been observed in animals and men during states of focused attention are the reflection of such an attention mechanism [131–134]. The observations that these field potential oscillations are only loosely related to the discharge probability of individual neurons, are coherent across different cortical areas, are particularly pronounced when the subjects are busy with tasks requiring integration of activity across different cortical areas, and stop immediately when the binding problem is solved — as witnessed by the execution of a well programmed motor act — are in agreement with such an interpretation.

#### 4.13 The Role of Synchrony in Synaptic Plasticity

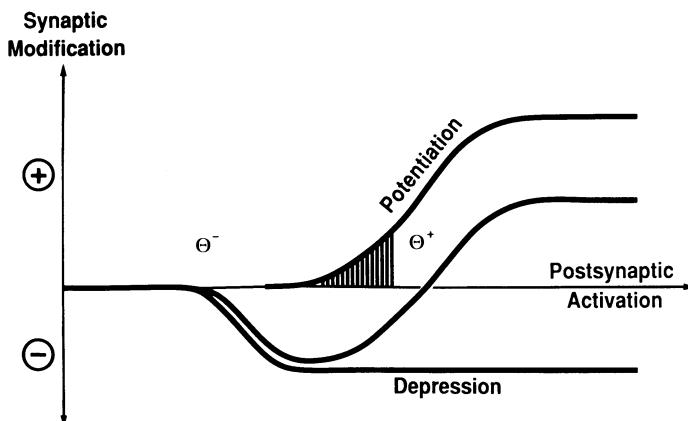
While it is still a matter of debate whether response synchronization is of functional relevance in the context of binding, there can be no doubt that synchronization of neuronal discharges is of eminent functional importance for neuronal plasticity. This holds both for the activity-dependent shaping of neuronal connectivity during development and use-dependent changes of synaptic gain in the mature brain [for review see 135]. The reason is that the processes that lead to modifications in synaptic connectivity and in

synaptic gain have thresholds that are voltage-dependent. Thus, it has been shown that experience-dependent modifications of connections in the developing visual cortex occur only if the target cells of the modifiable connections are activated beyond a critical threshold which appears to be related with the activation of NMDA receptor-gated  $\text{Ca}^{2+}$  conductances [136–138; for review of the extensive literature see 139]. For synaptic gain changes in the mature cortex, two thresholds have been identified. If neurons are driven above a first threshold, modifiable synapses undergo a long-term depression (LTD) and when a second threshold is reached, they undergo long-term potentiation (LTP) of their efficacy [140,141]. The first threshold appears to be related to the activation threshold of voltage-gated  $\text{Ca}^{2+}$  channels [140,142], while the second is related to the activation of NMDA-receptor-gated conductances [13,14]. Rather similar conditions are found for synaptic modifications in the hippocampus and a variety of other brain structures [for review see 143]. Thus, as summarized in Fig. 4.3, the probability of occurrence of a synaptic modification and its direction, increase or decrease of efficacy, depend on the level of postsynaptic depolarization and hence directly on the synchrony of discharges in the afferents converging onto a particular neuron. Synchronously active inputs are likely to become enhanced, while inputs that are active out-of-phase with other synchronous inputs are likely to undergo depression [see, e.g., 144]. Finally, if there is no synchrony between any of the converging afferents, modifications are unlikely to occur because the relatively high modification thresholds will not be reached. It follows from this that the temporal coordination of distributed neuronal responses is crucial not only for signal transmission, but also for synaptic plasticity.

#### 4.14 The Role of Oscillations in Synaptic Plasticity

The fact that the temporal correlation between converging inputs to a cell is a critical variable in use-dependent plasticity adds a further aspect to the putative functional significance of an oscillatory time structure in neuronal discharge patterns. The responses of neurons are usually rather long, extending over several 100 ms. Thus, most of the responses evoked by a visual scene in the visual cortex are overlapping in time. According to the established synaptic modification rules, this would in the long run increase the gain of most of the coupling connections in the visual cortex and hence lead to unselective and meaningless association of neurons. This superposition problem can be avoided if the responses of the connected neurons have an oscillatory time structure. In that case, the membrane potential alternates rapidly between de- and hyperpolarizing phases [145]. As a consequence, only those of the simultaneously active oscillating inputs will have

### Induction of Homosynaptic Modifications



**Fig. 4.3.** Schematic representation of the two different voltage-dependent thresholds ( $\Theta^-$  and  $\Theta^+$ ) for the induction of homosynaptic LTD and LTP. The ordinate indicates the direction of the synaptic gain change and the abscissa the membrane potential level of the postsynaptic neuron that is maintained during presynaptic activation. If the first threshold  $\Theta^-$  is reached a mechanism is activated which leads to a long-lasting depression of synaptic efficacy, and if the second threshold  $\Theta^+$  is reached another process is triggered which leads to the potentiation of the synapse (taken from [149]).

a chance to improve synaptic gain, which oscillate in precise synchrony with the target cell. Only then will the active synapses be able to activate voltage-sensitive NMDA receptor-gated conductances [146,147], which has been identified as a necessary prerequisite for homosynaptic potentiation of synaptic transmission in hippocampus [148] and neocortex [13,14]. Inputs whose oscillatory responses are out of phase with the oscillatory activity of the postsynaptic cell, or exhibit a fixed phase shift, will either become depressed or remain unchanged depending on their phase relation with the cell's activity. Direct support for this prediction comes from results obtained in the hippocampus [144]. The coincidence criterion for synaptic gain changes is then determined by the frequency of the oscillation of a response rather than by the overall duration of a response. For oscillations in the  $\gamma$ -frequency range, this implies that the temporal window for cooperative interactions between converging inputs narrows down to 10 ms and less. Thus, imposing an oscillatory burst and pause modulation on neuronal responses improves by an order of magnitude the temporal precision with which mechanisms that rely on coincidence detection and cooperativity can operate.

In a sense the functional role of temporally structured responses in synaptic plasticity is the same as that proposed for assembly coding. It serves to

resolve superposition problems by providing a fine grain temporal code to express relations. These considerations on assembly coding and on synaptic plasticity both lead to the same postulate of a temporal code that relies on synchrony at a millisecond time scale. This convergence is, of course, not unexpected because both the formation of assemblies and use-dependent synaptic modifications are intimately related. In order to assure that a particular constellation of features always leads to the emergence of the same assembly of coherently active neurons, an architecture of synchronizing connections has to be developed that assures preferential coupling between the neurons constituting this assembly. Hence, connections between cells that have repeatedly been part of an assembly should strengthen and consolidate. According to the concept of temporal coding, the signature for cells belonging to an assembly is the synchronization of their temporally structured responses. As synchrony between pre- and postsynaptic activation is at the same time the condition that favors strengthening of synaptic connections, the same temporal code that serves to distinguish the neurons of an assembly can thus be used to distinguish the connections that need to be reinforced to stabilize an assembly.

In conclusion, several independent lines of argument concerning the nature of neuronal representations, the constraints for signal transmission, and conditions of use-dependent synaptic plasticity all lead to the postulate of a temporal code and all emphasize the significance of synchrony at a millisecond time scale.

## 4.15 Outlook

In order to obtain direct evidence in support of or against the hypothesis that response synchronization serves as a code for neuronal processing, experiments are needed in which causal relation can be established between the occurrence of response synchronization in defined subgroups of neurons and particular functions that need to be assessed at the behavioral level. This will require simultaneous recordings from several selected areas in the brain of awake behaving animals with techniques that enable assessment of response synchronization with high temporal resolution. With the techniques that are currently available, the number of recording sites that can be examined simultaneously is bound to remain small. This poses a severe sampling problem. For the brain, one or two bursts may be a highly significant event if they occur simultaneously in a large number of cells. Hence, for nonambiguous conditions, the episodes characterized by synchronization can be kept very short and in extremis can be restricted to one or two bursts. For the experimenter, however, who can only look at a few cells, such brief events of synchrony may pass undetected, and may be deemed significant only if they recur. Thus, until new techniques become available, the relationship between synchronization and behavior

will be detectable only for conditions where synchronization is maintained long enough to be observed, even if only a few cells can be looked at simultaneously. This may confine the behavioral conditions suitable for such analyses to problem-solving tasks that are difficult, fraught with ambiguity, and require long periods of sustained, focused attention. It might be rewarding in this context to reconsult the EEG literature for the identification of behavioral states that are likely to be associated with response synchronization in the  $\gamma$ -frequency band.

## 4.16 Concluding Remarks

The data reviewed in this chapter indicate that neocortical neurons can under certain conditions synchronize their discharges with a precision in the range of milliseconds. This certainly has consequences for the transmission of signals since the cortical network is characterized by low safety factors of synaptic interactions between individual elements. Synchronous states are with all likelihood also of significance for the induction of use-dependent synaptic modifications such as are thought to occur during learning processes in the mature system and they certainly play an important role in the experience-dependent shaping of connectivity that occurs during the development of cortical architectures. Available data on dynamic and stimulus-dependent variations of synchronization probability are further compatible with the hypothesis that response synchronization could serve to tag responses which need to be bound together during sensory processing and the preparation of motor programs. But direct evidence that the nervous system actually uses synchrony to establish flexible relations between distributed neuronal responses is still lacking. To prove this conjecture requires the documentation of a causal relation between the occurrence of specific synchronization patterns on the one hand and behavioral correlates on the other. This goal can only be attained if multielectrode recordings are performed in awake animals that are trained to accomplish special recognition tasks or motor acts.

## REFERENCES

- [1] A.P. Georgopoulos (1990) Cold Spring Harbor Symp. Quant. Biol. **55**:849
- [2] F.A. Mussa-Ivaldi, S.F. Giszter, E. Bizzi (1990) Cold Spring Harbor Symp. Quant. Biol. **55**:827
- [3] D.L. Sparks, C. Lee, W.H. Rohrer (1990) Cold Spring Harbor Symp. Quant. Biol. **55**:805
- [4] J.H. Maunsell, W.T. Newsome (1987) Ann. Rev. Neurosci. **10**:363
- [5] D.J. Felleman, D.C. van Essen (1991) Cerebral Cortex **1**:1

- [6] R. Desimone, L.G. Ungerleider (1989) *Handbook of Neuropsychology* (Elsevier, Amsterdam), Vol. 2, p. 267
- [7] P.M. Milner (1974) *Psycholog. Rev.* **81**:521
- [8] C. von der Malsburg (1985) *Ber. Bunsenges. Phys. Chem.* **89**:703
- [9] C. von der Malsburg, W. Schneider (1986) *Biol. Cybern.* **54**:29
- [10] H. Shimizu, Y. Yamaguchi, I. Tsuda, M. Yano (1986) In: *Complex Systems — Operational Approaches*, p. 225
- [11] V. Braintenberg, A. Schütz (1991) *Anatomy of the Cortex* (Springer, Berlin)
- [12] C.F. Stevens (1987) *Synaptic Function* **24**:699
- [13] A. Artola, W. Singer (1987) *Nature* **330**:649
- [14] A. Artola, W. Singer (1990) *Eur. J. Neurosci.* **2**:254
- [15] H.B. Barlow (1972) *Perception* **1**:371
- [16] C.G. Gross, E.C. Rocha-Miranda, D.B. Bender (1972) *J. Neurophysiol.* **35**:96
- [17] E.T. Rolls (1991) *Curr. Opin. Neurobiol.* **1**:274
- [18] R. Desimone, T.D. Albright, C.G. Gross, C. Bruce (1984) *J. Neurosci.* **4**:2051
- [19] R. Desimone, S.J. Schein, J. Moran, L.G. Ungerleider (1985) *Vision Res.* **24**:441
- [20] D.I. Perrett, A.J. Mistlin, A.J. Chitty (1987) *Trends in Neurosci.* **10**:358
- [21] Y. Miyashita (1988) *Nature* **335**:817
- [22] K. Sakai, Y. Miyashita (1991) *Nature* **354**:152
- [23] S. M. Zeki (1973) *Brain Res.* **53**:422
- [24] S.M. Zeki, J.D.G. Watson, C.J. Lueck, K.J. Friston, C. Kennard, R.S.J. Frackowiak (1991) *J. Neurosci.* **11**:641
- [25] J.L. Gallant, J. Braun, D.C. van Essen (1993) *Science* **259**:100
- [26] K. Tanaka, H. Saito, Y. Fukuda, M. Moriya (1991) *J. Neurophys.* **66**:170
- [27] G.C. Baylis, E.T. Rolls, C.M. Leonard (1985) *Brain Res.* **342**:91
- [28] M. Abeles (1991) *Corticonics* (Cambridge University, Cambridge)
- [29] V. Braintenberg (1978) *Lecture Notes in Biomathematics. 21. Theoretical Approaches to Complex Systems* (Springer, Berlin), p. 171
- [30] F. Crick (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**:4586
- [31] G.M. Edelman (1987) *Neural Darwinism: The Theory of Neuronal Group Selection* (Basic Books, New York)

- [32] G.M. Edelman (1989) *The Remembered Present* (Basic Books, New York)
- [33] G.M. Edelman, V.B. Mountcastle (1978) *The Mindful Brain* (MIT, Cambridge)
- [34] S. Grossberg (1980) *Psychol. Rev.* **87**:1
- [35] D.O. Hebb (1949) *The Organization of Behavior* (Wiley, New York)
- [36] G. Palm (1982) *Neural Assemblies* (Springer, Berlin)
- [37] G. Palm (1990) *Concepts in Neurosci.* **1**:133
- [38] W. Singer (1985) *Models of the Visual Cortex* (Wiley, Chichester) p. 123
- [39] W. Singer (1990) *Concepts in Neurosci.* **1**:1
- [40] W. Singer (1979) *The Neurosciences, Fourth Study Program* (MIT, Cambridge) p. 1093
- [41] A.R. Damasio (1990) *Sem. Neurosci.* **2**:287
- [42] F. Crick, C. Koch (1990) *Sem. Neurosci.* **2**:263
- [43] C.M. Gray, W. Singer (1987) *Soc. Neurosci. Abstr.* **13**:404.3
- [44] C.M. Gray, W. Singer (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**:1698
- [45] R. Eckhorn, R. Bauer, W. Jordan, M. Brosch, W. Kruse, M. Munk, H.J. Reitboeck (1988) *Biol. Cybern.* **60**:121
- [46] C.M. Gray, A.K. Engel, P. König, W. Singer (1990) *Eur. J. Neurosci.* **2**:607
- [47] C. Schwarz, J. Bolz (1991) *J. Neurosci.* **11**:2995
- [48] M.S. Livingstone (1991) *Soc. Neurosci. Abstr.* **17**:73.3
- [49] R. Eckhorn, T. Schanze, M. Brosch, W. Salem, R. Bauer (1992) In: *Induced Rhythms in the Brain* (Birkhäuser, Boston), p. 47
- [50] A.K. Engel, A.K. Kreiter, P. König, W. Singer (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**:6048
- [51] A. Raether, C.M. Gray, W. Singer (1989) *Eur. Neurosci. Abstr.* **12**:72.5
- [52] C.M. Gray, G. Viana di Prisco, *Soc. Neurosci. Abstr.*, in press
- [53] S. Neuenschwander, F.J. Varela (1990) *Soc. Neurosci. Abstr.* **16**:47.6
- [54] A.K. Kreiter, W. Singer (1992) *Eur. J. Neurosci.* **4**:369
- [55] A.K. Engel, P. König, C.M. Gray, W. Singer (1990) *Eur. J. Neurosci.* **2**:588
- [56] C.M. Gray, P. König, A.K. Engel, W. Singer (1989) *Nature* **338**:334
- [57] A.K. Kreiter, A.K. Engel, W. Singer (1992) *Eur. Neurosci. Assoc. Abstr.* **15**:1076
- [58] A.K. Engel, P. König, W. Singer (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**:9136

- [59] J.I. Nelson, P.A. Salin, M.H.J. Munk, M. Arzi, J. Bullier (1992) Vis. Neurosci. **9**:21
- [60] A.K. Engel, P. König, A.K. Kreiter, W. Singer (1991) Science **252**:1177
- [61] P. König, A.K. Engel, S. Löwel, W. Singer (1993) Eur. J. Neurosci. **5**:501
- [62] D. Ts'o, C. Gilbert, T.N. Wiesel (1986) J. Neurosci. **6**:1160
- [63] A.M.H.J. Aertsen, G.L. Gerstein (1985) Brain Res. **340**:341
- [64] M.H.J. Munk, L.G. Nowak, G. Chouvet, J.I. Nelson, J. Bullier (1992) Eur. J. Neurosci. Suppl. **5**:21
- [65] J.I. Nelson, L.G. Nowak, G. Chouvet, M.H.J. Munk, J. Bullier (1992) Soc. Neurosci. Abstr. **18**:11.8
- [66] M.J. Bullier, M.H.L. Munk, L.G. Nowak (1992) Soc. Neurosci. Abstr. **18**:11.7
- [67] A.W. Roe, D.Y. Ts'o (1992) Soc. Neurosci. Abstr. **18**:11.4
- [68] G.L. Gerstein, D.H. Perkel (1972) Biophys. J. **12**:453
- [69] R. Llinas, U. Ribary, Proc. Natl. Acad. Sci. U.S.A., in press
- [70] P.O. Bishop, W.R. Levick, W.O. Williams (1964) J. Physiol. **170**:598
- [71] R.W. Doty, D.S. Kimura (1963) J. Physiol. **168**:205
- [72] J.M. Fuster, A. Herz, O.D. Creutzfeldt (1965) Arch. Ital. Biol. **103**:159
- [73] D.W. Arnett (1975) Exp. Brain Res. **24**:111
- [74] J. Munemori, K. Hara, M. Kimura, R. Sato (1984) Biol. Cybern. **50**:167
- [75] M. Ariel, N.W. Daw, R.K. Rader (1983) Vis. Res. **23**:1485
- [76] G.M. Ghose, R.D. Freeman (1990) Soc. Neurosci. Abstr. **16**:523.4
- [77] G.M. Ghose, R.D. Freeman (1992) J. Neurophys. **68**:1558
- [78] M. Steriade, R. Curro-Dossi, D. Paré (1991) G. Oakson, Proc. Natl. Acad. Sci. U.S.A. **88**:4396
- [79] D. Pinault, M. Deschênes (1992) Neuroscience **51**:245
- [80] D. Pinault, M. Deschênes (1992) Neuroscience **51**:259
- [81] M. Steriade, R. Curro-Dossi, D. Contreras, Neuroscience, in press
- [82] P. König, T.B. Schillen (1991) Neural Comput. **3**:155
- [83] T.B. Schillen, P. König (1990) In: *Parallel Processing in Neural Systems and Computers* (Elsevier, Amsterdam), p. 139
- [84] H.G. Schuster, P. Wagner (1990) Biol. Cybern. **64**:77
- [85] H.G. Schuster, P. Wagner (1990) Biol. Cybern. **64**:83

- [86] S. Löwel, W. Singer (1992) *Science* **255**:209
- [87] E.M. Callaway, L.C. Katz (1990) *J. Neurosci.* **10**:1134
- [88] G.M. Innocenti (1981) *Science* **212**:824
- [89] H.J. Luhmann, L. Martinez-Millan, W. Singer (1986) *Exp. Brain Res.* **63**:443
- [90] D.J. Price, C. Blakemore (1985) *J. Neurosci.* **5**:2443
- [91] E.M. Callaway, L.C. Katz (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**:745
- [92] G.M. Innocenti, D.O. Frost (1979) *Nature* **280**:231
- [93] H.J. Luhmann, W. Singer, L. Martinez-Millan (1990) *Eur. J. Neurosci.* **2**:344
- [94] D.J. Price, C. Blakemore (1985) *Nature* **316**:721
- [95] D.H. Hubel, T.N. Wiesel (1965) *J. Neurophysiol.* **28**:1041
- [96] P. König, A.K. Engel, S. Löwel, W. Singer (1990) *Soc. Neurosci. Abstr.* **16**:523.2
- [97] G.K. von Noorden (1990) In: *Binocular Vision and Ocular Motility, Theory and Management of Strabismus* (C.V. Mosby, St. Louis)
- [98] D.P. Crewther, S.G. Crewther (1990) *Exp. Brain Res.* **79**:615
- [99] C. Blakemore, F. Vital-Durand (1992) *Ophthal. Physiol. Opt.* **12**
- [100] P.R. Roelfsema, P. König, A.K. Engel, R. Sireteanu, W. Singer, (1994) *Eur. J. Neurosci.*, in press
- [101] W.J. Freeman, C.A. Skarda (1985) *Brain Res. Rev.* **10**:147
- [102] A.K. Engel, P. König, A.K. Kreiter, T.B. Schillen, W. Singer (1992) *Trends in Neurosci.* **15**:218
- [103] P. König, T.B. Schillen (1990) In: *Parallel Processing in Neural Systems and Computers* (Elsevier, Amsterdam), p. 117
- [104] T.B. Schillen, P. König (1992) In: *Computation and Neural Systems* (Kluwer Academic Press, Dordrecht)
- [105] C. Koch, H.G. Schuster (1992) *Neural Comput.* **4**:211
- [106] J. Deppisch, H.-U. Bauer, T.B. Schillen, P. König, K. Pawelzik, T. Geisel (1992) *Artificial Neural Networks* **2**:921
- [107] O. Sporns, G. Tononi, G.M. Edelman (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**:129
- [108] R.R. Llinas (1988) *Science* **242**:1654
- [109] R.R. Llinas (1990) *Cold Spring Harbor Symp. Quant. Biol.* **55**:933
- [110] D. Hansel, H. Sompolinsky (1992) *Phys. Rev. Lett.* **68**:718

172 4. Neocortical Processing and Synaptic Plasticity

- [111] P. König, B. Janosch, T.B. Schillen (1992) *Neural Comput.* **4**:666
- [112] T.B. Schillen, P. König (1991) *Neural Comput.* **3**:167
- [113] W.J. Freeman (1975) *Mass Action in the Nervous System* (Academic, New York)
- [114] C.M. Gray, A.K. Engel, P. König, W. Singer (1992) *Vis. Neurosci.* **8**:337
- [115] J.S. Lund, G.H. Henry, C.L. MacQueen, A.R. Harvey (1979) *J. Comp. Neurol.* **184**:599
- [116] K.A.C. Martin (1984) *Cerebral Cortex, Vol. 2. Functional Properties of Cortical Cells* (Plenum, New York)
- [117] H.R. Wilson, J.D. Cowan (1972) *Biophys. J.* **12**:1
- [118] W.A. Wilson, J.A. Bower (1991) *Neural Comput.* **3**:498
- [119] P. Bush, R.J. Douglas (1991) *Neural Comput.* **3**:19
- [120] D.A. McCormick, H. Pape (1990) *J. Physiol.* **431**:291
- [121] M. Steriade, E.G. Jones, R.R. Llinas (1990) *Thalamic Oscillations and Signaling* (Wiley, New York)
- [122] T. Bal, D. McCormick, *J. Physiol.*, submitted
- [123] R.R. Llinas, A.A. Grace, Y. Yarom (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**:897
- [124] L.R. Silva, Y. Amitai, B.W. Connors (1991) *Science* **251**:432
- [125] B. Jagadeesh, C.M. Gray, D. Ferster (1992) *Science* **257**:552
- [126] J.M. Tovee, E.T. Rolls (1992) *Trends in Neurosci.* **15**:387
- [127] C.M. Gray, A.K. Engel, P. König, W. Singer (1992) In: *Induced Rhythms in the Brain* Brain Dynamic Series (Birkhäuser, Boston), p. 29
- [128] G.L. Gerstein, D.H. Perkel, J.E. Dayhoff (1985) *J. Neurosci.* **5**:881
- [129] G.L. Gerstein, A.M.H.J. Aertsen (1985) *J. Neurophysiol.* **54**:1513
- [130] A. Aertsen, E. Vaadia, M. Abeles, E. Ahissar, I.I. Bergmann, B. Carmon, Y. Lavner, E. Margalit, I. Nelken, S. Rotter (1991) *J. Hirnforschung* **32**:735
- [131] V.N. Murthy, E.E. Fetz (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**:5670
- [132] A. Rougeul, J.J. Bouyer, L. Dedet, O. Debray (1979) *Electroenceph. Clin. Neurophysiol.* **46**:310
- [133] D.E. Sheer (1984) In: *Self-regulation of the Brain and Behavior* (Springer, Berlin), p. 66
- [134] D.E. Sheer (1989) In: *Brain Dynamics* (Springer, Berlin), p. 338
- [135] J.E. Cook (1991) *Trends Neurosci.* **14**:397

- [136] H.T. Cline, E. Debski, M. Constantine-Paton (1987) Proc. Natl. Acad. Sci. U.S.A. **84**:4342
- [137] M. Constantine-Paton (1990) Cold Spring Harbor Symp. Quant. Biol. **55**:431
- [138] A. Kleinschmidt, M.F. Bear, W. Singer (1987) Science **238**:355
- [139] W. Singer (1990) J. Exp. Biol. **153**:177
- [140] A. Artola, S. Bröcher, W. Singer (1990) Nature **347**
- [141] J.C. Hirsch, F. Crepel (1990) J. Physiol. **427**:31
- [142] S. Bröcher, A. Artola, W. Singer (1992) Proc. Natl. Acad. Sci. U.S.A. **89**:123
- [143] B. Gustafsson, H. Wigstrom (1990) Sem. Neurosci. **2**:317
- [144] P.K. Stanton, T.J. Sejnowski (1989) Nature **339**:215
- [145] B. Jagadeesh, D. Ferster, C.M. Gray (1991) Soc. Neurosci. Abstr. **17**:73.2
- [146] M.L. Mayer, G.L. Westbrook, P.B. Guthrie (1984) Nature **309**:261
- [147] L. Nowak, P. Bregestovski, P. Ascher, A. Herbet, A. Prochiantz (1984) Nature **307**:462
- [148] G.L. Collingridge, S.J. Kehl, H. McLennan (1983) J. Physiol. **334**:33
- [149] A. Artola, W. Singer, *The NMDA Receptor*, 2nd Ed., in press

# 5

# Associative Binding and Segregation in a Network of Spiking Neurons

Raphael Ritz, Wulfram Gerstner, and J. Leo van Hemmen<sup>1</sup>

with 17 figures

**Synopsis.** A model of an associative network of *spiking* neurons (the *Spike Response Model*) with stationary states, globally locked oscillations, and weakly locked oscillatory states is presented and analyzed. The network is close to biology in the following sense. First, the neuron spikes and our model includes an absolute refractory period after each spike. Second, we consider a distribution of axonal delay times. Finally, we describe synaptic signal transmission by excitatory and inhibitory potentials (EPSP and IPSP) with a realistic shape, that is, through a response kernel. The patterns have been learned by an asymmetric Hebbian rule that can handle a low activity which may vary from pattern to pattern. During retrieval of a pattern all active neurons exhibit periodic spike bursts which may or may not be synchronized (“locked”) into a coherent oscillation. We derive an analytical condition of locking and calculate the period of collective activity during oscillatory retrieval. It is argued that in a biological network an intermediate scenario of “*weak locking*” is most likely. In this regime, we discuss applications to feature linking and pattern segmentation as well as the problem of context sensitive binding that can be solved in a layered structure *including feedback*. In addition, we address the question of synchronization between the two hemispheres of the brain.

## 5.1 Introduction

Whereas associative retrieval of stationary patterns has attracted a lot of theoretical attention during the last 10 years [3,4,41,42]; for a review see

---

<sup>1</sup>Physik-Department der TU München, D-85 747 Garching bei München, Germany; fax:+49(89) 3209 2296, e-mail: lvh@Physik.TU Muenchen.DE.

[2] and [15]), emphasis has now shifted towards the problem of collective oscillations after evidence of a coherent activity has been found in the visual and sensorimotor cortex [16,21,29,30,60]. This shift of interest is mainly due to the fact that synchronism of firing might be used by the brain as a code to link features that belong to the same pattern and to separate them from the rest [16,79]. Coherence could thus solve two old questions of information processing in the brain which are known as the problems of *feature linking* and *pattern segmentation*.

In most network models of collective oscillations, the applications to image processing have been the center of interest [44,49,65,67,70,80] and the biology of the experimental systems has been incorporated at a rather crude phenomenological level. In particular, the basic unit in most model networks is a set of ad hoc differential equations for amplitudes or phases of the excitation of an *ensemble* of neurons, for example, one column [5,49,66,80,82]. This is to be contrasted with traditional associative networks where the basic unit is taken to describe the mean firing rate of a *single* neuron [41,42]. It is thus an interesting question, which will be perused here, how both approaches — the single neuron Ansatz used in associative networks and the ensemble Ansatz used in oscillator networks — can be combined.

To unify the two approaches we return to biology and include some details of neuronal signal transmission which are usually neglected. In so doing we focus on a *single spike* as the important phenomenon and describe its generation at the soma, transmission along the axon, and transfer across the synapse by plausible model assumptions. We therefore call our model the *spike response model* (SRM) [24,25,34]. Within the SRM, we derive exact conditions under which stationary or oscillatory activity occurs.

There have been, of course, several other models that focus on spikes and biological realism. Buhmann and Schulten [12] and Horn and Usher [45] discuss associative networks of spiking units similar to our model neuron but concentrate on stationary retrieval. Results on collective oscillations in large networks of spiking neurons have been reported by Buhmann [11], Hansel and Sompolinsky [31], Kurrer et al. [53], and Horn et al. [44]. On the other hand, Bush and Douglas [13] consider a cortical “network” of 10 bursting pyramidal neurons and one basket cell and show in simulations with a set of realistic neuronal parameters and connections that a burst synchronization is possible. A similar, but larger network of compartmental model neurons with local connectivity which is intended to describe an area of the visual cortex is considered by Wilson and Bower [81]. A rather complex but interesting model that includes columnar organization of a large number of neurons has been studied by Sporns et al. [72–74]; in particular, its application to the problem of figure-ground segregation is discussed in detail.

In all cases, however, the results are mainly based on computer simulations of the respective model. In contrast to this, our model of a “biological”

network is simple enough to allow analytical solutions which are exact in the limit of a large number of noiseless neurons and realistic enough to provide insight in the behavior of biological systems. Furthermore, we are able to derive conditions for the existence and stability of a collective oscillation and to determine its period. The analytical results can then be compared with computer simulations of a large but finite system at a nonzero noise level. The present network is organized in two layers, viz., inhibitory and “pyramidal” neurons, and is a generalization of an earlier homogeneous model [26].

In Sec. 5.2, we define the model network based on a simplified description of neurobiological signal transmission. In Sec. 5.3, we present an analytical solution of the macroscopic network states in terms of the dynamic evolution of so-called overlaps. We calculate the retrieval overlap in a stationary state, the oscillation period of an oscillatory state, and derive a condition of “locking” into a collective oscillation. If the locking condition is not fulfilled, *weak locking* may result. Section 5.4 is devoted to computer simulations of a network consisting of two layers of 4000 neurons. The simulations visualize typical network states in three different scenarios. Classification of the network states yields a phase diagram in parameter space. In Sec. 5.5 we demonstrate how the SRM is able to perform feature linking and pattern segmentation. Finally, in Sec. 5.6 the even harder task of context sensitive binding is solved in a layered network with feedback, which is omnipresent in the cortex [22,84].

## 5.2 Spike Response Model

### 5.2.1 STARTING POINT

Before a network model can be defined, a couple of basic decisions concerning the desired level of description and the treatment of biological details have to be made. Some of the most important and intriguing problems of theoretical neurobiology concern the coding procedure through which a biological system maps data from the outside world onto the network states. This includes the question whether the information concerning the environment is contained in the mean firing rates of the neurons or rather in complex temporal spike patterns. Or is it even the exact voltage trace and shape of a single spike which is important? Furthermore, are there ensembles of neurons (e.g., columns) which share the same information processing tasks, or should we take the spiking of each single neuron into account? It is our opinion that a model network should incorporate as few assumptions as possible concerning the above questions. We therefore take a single spiking neuron as the basic unit of our network. Averaging over time (to get the mean firing rate) or over space (to define the ensemble average or global activity) can then, if appropriate, be done at a later stage during the the-

oretical analysis of the network — but not beforehand. Averaging should not be an underlying assumption of the model definitions either.

It is, of course, impossible to include all known details of a specific neural system, for example, the visual cortex of the cat, into a general artificial network. Our approach neglects all details of processes at the level of neurotransmitters and ion channels as well as the branching structure of axons and dendritic trees. A microscopic Ansatz including these effects would yield the exact shape and velocity of spikes moving along the axon to the synapses [19,40] and of the postsynaptic potentials spreading over the dendritic tree [1,47,64]. It is, however, our assumption that the exact shape of the spikes and dendritic signals conveys no information in addition to that of the spiking *event* itself which is already contained in much simpler models. It is thus convenient to adopt a more phenomenological approach and consider *formal* spikes that are generated by a threshold process and transmitted along the axon to the synapses. The signal arrives there after some delay  $\Delta^{\text{ax}}$  and evokes an excitatory or inhibitory postsynaptic potential (EPSP or IPSP). The EPSP and IPSP which change on a time scale much longer than a single spike are modeled by a response function with a realistic time course. Finally, the postsynaptic contributions of all active neurons are added in a linear fashion and compared to the firing threshold so as to determine the firing probability during the next time step.

As to the connection topology, it is known that the pyramidal neurons make both long-ranged and short-ranged connections whereas inhibition is primarily local. Here we take the point of view of a theoretician and model the extreme and most transparent case, that is, full connectivity between pyramidal neurons and strictly local inhibition by inhibitory partner neurons. An exact definition of the above ideas is given in the next subsection.

### 5.2.2 NEURONS AND SYNAPSES

A phenomenological model of a single neuron  $i$  can be built on two measurable parameters, both amenable to experiment, viz., the firing threshold  $\theta$  and the absolute refractory period  $\tau_{\text{ref}}$ . Since we do not want to describe the shape of a single spike, we model spikes by a formal variable  $S_i(t) \in \{0, 1\}$  and take a typical spike width of 1 ms as the basic time step of our model ( $\Delta t = 1$  ms). The spiking dynamics of neuron  $i$  is defined by the probability of firing during a time step, given a membrane potential  $h_i$ ,

$$P_F[S_i(t + \Delta t) = 1 | h_i(t)] = (1/2)\{1 + \tanh[\beta(h_i(t) - \theta)]\}, \quad (5.1)$$

where  $\beta$  is a parameter that takes care of the internal noise of a neuron. The membrane potential  $h_i(t)$  consists of three components:

$$h_i(t) = h_i^{\text{syn}}(t) + h_i^{\text{ext}}(t) + h_i^{\text{ref}}(t), \quad (5.2)$$

where  $h_i^{\text{syn}}(t)$  is the sum of the synaptic inputs from all other neurons in the network (see below),  $h_i^{\text{ext}}(t)$  is some external input, and  $h_i^{\text{ref}}(t)$  is a

formal contribution to describe the refractoriness of the neuron. If we take

$$h_i^{\text{ref}}(t) = \begin{cases} -R & \text{for } t_F \leq t \leq t_F + \tau_{\text{ref}} \\ 0 & \text{otherwise} \end{cases} \quad (5.3)$$

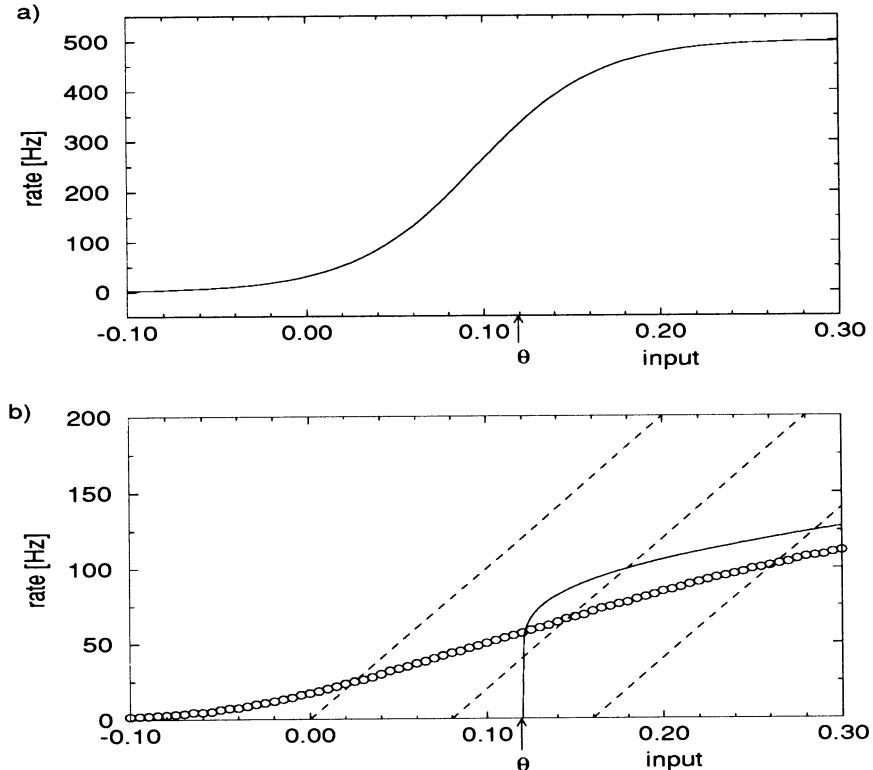
and  $R \gg 1$ , then firing is prevented during a time  $\tau_{\text{ref}}$  after emission of a spike at  $t = t_F$ . In all simulations we take the absolute refractory time  $\tau_{\text{ref}}$  equal to 1 ms, a value that is commonly reported (see any textbook on neurobiology, e.g., [51]). This choice of  $\tau_{\text{ref}}$  limits the maximum spiking frequency to 500 Hz which is a reasonable value during short activity bursts. Relative refractoriness and adaptation are excluded in our present model, but the Ansatz (5.3) for  $h_i^{\text{ref}}(t)$  can be generalized easily so as to include both effects [26].

A single neuron, separated from the rest of the network [ $h_i^{\text{syn}}(t) = 0$ ] and driven by a constant input  $h_i^{\text{ext}}(t) = \gamma$ , has the input/output characteristics shown in Fig. 5.1(a). The graph exhibits the typical sigmoidal dependence of the mean firing rate  $f = n_F/T$  (number of spikes during a measurement interval  $T$ ) upon the input  $\gamma$ . With a simple argument based on the spiking probability per time step (5.1) and the duration of absolute refractoriness (5.3) it can be shown [26] that  $f(\gamma) = P_F(\gamma)/[\tau_{\text{ref}} + P_F(\gamma)]$  where  $P_F(\gamma)$  is the firing probability for the potential  $h_i = \gamma$ ; cf. Eq. (5.1). The gain function  $f$  is plotted in Fig. 5.1(a). The set of parameters ( $\tau_{\text{ref}} = 1$  ms,  $\theta = 0.12$ ,  $\beta = 15$ ) is the same as in the simulations of a network of neurons in Secs. 5.4–5.6. The maximum spiking rate of a single neuron equipped with these parameters is 500 Hz. If, however, the model neuron is combined with an inhibitory partner neuron — as in the network introduced below — then the mean firing rate at high input levels is reduced to less than 150 Hz (Fig. 5.1(b)). Let us therefore turn to the structure of the network.

The neurons are connected in a scheme of  $N$  elements (Fig. 5.2), each element consisting of a pyramidal neuron (top layer) and an inhibitory partner neuron (bottom layer). The pyramidal neurons form a fully connected network which is able to store  $q$  patterns by adjusting the synaptic weights according to a “Hebbian” learning rule [32,38,39]. A pattern  $\mu$  is defined as a set of independent, identically distributed random variables  $\xi_i^\mu$  with  $1 \leq \mu \leq q$  and  $1 \leq i \leq N$ , which assume the values  $\xi_i^\mu = \pm 1$  with probability  $(1 \pm a)/2$  where  $a$  is the mean. The patterns are stationary and learned by adapting the synaptic efficacies  $J_{ij}$  of signal transmission from neuron  $j$  to neuron  $i$  [33] according to the rule

$$J_{ij} = \frac{2}{N(1 - a^2)} \sum_{\mu=1}^q \xi_i^\mu (\xi_j^\mu - a). \quad (5.4)$$

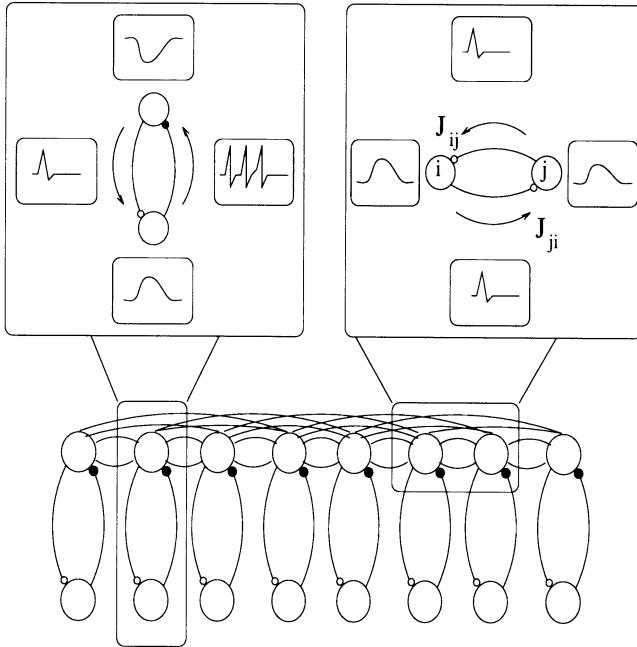
The neurons in the network communicate via the exchange of spikes which are generated according to the threshold firing process (5.1). If a neuron  $j$  in the Hebbian layer fires, the spike is transmitted to all neurons of the Hebbian layer as well as to the partner neuron in the inhibitory layer. Let



**Fig. 5.1. Gain functions.** (a) Gain function of a single neuron as predicted by the spike response model. The mean firing rate ( $y$  axis) displays a sigmoid dependence upon the applied input ( $x$  axis, arbitrary units). All parameters (noise  $\beta = 15$ , threshold  $\theta = 0.12$ , absolute refractory period  $\tau_{\text{ref}} = 1$  ms) are the same as used later on during the simulations of a large network of model neurons; cf. Figs. 5.6–5.17. (b) Gain function of the very same neuron, after it has been connected to its inhibitory partner neuron (circles:  $\beta = 15$ ; solid line:  $\beta = \infty$ ). The mean firing rate of the neuron pair is much lower than that of a single neuron. Note the different scale in (a) and (b). The intersection of the gain function with the dashed lines (from left to right,  $\gamma = 0.00, 0.08$ , and  $0.16$ ) yields the stationary mean firing rate at different input levels, if the neuron pair is part of a large network; see Sec. 5.3.2 for details. Taken from [24].

us now analyze the effect on a neuron  $i$  in the Hebbian layer first (Fig. 5.2, inset top right).

A postsynaptic neuron  $i$  that is connected to neuron  $j$  through a synapse of efficacy  $J_{ij}$  will receive the spike of neuron  $j$  only after some axonal transmission delay  $\Delta_i^{\text{ax}}$  (this is the model) which is taken to vary stochastically in a range of  $\Delta_{\min}^{\text{ax}} \leq \Delta_i^{\text{ax}} \leq \Delta_{\max}^{\text{ax}}$ . After this delay, an excitatory (or inhibitory) potential is induced at the axon hillock of the postsynaptic neuron. Its time course is described by a response function  $\epsilon(\tau)$ . Its am-



**Fig. 5.2. Network structure.** The network to be studied below consists of two layers of 4000 neurons (only 8 are shown) which communicate via the exchange of spikes. The neurons in the top layer (pyramidal neurons) are fully connected by Hebbian synapses  $J_{ij}$  that store a fixed set of patterns. Each pyramidal neuron is also connected to an inhibitory partner neuron (bottom). If one of the neurons fires, a spike is transmitted along the axon and, after some delay, evokes an excitatory (EPSP) or inhibitory (IPSP) postsynaptic potential at the receiving neuron (inset top left and top right). Taken from [24].

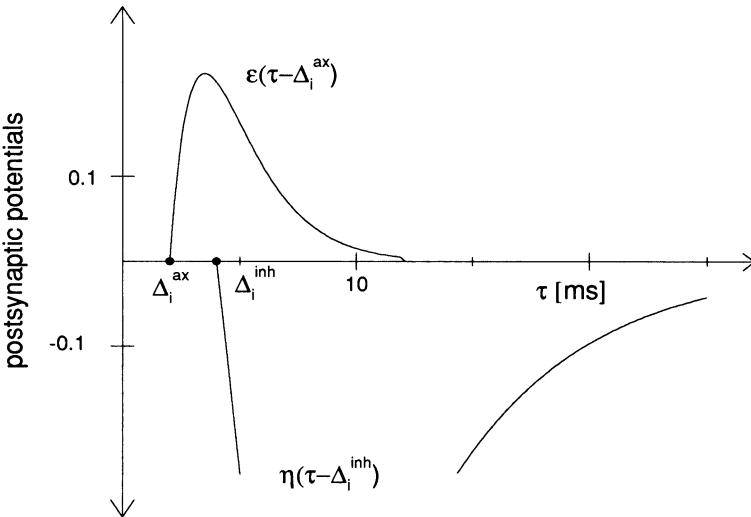
plitude, however, is determined by the synaptic efficacy  $J_{ij}$  given by Eq. (5.4). The response function  $\epsilon(\tau)$  of an excitatory synapse is simply the excitatory postsynaptic potential (EPSP) which has been determined experimentally (for a review see, e.g., [58]). We approximate this function by a discrete version of the  $\alpha$  function [47]

$$\epsilon(t) = (t/\tau_e^2) \exp(-t/\tau_e) \quad (5.5)$$

which is shown in Fig. 5.3.

Summing the contributions of all neurons which send signals to neuron  $i$  yields the synaptic part of the membrane potential (5.2),

$$h_i^{\text{syn}}(t) = \sum_{j=1}^N J_{ij} \sum_{\tau=0}^{\infty} \epsilon(\tau) S_j(t-\tau-\Delta_i^{\text{ax}}) + h_i^{\text{inh}}(t), \quad (5.6)$$



**Fig. 5.3. Excitatory (EPSP) and inhibitory (IPSP) postsynaptic potential.** If a pyramidal neuron  $i$  fires at  $\tau = 0$ , a spike is transmitted to all other pyramidal neurons to which it is connected. After some transmission delay  $\Delta_i^{ax}$  it evokes an EPSP at the soma of the postsynaptic neurons. The pyramidal neuron also sends a signal to the inhibitory partner neuron, which in turn sends a spike back so as to induce — after a delay  $\Delta_i^{inh}$  — a strong IPSP; cf. Fig. 5.2. The time course of an EPSP is modeled by the function  $\epsilon(t) = (t/\tau_\epsilon^2) \exp(-t/\tau_\epsilon)$  with  $\tau_\epsilon = 2$  ms (top). Inhibition is described by a sharp rise and an exponential decay with a time constant  $\tau_\eta = 6$  ms (bottom). Taken from [24].

where  $h_i^{inh}(t)$  is the contribution of the inhibitory partner neuron of neuron  $i$ .

The inhibitory contribution can be determined as follows (Fig. 5.2, top left). If neuron  $i$  starts firing at time  $t_F$ , a couple of spikes are transmitted to the inhibitory partner neuron where they evoke — after a delay  $\delta_1$  — an EPSP. This excites the inhibitory neuron to send a series of spikes back to neuron  $i$  where they generate — after another delay  $\delta_2$  — an inhibitory signal, that is, an IPSP. The net effect is a strong inhibitory feedback to the active neuron  $i$ ,  $\eta(\tau)$ , which sets in after a total delay of  $\Delta_i^{inh} = \delta_1 + \delta_2$  with a steep increase and decays exponentially with a time constant  $\tau_\eta = 6$  ms afterwards; see Fig. 5.3. The delay  $\Delta_i^{inh}$  is assumed to vary for different neurons in a range of  $\Delta_{min}^{inh} \leq \Delta_i^{inh} \leq \Delta_{max}^{inh}$  with equal probability. In view of the above considerations we obtain

$$h_i^{inh}(t) = \sum_{\tau=0}^{\tau_{max}} \eta(\tau) S_i(t - \tau - \Delta_i^{inh}) \quad (5.7)$$

with

$$\eta(t) = J^{inh} \exp(-t/\tau_\eta). \quad (5.8)$$

To include saturation effects at strong inhibition levels, we take the upper bound  $\tau_{\max}$  in Eq. (5.7) as a flexible limit and stop the summation after the first nonvanishing summation term. That is, inhibitory potentials which arise after a bursting episode of neuron  $i$  do not add up linearly, but only the most recent contribution matters.

## 5.3 Theory of Locking

The network as defined in the preceding section combines the features of association networks with the properties of oscillation networks (see also [5,11]). In particular, we find retrieval states that have a macroscopic overlap with a single pattern. Depending on the time structure of excitation and inhibition, the *overlap* (see below) is either stationary or oscillatory. In contrast to most oscillation models where a neuron — or a whole set of neurons — is modeled by a mere nonlinear oscillator with some limit cycle, our model of a spiking neuron allows one to determine the period and stability of a collective oscillation in terms of biological parameters, notably, the shape of EPSP and IPSP. This can be done in spite of the fact that we have a *distribution* of delay times  $\Delta_i^{\text{ax}}$  and  $\Delta_i^{\text{inh}}$ .

The analysis of a large network ( $N \rightarrow \infty$ ) in the low-noise limit ( $\beta \rightarrow \infty$ ) is done in three steps. First we derive the dynamic equation for the evolution of the network states (Sec. 5.3.1). Then we calculate the overlap in a stationary retrieval state (Sec. 5.3.2) and the oscillation period in an oscillatory retrieval state (Sec. 5.3.3). Finally, we present a condition for locking of neuronal activity into a collective oscillation (Sec. 5.3.4), a stability analysis.

### 5.3.1 EQUATION OF MOTION

To simplify notation we introduce the overlap  $m_\mu(t)$  as a measure of the correlations in the firing pattern of the neurons with pattern  $\mu$ :

$$m_\mu(t) = \frac{2}{N(1-a^2)} \sum_{j=1}^N (\xi_j^\mu - a) S_j. \quad (5.9)$$

Here  $a = <\xi_j^\mu>$  is the mean activity of all the random patterns. The overlap  $m_\mu(t)$  takes a maximum value, if all neurons  $i$  in the “foreground” of pattern  $\mu$  ( $\xi_i^\mu = +1$ ) fire synchronously during one time step while all “background” neurons  $j$  ( $\xi_j^\mu = -1$ ) stay quiescent. It vanishes, if the firing of foreground as well as background neurons occurs stochastically and  $N$  is large ( $N \rightarrow \infty$ ).

Substituting Eqs. (5.4), (5.7), and (5.9) into Eq. (5.6) we obtain the

postsynaptic potential

$$h_i^{\text{syn}}(t) = \sum_{\mu=1}^q \xi_i^\mu \sum_{\tau=0}^{\infty} \epsilon(\tau) m_\mu(t-\tau-\Delta_i^{\text{ax}}) + \sum_{\tau=0}^{\tau_{\max}} \eta(\tau) S_i(t-\tau-\Delta_i^{\text{inh}}). \quad (5.10)$$

At this point, it is convenient to group the neurons into ensembles of those neurons that have identical “properties.” To this end we introduce *sublattices* [35–37]

$$L(\mathbf{x}, D^{\text{ax}}, \eta, \rho) = \{i | \xi_i^\mu = \mathbf{x}, \Delta_i^{\text{ax}} = D^{\text{ax}}, h_i^{\text{inh}}(t) = -\eta, h_i^{\text{ref}}(t) = -\rho\} \quad (5.11)$$

that gather neurons with the same storage vector ( $\xi_i^\mu, 1 \leq \mu \leq q$ ), with an identical axonal delay  $\Delta_i^{\text{ax}}$ , the same (momentary) inhibition strength  $h_i^{\text{inh}}(t)$  and refractory field  $h_i^{\text{ref}}(t)$  into a common class. If  $p(\mathbf{x}, D^{\text{ax}}, \eta, \rho; t)$  is the portion of neurons that belong to the sublattice  $L(\mathbf{x}, D^{\text{ax}}, \eta, \rho)$  at time  $t$ , then the overlap can be written [38,39]

$$m_\mu(t+1) = [N(1-a^2)]^{-1} \sum_{\mathbf{x}} \sum_{D^{\text{ax}}=\Delta_{\min}^{\text{ax}}}^{\Delta_{\max}^{\text{ax}}} \sum_{\eta=0}^{\eta_{\max}} \sum_{\rho=0}^R p(\mathbf{x}, D^{\text{ax}}, \eta, \rho; t) x^\mu \{1 + \tanh[\beta(h(\mathbf{x}, D^{\text{ax}}, \eta, \rho, t) - \theta)]\} \quad (5.12)$$

with

$$h(\mathbf{x}, D^{\text{ax}}, \eta, \rho, t) = \sum_{\mu=1}^q x^\mu \sum_{\tau=0}^{\infty} \epsilon(\tau) m_\mu(t-\tau-D^{\text{ax}}) - \eta - \rho + h^{\text{ext}}(\mathbf{x}). \quad (5.13)$$

For the sake of simplicity we have assumed that the external signal depends on the storage vector  $\mathbf{x}$  only. Equation (5.12) is a simple consequence of the strong law of large numbers [54] applied to the dynamics (5.1).

Equations (5.12) and (5.13) describe the dynamic evolution of the macroscopic network states and contain all possible solutions of our model network. The derivation of these results is based on a time-dependent mean-field theory which is exact in the case of a finite number of patterns in a large network ( $N \rightarrow \infty$ ). See Gerstner and van Hemmen [26] for details of the derivation.

Two basic types of solution — stationary and oscillatory overlaps — may occur, which we study in the limit of noiseless threshold neurons ( $\beta \rightarrow \infty$ ). Intermediate between both cases is a regime of *weak locking* which is discussed in Sec. 5.3.5.

### 5.3.2 STATIONARY STATES

A stationary state is defined by the condition  $m_\mu(t) \equiv m_\mu$  for all patterns  $1 \leq \mu \leq q$ . In this case Eq. (5.13) can be simplified,

$$h(\mathbf{x}, D^{\text{ax}}, \eta, \rho, t) = \sum_{\mu=1}^q x^\mu m_\mu - \eta - \rho + h^{\text{ext}}(\mathbf{x}), \quad (5.14)$$

where we have used the normalization condition  $\sum_{\tau=0}^{\infty} \epsilon(\tau) = 1$ . Equation (5.14) shows that the interaction of neuron  $i \in L(\mathbf{x}, D^{\text{ax}}, \eta, \rho)$  with all of the Hebbian neurons and the external signal yields a *constant* input  $h_{\text{const}}(\mathbf{x}) = \sum_{\mu=1}^q x^\mu m_\mu + h^{\text{ext}}(\mathbf{x})$ . The only time-dependent part in the postsynaptic potential of a neuron  $i$  belonging to  $L(\mathbf{x}, D^{\text{ax}}, \eta, \rho)$  stems from the interaction with its inhibitory partner neuron and its own refractory field. The effect of inhibition and refractoriness can be understood easily.

If  $h_{\text{const}}(\mathbf{x}) < \theta$ , a neuron  $i \in L(\mathbf{x}, D^{\text{ax}}, \eta, \rho)$  stays quiescent and generates no spikes ( $n_B = 0$ ). If, however,  $h_{\text{const}}(\mathbf{x}) > \theta$ , then neuron  $i$  fires with maximum frequency<sup>2</sup> and emits a burst of spikes until the inhibitory feedback becomes strong enough to end the spiking. Only after another time  $\tau^0$  when the inhibitory signal has declined sufficiently, that is,  $\eta(\tau^0 - 1) < \theta - h_{\text{const}}(\mathbf{x})$  and  $\eta(\tau^0) \geq \theta - h_{\text{const}}(\mathbf{x})$ , can the firing start again.

This result can be used to simplify Eq. (5.12). Let us assume that the IPSP  $\eta(\tau)$  is a sharply rising function and blocks firing immediately when it sets in. Then a burst of neuron  $i$  that has started at  $t = 0$  is ended at  $t = \Delta_i^{\text{inh}}$  when the inhibitory feedback due to the first spike arrives. This leaves the neuron time to fire a total of  $n_B(\Delta_i^{\text{inh}}) = 1 + \text{Int}[\Delta_i^{\text{inh}} / (\tau_{\text{ref}} + 1)]$  spikes where  $(\tau_{\text{ref}} + 1)^{-1}$  is the firing frequency during the burst and  $\text{Int}[ \cdot ]$  denotes the integer part of the term in square brackets. Thus spiking occurs at times  $t_i = 0, (\tau_{\text{ref}} + 1), \dots, [n_B(\Delta_i^{\text{inh}} - 1)](\tau_{\text{ref}} + 1)$  and stops afterwards due to the inhibitory feedback. Only if the inhibition that is due to the last spike of the burst has declined sufficiently can the neuron fire again and start the next burst. The bursting period  $T_B$  is therefore given by

$$T_B(\Delta_i^{\text{inh}}) = [n_B(\Delta_i^{\text{inh}}) - 1](\tau_{\text{ref}} + 1) + \tau^0, \quad (5.15)$$

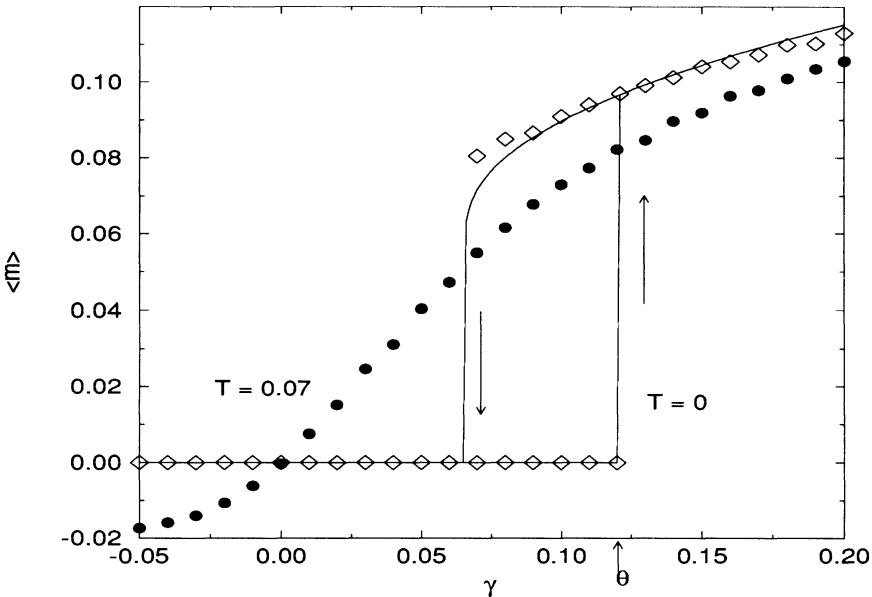
$\tau^0$  being the recovery time after inhibition. We see that the result (5.15) only depends on the inhibitory delay loop  $\Delta_i^{\text{inh}}$ , so it is independent of the axonal delay times  $\Delta_i^{\text{ax}}$ . Instead of a classification of the neurons by sublattices  $L(\mathbf{x}, D^{\text{ax}}, \eta, \rho)$  it is now more convenient to introduce new sublattices defined by  $L(\mathbf{x}, D^{\text{inh}}) = \{i | \xi_i^\mu = \mathbf{x}, \Delta_i^{\text{inh}} = D^{\text{inh}}\}$ . If  $p(\mathbf{x}, D^{\text{inh}})$  is the portion of neurons belonging to  $L(\mathbf{x}, D^{\text{inh}})$ , a simple probability, we find for the overlap

$$m_\mu = \frac{2}{1 - a^2} \sum_{\mathbf{x}} \sum_{D^{\text{inh}} = \Delta_{\min}^{\text{inh}}}^{\Delta_{\max}^{\text{inh}}} x^\mu p(\mathbf{x}, D^{\text{inh}}) \frac{n_B(D^{\text{inh}})}{T_B(D^{\text{inh}})}. \quad (5.16)$$

The quantity  $[n_B(D^{\text{inh}})/T_B(D^{\text{inh}})]$  is the mean firing rate of neurons with a fixed inhibitory delay loop  $\Delta_i^{\text{inh}} = D^{\text{inh}}$ . This result, Eq. (5.16), is in accordance with a much more general theorem that in a stationary state

---

<sup>2</sup>This does not hold any longer as soon as we include *relative* refractory behavior [51, p. 152].



**Fig. 5.4.** *Stationary retrieval  $\langle m \rangle$  as a function of the external stimulus.* Theoretical results of a large ( $N \rightarrow \infty$ ) and noiseless ( $\beta \rightarrow \infty$ ) network are shown by the solid lines. Simulation results for a network of  $N = 4000$  neurons pairs are plotted as open diamonds ( $\beta = \infty$ ) and filled circles ( $\beta = 15$ ). If the external stimulus  $\gamma$  is increased, the noiseless system passes a regime of bistability which ends at  $\gamma = \theta = 0.12$ . There the overlap jumps to the upper branch (pattern retrieval). The hysteresis loop has been indicated by arrows. At finite temperature ( $\beta = 15$ ) the transition to the retrieval state is continuous. These results can be understood with the aid of a graphical solution, indicated in Fig. 5.1(b), using the gain function of a neuron pair. Taken from [24].

of a fully connected network storing a finite number of patterns the *mean firing rate* of the neurons is the only important quantity [27].

Equations (5.14)–(5.16) allow a straightforward numerical solution for the stationary overlap  $m_\mu$ . In particular, it is possible to find the retrieval states  $m_\mu = m\delta_{\mu\nu}$  as a function of the external signal  $h^{\text{ext}}(\mathbf{x}) = \gamma(x^\nu + 1)/2$ . To simplify the analysis, we now consider time as a continuous variable and calculate  $\tau_0$  from the exact threshold equation  $\eta(\tau_0) = \theta - \gamma$ . In Fig. (5.4) the theoretical predictions for a large and noiseless network are compared with simulation results of a network consisting of two layers of 4000 neurons with  $\Delta_{\min}^{\text{inh}} = 3$  ms and  $\Delta_{\max}^{\text{inh}} = 6$  ms. Whereas at  $T = 0$  the transition is sharp and occurs exactly at  $\gamma = 0.12$  (open diamonds), it is smoothed to a continuous transition at a finite noise level ( $\beta = 15$ , filled circles). These results can be understood with the help of Fig. 5.1(b).

According to Eqs. (5.14)–(5.16), the stationary retrieval overlap is given by  $m = \bar{f}(m + \gamma)$  where  $\bar{f}$  is the mean firing rate [in kHz] averaged over

all pairs of neuron and inhibitory partner neuron. Suppose for the moment that  $\Delta_{\min}^{\text{inh}} = \Delta_{\max}^{\text{inh}}$ . Then the intersection of a typical gain function of a neuron pair (Fig. 5.1(b), solid line — noiseless; circles for  $\beta = 15$ ) with a straight line of slope one (dashed) yields the stationary retrieval overlap. The external signal  $\gamma$  shifts the dashed line parallel to the  $x$  axis. Using this graphical construction we can explain the nature of the solution in Fig. 5.4. In the noiseless case we have the trivial solution for small  $\gamma$ . With increasing external signal, we suddenly obtain a nonzero firing rate, move through a regime of bistability, and then arrive at a regime of large retrieval overlap only. In the case of finite but high enough temperature the gain function does not show a steep increase at threshold and the overlap increases continually with the external input.

The agreement between theory and numerical simulation is fair. The difference in Fig. 5.4 in the noiseless case is apart from finite-size effects also due to a *small* oscillation in the overlap. The stationarity condition is thus a good approximation, but it is not exactly fulfilled. This is not too surprising since we had taken the stability of the stationary solution for granted. It may, and in general will, undergo a Hopf bifurcation with small amplitude [28]. Oscillations with a large amplitude are considered in the next subsection.

### 5.3.3 OSCILLATORY STATES

In addition to stationary retrieval states with constant overlap  $m_\mu(t) \equiv m\delta_{\mu\nu}$  there are also collective oscillation states:

$$m_\mu(t) = \delta_{\mu\nu} f(t + nT_{\text{osc}}), \quad n \in \mathbb{Z}, \quad (5.17)$$

where  $T_{\text{osc}}$  is the period of a collective oscillation. Here  $f$  is a  $T_{\text{osc}}$  periodic function and  $\mathbb{Z}$  represents the integers.

We have seen in the preceding subsection that, given a constant input, neurons with different delay loops  $\Delta_i^{\text{inh}}$  have a different intrinsic bursting period  $T_B(\Delta_i^{\text{inh}})$ . The above Ansatz (5.17) presupposes that a collective oscillation with a *common* period  $T_{\text{osc}}$  is nevertheless possible. In other words, Eq. (5.17) implies “locking” of oscillatory elements with variable intrinsic frequencies. In this subsection we show that these “locked” oscillations indeed exist. The next subsection is devoted to the question under what condition they are *stable* solutions of the system’s dynamics.

In order to check whether oscillatory solutions are possible, we assume that all neurons in the foreground of a specific pattern ( $\xi_i^\nu = 1$ ) have fired regularly in a coherent oscillation for all times in the past ( $t < 0$ ). To be more explicit, we assume that there have been collective bursts at times  $T_F = -nT_{\text{osc}}$ , where  $n \in \mathbb{N}$  denotes the positive integers and  $T_{\text{osc}}$  is the as yet unknown bursting period which we are going to determine self-consistently. Despite the common period  $T_{\text{osc}}$ , groups of neurons with different delay times will not burst exactly at the same time since neurons

with short delays start a little too early in relation to the collective firing time  $T_F$  and neurons with long delays fire a little too late. We denote the delay of the burst's start for a neuron with axonal and inhibitory delays ( $D^{\text{inh}}, D^{\text{ax}}$ ) as compared to the collective firing time  $T_F$  by  $\delta(D^{\text{inh}}, D^{\text{ax}})$ . If we now recall from the preceding subsection that each burst contains  $n_B(\Delta_i^{\text{inh}})$  spikes fired with a frequency  $(\tau_{\text{ref}} + 1)^{-1}$ , then we can write the spiking times for a neuron  $i \in L(D^{\text{inh}}, D^{\text{ax}})$  where  $L(D^{\text{inh}}, D^{\text{ax}}) = \{i | \Delta_i^{\text{inh}} = D^{\text{inh}}, \Delta_i^{\text{ax}} = D^{\text{ax}}\}$  and  $(\xi_i^\nu = 1)$  in the form

$$t_i = -nT_{\text{osc}} + \delta(D^{\text{inh}}, D^{\text{ax}}) + (k-1)(\tau_{\text{ref}} + 1), \quad (5.18)$$

where  $T_F = -nT_{\text{osc}}$  with  $n \in \mathbb{N}$  denotes the hypothetical collective firing times,  $\delta(D^{\text{inh}}, D^{\text{ax}})$  is the shift in the burst's start of a neuron  $i \in L(D^{\text{inh}}, D^{\text{ax}})$ , and  $k$  with  $1 \leq k \leq n_B(D^{\text{inh}})$  denotes the  $k$ th spike in the burst. The origin of time is chosen in such a way that the next supposedly collective burst ( $n = 0$ ) occurs at  $t = T_F = 0$ . In the following we are going to determine the unknown quantities  $T_{\text{osc}}$  and  $\delta(D^{\text{inh}}, D^{\text{ax}})$ .

To simplify the notation we define a summation over all past EPSP's:

$$\tilde{\epsilon}(D^{\text{inh}}; T, t) = \sum_{n=1}^{\infty} \sum_{k=1}^{n_B} \epsilon[t + nT - (k-1)(\tau_{\text{ref}} + 1)]. \quad (5.19)$$

$T$  and  $t$  are considered as free variables, whereas  $D^{\text{inh}}$  is a parameter. The quantity  $\tilde{\epsilon}$  is a function of  $T$  and can be used to determine the oscillation period  $T_{\text{osc}}$  in the following way. Because of the regular firing in the past, viz., Eq. (5.18), the next burst of a neuron  $i \in L(D^{\text{inh}}, D^{\text{ax}})$  is due at time  $\delta(D^{\text{inh}}, D^{\text{ax}})$ . This implies that the threshold condition  $h_i^{\text{syn}}(t) + h_i^{\text{ext}}(t) \geq \theta$  is fulfilled at time  $t = \delta(D^{\text{inh}}, D^{\text{ax}})$ , but not yet at  $t = \delta(D^{\text{inh}}, D^{\text{ax}}) - 1$ . If we replace the inequality by an equality and insert the explicit expression for  $h_i^{\text{syn}}(t)$ , we can write the threshold condition

$$\begin{aligned} \theta &= \sum_{\substack{\Delta_{\max}^{\text{inh}} \\ D_0^{\text{inh}} = \Delta_{\min}^{\text{inh}}}}^{\Delta_{\max}^{\text{inh}}} \sum_{\substack{\Delta_{\max}^{\text{ax}} \\ D_0^{\text{ax}} = \Delta_{\min}^{\text{ax}}}}^{\Delta_{\max}^{\text{ax}}} \tilde{\epsilon}[D_0^{\text{inh}}; T, \delta(D^{\text{inh}}, D^{\text{ax}}) - D^{\text{ax}} - \delta(D_0^{\text{inh}}, D_0^{\text{ax}})] \\ &\quad + \eta\{T - [n_B(D^{\text{inh}}) - 1](\tau_{\text{ref}} + 1) - D^{\text{inh}}\} + h^{\text{ext}}(\mathbf{x}). \end{aligned} \quad (5.20)$$

As before we take  $h^{\text{ext}}(\mathbf{x}) = \gamma(x^\nu + 1)/2$ . Simultaneous solution of Eq. (5.20) for all groups  $L(D^{\text{inh}}, D^{\text{ax}})$  yields the oscillation period  $T_{\text{osc}} = \text{Int}[T + 1]$  and the shift in the bursts start  $\delta(D^{\text{inh}}, D^{\text{ax}})$ .

To simplify expression (5.20) we now assume that  $\tilde{\epsilon}$  and  $\eta$  vary only slowly during one time step and that differences in the delays  $D^{\text{inh}}$  and  $D^{\text{ax}}$  are small and also result in a spread of firing times  $\delta(D^{\text{inh}}, D^{\text{ax}})$ . In this case,  $\tilde{\epsilon}$  and  $\eta$  can be taken continuous and differentiable. Expansion around the mean  $\bar{D}^{\text{inh}}$  and  $\bar{D}^{\text{ax}}$  yields after some algebra the approximate formula

$$\theta \approx \eta(T - 2\bar{D}^{\text{inh}}) + \tilde{\epsilon}(\bar{D}^{\text{inh}}; T, -\bar{D}^{\text{ax}}) + h^{\text{ext}}(\mathbf{x}), \quad (5.21)$$

where  $\bar{D}^{\text{inh}} \approx [n_B(\bar{D}^{\text{inh}}) - 1](\tau_{\text{ref}} + 1)$  has been used. This allows a solution for the oscillation period  $T_{\text{osc}}$ .

Within the same approximation, the starting times of the bursts depend on the inhibitory delay through

$$\delta(D^{\text{inh}}, D^{\text{ax}}) = 2 \frac{\frac{d}{dt} \eta(T_{\text{osc}} - 2\bar{D}^{\text{inh}})}{\frac{d}{dt} \tilde{\epsilon}(\bar{D}^{\text{inh}}, T_{\text{osc}}, -\bar{D}^{\text{ax}})} (D^{\text{inh}} - \bar{D}^{\text{inh}}). \quad (5.22)$$

If we assume realistic parameters for the shape of the EPSP, the response function  $\epsilon(\tau)$  shows a pronounced maximum followed by a fast decay. Then  $\tilde{\epsilon}(\bar{D}^{\text{inh}}; T, -\bar{D}^{\text{ax}})$  can be approximated by

$$\tilde{\epsilon}(\bar{D}^{\text{inh}}; T, -\bar{D}^{\text{ax}}) \approx \epsilon(T - \bar{D}^{\text{ax}}) n_B(\bar{D}^{\text{inh}}). \quad (5.23)$$

This equation allows a simple graphical interpretation of the above result; see Fig. 5.5. The firing period  $T_{\text{osc}}$  is given by the intersection of the appropriately scaled and delayed graphs of  $\eta(\tau)$  and  $\epsilon(\tau)$ ; cf. Eq. (5.21). The spread of the starting times of the bursts is determined by the slope of the graphs at the intersection point; see Eq. (5.22).

In a reasonable oscillation scenario, we expect that the fast neurons with  $\Delta_i^{\text{inh}} < \bar{D}^{\text{inh}}$  come a bit too early ( $\delta_i < 0$ ) in relation to the collective firing, whereas slow neurons  $\Delta_i^{\text{inh}} > \bar{D}^{\text{inh}}$  are a bit too late ( $\delta_i > 0$ ). Equation (5.22) tells us that this is true only if  $d\tilde{\epsilon}/dt$  has the same sign as  $d\eta/dt$ . Since  $(d\eta/dt)(T_{\text{osc}} - 2\bar{D}^{\text{inh}}) > 0$ , it follows that  $\tilde{\epsilon}$  must have a *positive slope* at the firing point. We will see in the following paragraph that this is also the necessary condition for a stable oscillation.

### 5.3.4 A LOCKING CONDITION

So far we have assumed that a coherent oscillation exists and — based on the assumption that the collective bursting has been stable for a long time in the past [cf. Eq. (5.18)] — we have calculated the period  $T_{\text{osc}}$  of the oscillation and the delay  $\delta(D^{\text{inh}}, D^{\text{ax}})$  of the burst's start for the various groups of neurons.

In this subsection, we derive a condition that must be fulfilled to ensure that an oscillation is indeed stable. To this end, we assume that one of the neurons, let us say neuron  $j \in L(D^{\text{inh}}, D^{\text{ax}})$ , is not perfectly in time with its group, but starts bursting a little too early or too late, in other words, its firing is shifted by a short time  $\Delta t$ . The oscillation is stable, if the neuron is “drawn back” into its group, that is, if the time shift  $\tilde{\Delta}t$  is reduced during the next cycle.

The synaptic part of the membrane potential has two contributions:

$$h_j^{\text{syn}}(t) = h_j^\epsilon(t) + h_j^\eta(t). \quad (5.24)$$

The first term,

$$h_j^\epsilon(t) = \sum_{D_0^{\text{inh}}, D_0^{\text{ax}}} \tilde{\epsilon}[D_0^{\text{inh}}; T, \delta(D^{\text{inh}}, D^{\text{ax}}) - D^{\text{ax}} - \delta(D_0^{\text{inh}}, D_0^{\text{ax}})], \quad (5.25)$$

is induced by the signals of the other neurons in the net and is unchanged compared to the unperturbed system. The inhibitory term, however, is shifted due to the perturbation [cf. Eq. (5.20)]

$$h_j^\eta(t) = \eta\{t - [n_B(D^{\text{inh}})-1](\tau_{\text{ref}}+1) - D^{\text{inh}} - \Delta t\}, = h_0^\eta(t - \Delta t), \quad (5.26)$$

where the subscript 0 indicates the unperturbed system. The next firing occurs at  $t_0 + \tilde{\Delta}t$  where  $t_0$  is the firing time of the unperturbed network. In the limit of continuous time, the threshold condition is

$$h_j^\epsilon(t_0 + \tilde{\Delta}t) + h_0^\eta(t_0 + \tilde{\Delta}t - \Delta t) = \theta - h_i^{\text{ext}}. \quad (5.27)$$

Expansion around  $t_0$  and using the fact that  $h_j^\epsilon(t_0) + h_0^\eta(t_0) = \theta - h_i^{\text{ext}}(t)$  yields

$$\tilde{\Delta}t = \frac{d}{dt} h_0^\eta(t_0) \left[ \frac{d}{dt} h_0^\eta(t_0) + \frac{d}{dt} h_j^\epsilon(t_0) \right]^{-1} \Delta t. \quad (5.28)$$

Since  $(dh_0^\eta/dt)(t_0) > 0$ , a condition for locking is  $(dh_j^\epsilon/dt)(t_0) > 0$ . In other words, a global oscillation is stable only if a burst starts while excitation is still growing.

### 5.3.5 WEAK LOCKING

Equation (5.28) gives a necessary condition for a *globally* locked oscillation. But what happens if some of the neurons, let us say all neurons in  $L(D^{\text{inh}}, D^{\text{ax}})$ , do not fulfill Eq. (5.28)? If so, they will remain out of phase, once they have dropped out of the collective oscillation. This does not imply that the oscillation is destroyed completely. It is also possible that only a small group of neurons with very long or very short axonal delay  $\Delta_i^{\text{ax}}$  drop out of the collective oscillation whereas all other neurons remain in an oscillatory state. In this case, the macroscopic overlap is — as before — a T-periodic function of time, but not all neurons fire exactly T-periodic. Some of the neurons (i.e., neurons in  $L(D^{\text{inh}}, D^{\text{ax}})$ ) occasionally slip through and “miss” the collective firing or just do not care at all. We call this a state of *weak locking*.

To find a lower bound of the regime of “weakly locked” solutions we can adopt an iterative approach. First, we calculate the globally locked solution, and determine the stability of locking Eq. (5.28) for all neurons of all sublattices  $L(D^{\text{inh}}, D^{\text{ax}})$ . If Eq. (5.28) is not fulfilled for some neurons, we *assume* that they are completely out of phase and give — on the average — no contribution to the T-periodic Ansatz (5.17), providing only a stationary background. Based on this assumption we repeat the procedure, calculate a

new period, check whether all remaining neurons are stably locked, and so on. If the iterative procedure converges to a state with “nonparticipating” neurons, then a “weakly locked” solution is possible. This criterion, however, is only a necessary condition, since simulations show that the neurons which have dropped out of the collective oscillation are not completely out of phase, but only slip through occasionally. Due to the remaining oscillations in the overlap which acts as a periodic stimulus on those neurons, they have a tendency to spike in phase with the rest of the network. Thus they are, in a true sense, *weakly locked*.

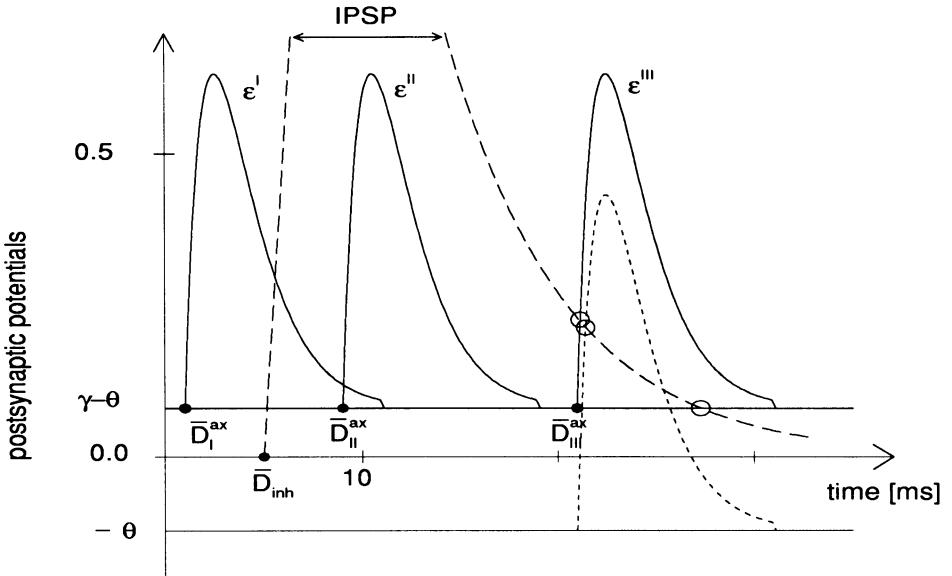
The above arguments are valid in the noiseless case ( $T = 0$ ). The regimes of stationary states, global oscillations, and weakly locked oscillatory states are determined only by the width and position of the distributions of the delays  $\Delta_i^{\text{inh}}$  and  $\Delta_i^{\text{ax}}$ . If noise is added, each neuron fires stochastically, and the spikes may come a little too early or too late in comparison with the noiseless case. It is therefore to be expected that adding noise lowers the oscillation amplitude and drives the system from global oscillation to weak locking or stationarity. Simulations of a network at a finite noise level which show this behavior are presented in the next section.

## 5.4 Simulation Results

In the preceding section we have shown that two basic types of solution are possible: Oscillatory and stationary retrieval of patterns. Oscillatory solutions can be divided in globally and weakly locked oscillations. The theoretical arguments have demonstrated that it is the timing of the EPSP in relation to the IPSP which determines the solution type. In this section, we test the validity of the theoretical results in simulations of a network consisting of two layers of 4000 neurons at a *finite* noise level ( $\beta = 15$ ) with  $q = 5$  patterns and  $a = -0.8$ . In passing we note that extensively many patterns ( $q = 200$  in a system of  $N = 4000$ ) will do as well but are computationally more expensive. The EPSP and IPSP are modeled with a biological shape as shown in Fig. 5.3. The delay in the inhibitory loop of a specific neuron pair is chosen once and for all from a block-shaped distribution that varies between  $\Delta_{\min}^{\text{inh}} = 3$  ms and  $\Delta_{\max}^{\text{inh}} = 6$  ms. The only remaining free parameter then is the duration of the axonal transmission delay  $\Delta_i^{\text{ax}}$ . To test its influence we have considered three different scenarios: see Fig. 5.6.

### 5.4.1 THREE SCENARIOS

In all scenarios we start with a randomly chosen state of mean activity  $a$ . For the first 200 ms no external signal is applied, then a weak signal is switched on and kept on a constant level of  $h_i^{\text{ext}}(t) = \gamma(\xi_i^\mu + 1)/2$  with

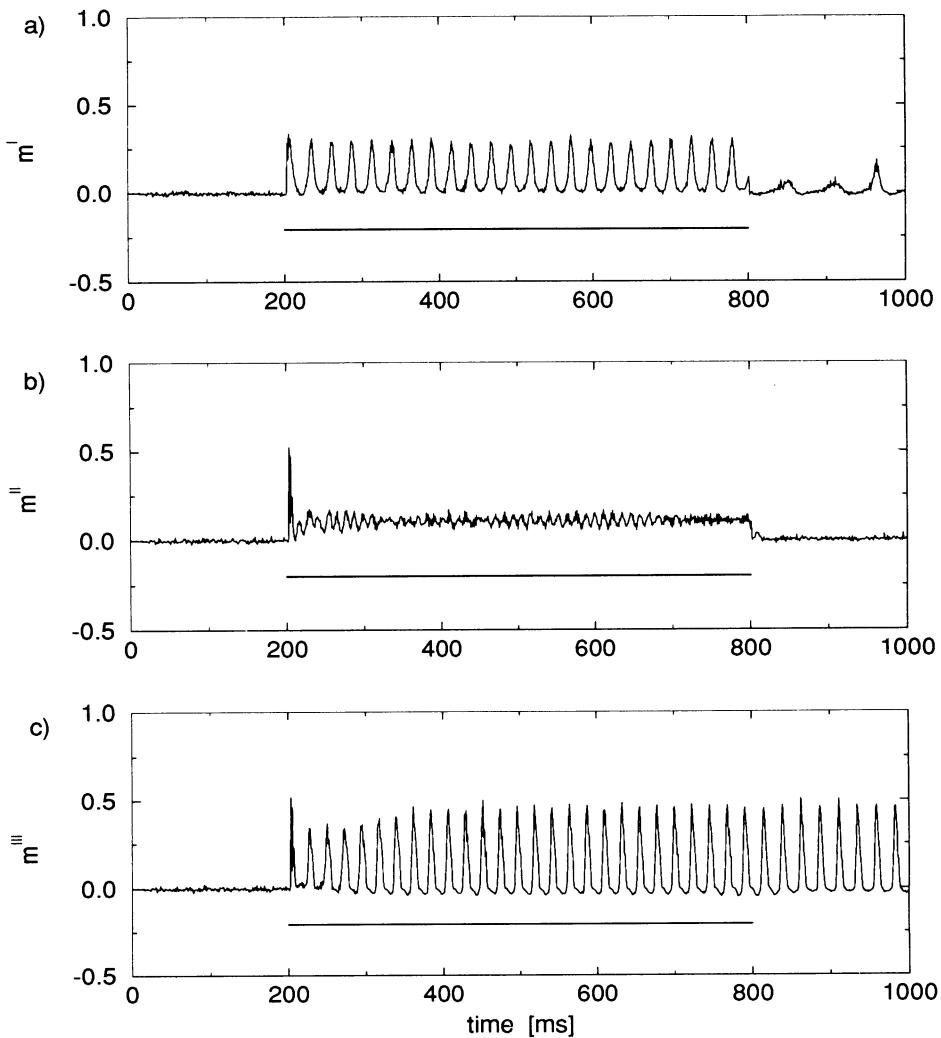


**Fig. 5.5.** Relative position of EPSP (solid line) and IPSP (long dashes) in three different scenarios. We assume that all neurons have fired coherently at  $t = 0$ . Scenario I: Short axonal delays between 0 and 2 ms; mean value  $\bar{D}_I^{\text{ax}} = 1$  ms. Scenario II: medium axonal delays (8–10 ms,  $\bar{D}_{II}^{\text{ax}} = 9$  ms). Scenario III: long axonal delays (20–22 ms,  $\bar{D}_{III}^{\text{ax}} = 21$  ms). The intersection of  $\epsilon^{I,III} = n_B \epsilon(\tau - \bar{D}_{I,III}^{\text{ax}}) + (\gamma - \theta)$  with the function  $-\eta(\tau - 2\bar{D}^{\text{inh}})$  yields the oscillation period in scenario I ( $T_{\text{osc}}^I = 27$  ms) and scenario III ( $T_{\text{osc}}^I = 23$  ms); see Eqs. (1.21) and (1.23), and the text for details. Here  $\Theta$  is the threshold and  $\gamma = 0.2 \ll 1$  is the amplitude of the input (stimulus). In scenario III an oscillation is possible even after the external signal has been turned off ( $\gamma = 0$ , dashed line). Scenario II does not allow oscillatory solutions. Taken from [24].

$\gamma = 0.2$ , which is only slightly above threshold. After another 600 ms the signal is switched off again.

In scenario I (Fig. 5.6(a)), we have assumed that axonal delays  $\Delta_i^{\text{ax}}$  are short and distributed in a range between 0 and 2 ms with equal probability. With this delay distribution, an oscillation starts, if the signal is switched on, and it continues *as long as the signal is active*. It stops immediately after the signal is turned off. This is consistent with the theoretical results (5.20) and (5.28) — and with experiment [69].

In scenario II, the axonal delays are slightly longer, distributed in a range  $8 \text{ ms} \leq \Delta_i^{\text{ax}} \leq 10 \text{ ms}$ . In this case, excitation is growing while inhibition is still dominant (see Fig. 5.5, curve “II”) and an oscillation cannot be stable; cf. Eq. (5.28). This is confirmed in the simulation (Fig. 5.6(b)) where a constant overlap develops after initialization of the signal. The magnitude of the overlap is consistent with the theoretical result (5.16). If the signal



**Fig. 5.6. Response of a network to a stimulus.** Overlaps with a specific pattern are shown as a function of time for the three different scenarios. The pattern is supported by a weak external signal for a time between 200 and 800 time steps (ms) as indicated by the horizontal bar. Top: scenario I, oscillations occur only while an external stimulus is applied. Middle: scenario II, transient behavior to a stationary retrieval state. Bottom: scenario III, oscillations remain even after the stimulus has been turned off. Taken from [24].

is turned off, the overlap vanishes and the network returns to the quiescent state.

All this is different in scenario III where we have long axonal delays ( $20 \text{ ms} \leq \Delta_i^{\text{ax}} \leq 22 \text{ ms}$ ). Figure 5.6(c) shows that a collective oscillation develops which is stable even *after* the signal has been turned off. We can understand this result by analyzing the graphical interpretation of Eqs. (5.21), (5.23), and (5.28) as shown in Fig. 5.5. The intersection of the graph of  $-\eta(\tau)$  with the scaled version of  $\epsilon(\tau)$  (curve “III”) yields an oscillation period  $T_{\text{osc}} \approx 23 \text{ ms}$  that matches well with the simulation result, Fig. 5.6(c). In both cases — with or without signal — a collective oscillation is stable, since bursts start while excitation is still growing; cf. Eq. (5.28).

Scenarios I–III are typical examples of a network behavior that varies continuously from stationarity to full locking. The transition between the different regimes can be seen in the “phase diagram” (Fig. 5.7) where the amplitude of the oscillation is shown by a scale of grey values. Dark areas indicate large oscillations, white area negligible oscillations. Depending on the minimal duration of axonal transmission delays ( $y$  axis) and the width of their distribution ( $x$  axis) the amplitude varies from below 0.1 (stationary) to above 0.3 (perfect locking). Whereas a broader distribution (larger  $x$  values) always lowers the amplitude, a change in the  $y$  values may switch the system from one regime to the next and either increase or decrease the amplitude, as discussed above.

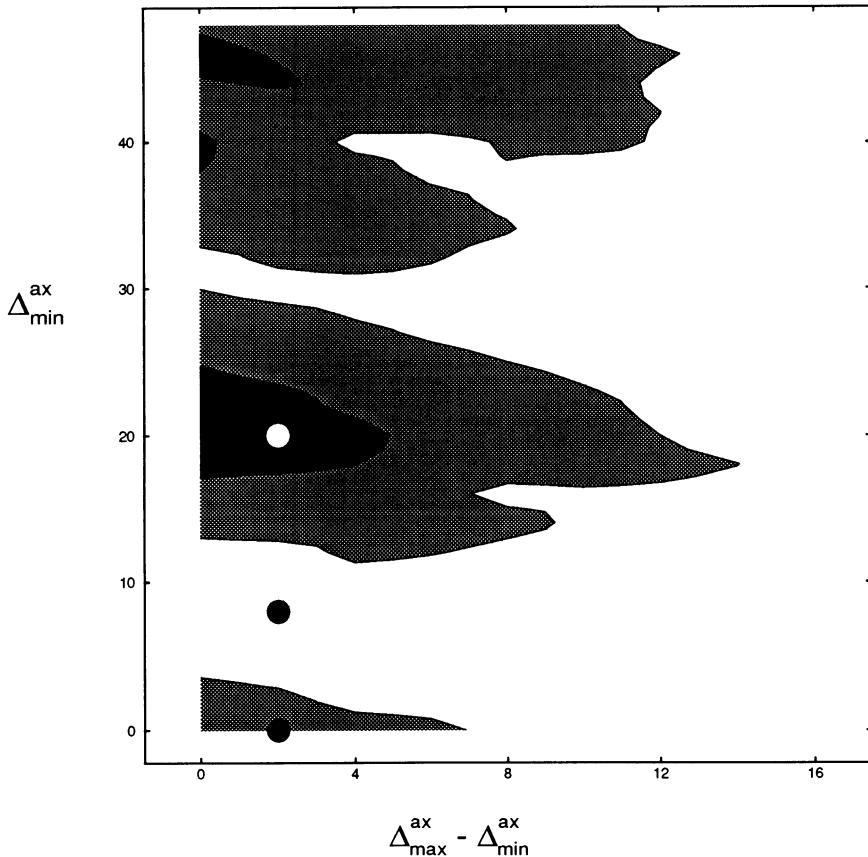
#### 5.4.2 INTERPRETATION

The most interesting behavior is seen in scenario I where the network oscillates only *as long as* the external signal is applied. If the signal is removed, the system returns to the stationary resting state. This behavior in scenario I compares favorably to experiments in the visual cortex of the cat where oscillations are seen only during stimulation [69]. Otherwise a stationary background activity is found.

Scenario I is also the most realistic one, at least for the cat. It is known that axonal delays within one area of the cortex are rather short. Delay times distributed in a range between 0 and 2 ms seem to be a reasonable assumption [51]. The other parameters (rise time of EPSP  $\tau_\epsilon \approx 3 \text{ ms}$ , rise time of IPSP  $\approx 2 \text{ ms}$ , decay time  $\tau_\eta \approx 6 \text{ ms}$ , inhibition loop delay  $3 - 6 \text{ ms}$ ) are also taken in a biologically plausible range. With this set of parameters we find a period of oscillation of approximately  $20 - 25 \text{ ms}$  which is consistent with experiments where oscillations with  $40 - 50 \text{ Hz}$  have been found [16,30].

The theoretical argument (1.28) shows that the oscillation is only weakly stable since  $\frac{d}{dt} h^\epsilon(t_0) \approx 0$  for the “fast” neurons. This may be an explanation of the fact that in the noisy environment of real systems oscillations are barely visible [62,63].

Compared to the experimental observations where the oscillations only



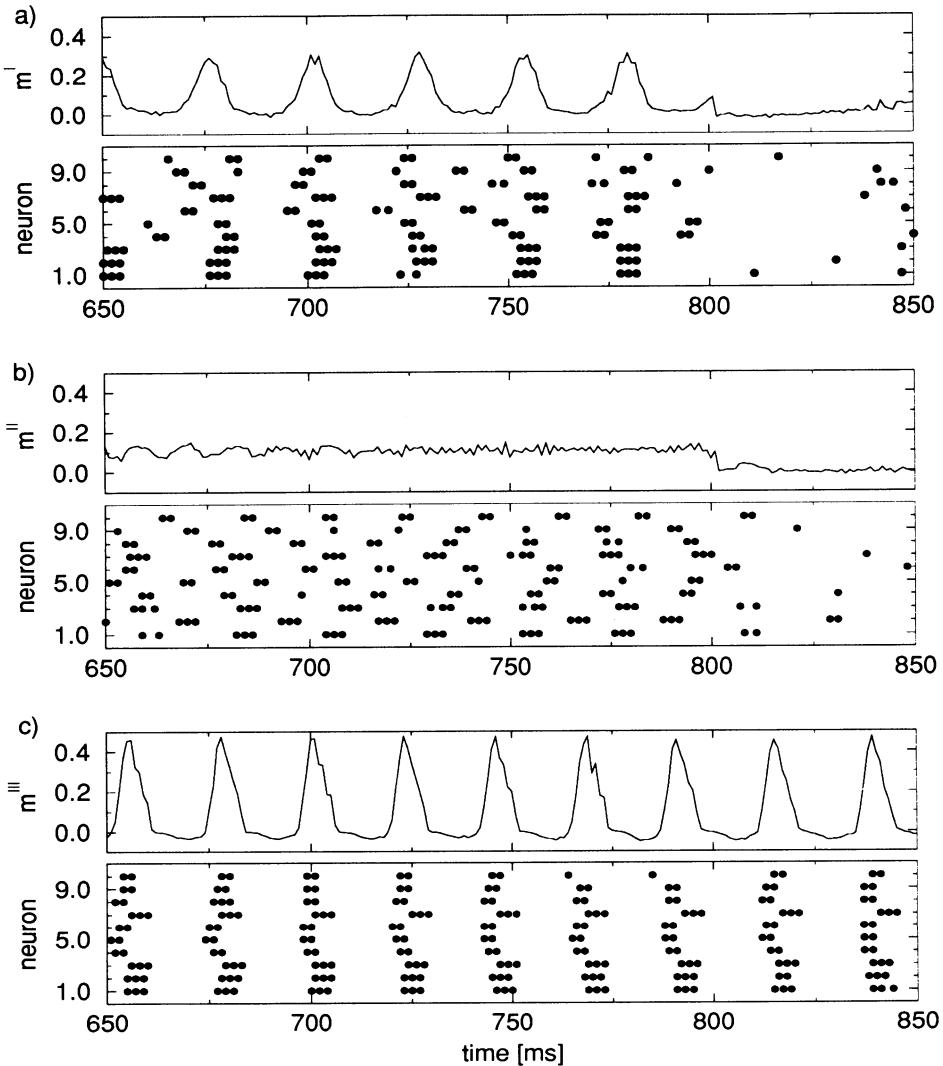
**Fig. 5.7. Phase diagram.** The amplitude of an oscillation  $m_{\text{osc}} = \langle m_{\max} - m_{\min} \rangle$  is plotted as a function of both the minimal axonal delay  $\Delta_{\min}^{\text{ax}}$  ( $y$  axis) and the width  $\Delta_{\max}^{\text{ax}} - \Delta_{\min}^{\text{ax}}$  of the delay distribution ( $x$  axis). Dark areas indicate a large amplitude oscillation ( $m_{\text{osc}} > 0.3$ ), white areas indicate a stationary overlap ( $m_{\text{osc}} < 0.1$ ) while gray shaded areas define the region of weak locking ( $0.1 < m_{\text{osc}} < 0.3$ ). The three scenarios of Fig. 5.6 lie along the line ( $x = \Delta_{\max}^{\text{ax}} - \Delta_{\min}^{\text{ax}} = 2$  ms) parallel to the  $y$  axis and are marked by black and white circles. Whereas for a sharp distribution of delay times ( $x = 0$ ), the oscillation amplitude depends critically on the transmission delay, a broad distribution ( $x > 16$ ) always leads to a stationary solution. Note that scenario I, the weak locking scenario, lies in a regime intermediate between strong oscillations and stationarity (grey shaded area). Taken from [24].

last for a short time (150–200 ms, [18]) — they are packed in spindles — our network oscillates too clearly and persistently. There have been some suggestions how a more realistic time structure of the neural response can be achieved as well in simulations. For example, Deppisch et al. [14] propose a rather noisy input which tends to desynchronize the neural activity and, thus, is in conflict with a synchronizing synaptic interaction in the network. Bauer et al. [6,7] use nonlinear synaptic couplings to construct a bistable regime where it is possible to get noise-induced switching between oscillatory and stochastic network activity.

Up to now, however, it is also possible that other mechanisms such as the influence of “higher” areas play a role in determining the temporal structure of the neural signal and we are far from understanding all the details.

### 5.4.3 SPIKE RASTER

In contrast to most other models of oscillatory neural networks our approach yields not only global activities and time-averaged firing rates but also *spike trains of single neurons* and their correlations. The full information on the network behavior is contained in the spike raster which shows all spikes of an arbitrary set of neurons in a time resolved plot. A spike raster is thus equivalent to experimental multielectrode recordings where the activity of several neurons is measured at the same time. The spike raster of Fig. 5.8 shows the spikes of 10 neurons in scenarios I to III. All neurons from which recordings are made are numbered and their labels are plotted along the  $y$  axis. Time is plotted in the  $x$  direction and the spikes of each neuron are marked as a dot along a horizontal line parallel to the  $x$  axis. Coherent activity can then be visualized as vertical columns of dots; see, for example, at  $t = 700$  ms in scenario I (Fig. 5.8(a)) and III (Fig. 5.8(c)). The burst structure of any single neuron (e.g., bursts of two or three spikes) is clearly visible if we follow the spike dots along a horizontal line. A closer look at the spike raster of scenario III reveals that, while bursting is globally synchronized for all active neurons, the bursts of different neurons do not start exactly at the same time. Neurons with a short inhibitory delay loop (e.g., # 4 – 6) start firing earlier than those with a long delay loop (e.g., # 3 and 7), as predicted by Eq. (5.22). Both groups, however, fast neurons as well as slow neurons, have the same period of global collective oscillation ( $T_{\text{osc}} \approx 23$  ms). In scenario II, each neuron exhibits bursting activity, but the bursts occur stochastically and are not synchronized. An intermediate case is shown in the spike raster of scenario I where some of the neurons (e.g. # 6 and 9) slip through the “oscillation maze” occasionally, that is, are *weakly locked*. Nevertheless, global coherence is still dominant as is indicated by the oscillation of the overlap  $m$ . The spike raster, however, enables us to reveal additional information that is not obvious, if we look at the overlap only.



**Fig. 5.8.** *Spike raster in the three scenarios.* The upper curve in (a)–(c) shows a small portion of the overlap function of Fig. 5.6 with higher time resolution. The stimulus is active up to  $t = 800$  ms and is turned off afterwards. The spike raster below shows the spikes of 10 active neurons. Neurons are numbered and plotted along the  $y$  axis, time is plotted along the  $x$  axis. Each spike of a neuron is marked by a dot. (a) In scenario I, coherent activity — visible as vertical columns of spikes — is dominant but some of the neurons (e.g., numbers 6 and 9) occasionally slip through: weak locking. (b) In scenario II, all neurons burst asynchronously as is confirmed by the stationary overlap. (c) In scenario III, locking is perfect. Note that neurons with short delays (e.g., numbers 4–6) tend to come earlier than neurons with long delays (e.g., numbers 3 and 7), as predicted by the theory. Taken from [24].

#### 5.4.4 SYNCHRONIZATION BETWEEN TWO HEMISPHERES

Considering the possible role of collective oscillations in the cortex as a clue to representing objects, we have to face the question whether neural activity can synchronize between the two cortical hemispheres. Due to the longer pathway via the corpus callosum one may assume that the main difference as compared to intrahemispheric interactions consists in longer axonal delays tending to deteriorate a synchronization.

Nevertheless, Engel et al. [20] have shown experimentally that a synchronization of neuronal activity is indeed possible throughout the hemispheres despite the interhemispheric delays. They were able to destroy the synchronization by cutting the corpus callosum. We now address the question of whether synchronization is influenced by longer delays. In so doing we use the SRM.

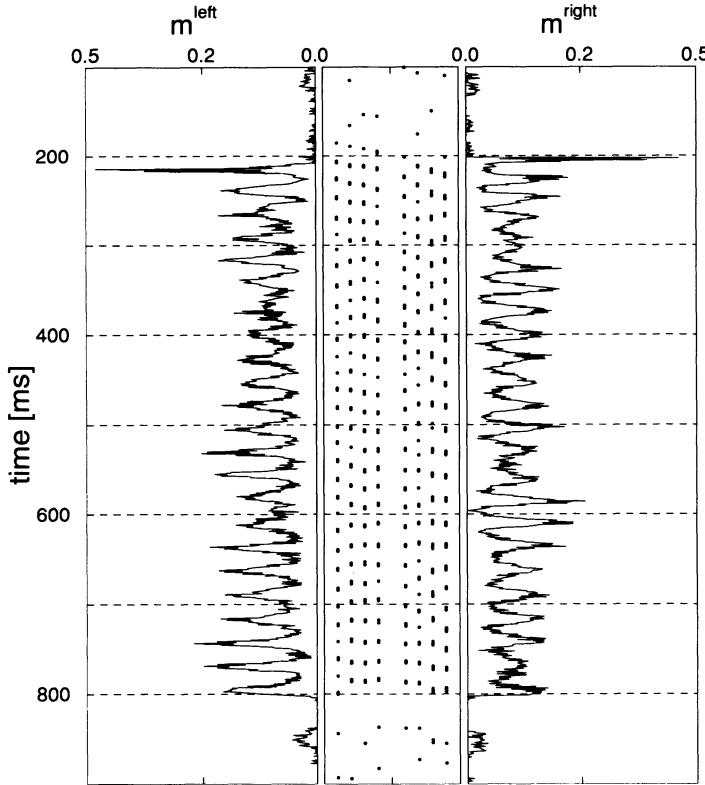
The network consists of two parts representing the two hemispheres. Each of them has the same structure as considered before in scenario I. The interactions in and between the two hemispheres are all to all (every neuron gets input from every neuron) but it takes more time for a spike to travel from one hemisphere to the other. We model this by introducing for each neuron an interhemispheric delay  $\Delta_i^{\text{hem}}$  that we sample from a uniform distribution between  $\Delta_{\min}^{\text{hem}}$  and  $\Delta_{\max}^{\text{hem}}$ . Using the overlaps, which this time have to be calculated for the two hemispheres separately, we get for the local field of neuron  $i$  [cf. Eq. (5.10)]:

$$\begin{aligned} h_i^{\text{syn}}(t) &= h_i^{\text{inh}}(t) \\ &+ \sum_{\mu=1}^q \xi_i^{\mu} \sum_{\tau=0}^{\infty} \epsilon(\tau) [m_{\mu}^{\text{left}}(t - \tau - \Delta_i^{\text{ax}}) + m_{\mu}^{\text{right}}(t - \tau - \Delta_i^{\text{hem}})] \end{aligned} \quad (5.29)$$

if this neuron belongs to the left hemisphere; otherwise  $m^{\text{left}}$  and  $m^{\text{right}}$  have to be interchanged. All other parameters are as before in scenario I.

To test the stability of a synchronized oscillation, we stimulate the network through a learned pattern that is represented by neurons of both hemispheres. The onset of the stimulus for the two hemispheres differs, however, by 12 ms, which is roughly half of the period of the expected oscillation. So the initial condition tends to evoke an oscillation with a phase shift  $\pi$  between the two hemispheres. Now it depends on the choice of  $\Delta_{\min}^{\text{hem}}$  and  $\Delta_{\max}^{\text{hem}}$  whether they continue oscillating phase-shifted or whether they synchronize.

To give an example of the network behavior we show in Fig. 5.9 the time evolution of the overlaps in the two hemispheres together with a spike raster of randomly chosen, active, neurons from both hemispheres. The interhemispheric delays vary between  $\Delta_{\min}^{\text{hem}} = 3$  ms and  $\Delta_{\max}^{\text{hem}} = 7$  ms. Despite the phase shift in the onset of the stimulus a synchronized oscillation evolves after a few cycles as can be seen by a closer look at the overlaps — but



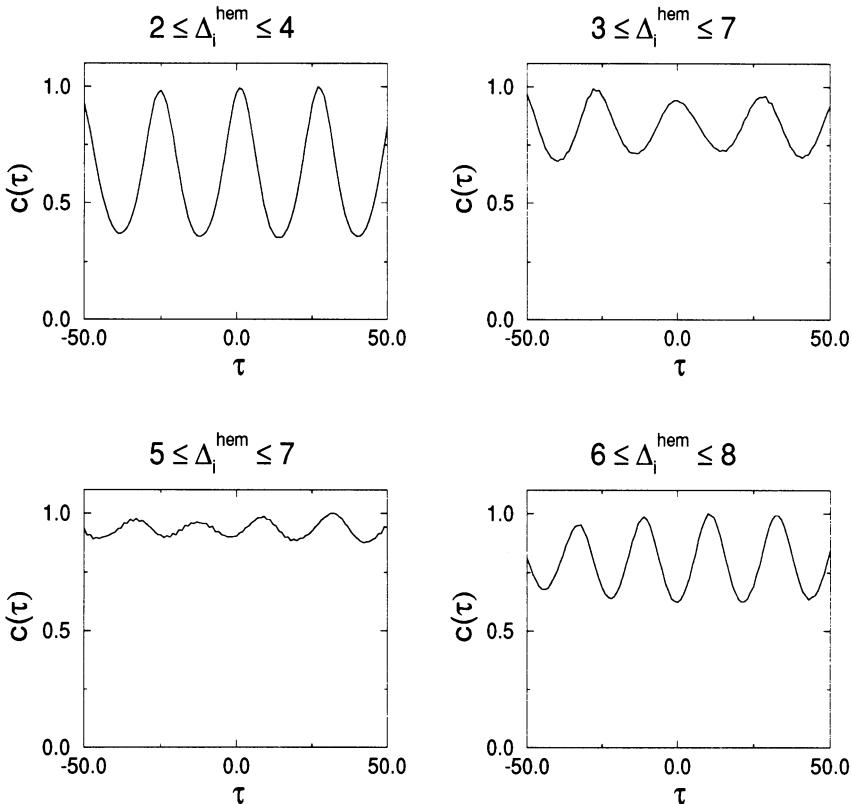
**Fig. 5.9.** *Synchronization between two hemispheres.* The network of scenario I now consists of two hemispheres. In each of them, the connection topology is the same as in Fig. 5.2 and so are the intrahemispheric delays. The transmission delays between the two hemispheres, however, are (expected to be) longer and we therefore sample them from a uniform distribution between 3 and 7 ms. The figure shows the overlap of the left and the right hemisphere with a pattern supported by an external signal, which lasts for 600 ms, as well as a spike raster of randomly chosen, active, neurons from both hemispheres. A synchronized oscillation evolves in the two hemispheres despite the fact that the onset of the stimulus differs by 12 ms for the two parts so that the hemispheres are initiated with a phase shift corresponding to roughly half of the period.

hardly in the spike raster. To make this synchronous activity more visible we compute a correlation function of the two overlaps:

$$\tilde{c}(\tau) = \sum_{t=t^{\text{start}}}^{t=t^{\text{stop}}} m^{\text{left}}(t) m^{\text{right}}(t - \tau) \quad (5.30)$$

and normalize so that the maximum is one, that is, with

$$n = \max_{\tau} \tilde{c}(\tau) \quad (5.31)$$



**Fig. 5.10.** Correlation between the two hemispheres for different delays. The dependence of the stability of the synchronous oscillation upon the interhemispheric delays  $\Delta_i^{\text{hem}}$  is shown in four correlograms. The scenario is the same as in Fig. 5.9 but the range from which the  $\Delta_i^{\text{hem}}$  are sampled varies. As long as the averaged delay  $\langle \Delta_i^{\text{hem}} \rangle$  is small enough ( $\leq 5$  ms) a synchronous oscillation evolves (upper part; peak at  $\tau = 0$ ) whereas for longer delays ( $\langle \Delta_i^{\text{hem}} \rangle \geq 6$  ms) the activity of the two hemispheres oscillates phase-shifted by  $\pi$  (lower part; minimum at  $\tau = 0$ ). The amplitude of the oscillations in the correlograms reflects the quality of the synchronization between the two hemispheres.

and

$$c(\tau) = \frac{1}{n} \tilde{c}(\tau). \quad (5.32)$$

This correlation function is plotted in Fig. 5.10 for four different choices of  $\Delta_{\min}^{\text{hem}}$  and  $\Delta_{\max}^{\text{hem}}$  as indicated in the figure, with  $t^{\text{start}} = 400$  ms and  $t^{\text{stop}} = 800$  ms. As can be seen clearly, the two hemispheres oscillate in synchrony as long as the interhemispheric delays are not too long on the average:  $\langle \Delta_i^{\text{hem}} \rangle \leq 5$  ms. However, as soon as the average exceeds a certain value (between 5 and 6 ms) the two parts stay in the phase-shifted oscillation.

The observed coherent activity in the neuronal response throughout the two hemispheres may thus be interpreted so that the axonal delays between the two parts of the brain do not exceed 5 ms *on the average*. It would be quite interesting to verify this experimentally.

## 5.5 Application to Binding and Segmentation

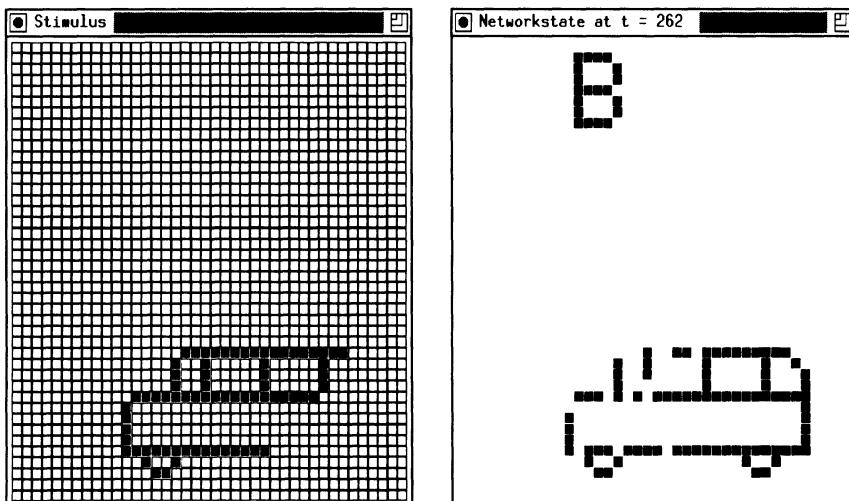
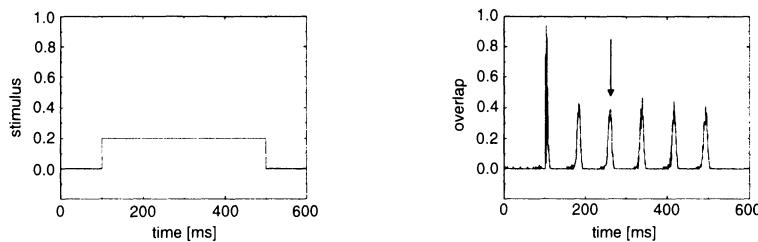
Now we choose the regime of weak locking which seems to be the state which is biologically most relevant and also most powerful in handling the binding problem — as we will show in this section. A network of 40x42 neurons is trained with  $q = 7$  patterns, each pattern consisting of a simple contour labeled by a letter; cf. Fig 5.11. The number of neurons which are active in a pattern varies in a range between 4% and 7%. The mean activity averaged over all neurons and all patterns is  $a = -0.9$ . Taking many more patterns is allowed but computationally expensive. We have therefore refrained from doing so.

### 5.5.1 FEATURE LINKING

First of all, it is worth noticing that the associative capabilities of the Hopfield model can be found in this network of *spiking* neurons as well. The neurons are represented by pixels to allow an easy interpretation to the human reader. This does not mean that the network has been arranged according to some two-dimensional topology. Neurons which have fired one or more spikes during the last five time steps are coded by black pixels while the rest is white. Figure 5.11 shows on the left-hand side the stimulus which is applied to the network. It is switched on at  $t = 100$  and kept constant up to  $t = 500$ . After some time, the network has reconstructed the pattern, as seen on the right-hand side of the figure. We also note that the system has accomplished some sort of heteroassociation task by assigning a “label,” here a B, to the retrieved pattern. We would like to stress that this kind of pattern retrieval is time-dependent, as can be seen in Fig. 5.11 top right, because, due to inhibition, the neurons are not able to fire all the time. So, feature *linking* is done through the simultaneous activity of all neurons coding the same object.

### 5.5.2 PATTERN SEGMENTATION

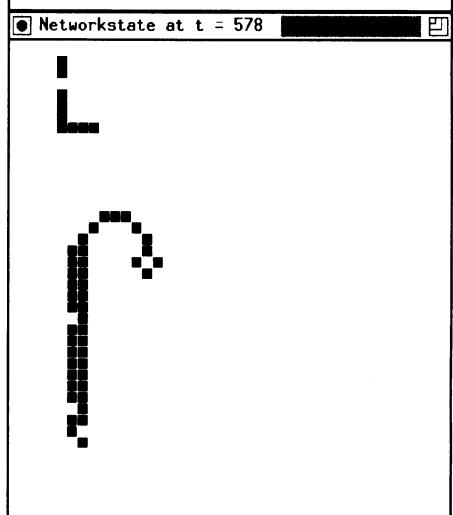
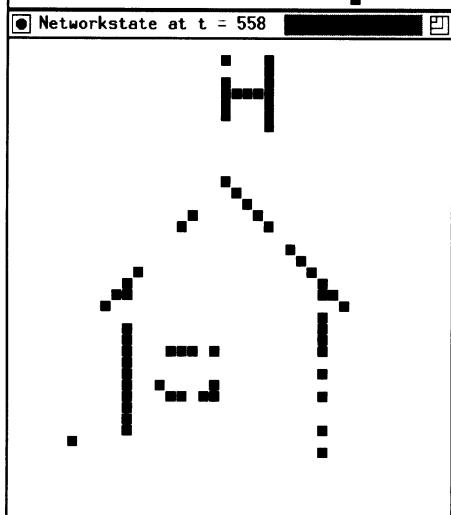
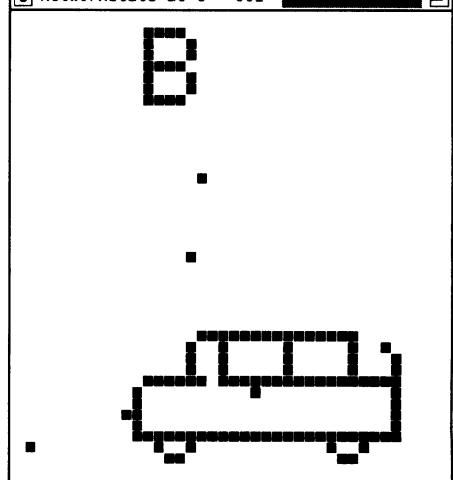
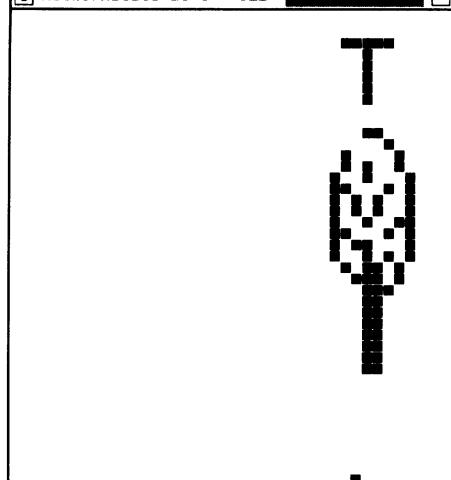
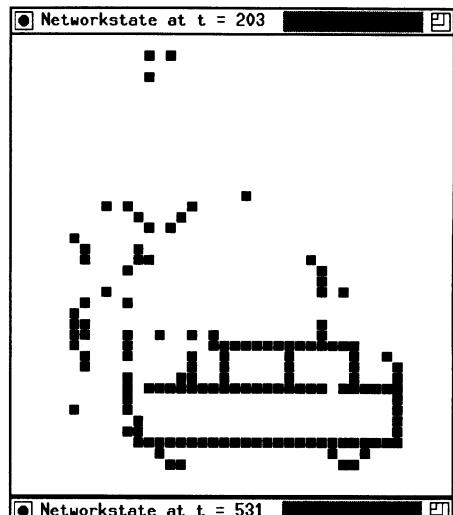
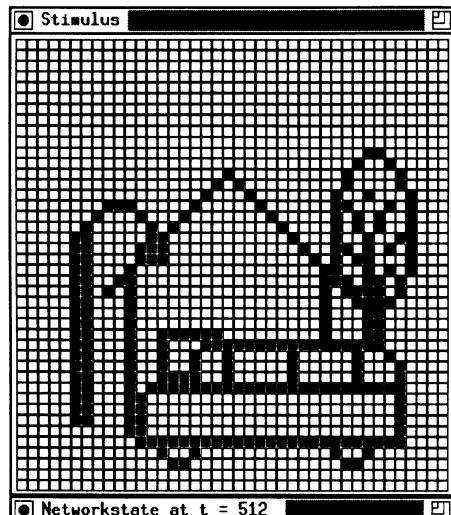
On the other hand, intermittent firing opens the possibility of exciting several patterns one after the other. This is exactly what is done when a superposition of *several* patterns is presented to the network. Figure 5.12 (top left to bottom right) shows the stimulus, then the network 100 time steps after the onset of the stimulus, and finally the system as it has evolved to a weakly locked oscillatory behavior where the four simultane-



**Fig. 5.11.** *Associative recall and feature linking.* Reconstruction of the bus which has been “learned” before (left: stimulus; right: momentary retrieval state). Associative retrieval and “labeling” of patterns are performed by a network of  $40 \times 42$  spiking neurons, which had learned the patterns before. In the upper part, the time course of the stimulus and overlap with the pattern “bus” are shown. The momentary time step is indicated by an arrow (top right). Note that the network evolves to a weakly locked oscillatory state which is interpreted as linking the different neurons coding the bus. Neurons which belong to the bus but are not stimulated are excited via synaptic inputs from the rest of the network. This leads to associative retrieval of the *full* pattern including the label B.

---

**Fig. 5.12.** *Pattern segmentation.* Four patterns which have been taught to the network as separate entities are now presented in a superposition (top left). After an intermediate phase the network nicely completes and separates the original patterns (T-B-H-L) through temporally disjoint, coherent spiking — in short, through coherent oscillations.



ously presented patterns are active *one after the other*. The frequency of this oscillation mainly depends on the strength  $J_{\text{inh}}$  and duration  $\tau_\eta$  of the local inhibition as can be seen in the following argument. A first estimate for the frequency of the bursts can be given in the noiseless case for a neuron which is only connected to its inhibitory partner neuron and receives a constant input higher than the activation threshold  $\theta$ . Then Eq. (5.1) reduces to an exact firing condition. With the local field given in Eqs. (5.7) and (5.8) one gets for a single neuron  $i$ :

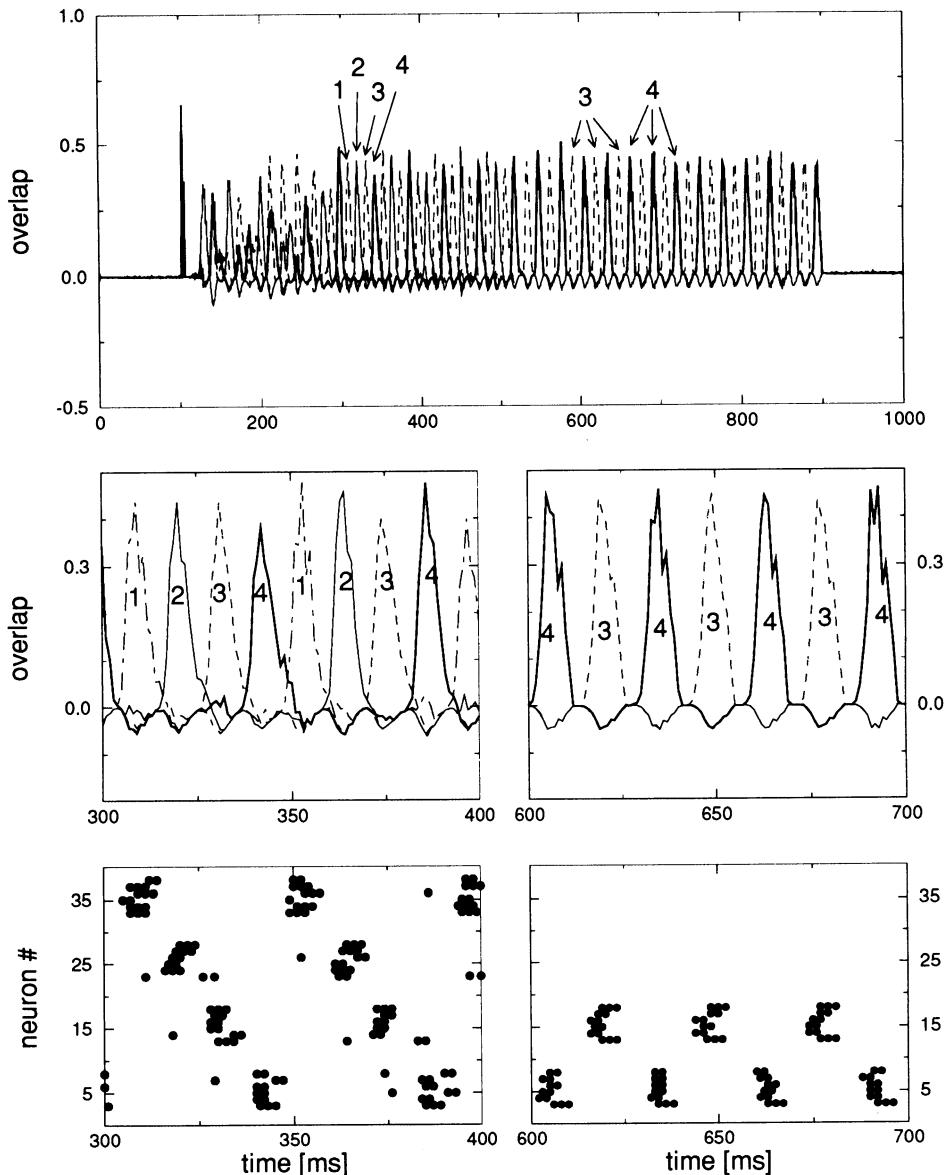
$$t_i^{\text{burst}} = \Delta_i^{\text{inh}} + \tau_\eta \ln \frac{|J_{\text{inh}}|}{\gamma - \theta} \quad \text{for } \gamma > \theta. \quad (5.33)$$

Here  $\gamma$  denotes the strength of the external stimulus ( $\gamma > \theta$ ). So, choosing  $J_{\text{inh}} = -2.5$  and  $\tau_\eta = 25$  ms we end up with a frequency about 11 Hz which is lower than the frequencies found in the cortex but convenient for the simulation. Note that parts of the patterns hidden in the stimulus are reconstructed and “labeling” is performed by assigning different letters to the different patterns (T-B-H-L). Due to noise, locking is not perfect in that some neurons drop out of the picture as they come a little too early or too late. It is fair to say, though, that the network can perform both the segmentation of the presented scene and the completion of the objects therein. We also note that in principle a superposition of arbitrarily many low activity patterns can be segmented by choosing the “hardware” suitably.

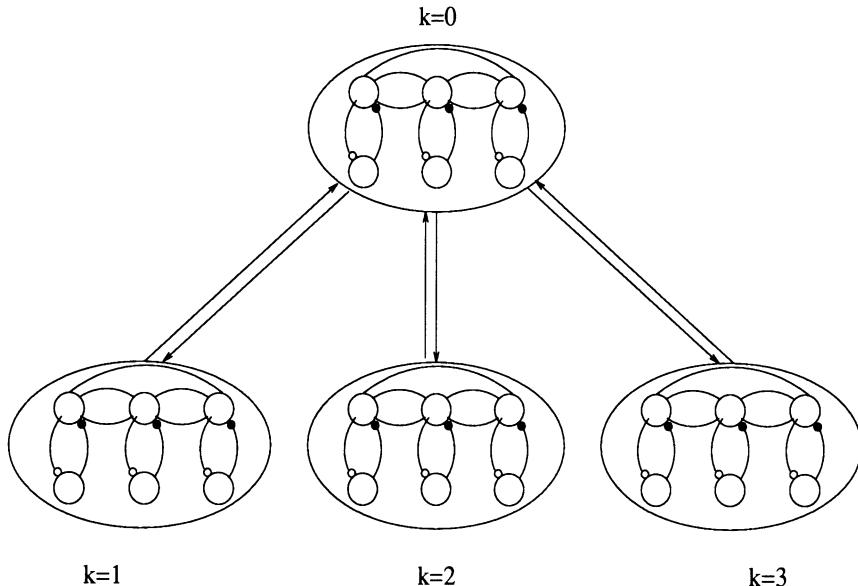
### 5.5.3 SWITCHING BETWEEN PATTERNS

The next picture (Fig. 5.13) shows the time structure of such a segmentation run in more detail. The overlaps with the presented patterns have been plotted during the whole simulation run (top) while in the middle two sections of their time course can be seen at a better time resolution. We also show a spike raster of some representative neurons from each of the stimulated patterns (bottom). In contrast to the example given above, all microscopic parameters are now in a biologically reasonable range ( $J_{\text{inh}} = -1$  and  $\tau_\eta = 6$  ms). Here we see that the maximum number of patterns that can be distinguished at the same time is restricted to *four* which has some support from psychophysical data [71,75]. In principle this number is arbitrary and depends mainly on the strength and duration of the local inhibition.

In addition, the network is able to change its global behavior on a fast time scale (ms range) as can be seen in the upper part of Fig. 5.13. At  $t = 500$  two of the four presented patterns are removed from the stimulus and the corresponding overlaps immediately stop oscillating. This can be observed once again at  $t = 900$  when the stimulus is removed completely. Plainly, such a fast reaction is an interesting effect since it allows the network to react immediately to a change in the environment.



**Fig. 5.13.** *Overlap and spike raster during pattern segmentation.* A superposition of four patterns has been presented to a network of 4000 neurons. The overlap with the four different patterns during a segmentation run with varying stimulus is shown as a function of time (top). All microscopic parameters are now in a biologically realistic range. After 200 ms the four patterns have been separated completely. The network reacts even faster, if patterns 1 and 2 are removed from the stimulus at  $t = 500$  ms (middle). Some typical spike rasters (bottom) before and after  $t = 500$  ms have also been shown. Since we have *weak locking* there is no strict periodicity.



**Fig. 5.14.** *Structure of the layered network.* Four columns are organized in two layers. Each column consists of 4000 fully connected neurons and 4000 inhibitory local neurons. Every neuron of the lower layer is connected to every neuron of the upper layer (and vice versa, if we allow feedback connections), but there are no connections between neurons of different columns of the same layer.

## 5.6 Context Sensitive Binding in a Layered Network with Feedback

As a last example of applications we demonstrate how the segmentation and binding capabilities of the SRM can be combined so as to solve the even harder task of context sensitive binding [23]. To do so, we modify the structure of our network. We now consider several groups of neurons (let us call them columns) organized in a layered structure; see Fig. 5.14. Every column consists of 4000 neurons which are fully connected and 4000 local inhibitory neurons, just as before in Sec. 5.4. The columns in the lower layer all have feedforward connections to the upper layer but there are no connections between the columns of the lower layer. The delays have been chosen once and for all from a uniform distribution ranging from 1 to 3 ms for the intracolumnar delays (as in scenario I before) while the delays between neurons of different layers are taken to vary between 8 and 10 ms.

The patterns are learned in the following way. Each column is trained on static low activity patterns resulting in intracolumnar couplings of the

same type as before (cf. Eq. (5.4))

$$J_{ki,kj} = g_k \frac{2}{N(1-a^2)} \sum_{\mu=1}^q \xi_{ki}^\mu (\xi_{kj}^\mu - a) \quad \begin{array}{l} 0 \leq k \leq K \\ 1 \leq i, j \leq N \end{array} \quad (5.34)$$

where  $k$  labels the columns and  $i$  the neurons in each column. In contrast to this, the feedforward and feedback couplings connecting the two layers store combinations of patterns in the lower layer by associating them with a specific pattern in the upper layer [76]. This results in intercolumnar couplings for the feed-forward connections

$$J_{0i,kj} = g_0 \frac{2}{N(1-a^2)} \sum_{\langle \mu, \mu' \rangle} \xi_{0i}^\mu (\xi_{kj}^{\mu'} - a), \quad \begin{array}{l} 1 \leq k \leq K \\ 1 \leq i, j \leq N \end{array} \quad (5.35)$$

and the feedback connections, if available, are given by

$$J_{ki,0j} = g_k \frac{2}{N(1-a^2)} \sum_{\langle \mu', \mu \rangle} \xi_{ki}^{\mu'} (\xi_{0j}^\mu - a), \quad \begin{array}{l} 1 \leq k \leq K \\ 1 \leq i, j \leq N \end{array} . \quad (5.36)$$

Here  $\langle \mu, \mu' \rangle$  denotes a combination of patterns  $\mu'$  in the different columns of the lower layer that is connected to pattern  $\mu$  in the upper layer. The normalization  $g_k$  is taken so that  $N/g_k$  equals the number of connections to every neuron in column  $k$ , so

$$g_k = \begin{cases} (1+K)^{-1} & \text{for } k = 0 \\ 1 & \text{for } 1 \leq k \leq K \text{ without feedback} \\ 0.5 & \text{for } 1 \leq k \leq K \text{ with feedback} \end{cases} . \quad (5.37)$$

Putting everything together we get for the local field of neuron  $i$  in column  $k$ :

$$h_{ki}^{\text{syn}} = g_k \sum_{\mu=1}^q \xi_{ki}^\mu \bar{m}_\mu^k(t - \Delta_{ki}) + g_k \sum_{\langle \mu', \mu \rangle} \xi_{ki}^{\mu'} \bar{m}_\mu^0(t - \Delta_{ki}^{\text{ar}}) \quad (5.38)$$

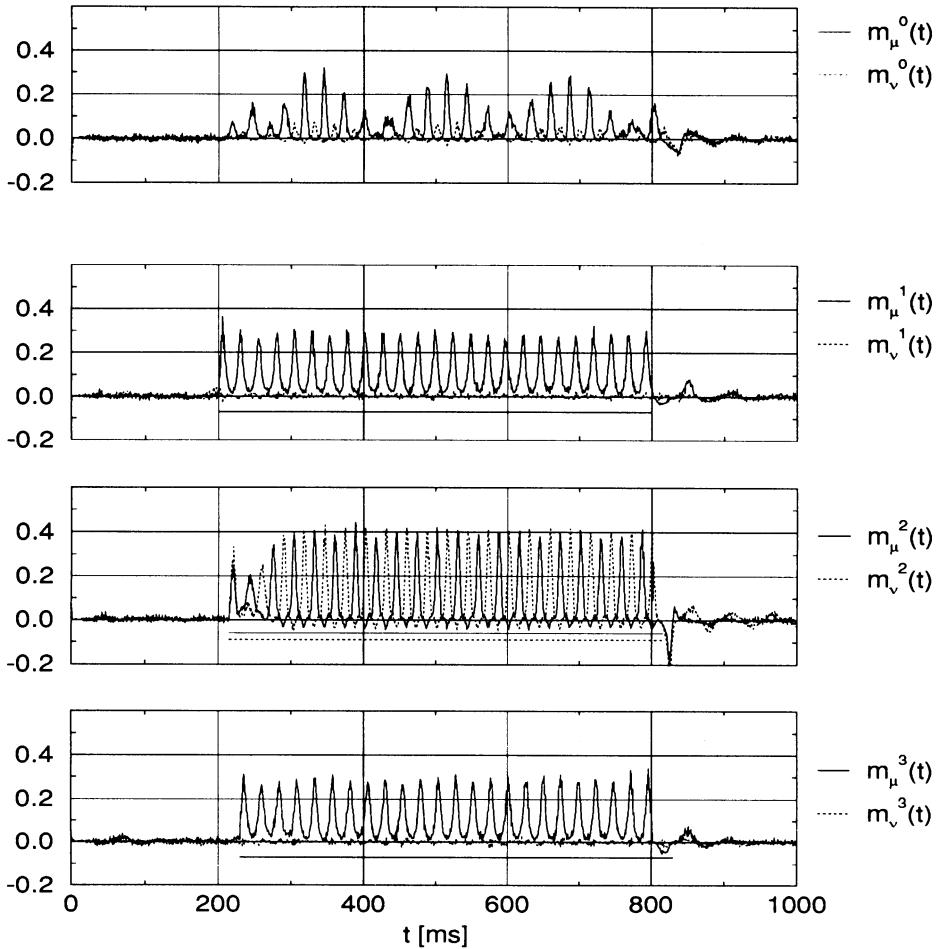
for  $1 \leq k \leq K$  and

$$h_{0i}^{\text{syn}} = g_0 \sum_{\mu=1}^q \xi_{0i}^\mu \bar{m}_\mu^0(t - \Delta_{0i}) + g_0 \sum_{\langle \mu, \mu' \rangle} \sum_{k=1}^K \xi_{0i}^\mu \bar{m}_{\mu'}^k(t - \Delta_{0i}^{\text{ar}}), \quad (5.39)$$

where we have introduced

$$\bar{m}_\mu^k(t) = \frac{2}{N(1-a^2)} \sum_{\tau=0}^{\infty} \epsilon(\tau) \sum_{j=1}^N (\xi_{kj}^\mu - a) S_{kj}(t - \tau), \quad 0 \leq k \leq K. \quad (5.40)$$

To demonstrate the capabilities of this network we consider the following task. The lower layer gets patterns as input that correspond to a learned



**Fig. 5.15.** Without feedback the binding task cannot be solved. The lower layer is stimulated by a learned combination of patterns but column  $k = 2$  also gets an additional pattern as input. The onset of the stimulation varies from column to column. We plot the time evolution of the overlaps with the stimulated patterns for columns 1 to 3 (lower three parts). On top, the overlap with the associated pattern in the upper layer (column  $k = 0$ ) is shown; note the beat. Without feedback the activities in the lower layer do not synchronize. Thus the network is not able to solve the binding task.

combination in the upper layer but column 2 gets an additional pattern as input so that it is confronted with a superposition of two patterns. The onset of the stimulus is taken to vary for the different columns. Due to the weak locking scenario which we have chosen, this results in oscillating overlaps with the presented patterns. In column 2 the patterns are separated through a phase shift in the activity of the corresponding neurons. If we do not allow any feedback, the oscillations in the different columns will not synchronize and the oscillation frequency in column 2 will be slightly lower due to the two concurrent patterns. This results in a beat in the time evolution of the overlap with the associated pattern in the upper layer (Fig. 5.15).

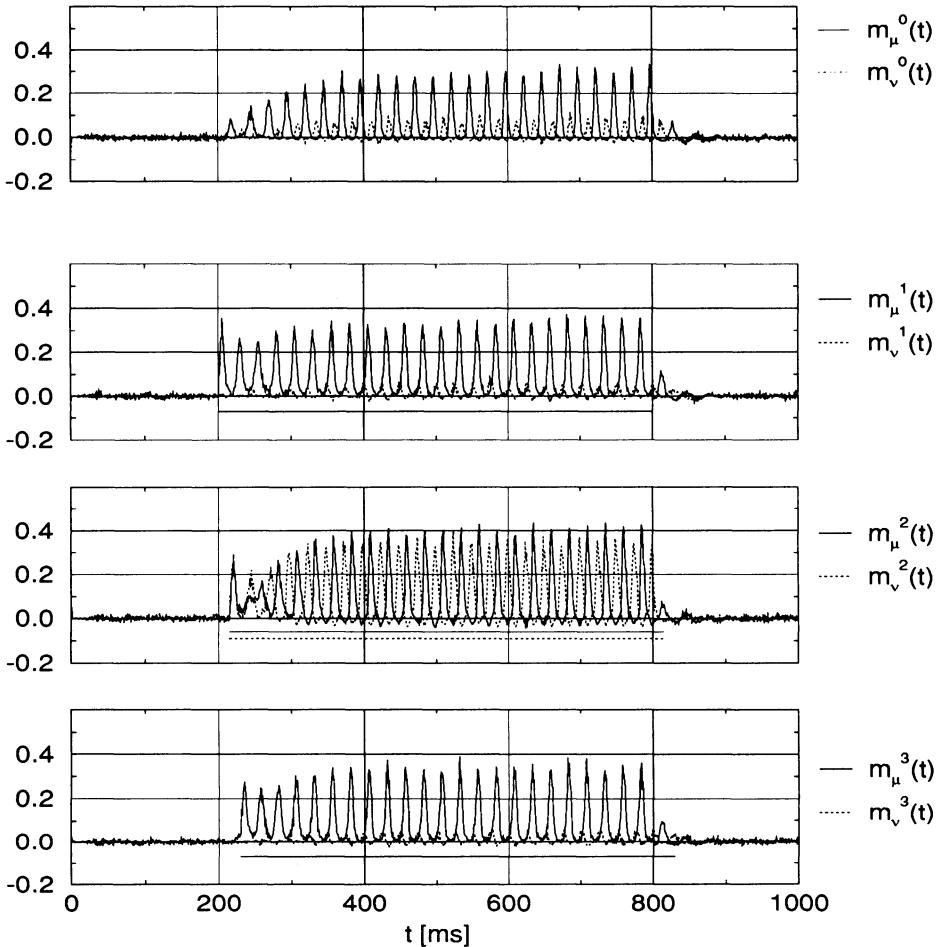
If we allow *feedback*, however, the situation changes. Now the activity of the recalled *combination* of patterns synchronizes throughout the lower layer, while the overlap with the additional pattern that is stimulated in column 2 oscillates phase-shifted by  $\pi$  with respect to the others. In the upper layer the overlap oscillates with stationary amplitude (Fig. 5.16).

To get a better idea of how this can be interpreted and used for pattern recognition we finally consider the same task as before but with a changing stimulus. In Fig. 5.17 we plot the stimuli and selected states of the network. For convenience only, the patterns have been taken to represent “letters,” so one may identify the learned combinations of patterns with “words.” In column 2 an incomplete superposition of A and H is given as stimulus. Depending on the stimulus in the other columns the H can be completed and synchronized with the T and the E to represent the learned word THE while the A is still present but phase-shifted by  $\pi$  with respect to the other letters. On the other hand, if the stimulus in columns 1 and 3 is changed to C and T (in the very same simulation run), the network now combines the A with the other patterns to represent the learned word CAT, while the H is activated with some delay.

So it is fair to say that the network is able to do context sensitive binding in a way that the very same stimulus (here for column 2) is interpreted in two different ways, depending on the stimulation of the other columns. In addition, this gives an idea what feedback connections that are found almost everywhere in the cortex may be good for.

## 5.7 Discussion

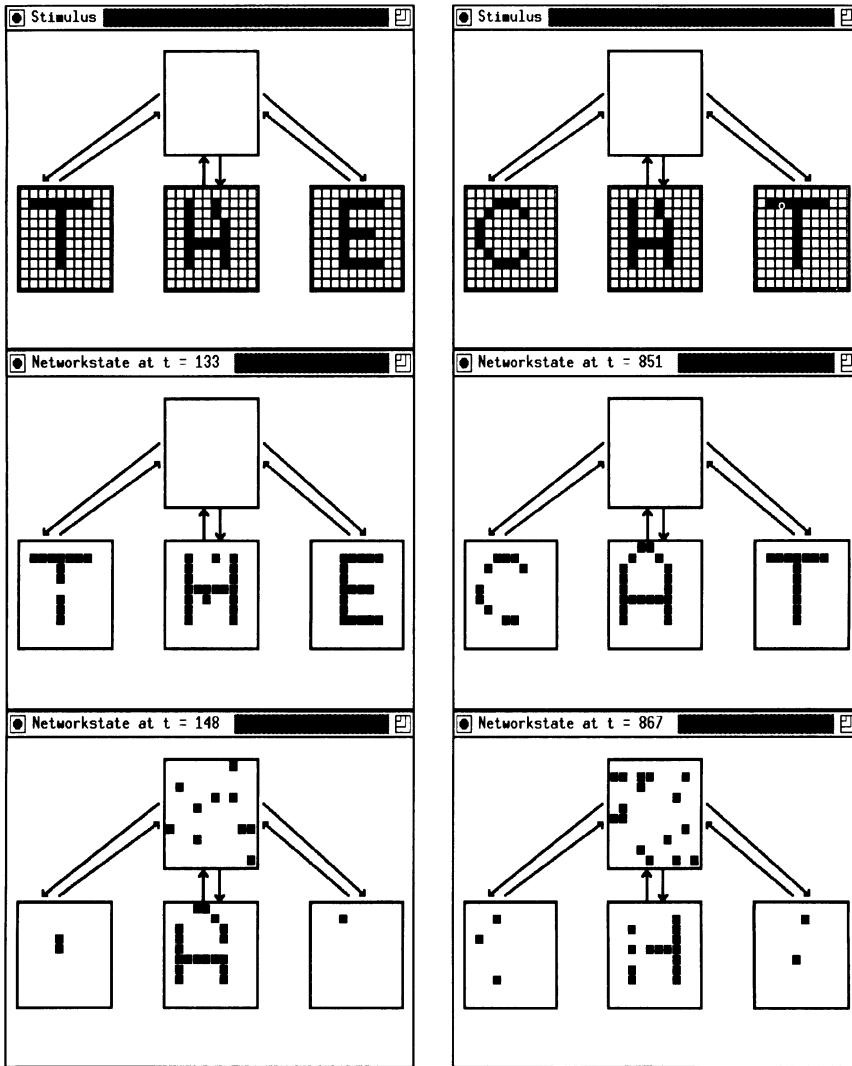
We have incorporated a couple of often neglected biological details into formal associative networks and showed their relevance. In particular, we have included an absolute refractory period of the model neuron, a distribution of axonal transmission delays, and, finally, a realistically shaped postsynaptic response (EPSP and IPSP). Our biological approach allows us to model network behavior in more temporal detail than usual and reveals



**Fig. 5.16.** Including feedback the network performs binding. We consider the same task as before (cf. Fig. 5.15) but we now allow feedback from the upper to lower layer. This results in a synchronization of the corresponding overlaps in the lower layer. An oscillation with stationary amplitude arises in the upper layer in contrast to the beat seen in Fig. 5.15.

the spiking behavior of single neurons and correlations between different neurons as well as the global activities.

We have shown that apart from the quiescent state at least two basic types of solution are possible: stationary and oscillatory retrieval of the learned patterns. It depends on the exact timing of the EPSP in relation to the IPSP whether the system relaxes to an oscillatory or stationary retrieval state. In the *stationary* state, all neurons in the foreground of a pattern are active, but the activity bursts of different neurons are not synchronized. The overlap — which is a global measure of the retrieval



**Fig. 5.17. Illustration of context sensitive binding.** Column 2 of our network is confronted with a superposition of the two letters “A” and “H.” This results in synchronous firing of all neurons representing the same letter but phase-shifted with respect to the neurons representing the other letter. When the “T” and the “E” is stimulated in the other columns, then the H synchronizes with them to represent “THE” (left-hand side), while in case of stimulating “C” and T in columns 1 and 3 the A synchronizes thus recalling “CAT” (right-hand side). The notions THE and CAT were stored in the upper layer and the very same stimulus (for column 2) is interpreted in two different ways, depending on the stimulation of the other columns. Therefore we call this response *context-sensitive binding*.

quality of a pattern — takes a constant value proportional to the mean firing rate of the neurons. Stationary retrieval occurs for medium delays in the axonal transmission (scenario II).

For short (scenario I) or long (scenario III) transmission delays *oscillatory* retrieval is dominant. The period of collective oscillations can be calculated if transmission delays and the shape of EPSP and IPSP are known. An important conclusion of our analysis is that locking of different neurons into a collective oscillation is possible *only if the total excitatory potential is still growing at the time of firing*. If it is declining, the collective oscillation is unstable and the neurons go into an unsynchronized firing state. In view of the *distribution* of axonal and inhibitory delays the occurrence of locking might seem surprising. We could show both analytically and in the simulations of scenario III that a collective oscillation with a common period is nevertheless possible. Our analysis also reveals that neurons with shorter delays spike always slightly earlier than those with longer delays.

In scenario I where the neurons have axonal delay times in a biologically plausible range, coherent oscillations occur *but locking is only weak* in that neurons with short delays occasionally slip through and miss the collective firing times. It can thus be speculated that biological networks operate in a regime intermediate between strict locking and unsynchronized firing — in agreement with available biological data [16,50]. This agreement is even better once one realizes that a prediction of the weak locking scenario [24], viz., a tripartition of the neurons in well locked, weakly locked, and not locked at all, has been confirmed a posteriori by Eckhorn and Obermueller [17].

Locking as well as the phase shift between fast and slow units are a much more general phenomenon and occur in many other circumstances [52,68,83]. But whereas most other models of collective oscillations are based on abstract differential equations, our approach relies on a specific model of neuronal signal transmission. Despite the fact that many biological phenomena have been included in the description, an analytical solution of our model in the limit of a large and noiseless network is nevertheless possible.

There are, of course, a couple of biological details which are still neglected in our approach. As we have explained in Sec. 5.2.1, we disregard all microscopic phenomena on the level of ion channels and the structure of dendritic trees. Furthermore, we assume throughout the paper that dendritic summation is linear and we neglect all effects of adaptation and relative refractoriness. It should be checked in future work how an incorporation of the above phenomena changes the network behavior. Preliminary investigations suggest that it does not.

In our model of synaptic transmission, we have assumed that a mathematical description can be separated into two parts: A response function which models the time course of the synaptic response and a synaptic efficacy factor which determines the amplitude of the response. Experi-

ments on hippocampal brain slices suggest that such a separation might be possible. Indeed, it is the amplitude of the postsynaptic response which changes during long-term potentiation while the time course remains unaffected (e.g., [55,57]). We have made no attempt to model the electrical and chemical processes that lead to long-term potentiation by correlation of pre- and postsynaptic activity ([32]; see [8–10] for a review of experiments [38,48,56,59] on modeling). Instead, we have taken an effective Ansatz (1.4) of Hebbian efficacies for random patterns. This allows us to compare our results with the retrieval properties of traditional formal networks [41]. We do not claim, however, that the visual cortex is an associative network that stores random patterns by the above rule (5.4). In fact, this rule is rather implausible, since the pyramidal model neurons make excitatory as well as inhibitory connections. In principle, the inhibitory synapses in the pyramidal layer can be removed if the threshold is adjusted at the same time. For technical reasons, however, we prefer to take the rule (5.4), which makes analysis more transparent.

On a more global level, the connection topology of our model may be criticized. The “extreme locality” of inhibition by partner neurons is certainly not a realistic model of biological systems. In the visual cortex of the cat, for example, it is known that only a small (20%) fraction of the neurons are inhibitory stellate cells and each cell makes local connections to a group of excitatory cells. Simulations of a more realistic model that preserves the characteristics of the present one but allows a reduced number of inhibitory neurons and local connections to a group of nearby pyramidal neurons show that the basic network behavior is unchanged [77]. In fact, it can be argued that the relaxed locality condition tends to synchronize firing and to stabilize oscillations. Our approach of extreme locality is thus not only the most transparent, but also the most difficult, in a sense the “worst” case, if we aim at collective oscillations. Regarding the pyramidal neurons, we have taken the other extreme and assumed full connectivity. This would be a bad assumption, if we wanted to model the cortex as a whole, but as a model of a single column or a very small area in the visual cortex of a cat or monkey it is a fair approximation.

In addition, the use of fixed couplings may be subject to criticism because it leads to the fact that only those patterns can be recalled and separated that have been learned before. This is true, however, for any kind of associative memory, independent of other details. It is rather unlikely that association takes place in the input layers of the primary visual cortex. On the other hand, binding and segregation should also be accomplished in higher areas and there they could be achieved simultaneously with association — as is shown in our model. Note that we do not need any change in our couplings to perform the segmentation task — in contrast to the dynamic link architecture used by Sporns et al. [74] and von der Malsburg and Buhmann [78].

As to the invariance problem, we did not make any effort to solve this

in our network as it boils down to suitable preprocessing. Neven and Aertsen [61] have shown how a solution of the binding problem based on an oscillatory network of the Kuramoto type [52] can be integrated into a hierarchical network that also deals with the problem of translation invariance. In a hierarchically organized system, where invariance is achieved in the first step, our network could solve the association and binding task in the second step. Thus it could replace layers 2 and 3 of Neven and Aertsen by a single layer.

In many respects our approach to pattern segmentation is similar to the model proposed by Horn and Usher [43,44,46]. These authors, however, only solve the macroscopic equations for the overlaps. Thus they lose all microscopic information about the underlying dynamics of single neurons which makes it almost impossible to compare their results with the experiments in detail. Our approach, in contrast, allows one to extract spike rasters similar to experimental multielectrode recordings. This allows, in principle, a direct comparison of our results with experimental data.

## 5.8 Conclusions

Incorporating a *spiking* model neuron and a *distribution* of transmission delays into an associative neural network opens up the way to a realization of both feature linking and pattern segmentation in the very same network. The number of patterns that can be segregated is in principle arbitrary and bounded only by the inverse activity of the patterns on the one hand and strength and duration of the local inhibition on the other hand. A biologically realistic set of parameters leads to the result that *up to four* patterns can be separated in a biological network simultaneously. The even harder task of context sensitive binding can be solved in a layered structure provided feedback connections are present.

*Acknowledgments.* The authors thank N.S. Sutherland (Brighton) for drawing their attention to the beautiful work of Sperling [71]. The present work has been supported by the Deutsche Forschungsgemeinschaft (Grant number HE 1729/2-1). R.R. holds a scholarship of the Freistaat Bayern.

## REFERENCES

- [1] Abbott LF (1991) Realistic synaptic inputs for model neural networks. *Network* 2:245–258
- [2] Amit DJ (1989) *Modeling Brain Function* (Cambridge University Press, Cambridge)

- [3] Amit DJ, Gutfreund H, Sompolinsky H (1985) Spin-glass models of neural networks. *Phys. Rev. A* **32**:1007–1032
- [4] Amit DJ, Gutfreund H, Sompolinsky H (1987) Statistical mechanics of neural networks near saturation. *Ann. Phys. (NY)* **173**:30–67
- [5] Baird B (1986) Nonlinear dynamics of pattern formation and pattern recognition in the rabbit olfactory bulb. *Physica D* **22**:150–175
- [6] Bauer HU, Pawelzik K, Geisel T (1993) Emergence of transient oscillations in an ensemble of neurons. In: *Proceedings of the ICANN'93*, S. Gielen and B. Kappen (Eds.) (Springer, Berlin), pp. 136–141
- [7] Bauer HU, Pawelzik K (1993) Alternating oscillatory and stochastic dynamics in a model for a neuronal assembly. *Physica D*, to appear
- [8] Bindman L, Christofi G, Murphy K, Nowicky A (1991) Long-term potentiation (LTP) and depression (LTD) in the neocortex and hippocampus: An overview. In: *Aspects of synaptic transmission*, T.W. Stone (Ed.) (Taylor Francis, London), Vol. 1, pp. 3–25
- [9] Brown TH (1994) Hebbian synaptic plasticity. This volume
- [10] Brown TH, Ganong AH, Kairiss EW, Keenan CL, Kelso SR (1989) Long-term potentiation in two synaptic systems of the hippocampal brain slice. In: *Neural Models of Plasticity*, J.H. Byrne and W.O. Berry (Eds.) (Academic Press, San Diego), pp. 266–306
- [11] Buhmann J (1989) Oscillations and low firing rates in associative memory neural networks. *Phys. Rev. A* **40**:4145–4148
- [12] Buhmann J, Schulten K (1986) Associative recognition and storage in a model network with physiological neurons. *Biol. Cybern.* **54**:319–335
- [13] Bush PC, Douglas RJ (1991) Synchronization of bursting action potential discharge in a model network of neocortical neurons. *Neural Comput.* **3**:19–30
- [14] Deppisch J, Bauer HU, Schillen T, König P, Pawelzik K, Geisel T (1993) Alternating oscillatory and stochastic states in a network of spiking neurons. *Network* **4**:243–257
- [15] Domany E, van Hemmen JL, Schulten K, Eds. (1991) *Models of Neural Networks* (Springer, Berlin)
- [16] Eckhorn R, Bauer R, Jordan W, Brosch M, Kruse W, Munk M, Reitboeck HJ (1988) Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biol. Cybern.* **60**:121–130
- [17] Eckhorn R, Obermueller A (1993) Single neurons are differently involved in stimulus-specific oscillations in cat visual cortex. *Exp. Brain Res.* **95**:177–182
- [18] Eckhorn R, Reitboeck HJ, Arndt M, Dicke P (1990) Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex. *Neural Comput.* **2**:293–307

- [19] Ekeberg O, Wallen P, Lansner A, Traven H, Brodin L, Grillner S (1991) A computer based model for realistic simulations of neural networks. *Biol. Cybern.* **65**:81–90
- [20] Engel AK, König P, Kreiter K, Singer W (1991) Interhemispheric synchronization of oscillatory neural responses in cat visual cortex. *Science* **252**:1177–1179
- [21] Engel AK, König P, Singer W (1991) Direct physiological evidence for scene segmentation by temporal coding. *Proc. Natl. Acad. Sci. U.S.A.* **88**:9136–9140
- [22] Felleman DJ, van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* **1**:1–47
- [23] Fuentes U (1993) Einfluß der Schicht- und Arealstruktur auf die Informationsverarbeitung im Cortex. Master's thesis, Technische Universität München
- [24] Gerstner W, Ritz R, van Hemmen JL (1993) A biologically motivated and analytically soluble model of collective oscillations in the cortex I. Theory of weak locking. *Biol. Cybern.* **68**:363–374
- [25] Gerstner W, Ritz R, van Hemmen JL (1993) Why spikes? Hebbian learning and retrieval of time resolved excitation patterns. *Biol. Cybern.* **69**:503–515
- [26] Gerstner W, van Hemmen JL (1992) Associative memory in a network of “spiking” neurons. *Network* **3**:139–164
- [27] Gerstner W, van Hemmen JL (1992) Universality in neural networks: The importance of the mean firing rate. *Biol. Cybern.* **67**:195–205
- [28] Gerstner W, van Hemmen JL (1993) Coherence and incoherence in a globally coupled ensemble of pulse emitting units. *Phys. Rev. Lett.* **71**:312–315
- [29] Gray CM, König P, Engel AK, Singer W (1989) Oscillatory responses in cat visual cortex exhibit intercolumnar synchronization which reflects global stimulus properties. *Nature* **338**:334–337
- [30] Gray CM, Singer W (1989) Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* **86**:1698–1702
- [31] Hansel D, Sompolinski H (1992) Synchronization and computation in a chaotic neural network. *Phys. Rev. Lett.* **68**:718–721
- [32] Hebb DO (1949) *The Organization of Behavior* (Wiley, New York)
- [33] van Hemmen JL, Gerstner W, Herz AVM, Kühn R, Sulzer B, Vaas M (1990) Encoding and decoding of patterns which are correlated in space and time. In: *Konnektionismus in Artificial Intelligence und Kognitionsforschung*, G. Dorffner (Ed.) (Springer, Berlin), pp. 153–162
- [34] van Hemmen JL, Gerstner W, Ritz R (1992) A “microscopic” model of collective oscillations in the cortex. In: *Neural Network Dynamics*, J.G. Taylor et al. (Eds.) (Springer, Berlin), pp. 250–257
- [35] van Hemmen JL, Grensing D, Huber A, Kühn R (1986) Elementary solution of classical spin glass models. *Z. Phys. B* **65**:53–63

- [36] van Hemmen JL, Grensing D, Huber A, Kühn R (1988) Nonlinear neural networks I and II. *J. Stat. Phys.* **50**:231–257 and 259–293
- [37] van Hemmen JL, Kühn R (1986) Nonlinear neural networks. *Phys. Rev. Lett.* **57**:913–916
- [38] Herz AVM, Sulzer B, Kühn R, van Hemmen JL (1988) The Hebb rule: Representation of static and dynamic objects in neural nets. *Europhys. Lett.* **7**:663–669
- [39] Herz AVM, Sulzer B, Kühn R, van Hemmen JL (1989) Hebbian learning reconsidered: Representation of static and dynamic objects in associative neural nets. *Biol. Cybern.* **60**:457–467
- [40] Hodgkin AL, Huxley AF (1952) A quantitative description of ion currents and its applications to conduction and excitation in nerve membranes. *J. Physiol. (London)* **117**:500–544
- [41] Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci. U.S.A.* **79**:2554–2558
- [42] Hopfield JJ (1984) Neurons with graded response have computational properties like those of two-state neurons. *Proc. Natl. Acad. Sci. U.S.A.* **81**:3088–3092
- [43] Horn D, Sagi D (1991) Parallel activation of memories in an oscillatory neural network. *Neural Comput.* **3**:31–43
- [44] Horn D, Sagi D, Usher M (1991) Segmentation, binding, and illusory conjunctions. *Neural Comput.* **3**:510–525
- [45] Horn D, Usher M (1989) Neural networks with dynamical thresholds. *Phys. Rev. A* **40**:1036–1040
- [46] Horn D, Usher M (1992) Oscillatory model of short-term memory. In: *Advances in Neural Information Processing Systems 4*, J.E. Moody et al. (Eds.) (Morgan Kaufmann, San Mateo, CA), pp. 125–132
- [47] Jack JJB, Noble D, Tsien RW (1975) *Electric Current Flow in Excitable Cells* (Clarendon Press, Oxford)
- [48] Kitajima T, Hara K (1990) A model of the mechanisms of long-term potentiation in the hippocampus. *Biol. Cybern.* **64**:33–39
- [49] König P, Schillen TB (1991) Stimulus-dependent assembly formation of oscillatory responses. I. Synchronization. *Neural Comput.* **3**:155–166
- [50] Kreiter AK, Singer W (1992) Oscillatory neuronal responses in the visual cortex of the awake macaque monkey. *Eur. J. Neurosci.* **4**:369–375
- [51] Kuffler SW, Nicholls JG, Martin AR (1984) *From Neuron to Brain*, 2nd Ed. (Sinauer, Sunderland, Mass.)
- [52] Kuramoto Y (1984) *Chemical Oscillations, Waves, and Turbulence* (Springer, Berlin), pp. 68–77

- [53] Kurrer C, Nieswand B, Schulten K (1990) A model for synchronous activity in the visual cortex. In: *Self-Organization, Emerging Properties, and Learning*, A. Babloyantz (Ed.) (Plenum Press), pp. 81–85
- [54] Lamperti J (1966) *Probability* (Benjamin, New York), Chap. 7
- [55] Larson J, Lynch G (1986) Induction of synaptic potentiation in hippocampus by patterned stimulation involves two events. *Science* **232**:985–988
- [56] Lisman J (1989) A mechanism for Hebb and anti-Hebb processes underlying learning and memory. *Proc. Natl. Acad. Sci. U.S.A.* **86**:9574–9578
- [57] Malinow R, Miller JP (1986) Synaptic hyperpolarization during conditioning reversibly blocks induction of long-term potentiation. *Nature* **320**:529–530
- [58] McCormick DA (1990) Membrane properties and neurotransmitter actions. In: *The Synaptic Organization of the Brain*, G.M. Sheperd (Ed.) (Oxford University Press)
- [59] Miller KD (1990) Derivation of linear Hebbian equations from a nonlinear mode of synaptic plasticity. *Neural Comput.* **2**:321–333
- [60] Murthy VN, Fetz EE (1992) Coherent 25 to 35 Hz oscillations in the sensorimotor cortex of awake behaving monkeys. *Proc. Natl. Acad. Sci. U.S.A.* **89**:5670–5674
- [61] Neven H, Aertsen A (1992) Rate coherence and event coherence in the visual cortex: A neural model of object recognition. *Biol. Cybern.* **67**:309–322
- [62] Pawelzik K (1994) Detecting coherence in neuronal data. This volume
- [63] Pawelzik K (1991) *Nichtlineare Dynamik und Hirnaktivität* (Verlag Harri Deutsch, Frankfurt)
- [64] Rall W (1964) Theoretical significance of dendritic trees for neuronal input-output relations. In: *Neural Theory and Modeling*, R.F. Reiss (Ed.) (Stanford University Press), pp. 73–97
- [65] Schillen TB, König P (1991) Stimulus-dependent assembly formation of oscillatory responses II. Desynchronization. *Neural Comput.* **3**:167–178
- [66] Schuster HG, Wagner P (1990) A model for neuronal oscillations in the visual cortex. 1. Mean-field theory and derivation of the phase equations. *Biol. Cybern.* **64**:77–82
- [67] Schuster HG, Wagner P (1990) A model for neuronal oscillations in the visual cortex. 2. Phase description and feature-dependent synchronization. *Biol. Cybern.* **64**:83–85
- [68] Shiino M, Frankowicz M (1989) Synchronization of infinitely many coupled limit cycle oscillators. *Phys. Lett. A* **136**:103–108
- [69] Singer W (1994) The role of synchrony in neocortical processing and synaptic plasticity. This volume
- [70] Sompolinsky H, Golomb D, Kleinfeld D (1990) Global processing of visual stimuli in a neural network of coupled oscillators. *Proc. Natl. Acad. Sci. U.S.A.* **87**:7200–7204

- [71] Sperling G (1960) The information available in brief visual presentations. *Psychol. Monogr.* **74** (11 Whole No. 498):1–29
- [72] Sporns O, Tononi G, Edelman GM (1994) Reentry and dynamical interactions of cortical networks. This volume
- [73] Sporns O, Gally JA, Reeke GN, Edelman GM (1989) Reentrant signaling among simulated neuronal groups leads to coherency in their oscillatory activity. *Proc. Natl. Acad. Sci. U.S.A.* **86**:7265–7269
- [74] Sporns O, Tononi G, Edelman GM (1991) Modeling perceptual grouping and figure-ground segregation by means of active reentrant connections. *Proc. Natl. Acad. Sci. U.S.A.* **88**:129–133
- [75] Sutherland S (1991) Only four possible solutions. *Nature* **353**:389–390
- [76] Sutton JP, Beis JS, Trainor LEH (1988) Hierarchical model of memory and memory loss. *J. Phys. A* **21**:4443–4454
- [77] Trefz T (1991) Oszillationen im Cortex, Diplomarbeit, Technische Universität, München
- [78] von der Malsburg C, Buhmann J (1992) Sensory segmentation with coupled neural oscillators. *Biol. Cybern.* **67**:233–242
- [79] von der Malsburg C, Schneider W (1986) A neural cocktail party processor. *Biol. Cybern.* **54**:29–40
- [80] Wang D, Buhmann J, von der Malsburg C (1990) Pattern segmentation in associative memory. *Neural Comput.* **2**:94–106
- [81] Wilson AM, Bower JM (1991) A computer simulation of oscillatory behavior in primary visual cortex. *Neural Comput.* **3**:498–509
- [82] Wilson HR, Cowan JD (1972) Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J.* **12**:1–24
- [83] Yamaguchi Y, Shimizu H (1984) Theory of self-synchronization in the presence of native frequency distribution and external noises. *Physica D* **11**:212–226
- [84] Zeki S, Shipp S (1988) The functional logic of cortical connections. *Nature* **335**:310–317

# 6

## Modeling the Sensory Computations of the Olfactory Bulb

Zhaoping Li<sup>1</sup>

with 5 figures

**Synopsis.** Two models are presented as examples to discuss the computational tasks, sensory environments, and computational mechanisms in early olfactory processing. One model proposes that the odor inputs during a sniff cycle are detected and coded by the emergence of odor-dependent and coherent neural oscillatory patterns observed in the mammalian bulb, and that the odor mixtures are segmented by the selective olfactory adaptation to the preexisting and detected odors. The other model argues that the early olfactory processing is to separate the odor mixtures in the input environment into individual odor sources before the odor identities are known, and proposes a fast synaptic learning algorithm that uses the temporal fluctuation structures in the receptor cells for source separation.

### 6.1 Introduction

The olfactory system is a phylogenetically primitive part of the cerebral cortex [1]. In lower vertebrates, the olfactory system is the largest part of the telencephalon. It also has an intrinsically simple cortical structure, which in modified form is used in other parts of the brain [1]. The odor molecules of the distal objects to be detected are bound to and crudely recognized by receptor proteins, giving a relatively well defined input signal, compared to the photoreceptor and hair cell activities in vision and audition, for object identities. Hence the olfactory system conceivably handles a relatively simpler computational problem. Having phylogenetic importance and computational simplicity, the olfactory system promises to yield insight on the principles of sensory information processing.

---

<sup>1</sup>The Rockefeller University, 1230 York Avenue, New York, NY 10021.

What is the olfactory system trying to compute from its sensory input? It should try to perform some or all of the following tasks: (1) odor detection; (2) if there is only one odor present, odor identification and concentration estimation [3]; (3) if there are multiple odors present, odor segmentation — before identification and concentration estimation can be performed [2,4]; (4) localization of the odor sources [2]. Tasks 1 and 2 should be performed by any olfactory system. Task 3 is often necessary for odor recognition and localization since the olfactory environment is frequently composed of odor mixtures, and any olfactory model should address it. Task 4 is especially important and necessary for animals, such as insects and nocturnal animals, who depend largely on chemical senses in their environment [2,5,6].

The olfactory bulb, the first processing stage after the sensory receptors, is likely to compute towards the computational goals above. It transforms the input signals, represented by the spatio-temporal activation patterns in the receptor cells, to input dependent ([7–10]) spatio-temporal oscillatory activities at the output, which are then sent directly to the olfactory cortex for mammals (Fig 6.1.)[1]. Hence the bulbar output carries the relevant information about the input odors. The models described in this paper assume that some of the computational goals above are first attempted by the processings in the bulb. The computation in the olfactory cortex cannot be understood without knowing the bulbar processing and outputs.

This article demonstrates the olfactory modeling approach by comprehensive description of two models of the olfactory bulb. One of them [3,4] suggests that the mammalian bulb detects and identifies an odor and its concentration by the amplitudes and phases of the oscillation across the bulb (tasks 1 and 2). Furthermore, it hypothesizes that the bulb detects and identifies a more recent odor input in the presence of another odor — odor segmentation (task 3) — by olfactory adaptation. It suggests that the adaptation should be viewed as a computational strategy instead of as olfactory fatigue. The odor localization task (task 4), however, is not addressed by this model. The other model [2] has a different perspective and emphasizes odor segmentation and localization (tasks 3 and 4), even when the odor objects are not familiar and many odors are simultaneously present. It hypothesizes that the bulb segments or identifies the individual odors by analyzing the correlations in the input temporal fluctuations. As a result, identities and fluctuations of individual odor components are identified from the input mixture. Each component fluctuation, which contains the information about the odor source location, can then be used by higher brain centers to localize the source. The problem is solved by dynamic changes of the synaptic strengths, which are proposed to encode the odor identity.

This article is thus not intended as a complete review of all the modeling and other relevant works in the field, which can be found in the literature [11]. Among other works are studies of odor transduction by receptors [12], detailed computer simulation using biological parameters to repro-

duce physiological observations [13,14], a model to study macroglomerulus neural response variations to behaviorally relevant input changes in insects [15], a model of efficient coding in the bulb [16], computational studies of olfactory cortex as a content-addressable memory and learning [17–19], and a discussion on nonlinear neural dynamical processing of odor information [20].

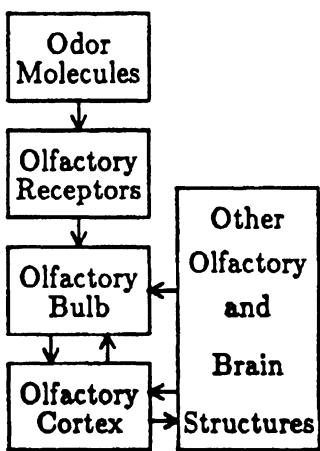
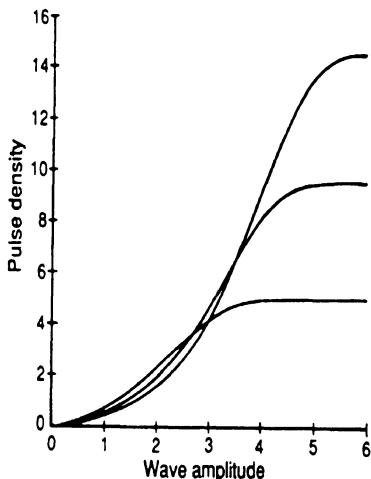
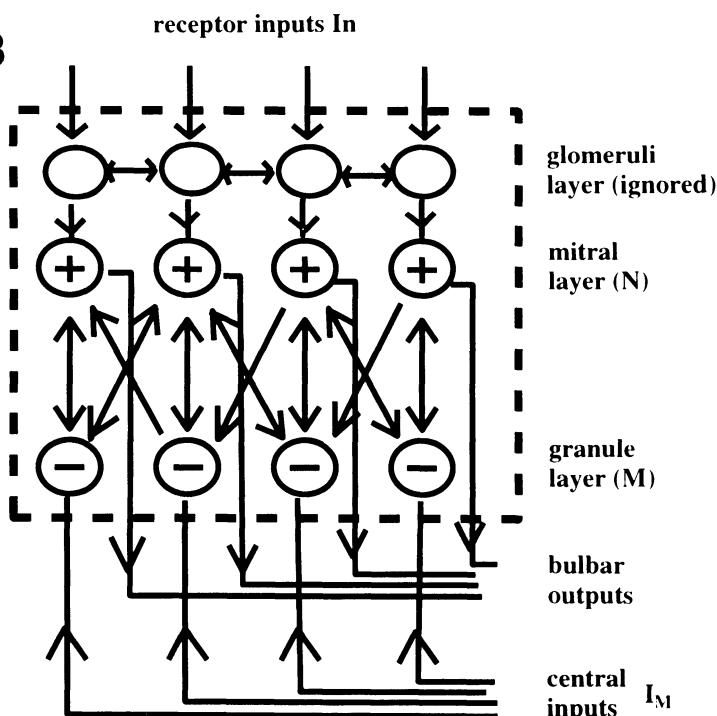
In the next section, some background is given on the olfactory bulb and its sensory receptor inputs. Section 6.3 presents a model of the neural activities observed in the mammalian bulb. Section 6.4 shows how this model codes and segments odors. Section 6.5 presents Hopfield's model [2] of odor identity and fluctuation segmentation in odor mixtures. In Sec. 6.6, we discuss the differences between these models, their relationships to the computational tasks and environments, and relate the computational approaches used in olfaction to those used in other sensory systems.

## 6.2 Anatomical and Physiological Background

The olfactory bulb contains sharply differentiated cell types located on different parallel lamina [1]. Each receptor sends a single unbranched axon to the topmost layer, terminating in one of the spherical regions of neuropil termed glomeruli (Figs. 6.1(A), (B)). The receptor axons ramify inside the glomeruli, synapsing on the excitatory mitral cells and the inhibitory short axon cells.

With increased odor concentration, the receptor cell firing rate in vertebrates [21] increases from the spontaneous background of 1–3 impulses/s, and may reach 10–60 impulses/s. With an odor pulse delivered to the mucosa, the receptor firing rate increases approximately linearly in time as long as the pulse is not too long, and then terminates quickly, after the odor pulse terminates [21]. More than 100 genes for odor receptors have been identified in mammals [22]. It is not known whether a single type or a few types are expressed in each receptor cell. Most receptor neurons in vertebrates respond in different degrees to a broad range of odor molecules, each response spectrum is unique [12]. Hence there is no clear specificity between odorants and receptor cells, except between pheromone molecules and a subgroup of specialist receptor neurons in insects [12]. There are also inputs from higher olfactory centers to the bulb, but little is known about them [1].

The bulbar cell organization is reviewed in detail by Shepherd [1]. The main cell types of the mammalian bulb are the excitatory mitral cells and the inhibitory granule cells at different cell layers (Fig. 6.1(B)). Each mitral cell receives inputs from one or several glomeruli [12]. The granule cells inhibit, and receive excitatory inputs from, the local mitral cells by local dendrodendritic synapses [1] activated by graded presynaptic depolarization. They also receive additional inputs from the mitral cell axon

**A****C****B**

**Fig. 6.1.** (A) Olfactory system block diagram. (B) Olfactory bulb structure, circles with "+" signs are excitatory mitral cells; with "-" signs are inhibitory granule cells. (C) Three examples of experimentally measured functions (taken from [20]), relating the pulse probability of single or small groups of mitral cells to the EEG wave amplitudes. This function is used to model the cell input-output relationships [20].

collaterals. The outputs of the bulb are carried by the mitral cell axons; while central feedbacks to the bulb are directed to the granule cells [1].

The activities in the glomeruli layer are odor dependent [9,23]. Stimulation with odors causes an onset of high-amplitude bulbar oscillations, which may be detected by surface EEG electrodes and which returns to low-amplitude oscillations upon the cessation of odor stimulus [7]. Since the odor stimulation is usually synchronized with respiration, the EEG [7,24] shows a high-amplitude oscillation arising during the inhalation and stopping early in the exhalation (Fig. 6.2(B)). However, respiration itself does not seem to have an effect on bulbar activities [25]. The oscillation is an intrinsic property of the bulb itself, persisting after central connections to the bulb are cut off [20,26]. However, Freeman and co-workers also reported that the central inputs [27] influence oscillation onset, which exists only in motivated animals. Nevertheless, oscillation disappears when the nasal air flow is blocked [24], although the same group reported that it can be present without odor inputs [7]. In invertebrates, the oscillation activities exist without odor inputs, but are modulated by odors [28]. The oscillation bursts have a peak frequency in the range of 35–90 Hz in mammals [24]. Different parts of the bulb have the same dominant frequency but different amplitudes and phases [8,24]. A specific odor input, subject to behavioral conditioning, sets a specific oscillation pattern [8,10].

There are roughly 1000 receptor axons and dendrites from 25 mitral cells in each glomerulus, while there are about 200 granule cells for each mitral cell [1]. A rabbit has about 50,000 mitral cells [1]. Both the mitral and granule cells have a nonlinear input–output relationship (Fig. 6.1(C), [27]), and a membrane time constant of 5–10 ms [29]. Very little is known about the strength of the synapses in the olfactory bulb.

### 6.3 Modeling the Neural Oscillations in the Olfactory Bulb

The fact that the oscillatory patterns in the bulb correlate with odor inputs and disappear when air flow is blocked indicates that the odor information is very likely carried in the neural oscillation activities. Before discussing how such activities code odors, we present a model to reproduce the physiologically observed oscillations. The computational issue of odor coding and segmentation will then become apparent by analyzing the dependence of oscillatory activities on the receptor and centrifugal inputs, as will be explained in the next section.

### 6.3.1 GENERAL MODEL STRUCTURE

This model [3] attempts to include enough physiological realism to contain the essential computational components and facilitate experimental comparison, yet, nevertheless retains sufficient simplicity to avoid superfluous details. Both  $M$ , the number of granule cells, and  $N$ , the number of mitral cells, as well as their ratio  $M/N$  are much reduced from reality in the simulations, although the mathematical analysis imposes no limit on their actual values. Excitation and inhibition are kept in balance by correspondingly increasing the strength of the granule cell (inhibitory) synapses.

Determining how to wire each receptor to the correct glomerulus and then to the mitral cells for optimal olfactory computation [12,16], is a major computational problem. However, this requires better experimental knowledge of receptor response spectrum to odors than is currently available. Our model thus ignores the glomeruli structure and regards receptor cells as effectively giving inputs  $I_i$  onto the  $i$ th mitral cell for  $1 \leq i \leq N$ . This input vector  $I$  is a superposition of a true odor signal  $I_{\text{odor}}$  and a background input  $I_{\text{background}}$ , that is,  $I = I_{\text{odor}} + I_{\text{background}}$ .  $I_{\text{odor}}$  is determined by odor pattern  $P_{\text{odor},i}$  for  $1 \leq i \leq N$ . We use the approximation, based on experimental findings [30], that for low odor intensity, the direction of vector  $P_{\text{odor}}$  depends only on the input odor identity, while its length or amplitude increases with the odor concentration.

Similarly, the central inputs to the granule cells are modeled as a vector  $I_c = I_{c,\text{background}} + I_{c,\text{control}}$ , where  $I_{c,\text{background}}$  is the background signal to the granule cells.  $I_{c,\text{control}}$  is the active central input that serves some computational purpose as will be discussed in the next section. For the moment, it is assumed that  $I_{c,\text{control}} = 0$ .

Each  $I_{\text{odor},i}$  is taken to be excitatory. We model the  $I_{\text{odor}}$  to increase in time during inhalation, as observed in experiment [21]. Because of the odor absorption by the lungs, as odors are necessarily highly soluble in water [31],  $I_{\text{odor}}$  is modeled to exponentially return toward the ambient during exhalation, see Fig. 6.2(A). Both  $I_{\text{background}}$  and  $I_{c,\text{background}}$  do not change during a sniff cycle, their scales being such that when  $I_{\text{odor}} = 0$ , most of the mitral and granule cells have their cell internal states just below maximum slope points on their input–output functional curves.

Each cell is modeled as one unit [3] under the assumption of short dendritic electrotonic lengths [1]. The state variables, modeling the membrane potentials, are, respectively,  $X = \{x_1, x_2, \dots, x_N\}$  and  $Y = \{y_1, y_2, \dots, y_M\}$  for mitral and granule cells. Their outputs are, respectively,  $G_x(X) = \{g_x(x_1), g_x(x_2), \dots, g_x(x_N)\}$  and  $G_y(Y) = \{g_y(y_1), g_y(y_2), \dots, g_y(y_M)\}$ . Both  $g_x$  and  $g_y$  model the probability of the cell firing or firing rate, their functional forms resemble their physiological correlate (Fig. 6.1(C)). They have a gain or slope that is approximately zero before the state when the cell response deviates from zero, modeling the firing threshold. This gain

becomes large immediately after the threshold. Such nonlinearity is essential for the proposed odor computations.

The dynamic equations for the cell population is modeled as

$$\begin{aligned} dX/dt &= -H_o G_y(Y) - \alpha X + I, \\ dY/dt &= W_o G_x(X) - \alpha Y + I_c, \end{aligned} \quad (6.1)$$

where  $\alpha = 1/\tau$ , and  $\tau = 7$  ms is the cell time constant, assumed to be the same for both cell populations for simplicity. Weak random noise with a correlation time of roughly 10 ms is added to  $I$  and  $I_c$  to simulate the fluctuations in cell potentials. The matrices  $H_o$  and  $W_o$  have non-negative elements and model the synaptic connections. Hence  $(H_o)_{ij} g_y(y_j)$  models the inhibition from the  $j$ th granule to the  $i$ th mitral cell, and  $(W_o)_{ji} g_x(x_i)$  is the reciprocal excitation. The cell indices  $i$  and  $j$  approximate cell locations in the bulb (assuming periodic boundary conditions: the  $N$ th mitral and the  $M$ th granule cells are next to the first mitral and granule cells, respectively). Local interaction implies that  $(H_o)_{ij}$  and  $(W_o)_{ji}$  are nonzero only when  $i \approx (N/M)j$ , that is,  $H_o$  and  $W_o$  are near-diagonal matrices for  $N = M$ .

### 6.3.2 THE OLFACTORY BULB AS A GROUP OF COUPLED NONLINEAR OSCILLATORS

Computer simulations demonstrate that our model captures the major known effects of the real bulb. In the example of 10 mitral and 10 granule cells, the model bulb exhibits the following. (1) Neural oscillations rise with inhalations and fall with exhalations; (2) all cells oscillate coherently with the same frequency which may depend on inputs; and (3) different odor inputs induce different amplitude and phase patterns in cells, and some, including the zero-odor input, do not induce any coherent oscillations. All of the above agree qualitatively with what are observed in physiology, as demonstrated in Fig. 6.2 ([7,24]). Furthermore, observation (3) demonstrates the model's capabilities as a pattern classifier.

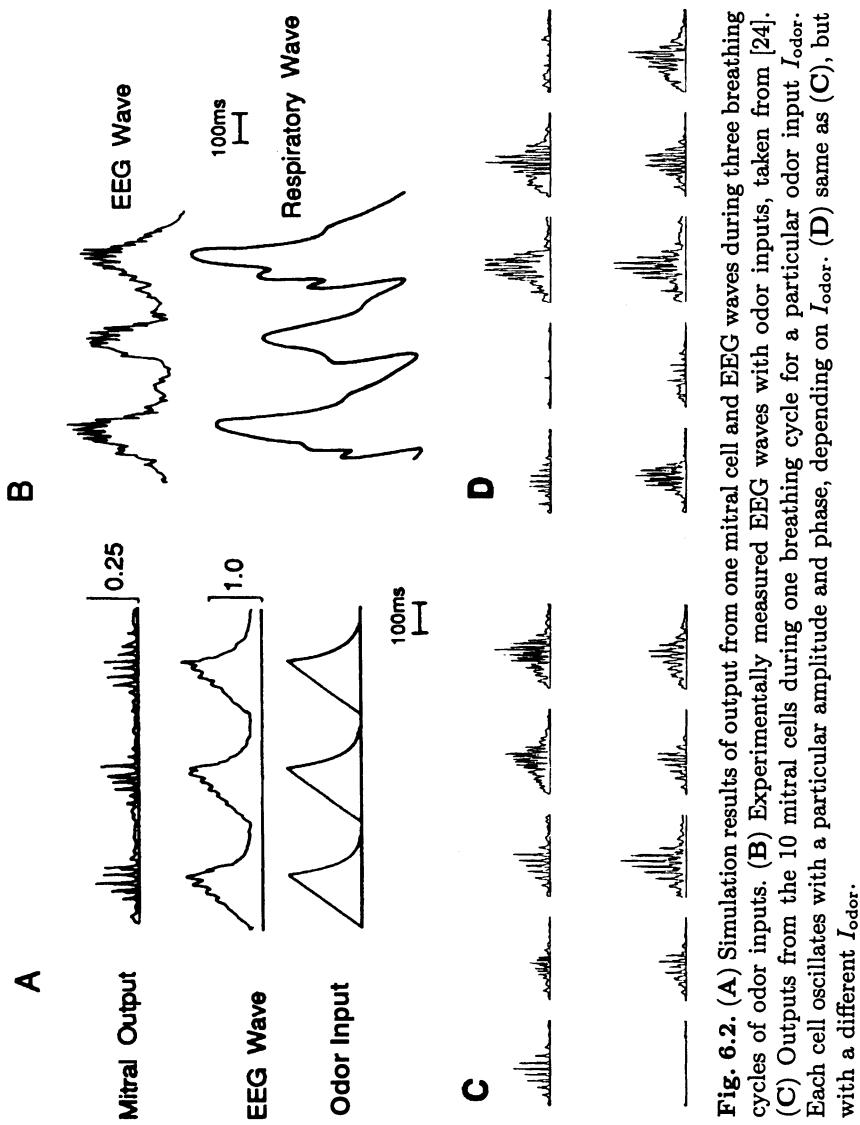
Such agreement between the model and real bulbs can be understood using the following analysis. First, concentrating on the oscillations, notice that a system such as

$$\begin{aligned} dx/dt &= -\omega y - \alpha x \quad \text{or} \quad d^2x/dt^2 + 2\alpha dx/dt + (\omega^2 + \alpha^2)x = 0 \\ dy/dt &= \omega x - \alpha y \end{aligned} \quad (6.2)$$

describes a damped oscillator of frequency  $\omega$  oscillating with time  $t$  as

$$x(t) = r_o e^{-\alpha t} \sin(\omega t + \phi),$$

where  $\alpha$  is the dissipation constant and  $r_o$ ,  $\phi$  the initial amplitude and phase. When  $\alpha = 0$ , the oscillator trajectory in the  $x-y$  space is a circle with radius  $r_o$ . This system resembles a coupled pair of mitral and granule



**Fig. 6.2.** (A) Simulation results of output from one mitral cell and EEG waves during three breathing cycles of odor inputs. (B) Experimentally measured EEG waves with odor inputs, taken from [24]. (C) Outputs from the 10 mitral cells during one breathing cycle for a particular odor input  $I_{\text{odor}}$ . Each cell oscillates with a particular amplitude and phase, depending on  $I_{\text{odor}}$ . (D) same as (C), but with a different  $I_{\text{odor}}$ .

cells, with external inputs  $i(t)$  and  $i_c(t)$ , respectively,

$$\begin{aligned} \frac{dx}{dt} &= -h \cdot g_y(y) - \alpha x + i(t), \\ \frac{dy}{dt} &= w \cdot g_x(x) - \alpha y + i_c(t). \end{aligned} \quad (6.3)$$

This is the scalar version of Eq. (6.1), with each upper case letter which represented a vector or matrix replaced by a lower case one representing a scalar. The differences between Eqs. (6.2) and (6.3) are mostly the non-linearity of  $g_x$  and  $g_y$  and the existence of the external inputs. Stationary external inputs can be eliminated by a shift of origin in  $x$ - $y$  space. Let the equilibrium point of this system be  $(x_o, y_o)$  such that

$$\begin{aligned} \frac{dx_o}{dt} &= -h \cdot g_y(y_o) - \alpha x_o + i = 0, \\ \frac{dy_o}{dt} &= w \cdot g_x(x_o) - \alpha y_o + i_c = 0. \end{aligned} \quad (6.4)$$

Define  $x' \equiv x - x_o$ ,  $y' \equiv y - y_o$ , then

$$\begin{aligned} \frac{dx'}{dt} &= -h [g_y(y) - g_y(y_o)] - \alpha x', \\ \frac{dy'}{dt} &= w [g_x(x) - g_x(x_o)] - \alpha y'. \end{aligned}$$

If  $g_y(y) - g_y(y_o) \propto y'$  and  $g_x(x) - g_x(x_o) \propto x'$ , a direct comparison with Eq. (6.2) can be made.  $h$  and  $w$  are accordingly related to the oscillation frequency and  $\alpha$  to the damping constant. Indeed, when  $\alpha = 0$ ,  $x$  and  $y$  oscillate around  $(x_o, y_o)$  in a closed curve:

$$R \equiv \int_{x_o}^{x_o+x'} w [g_x(s) - g_x(x_o)] ds + \int_{y_o}^{y_o+y'} h [g_y(s) - g_y(y_o)] ds = \text{constant} \geq 0$$

which is a circle if  $wg_x = hg_y$  are identical linear functions. With nonzero  $\alpha$  and monotonic  $g_x$  and  $g_y$ , the trajectory spirals toward  $(x_o, y_o)$  as  $R$  decreases to zero:

$$dR/dt = -\alpha w [g_x(x) - g_x(x_o)] (x - x_o) - \alpha h [g_y(y) - g_y(y_o)] (y - y_o) \leq 0,$$

that is, the oscillation is damped. So we see that a pair of coupled mitral and granule cells can be approximated as a nonlinear damped oscillator.

For small amplitudes, the frequency can be calculated by a linear approximation around the  $(x_o, y_o)$ :

$$\begin{aligned} \frac{dx}{dt} &= -h \cdot g'_y(y_o)y - \alpha x, \\ \frac{dy}{dt} &= w \cdot g'_x(x_o)x - \alpha y, \end{aligned} \quad (6.5)$$

where  $(x, y)$  is now the deviation from  $(x_o, y_o)$ . The solution is  $x(t) = r_o e^{-\alpha t} \sin(\omega t + \phi)$  with frequency  $\omega = \sqrt{hwg'_x(x_o)g'_y(y_o)}$ .

Under realistic conditions, the odor inputs change on the time scale of the breathing cycle (the same is true when central inputs are concerned,

see below). However, since this change is much slower than the oscillation period  $\approx 25$  ms, the external inputs can be seen as only adiabatically changing. In other words, the input dependent oscillation origin  $(x_o, y_o)$  shifts slowly with  $i$  and  $i_c$ . The oscillation frequency  $\omega$  shifts accordingly by its dependence on  $(x_o, y_o)$ . The frequency of this neural oscillator can thus be designed by the synaptic connections  $h$  and  $w$  and fine-tuned by the external inputs. This is how the scales of  $H_o$  and  $W_o$  are chosen to have the model oscillate in the physiologically observed frequency range.

Imagine a group of  $N$  such oscillators coupled together by synaptic interactions between cells in different oscillators; it is then a group of coupled nonlinear oscillators. This is exactly the case in the olfactory bulb. That there are many more granule cells than mitral cells only means that there is more than one granule cell in each oscillator. For small amplitude oscillations [27], this group can be mathematically approximated as linear oscillators. Proceeding analogously as from Eqs. (6.3) to (6.5), we can start from Eq. (6.1) to the analog of Eq. (6.5):

$$\begin{aligned} dX/dt &= -H_o G'_y(Y_o)Y - \alpha X \equiv -HY - \alpha X, \\ dY/dt &= W_o G'_x(X_o)X - \alpha Y \equiv WX - \alpha Y, \end{aligned} \quad (6.6)$$

where  $G'_x(X_o)$  and  $G'_y(Y_o)$  are diagonal matrices with elements:  $[G'_x(X_o)]_{ii} = g'_x(x_{i,o})$ ,  $[G'_y(Y_o)]_{jj} = g'_y(y_{j,o})$ . Since only the mitral cells send bulbar outputs, we concentrate on them by eliminating  $Y$ :

$$\ddot{X} + 2\alpha\dot{X} + (A + \alpha^2)X = 0, \quad (6.7)$$

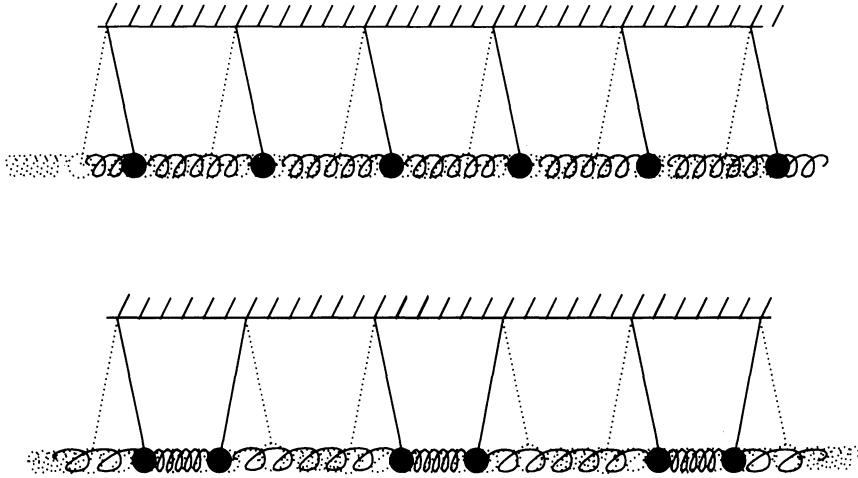
where  $A \equiv HW = H_o G'_y(Y_o)W_o G'_x(X_o)$ . This is the equation for a system of  $N$  coupled oscillators (cf. Eq. (6.2)). The  $i$ th oscillator follows the equation

$$\ddot{x}_i + 2\alpha\dot{x}_i + (A_{ii} + \alpha^2)x_i + \sum_{j \neq i} A_{ij}x_j = 0, \quad (6.8)$$

The first three terms are for a single ( $i$ th) oscillator (cf. Eq. (6.2)).  $A_{ij} = \sum_l (H_o)_{il} g'_y(y_{l,o}) (W_o)_{lj} g'_x(x_{j,o}) \geq 0$  is the coupling strength from the  $j$ th to the  $i$ th oscillator. It originates from the  $j$ th to the  $i$ th mitral cell via the intermediate granule cells in the connection path. Thus, local synaptic connections imply local oscillator connections, or in other words,  $A$  is near-diagonal. However, unlike a single oscillator, the group is no longer damped and can spontaneously oscillate, induced by the coupling  $A$ , as we will see next.

### 6.3.3 EXPLANATION OF BULBAR ACTIVITIES

Let us see how this model can explain the bulbar activities. First, the single oscillator analysis (Eqs. (6.2) and (6.5)) predicts that local mitral cells oscillate with a  $90^\circ$  phase lead over the local granule cells. This is observed in physiological experiments [32].



**Fig. 6.3.** Two example modes in a system of coupled pendulums. The upper one has all the pendulums oscillating with the same amplitude and phase, back and forth from right (solid positions) to left (dashed positions). Each pendulum in the mode of the bottom figure oscillates  $180^\circ$  out of phase from its neighbors.

Second, the model predicts that the oscillation should have the same dominant frequency everywhere in the bulb. One can see this by noticing that Eq. (6.7) has  $N$ -independent mode solutions. The  $k$ th mode is

$$X \propto X_k e^{-\alpha t \pm i\sqrt{\lambda_k}t}, \quad (6.9)$$

where  $X_k$  and  $\lambda_k$  for  $k = 1, 2, \dots, N$  are eigenvectors and eigenvalues of  $A$ . (Figure 6.3 gives examples of oscillation modes in a coupled oscillator system.) Each mode has frequency  $\text{Re}\sqrt{\lambda_k}$ , where  $\text{Re}$  means the real part of a complex number. If

$$\text{Re}(-\alpha \pm i\sqrt{\lambda_k}) > 0 \quad (6.10)$$

for some  $k$ , as can happen for complex  $\lambda_k$ , the amplitude of the  $k$ th mode will increase with time. Starting from an initial condition of arbitrarily small amplitudes, the mode satisfying Eq. (6.10) with the fastest growing amplitude will dominate the output, the nonlinear effect will suppress the other modes, and the final output will be a single “mode” in a nonlinear regime. In that case, although each oscillator has its own amplitude and phase described by the components of  $X_k$ , they all share the same oscillation frequency  $\text{Re}\sqrt{\lambda_k}$ , as observed in the experiments [8,24].

The third consequence is that the oscillation phase will have a nonzero gradient across the bulb. Each oscillator oscillates only when the neighboring oscillators give a driving force  $F_i = -A_{ij}x_j$  which overcomes the

damping force  $-2\alpha\dot{x}_i$  (cf. Eq. (6.8)). This means the neighbor oscillation  $x_j$  should have a component that is parallel to  $-\dot{x}_i$  or has a phase difference from  $x_i$ , generating a nonzero phase gradient field across the bulb, as observed physiologically [7,24]. This is not necessarily true if mitral-to-mitral or other couplings exist to change the oscillator coupling nature [3], however, there is still not much evidence for mitral-to-mitral connections [1].

The fourth consequence is the experimentally observed rise and fall of oscillations with inhalation and exhalation. This is because growing oscillations require large enough  $\lambda_k$  to satisfy  $\text{Re}(-\alpha \pm i\sqrt{\lambda}_k) > 0$  for some  $k$ . This in turn implies large matrix elements of  $A = H_o G'_y(Y_o) W_o G'_x(X_o)$ , that is, large gain  $G'$ 's. Thus  $(X_o, Y_o)$  should be near the high slope (gain) region of  $g_x$  and  $g_y$  functions past the threshold. As in the single oscillator case,  $(X_o, Y_o)$  depends on the odor input. When the central inputs are constant,

$$\begin{aligned} dX_o &\approx (\alpha^2 + HW)^{-1} \alpha dI, \\ dY_o &\approx (\alpha^2 + WH)^{-1} W dI. \end{aligned} \quad (6.11)$$

Before inhalation,  $(X_o, Y_o)$  is in the low gain region below the threshold so that none of the modes satisfies Eq. (6.10), and the equilibrium point  $(X_o, Y_o)$  is thus stable. Inhalation,  $dI > 0$ , pushes  $(X_o, Y_o)$  into the higher gain region, making inequality (6.10) possible for some  $k$ .  $(X_o, Y_o)$  then becomes unstable and the  $k$ th mode emerges from noise to visible oscillations across the bulb. Exhalation reverses this situation and the oscillations cease.

## 6.4 A Model of Odor Recognition and Segmentation in the Olfactory Bulb

Having successfully modeled the bulbar oscillation phenomena, we now discuss the cognitive value of this model for olfactory computation.

### 6.4.1 EMERGENCE OF OSCILLATIONS DETECTS ODORS: PATTERNS OF OSCILLATIONS CODE THE ODOR IDENTITY AND STRENGTH

Odor inputs  $I_{\text{odor}}$  influence the emergence and patterns of the bulbar oscillation via the path  $I_{\text{odor}} \rightarrow (X_o, Y_o) \rightarrow A \rightarrow (X_k, \lambda_k)$ . Therefore, this model proposes that the olfactory bulb codes the odor information in the following way. First, it detects the presence of any relevant odor by the presence of a global oscillation. Second, if it detects an odor, it determines the odor identity by the specific oscillation pattern, that is, the oscillation amplitudes and phases across the bulb, of the mode  $X_k$  that emerges. Third, the odor strengths can be roughly coded in the overall oscillation

amplitudes, at least for small odor concentrations. For large odor concentrations,  $I_{\text{odor}}$  is no longer proportional to its value at a small concentration [30]. As we sometimes perceive, the character of an odor may depend on its concentration.

Absence of oscillation is apparent with absence of odor  $I_{\text{odor}} = 0$ , as is the case before inhalation. But some inputs may still be irrelevant or “odorless” for the animal to detect, and the bulb may choose to ignore them even when they are large. Equation (6.10) suggests that only those modes with nonreal  $\lambda_k$  can emerge. Hence, only those  $I_{\text{odor}}$  that can make matrix  $A$  sufficiently asymmetric to have large nonreal  $\lambda_k$ , will cause the bulb to respond. For illustration, consider an example when both  $H_o$  and  $W_o$  are symmetric and uniform (cyclic):

$$H_o = \begin{pmatrix} h & h' & 0 & 0 \dots & 0 & h' \\ h' & h & h' & 0 \dots & 0 & 0 \\ 0 & h' & h & h' & 0 \dots & 0 \\ \vdots & \vdots & \ddots & & \vdots & \\ h' & 0 & \dots & 0 & h' & h \end{pmatrix}, \quad W_o = \begin{pmatrix} w & 0 & 0 & 0 \dots & 0 \\ 0 & w & 0 & 0 \dots & 0 \\ 0 & 0 & w & 0 \dots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & 0 & w \end{pmatrix}.$$

Each mitral cell only connects to the nearest granule cell with the same strength  $w$ ; each granule cell synapses on the nearest mitral cell with strength  $h$ , and, in addition, onto the mitral cells at the neighboring left and right with strength  $h'$ . If in addition all the mitral (granule) cells are identical and receive the same receptor (central) input strength  $I_i$  ( $I_{c,i}$ ), then by symmetry  $(X_o, Y_o)$  is uniform, that is, each component in  $X_o$  or  $Y_o$  is the same. The matrices  $G'_x(X_o)$  and  $G'_y(Y_o)$  will be proportional to the identity matrix,  $A = H_o G'_y(Y_o) W_o G'_x(X_o)$  is then symmetric:

$$A = \begin{pmatrix} a & b & 0 & 0 & \dots & 0 & b \\ b & a & b & 0 & \dots & 0 & 0 \\ 0 & b & a & b & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & & \dots & & \vdots \\ b & 0 & \dots & 0 & b & a \end{pmatrix}.$$

The  $N$  oscillation modes will be

$$\begin{pmatrix} \sin(k1) \\ \sin(k2) \\ \vdots \\ \sin(ki) \\ \vdots \\ \sin(kN) \end{pmatrix} e^{-\alpha t \pm i\sqrt{\lambda_k}t}, \quad \begin{pmatrix} \cos(k1) \\ \cos(k2) \\ \vdots \\ \cos(ki) \\ \vdots \\ \cos(kN) \end{pmatrix} e^{-\alpha t \pm i\sqrt{\lambda_k}t},$$

where  $k = 2\pi \frac{K}{N}$ ,  $K$  is an integer,  $0 \leq K < \frac{N}{2}$ ,  $\lambda_k = a + 2b \cos(k)$ . For  $b < a/2$  and  $\lambda_k > 0$ , all the modes will be damped oscillations with similar

frequencies close to  $\sqrt{a}$ . Inhalation of a “uniform odor” ( $I_{\text{odor},i} = I_{\text{odor},j}$  for all  $i, j$ ) only increases proportionally the values of  $a$  and  $b$ , and thus  $\lambda_k$ , but  $A$  remains symmetric and oscillations never emerge spontaneously.

The bulb can design its responsiveness to selected odors by designing its synaptic connection strengths. In the above example, the bulb ignores the uniform odor and only odors activating different receptors differently can possibly induce global oscillations. However, the bulb can choose to respond to this uniform odor by changing the synaptic design above, for example, by deleting the connection of each granule cell to the mitral cell to the left, that is,  $(H_o)_{i,i+1} \rightarrow 0$ . Everything else staying the same, we then have a nonsymmetric matrix  $A \propto H_o$  with complex  $\lambda_k$ 's (as can be seen on a  $3 \times 3$  matrix) even before inhalation, ready to be raised by the uniform odor to higher values to satisfy Eq. (6.10).

The same influence path  $I_{\text{odor}} \rightarrow (X_o, Y_o) \rightarrow A \rightarrow (X_k, \lambda_k)$  makes it apparent that each  $I_{\text{odor}}$  induces a specific output to code the information by amplitude and phase pattern  $X_k$ .  $X_k$  is selected from the pool of  $N$  modes. Such a pool is itself different for different odors via a different  $A$  or dynamic system (6.7). Thus, there is potentially a large number of oscillation modes or a large number of odor types that can be distinguished. In our example of  $N = 10$ , three odors were observed [3]. However, it is not known how the odor number scales with system size. In principle, the frequency  $\text{Re}\sqrt{\lambda}$  can also carry odor information. However, since the frequency is the same across the bulb, it only contributes to the code by one variable, negligible compared to the  $N$  variables each from the amplitudes and phases.

The bulb codes the odor strength as follows. During inhalation, the input increases at a rate proportional to the odor concentration. Hence higher odor concentrations cause  $\text{Re}(-\alpha \pm i\sqrt{\lambda}_k)$  to shift sooner from negative to positive values. Thus the mode  $X_k$  can grow into a higher amplitude oscillation, which can be interpreted by the olfactory cortex as a higher odor concentration. For some odors, the bulb requires a smaller concentration than for other odors to lead to an emergence of oscillation. The bulbar sensitivity can thus be higher for particular odors.

In principle, it should be possible to design the synaptic connections such that the bulb can reach a desired correspondence between odor stimuli and oscillation patterns, different sensitivities for different odors, and different resolutions to odor discrimination tasks. Here, low resolution to odor discrimination means similar oscillation pattern responses to different odor inputs. How to design the bulb connections remains an open problem.

### 6.4.2 ODOR SEGMENTATION IN THE OLFACTORY BULB — OLFAC TORY ADAPTATION

It is desirable to recognize the odor components in an odor mixture instead of simply judging the mixture as a new distinctive odor, as each odor component may convey a distinct message (consider, e.g., an odor mixture which contains odorants from two distinct sources, one a predator the other some food). Any realistic model for olfactory computation should also solve the complex problem of odor segmentation in odor mixtures. The present bulbar model does so by olfactory adaptation.

Since receptors responding to different (nonpheromonal) odors do not segregate to different groups [12], it is not possible to segment the odors by attending to different receptor groups. It was proposed [4] that, in the case of a two odor mixture for example, this problem is solved by olfactory adaptation to one odor and recognizing the other as if it were the only odor component present. This model suggests that the olfactory adaptation should not be understood as fatigue, but as an active mechanism to screen out the current and already detected odor inputs, so that the olfactory system can concentrate on detecting new odors superposed on the existing ones. Humans have difficulty in identifying components in odor mixtures [33]. For example, two substances odorous singly may be inodorous together — counteraction; or only one odor type is sensed when two are mixed — masking [31]. Without odor adaptation, the new input odor superposed on existing ones may be masked and counteracted. Since odor receptors do not show much sign of odor adaptation [31] and the bulb on the other hand does [34], it is reasonable to believe that the odor segmentation problem is first solved in the bulb. And since adaptation is odor specific and subjects adapted to one odor can still detect new odors [31], this suggests that adaptation may involve odor specific control signals from higher olfactory centers, which have feedback paths to the bulb [1].

In our bulbar model, the odor input  $I_{\text{odor}}$  controls the output by raising  $(X_o, Y_o)$ . Thus odor adaptation can be achieved by a central input  $I_{c,\text{control}}$  that cancels the effect of  $I_{\text{odor}}$  on  $(X_o, Y_o)$ . Equation (6.11) can be generalized when  $I_{c,\text{control}} \neq 0$  as

$$\begin{aligned} dX_o &\approx (\alpha^2 + HW)^{-1}(\alpha dI - HdI_c), \\ dY_o &\approx (\alpha^2 + WH)^{-1}(WdI + \alpha dI_c), \end{aligned} \quad (6.12)$$

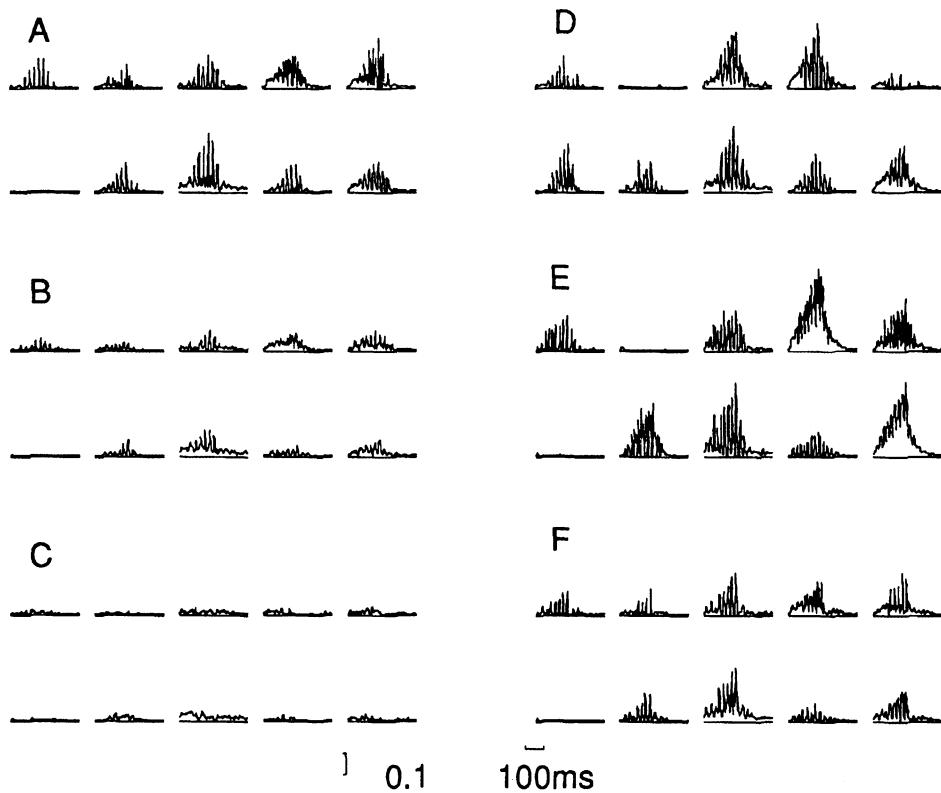
where  $I = I_{\text{background}} + I_{\text{odor}}$ ,  $I_c = I_{c,\text{background}} + I_{c,\text{control}}$ , and  $I_{\text{background}}$  and  $I_{c,\text{background}}$  are time invariant within a sniff cycle. The central feedback  $I_c$  is sent to the granule cells. It, like  $Y_o$ , is assumed to have dimension  $M$  (the granule cell number), while  $I$ , like  $X_o$ , has dimension  $N$ . A complete cancellation means a simultaneous satisfaction of the  $N + M$  equations  $dX_o = 0$  and  $dY_o = 0$ . This is generally impossible with only  $M$  control variables in  $I_c$ . Since mitral cells are the only bulbar output cells, one can loosen the demand and require only  $N$  equations  $dX_o = 0$ , which is possible

in the real bulb where  $M \gg N$ . Then the nonoscillatory output  $G_x(X_o)$  is the same as it is in the absence of odor. As a result, the oscillatory bulbar output will not exist either since its emergence requires *both*  $X_o$  and  $Y_o$  to rise above threshold to make  $\lambda$  large enough. We thus reach an effective means of odor adaptation.

$I_{c,\text{control}}$  is odor specific: from Eq. (6.12),  $dX_o = 0$  leads to  $dI_{c,\text{control}} = H^{-1}\alpha dI_{\text{odor}}$ . It is possible for the central brain to send such an odor-specific canceling, or adaptation, signal  $I_{c,\text{control}}$ , since adaptation occurs after  $I_{\text{odor}}$  has already been recognized by the brain from the previous sniff, and thus the higher olfactory centers have enough information about the odor to construct the appropriate  $I_{c,\text{control}}$ .

Computer simulations demonstrated successful adaptations by such a mechanism: an input odor can be fully canceled at the bulbar output as if no odor input existed. Reducing the adaptation signal  $I_{c,\text{control}}$  by half reduces the bulbar output amplitude considerably (from nonadapted situations) as if a much weaker odor were inhaled (Fig. 6.4). To test this model on odor segmentation, two inputs,  $I_{\text{odor}1}$  and  $I_{\text{odor}2}$ , representing two different odors are superposed linearly to give the odor mixture input as  $I_{\text{odor}} = I_{\text{odor}1} + I_{\text{odor}2}$ . Such linear approximation for receptor inputs is assumed valid for small odor concentrations. Without any adaptation signals  $I_{c,\text{control}}$ , the bulbar output pattern can be seen to resemble neither output when each odor is presented alone. Then,  $I_{\text{odor}2}$  is arbitrarily chosen as the preexisting and adapted odor, and the central feedback  $I_{c,\text{control}}$  is the adapting signal specific to  $I_{\text{odor}2}$ . As demonstrated in Fig. 6.4, the bulb clearly responds as if only  $I_{\text{odor}1}$  were present [4], achieving odor segmentation. The success of this model depends partly on the fact that the operating region of the bulb in normal odor environment is essentially linear, as suggested by physiological experiments for small oscillation amplitudes [27], although there is a small region of nonlinearity when breaking away from threshold for odor detection [4].

This adaptation model predicts that central feedback to the bulb should increase after initial exposures to an odor, and that the signal should vary on the same time scale as the breathing cycle  $\sim 0.2 - 1$  s, instead of the bulbar oscillatory time scale,  $\sim 25$  ms. The feedback signal should be odor-specific and directed to the granule cells in a distributed fashion. Experimentally, not much is known of such feedback signals except that feedback paths do exist anatomically [1]. It has also been demonstrated that when the central input to the bulb is blocked, the neural oscillatory activity induced by odor stimulus increases substantially [36] — supporting our proposed source of adaptation signals. A systematic experimental study on the central feedback to the bulb will provide a crucial test of this model.



**Fig. 6.4.** (A), (B), (C) Model response to  $I_{\text{odor}1}$  without adaptation, with half-strength adaptation, and with full strength adaptation to  $I_{\text{odor}1}$ , respectively. (D) Model response to  $I_{\text{odor}2}$ . (E) Model response to odor mixture of  $I_{\text{odor}1}$  and  $I_{\text{odor}2}$ . (F) Model response to the same odor mixture with adaptation to  $I_{\text{odor}2}$  — odor segmentation and detection of  $I_{\text{odor}1}$  in the mixture.

#### 6.4.3 OLFACTORY PSYCHOPHYSICS — CROSS-ADAPTATION, SENSITIVITY ENHANCEMENT, AND CROSS-ENHANCEMENT

The bulbar model presented above also explains other olfactory psychophysics. One such phenomenon is olfactory cross-adaptation. Experiments show that after sniffing one odor, another odor at next sniff smells less strong than it normally would and may even smell different [31]. The extent to which odor A is cross-adapted by B is different from that of B by A [35]. Our model explains such cross-adaptation naturally. After exposure to odor A, the central brain sends an adapting signal that is specific to A. As recovery from olfactory adaptation takes 1-3 min [31] after a preexisting odor is removed, the adapting signal will last for at least a few sniffs even after odor removal. Imagine that at the next sniff, the odor A is suddenly

replaced by odor B. Since the adapting signal is specific to A, the effect of odor B on the bulb cannot be completely canceled by the adapting signal, rather, the bulb's response to B will be distorted with the suppressive effect of the adaptation. Computer simulations confirm such a result [4].

This model can also enhance the bulb sensitivity to particular odors by reversing the adaptation signal. Sensitivity enhancement has been observed psychophysically in rats [37], but not yet in humans. Analogous to cross-adaptation is cross-enhancement, although it is unknown whether it exists psychophysically. Both enhancement and cross-enhancement have been demonstrated in computer simulations [4].

## 6.5 A Model of Odor Segmentation Through Odor Fluctuation Analysis

The olfactory bulb model described above applies most likely only to animals that do not depend on olfaction as the primary sense of the world. The computational tasks that model addresses are odor object detection, identification, and segmentation if multiple odors are initiated unsynchronously. Such goals are clearly not enough for animals that primarily depend on olfaction to explore their environment. These animals need olfaction to function as visual animals need vision, namely, they depend on olfaction for localizing, in addition to identifying and segmenting, odor sources with respect to their environment. A recent work by Hopfield [2] addressed this computational problem and proposed using temporal fluctuations in receptor inputs for such a purpose.

### 6.5.1 A DIFFERENT OLFACTORY ENVIRONMENT AND A DIFFERENT TASK

Hopfield [2] argued that in most olfactory environments, odors are brought to the nose by fluctuating and turbulent winds, such that the odor plume contains a complex spatial structure and is increasingly mixed with odors from other parts of the environment as time increases. Thus, if there is only one odor object present, the location of its source can be obtained by analyzing the fluctuations of the odor intensity with time and relating them to the local wind directions. On the other hand, if there are multiple odor sources, each at its own location, the odor plume to the nose will contain all components with different intensities fluctuating largely independently due to complex mixings by the wind before reaching the nose. Since receptor neurons activated by different odors overlap [12], different fluctuations of multiple odors are superposed in the receptor cell responses. A separation of odors and their fluctuation signals is needed before each odor location and identity can be decoded. This is so even when the odor objects are not

familiar to the animals [2]. Hopfield hypothesized that the function of the earliest part of the olfactory processing in highly olfactory animals is to achieve this separation task, such that the contributions of different odor sources to the receptor activities can be separated.

Both this model and the previous one presented in the last section address the odor segmentation task. However, there are major differences between the two. First, the olfactory environments are different. In Hopfield's model, the odor intensity fluctuates on a time scale of tens of milliseconds [38]. The previous model assumes an environment that changes slowly, on a time scale longer than the breathing cycle, before reaching the olfactory mucosa where the breathing modulates its effects on the receptors. Second, the segmentation tasks are different. The slowly changing environment allows the previous model to segment mixed odors by subtracting out the preexisting odors, but it cannot do so when each odor component in the mixture is initiated at the same time. Hopfield's model does not simply subtract out one odor in the mixture, as the fast fluctuations make it difficult. Rather, it *simultaneously* sorts out individual odor fluctuations from the mixed multisource input signals.

Behavioral evidence [39] demonstrates that olfactory animals such as the terrestrial mollusc *Limax maximus* [40] use the time fluctuations for olfactory tasks. When two odors are completely mixed and originate from the same spatial location, and thus have the same temporal fluctuations in their intensities, they are recognized as one distinct odor different from the two component odors. But as long as there are differences in temporal fluctuations in the two components — for example, placing the odor sources at even slightly different locations — the two odors are individually identified [40]. We will describe below how Hopfield's model achieves similar performances in cases of independent odor component fluctuations.

### 6.5.2 ODOR SEGMENTATION IN AN ADAPTIVE NETWORK

An idealization of olfactory bulb circuitry was used [2]. However, the odor information coding is qualitatively different from the previous model. First, the synaptic strengths between the cells adapt to the odor inputs, an essential feature for the functioning of Hopfield's model. Second, the odor identity information is coded by the synaptic strengths rather than by the cell activities, which instead code the instantaneous odor strengths.

A network of linear and mutually inhibitory neurons is used [2], each neuron is denoted as  $x_i$  in the state vector of  $X = \{x_1, x_2, \dots, x_N\}$ . The cell dynamics are (to compare with the previous model, the notations are modified from that of [2]):

$$\dot{X} = -TX - \alpha X + I. \quad (6.13)$$

This can be seen as a drastic simplification from Eq. (6.1), by assuming  $G_x(x) = x$  and  $G_y(y) = y$ , ignoring central inputs  $I_c$ , and assuming that

the granule cells instantaneously follow their inputs from mitral cells such that  $Y \propto W_o X$ . Substituting  $Y$  into Eq. (6.6) and assuming  $T \propto H_o W_o$ , we obtain Eq. (6.13) where the mitral cells effectively inhibit each other.

The odor input  $I$  differs from the previous model in its temporal courses to reflect the odor concentration fluctuations instead of the regular respiration. Each odor  $k$  at some unit concentration is assumed to excite the mitral cell  $i$  with strength  $S_{ki}$ . Its concentration is described by a fluctuating time function  $a_k(t)$ . Thus, the input to the  $i$ th mitral cell will also fluctuate with  $t$  as

$$I_i(t) = \sum_k a_k(t) S_{ki}. \quad (6.14)$$

Let us now look closer at the segmentation problem. There are two modules of the segmentation needed. The first is to segment the *odor concentrations*  $a_k(t)$ , and second is the *odor identity* defined by  $S_{ki}$ , for all odors  $k$ . Both of these signals are mixed in the inputs  $I$ . Because each odor excites many mitral cells, independent odor fluctuations  $a_k(t)$  for different odors  $k$  will result in *correlated* fluctuations in inputs  $I_i(t)$  to different cells  $i$ . It is these *correlations* that define the individual odor objects for the olfactory system [2]. This crucially requires that individual odor fluctuations themselves are uncorrelated, otherwise “olfactory illusions” should occur [2]. The idea is then to transform the correlated receptor inputs  $I$  to mitral outputs  $X$  such that first, the components of  $X$ , unlike  $I$ , are *uncorrelated*; second,  $X$  still carries all the temporal information in  $I$ . This is possible if each component  $x_i$  depends on no more than one odor component concentration  $a_k$ , which in turn is carried in the activity of only one output cell. This way, the first module of odor segmentation — to segment the individual odor concentration  $a_k$  from  $I$  — is achieved at the  $X$  level, and the higher olfactory centers only need to look at one component of  $X$  for each  $a_k$ .

Let us see if such decorrelation (segmentation of  $a_k$ ) also leads to the segmentation of  $S_{ki}$ , the second module of the segmentation task. To achieve this decorrelation at  $X$ , the synaptic connections  $T$  should satisfy some conditions depending on the odor, or in other words,  $T$  should have the knowledge of the individual odor components  $S_{ki}$ . It would be ideal if, in the case that mitral cell  $n$  ( $x_n$ ) carries concentration fluctuation of odor  $k$ , the odor identity information  $S_{ki}$  is carried in synapses  $T_{ni}$  or  $T_{in}$  associated with the *same* cell  $n$ . This would not only segment the odor identity  $S_{ki}$ , but also tag this piece of information to the same cell that carries this odor’s concentration. Using an example of a two odor mixture [2], and the simplification of small  $\tau \equiv 1/\alpha$  such that one can assume  $\dot{x}_i \approx 0$  for all  $t$ , we see that such segmentations are possible [2]: If, for example, the first column of matrix  $\alpha + T$  is the vector  $S_1$  and the second column  $S_2$ , then  $x_1 = a_1$ ,  $x_2 = a_2$ , and  $x_{i>2} = 0$ . Thus all cells are inactive except for two output cells which convey the odor strengths for the respective odors, and

their synaptic strengths to other cells, that is,  $T_{i1}$  and  $T_{i2}$ , respectively, carry the odor identities  $S_{1i}$  and  $S_{2i}$ , respectively.

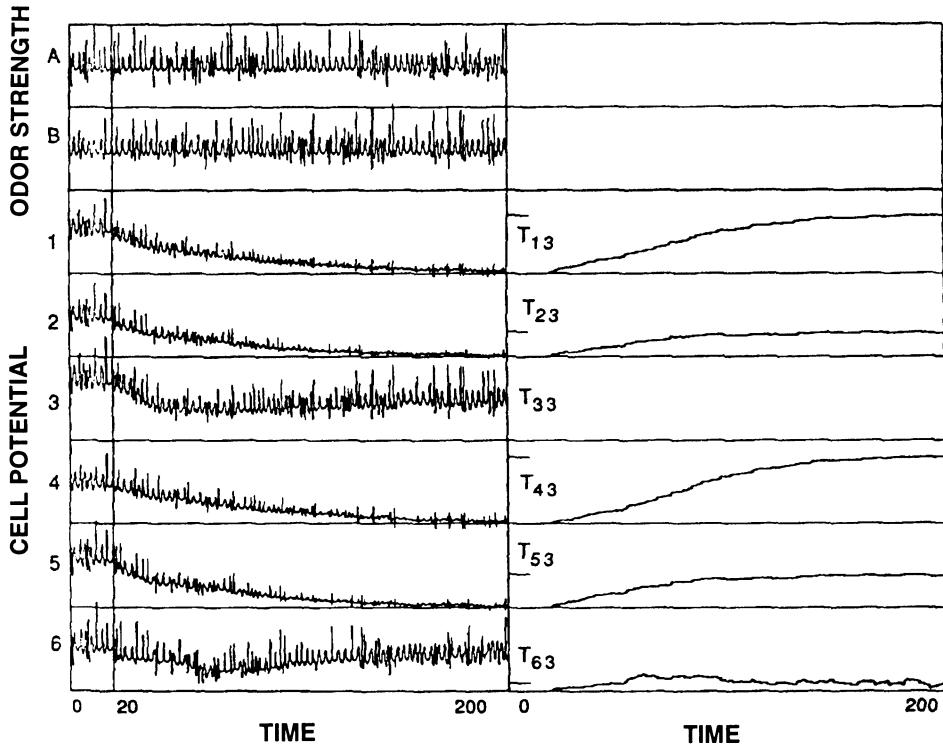
It is now clear that such tasks require  $T$  to adaptively change as the odor sources change with the environment. Hopfield [2] proposed the following synaptic dynamics in addition to the cell dynamics above:

$$\dot{T}_{ij} = fx_i \cdot fx_j [\delta + \epsilon(fx_j - \gamma fx_i)], \quad (6.15)$$

where  $fx_i$  is a high-pass filtered version of  $x_i$ ,  $\delta$ ,  $\epsilon$ , and  $\gamma$  are constants determining the learning speed, which should be much slower than the cell dynamic speed. Such a synaptic changing rule is local and depends only on the pre- and postsynaptic neurons. The synapses  $T$  will stop changing when each cell  $n$  is assigned to a different odor  $k$ ,  $x_n = a_k$ , or inactive,  $x_n = 0$ . In that case, all the terms on the right-hand side of Eq. (6.15) vanish under time average when odor fluctuations are independent. Thus Eqs. (6.13) and (6.15) give the desired performance. Furthermore, the desired solution for  $T$  is stable in the sense that if  $T$  is close to the solution, convergence is guaranteed [2].

This adaptive network demonstrates its performance in a computer simulation [2], with six neurons and a two odor mixture. Figure 6.5 shows that before synaptic learning, all six neurons fluctuate with the inputs, while after synaptic learning is turned on only neurons 3 and 6 keep their responses to follow odors  $B$  and  $A$ , respectively. These two neurons happen to be the recipients of the strongest components  $S_{ki}$  for the respective odors (i.e.,  $(k, i) = (B, 3)$  and  $(A, 6)$ ), since the inhibition between neurons resulted in the winner-take-all-type output configurations. The final synaptic strengths  $T_{i3}$  and  $T_{i6}$  from neurons 3 and 6 code the odor identities  $S_{Bi}$  and  $S_{Ai}$  for odors  $B$  and  $A$ , respectively. Such a network should also forget its synaptic strengths with vanished cell activity, so that it can adapt to new odor mixtures with new inputs from the environment. Here, the synaptic adaptation captures odor quality information  $S$ . One should notice that adaptation in this model does not mean to lose the perception to the odors as in the previous model.

The higher olfactory centers have to somehow query the synaptic strength  $T$  to get the information about the odor identity  $S_{ki}$ . They need to send a signal to the dendrites of the principle neuron that is reporting the concentration of the odor, and reads out the responses of the other neurons to that signal. Hopfield suggests one possible way to do this by multiplexing these signals on the same neuronal axons reporting the odor strengths. The low frequency component of the signals carries the odor concentration signal while the higher frequency component signals carry the odor identity  $S_{ki}$  information [2]. Another possible mechanism was suggested in [41]. One can have multiple copies of the network receiving the same odor inputs, each neuron in  $m$ th network inhibits its copies in all  $n$ th ( $n > m$ ) networks such that at the end of adaptation (Eq. (6.15)), each network has a different neuron reporting the corresponding odor fluctuation. Since the adaptation



**Fig. 6.5.** Simulation for a bulb of six neurons and two independent sources *A* and *B*. Synapse change begins at  $t = 20$ . On the left are odor strengths fluctuations and cell potential variations with time. On the right are synaptic strength development  $T_{n3}$  from neuron 3 to other neurons. These strengths carry the odor identity information  $S_B$ ; for odor *B* whose time variation is captured by neuron 3. Taken from Hopfield 1991 [2].

selects the winner neuron for each odor, extending the example of two odors above, the first network selects the two neurons, 3 and 6, most activated by the two odors, respectively; the second network selects the two second most activated, for example, 4 and 5, respectively, etc. The covariation between the neurons in different networks assigns them to the same odor. The odor concentration can be read out from the corresponding neuron in any network, while the identities of these neurons code the odor quality  $S_{ik}$ . A modification to make different networks receive different, maybe overlapping, subsets of receptor inputs, simulating glomeruli structure, could make the model more biologically plausible [41]. This mechanism makes it possible to have interesting psychophysics phenomena similar to cross-adaptation. If new odors are introduced before synaptic recovery from the previous odors, the winner neurons for the previous odors are more likely to remain winners by the advantage of their already developed inhibitory

synapses over others. Thus the winning neurons might be different from those if the new odors were introduced without prior odor exposures, resulting in distorted odor identity information. Preliminary simulations in a six neuron network exposed to two odors [42] confirmed such a possibility, although it does not happen most of the time.

The present model segments the odor sources even before the odor qualities are known [2]. This demonstrates the model capability to capture the invariants in the varying sensory inputs. In a typical olfactory environment, although the receptor inputs  $I$  fluctuate with time, the odor qualities  $S_{ki}$  are constant at least for a short period of time, and there should be a perceptual constancy. In most cases, the number of neurons  $N$  is larger than the number of odors  $K$ . The variation of the input  $I$  in the  $N$ -dimensional input space, ignoring noises, should be confined to a  $K$ -dimensional subspace, reflecting the regularities in the inputs carrying the constant odor qualities. This means the matrix  $R$ , defined by input covariance

$$R_{ij} \equiv <(I_i - \langle I_i \rangle)(I_j - \langle I_j \rangle)> = \sum_k (a_k - \langle a_k \rangle)^2 S_{ki} S_{kj},$$

has only  $K$  eigenvalues that are substantially nonzero. This input regularity enables the network to find the  $K$ -dimensional subspace, spanned by the  $K$  eigenvectors corresponding to the first  $K$  nonzero eigenvalues of  $R$  (i.e., to find the  $K$  principle components of the input variations). Furthermore, the network identifies the  $K$  axes or directions  $S_{ki}$  for  $k = 1, 2, \dots, K$ , not necessarily orthogonal to each other, in that subspace representing the odor qualities. As we know, except when there is only one odor  $K = 1$ , there are infinitely many choices of axes or directions to span a  $K > 1$  dimensional space.  $S_{ki}$  in most cases are not in the directions of the individual eigenvectors (principle components). Not being able to identify them would result in confused odor quality perceptions. However, the non-Gaussian property of the fluctuating concentration  $a_k(t)$ , for example, the nonzero third moments  $\langle a_k^3 \rangle$  about the mean, is another regularity of the inputs that enables the network to decipher the odor quality [2]. Such positive third moments are assumed [2] as the odor strength distribution is expected to have longer tails above the mean.

The network, including the learning rule, has no prior knowledge of the odor quality, other than the non-Gaussian statistical knowledge of the odor fluctuation  $a_k$ . The  $\epsilon$  term of the learning rule (6.15) takes advantage of this non-Gaussian property to capture the odor quality  $S_{ki}$ . The learning stops when both the covariance  $\langle f x_i f x_j \rangle$  and the correlation  $\langle (f x_i)^2 f x_j \rangle$  between neurons vanish. Since the third moment  $a_k^3 \neq 0$ , the elimination of third order correlation  $\langle (f x_i)^2 f x_j \rangle = 0$ , by the  $\epsilon$  term, is essential to ensure that the activity of each output neuron depends on no more than one odor fluctuation.

The present model thus gives an entirely different way of thinking about the synaptic connections [2]. Unlike conventional neural networks, the syn-

apses code the sensory information, and are more than just algorithms to compute. There are no matched filters for the odors in the connections until after the successful adaptation. In a sense, recognition can be understood as the capability of adaptation. This is also reflected in the previous model where adaptation, although of a different nature, was possible only when the higher olfactory centers have the odor information. Fast synaptic modulation has been advocated by others as a computational element [43]. It has also been recently used in another olfactory model where the inhibitory synapses adjust to transform redundant responses in the receptor population to a more efficient representation at the output neurons [16].

## 6.6 Discussion

Two models of the early olfactory computation have been presented. One [3,4] samples the environment by discrete breathing cycles; during each cycle the environment is assumed to change little. It codes the receptor inputs into a coherent oscillatory output pattern, to be identified as a single distinct odor. It does not follow the fluctuations of the odor strength on a time scale shorter than the breathing cycle, and thus cannot easily decode the odor source locations by the fluctuation analysis. It segments two coexisting odors when one of them is present before the other appears, by subtracting the earlier odor background and thus recognizing the later one. The other model [2] samples the environment continuously, following the details of the input fluctuations which have a shorter time scale than the odor recognition process. It segments odor mixtures without the need to subtract out individual odors and can thus achieve simultaneous recognition. The neural output carries the time fluctuations of the odors which may be used to compute source locations, while the odor identities have to be determined from the synaptic strengths of the output neurons.

The two models differ dramatically in their tasks and computational mechanisms. Leaving aside the difference in mechanisms, it is not certain if different tasks are emphasized in different animals or in different environmental conditions. Highly olfactory animals certainly need to emphasize more on the odor source localization task than visual animals, which may also rely heavily on olfaction when visual function alone is not enough, for example, when hunting at night or searching for plant seeds under heavy leaves [44]. Insects seem to have a much higher sensitivity to odors than humans do [31]. This enables them to detect small odor intensity variations which occur on a very short time scale, around 10–100 ms in the outdoor environment [38]. This is possible also because their receptor cells have a time constant around 10 ms [6]. The lower animals also have their olfactory reception decoupled from their respiratory mechanism, enabling them to follow the odor fluctuations undisturbed and undisrupted. The vertebrate receptor neurons, on the other hand, have a long time constant,

about 200 ms [12], comparable to the breathing time scale in mammals (e.g., rats [44], although we humans breath once every 4 s). Such a long time constant makes it difficult to detect odor fluctuations on a finer time scale, and may be a necessary strategy for an insensitive receptor neuron to integrate the signal in time to increase the signal-to-noise ratio. However, since there is also a long time scale fluctuation in odor concentration of natural environment [38] — periods of about a couple hundred milliseconds of relatively strong odor concentration separated by periods of weak odor concentration — it is conceivable that this longer time scale variation can also be used for odor orientation. In an indoor environment, like the underground burrows of rodents, air is much less turbulent and air flow speed is an order of 10 times slower than it is outdoor. (You can test this by blowing soap bubbles and watching them flow in the air.) The odor fluctuation will be much slower, as we notice in the kitchen of an apartment, for example. Such an environment will have a long lasting background odor, and an odor adaptation model [4] to subtract the background is conceivably used for odor segmentation.

However, behavioral evidences clearly point to olfactory cues used even in many mammals for object localization. For example, sunflower seeds and hazel nuts under the soil surface can be localized by mice; large cats can be observed to sniff the wind continually and usually approach their preys from downwind [44]. It is conceivable that many mammals also use temporal fluctuations of odors, although maybe on a longer time scale, to locate odor sources. Another possible odor localization mechanism may be to compare the odor concentrations between the two nostrils, tentacles, or antennas. Observations from humans [45] to honeybees [39] indicate that a 10% difference in odor concentrations between the two sides is enough to make bilateral behavioral judgement on the odor source direction. It was shown that damaging one tentacle or antenna induces the experimental animals to turn towards the undamaged side in search of the odor source [39,46], supporting the stereo-olfactory mechanism. Many animals, for example, hammerhead sharks, have their two nares widely separated [44] presumably for the purpose of odor source localization. The two olfactory bulbs, each receiving inputs from the corresponding epithelium, interact with each other by connections from the mitral cells to the contralateral granule cells via anterior olfactory nucleus [47]. Therefore, the two olfactory bulbs may be the first stage where the stereo-information gets processed.

On the other hand, it has also been observed that some animals can still localize odors after one tentacle is damaged [46], or after the animal overcomes the initial confusion caused by the damaged tentacle [39]. Temporal analysis of odor concentrations is presumably used and sufficient in many cases. It is conceivable that temporal fluctuation analysis and stereo-olfaction are used in parallel or complementarily in many animals. Imagine a situation, for instance, when there are two spatially separated sources with identical odor substance (e.g., two pheromone emitting females for a

male animal), they would be seen by present early olfaction model as one single odor source whose time fluctuation is a superposition of the fluctuations from the two sources. The higher olfactory centers need to decouple the fluctuating sequence into two separate ones, maybe by filtering the signal with two different temporal filters. This requires additional knowledge of the environment [2], for example, maybe by using wind sensor signals and/or assuming different temporal characteristic time constants for the two component fluctuations. This problem is just as difficult, if not more, if a stereo-olfaction mechanism alone were used to localize the two sources. However, combining the temporal analysis and stereo-olfaction mechanisms would certainly give an edge, by maybe analyzing the correlations between the signals in the two nares — if there were only one source, the correlation would be different. Some animals swing their heads from side to side, or take a zig-zag path towards an odor source, sampling odor concentration at different spatial *and* temporal locations [39]. Such a maneuver suggests the use of a combination of spatial and temporal analysis, and would be more important for animals of smaller body sizes.

The first bulb model [3,4], subject to interruption by exhalation, will change its oscillation patterns and amplitudes as the odor mixture inputs fluctuate. It is not clear how it can simultaneously segment component identities and fluctuations in the mixed input fluctuating with a short characteristic time. In a typical outdoor environment, odor concentrations fluctuate with pulses of strong intensities lasting about several hundred milliseconds, separated by interpulses of low odor intensities lasting about five times longer [38]. In the case of low sensitivity and long time constants of the receptor neurons, the bulb may perceive one odor at each sniff in an environment of two dominant odor mixtures fluctuating independently, the other odor during its interpulse and the distant faint odors would be unnoticed. In this case, the component segmentation is approximately achieved by separate sniffs, the bulb could then in principle follow each odor fluctuation on the time scale of the sniff cycle. To incorporate stereo-olfaction, interbulb interaction should be introduced into the bulb.

Hopfield's model [2] should be sufficient for most olfactory environments even without stereo-olfaction, provided that the higher olfactory centers could disentangle the superposed concentration fluctuations. (This is analogous to individuals of only one functional eye, they segment the visual objects by imposing other constraints of the visual environment, e.g., motion and shape from shading [2].) Since the biologically realistic bulbs oscillate, it remains to be seen whether the temporal fluctuation analysis model performs well when both the excitatory and inhibitory cell interactions are included.

Instead of using a static neural activity pattern, our first model uses oscillatory patterns to code odor information, starting from an odor input that is comparatively much less oscillatory. The olfactory cortex, which receives inputs from the olfactory bulb, is also intrinsically oscillatory. An

intriguing possible explanation for the oscillatory nature of the cortex is that the oscillations may provide a natural mechanism for decoding its inputs, for example, by selectively resonating with inputs of specific amplitude and phase patterns as well as frequency. The second model codes the odor identity by synaptic strengths instead of neural activities, extending the role of synaptic strengths from merely algorithms of computation to information carriers. Recently, a model of the specialist olfactory system of insects [15] suggested coding by response duration, latency, and other variations in the temporal response for the input mixture ratios, input temporal profiles, and frequency, etc. which are behaviorally relevant. Coding by synaptic strengths, oscillation patterns [48], complex neural temporal structures [49], and using signal correlations for object binding or recognition could be used in many other computational contexts [2]. Interacting groups of excitatory and inhibitory neurons are ubiquitous in other cortical areas [1] and can be easily modeled by extending Eq. (6.1) [3].

The computational approaches used in the models are likely to apply to more than the olfaction. Adaptation to preexisting stimuli also happens in vision right at the first stage — the retina. The blood vessels in our eyes form a permanent and stationary image on our retina, we ourselves usually do not see them (although we can if we quickly flash a bright light very near our eyes) since our eyes have adapted to them in order not to be distracted to see other visual inputs. Adaptation in this sense can be seen as a mechanism to eliminate the temporal redundancy in the input information, so that output information is dominated by more recent sensory events. In Hopfield's model, adaptation is a mechanism to eliminate spatial redundancies (correlation) in the sensory input, by constructing matched filters for odors in the synaptic connections, resulting in uncorrelated outputs. In vision, for instance, the retinal output ganglion cells have center-surround receptive fields structure to respond to input spatial contrast [50] instead of input luminance, since input luminance at nearby receptors are correlated. Hence the ganglion cells have much less correlated activities than the photoreceptors. The matched filters or receptive fields indicate the knowledge acquired by the sensory system about the environment [51,52]. In Hopfield's model, the knowledge is the odor identity at that moment, in the retina it is the power spectrum of the input image ensemble [52].

Similarly, the olfactory perceptual shifts, such as cross-adaptation, have their analogy in vision as illusions and perceptual changes in special situations. One example is the “waterfall effect” [53]: after staring for a long time at a waterfall, stationary objects appear to move upwards. Another example is color hue shifts in perception after human subjects are exposed to an unnatural chromatic environment [54]. Theoretically, it has been proposed that the sensory after-effects are the results of efficient coding schemes for a new environment [53,55], and quantitative predictions of the post-adaptation color hue shifts have been compared with the measurements in psychophysical experiments with reasonable agreements [56]. We hope

that similar progresses in olfactory areas will occur when more systematic quantification of odor molecules and receptor codings become available. In both olfaction and vision, adaptation should not be thought of as fatigue, but a computational mechanism to reduce the spatio-temporal redundancy inherent in the input signals [51,52].

Perceptual constancy, such as the odor quality ( $S_{ki}$ ) captured by the synapses  $T$  in Hopfield's model, enables the observer to recognize objects and remain oriented in the environment despite changes in the physical stimulus. The odor source location, conveyed by the temporal profile of the concentration changes, should also be perceptually invariant irrespective of the concentration scale. A recent olfactory model of insects [15] observed such model neuronal responses depending only on the input profiles. There are also observations that perceived odor intensity stays the same for gentle and strong sniffs, but changes with the flow rate of odorized air, generated by olfactometers, in the nasal cavity [57], suggesting subjective compensations for the observation mode to achieve odor concentration constancy. Analogous constancies occur in vision and audition. For instance, color perception (spectral reflectance) of objects stays the same irrespective of the illuminance spectrum [58] (color constancy), recognition of visual objects should not depend on the locations and distances from the observer [59], and speech can be understood independent of the pitches of the speakers.

In summary, two olfactory bulb models have been presented as examples to discuss the computational tasks, environments, and mechanisms in olfaction. Studying olfaction can hopefully give us insights on the principles of computation in other sensory modalities.

*Acknowledgments.* I wish to thank J. Atick, S. Frautschi, A. Herz, J. Hopfield, M. Lyons, and M. Newton for carefully reading the draft and manuscript of this paper, and for very helpful comments, especially by J. Hopfield and M. Lyons.

## REFERENCES

- [1] G.M. Shepherd *The Synaptic Organization of the Brain*, Second Ed. 1979, Third Ed. 1990
- [2] J.J. Hopfield (1991) Proc. Natl. Acad. Sci. U.S.A. **88**:6462–6466
- [3] Z. Li, J.J. Hopfield (1989) Biol. Cybern. **61**:379–392
- [4] Z. Li (1990) Biol. Cybern. **62**:349–361
- [5] R.H. Wright (1964) *The Science of Smell* (Basic Books, New York)
- [6] K.E. Kaissling *Handbook of Sensory Physiology IV. Chemical Senses Part I*, L.M. Beidler (Ed.) pp. 351–431
- [7] W.J. Freeman (1978) Electroencephalogr. Clin. Neurophysiol. **44**:586–605

- [8] W.J. Freeman and K.A. Grajski (1987) *Behav. Neurosci.* **101**:766–777
- [9] G.M. Shepherd (1993) *Microsc. Res.* **24**:106
- [10] E.D. Adrian (1950) *Br. Med. Bull.* **6**:330–333
- [11] J.L. Davis, H. Eichenbaum (1990) *Olfaction — A Model System for Computational Neuroscience* (MIT Press, Cambridge)
- [12] G.M. Shepherd (1990) In: *Olfaction — A Model System for Computational Neuroscience*, J.L. Davis, H. Eichenbaum (Eds.) (MIT Press, Cambridge), pp. 225–250.
- [13] M. Wilson, J.M. Bower In: *Methods in Neuronal Modeling: From Synapses to Networks*, C. Koch, I. Segev (Eds.) (MIT Press, Cambridge), pp. 291–334
- [14] H. Liljenstrom (1991) *Int. J. Neural Sys.* **2**:1–15
- [15] C. Linster, C. Masson, M. Kerszberg, L. Personnaz, G. Dreyfus (1993) *Neural Comput.* **5**:228–241
- [16] J-C. Fort, J-P Rospars (1992) *C. R. Acad. Sci. Paris* **315**:331–336
- [17] G. Lynch, R. Granger (1990) In: *Olfaction — A Model System for Computational Neuroscience*, J.L. Davis, H. Eichenbaum (Eds.) (MIT Press, Cambridge), pp. 141–166
- [18] M.E. Hasselmo (1993) *Neural Comput.* **5**:32–44
- [19] E. Fransen, A. Lansner, Hl. Liljenstrom (1992) In: *Computation and Neural Systems 1992* (Kluwer Academic Publ., Dordrecht)
- [20] W.J. Freeman, C.A. Skarda (1985) *Brain Res. Rev.* **10**:147–175
- [21] T.V. Getchell, G.M. Shepherd (1978) *J. Physiol.* **282**:521–540
- [22] L. Buck, R. Axel (1991) *Cell* **65**:175–187
- [23] D. Lancet, C.A. Greet, J.S. Kauer, G.M. Shepherd (1982) *Proc. Natl. Acad. Sci. U.S.A.* **79**:670–674
- [24] W.J. Freeman, W.S. Schneider (1982) *Psychophysiology* **19**:44–56
- [25] E.C. Sobel, D.W. Tank. Preprint, AT&T Bell Laboratory
- [26] A. Gelperin, D.W. Tank (1990) *Nature* **345**:437–440
- [27] W.J. Freeman (1979) *Biol. Cybern.* **33**:237–247
- [28] K.R. Delaney, A. Gelperin, M.S. Fee, J.A. Flores, R. Gervais, D.W. Tank, D. Kleinfeld (1993) Submitted for publication
- [29] G.M. Shepherd (1988) Private communication
- [30] B. Johnson, R. Viogt, C. Merrill, J. Atema (1991) *Brain Res. Bull.* **26**:327–331
- [31] R.W. Moncrieff (1967) *The Chemical Senses*, Third Ed. (CRC Press)

- [32] F.H. Eeckman (1988) Statistical correlations between unit-firing and cortical EEG. Thesis, University of California, Berkeley
- [33] D.G. Laing, A. Glemarec (1992) *Physiol. & Behavior* **52**:1047–1053
- [34] M.A. Chaput, H. Panhuber (1982) *Brain Res.* **250**:41–52
- [35] W.S. Cain, T. Engen (1969) In: *Olfaction and Taste*, C. Pfaffmann (Ed.) (Rockefeller Press, New York), pp. 127–141
- [36] C.M. Gray, J.E. Skinner (1988) *Exp. Brain Res.* **69**:378–386
- [37] D.G. Laing, A. Mackay-Sim (1975) In: *Olfaction and Taste, V*, D.A. Denton, J.P. Coghlans (Eds.) (Academic Press, New York), pp. 291–295
- [38] J. Murlis, C.C. Jones (1981) *Physiol. Entomology* **6**:71–86
- [39] W.J. Bell, T.R. Tobin (1982) *Biol. Rev.* **57**:219–260
- [40] J.F. Hopfield, A. Gelperin (1989) *Behav. Neurosci.* **103**:329–333
- [41] O. Hendin, D. Horn, J.J. Hopfield (1993) Preprint; D. Horn (1993) Private communication
- [42] Simulations done by O. Hendin and myself (1993)
- [43] C. von der Malsburg, E. Bienenstock (1979) In: *Disordered Systems and Biological Organization*, E. Bienenstock, E. Fogelman Soulie, F. Weisbuch (Eds.) (Springer, Berlin), pp. 247–252
- [44] D.M. Stoddart (1980) *The Ecology of Vertebrate Olfaction* (Chapman and Hall, London)
- [45] G. von Bekesy (1964) *J. Appl. Physiol.* **19**:369–373
- [46] R.P. Croll (1983) *Biol. Rev.* **58**:293–319
- [47] Schoenfeld, Macrides (1984) *J. Comput. Neurol.* **227**:121
- [48] C. Gray, W. Singer (1989) *Nature (London)* **338**:334–337
- [49] J.W. McClurkin, L.M. Optican, B.J. Richmond, T.J. Gawne (1991) *Science* **453**:675–677
- [50] E.R. Kandel, J.H. Schwartz (1984) *Principles of Neural Science* 2nd Ed. (Elsevier, Amsterdam)
- [51] H.B. Barlow (1961) In: *Sensory Communication*, W.A. Rosenblith (Ed.) (MIT Press, Cambridge), pp. 217–234; (1961) *Current Problems in Animal Behavior*, W.H. Thorpe, O.L. Zangwill (Eds.) (MIT, Cambridge)
- [52] J.J. Atick (1992) *Network* **3**:213–251
- [53] H.B. Barlow (1990) In: *Vision: Coding and Efficiency*, C. Blakemore (Ed.) (Cambridge University Press, Cambridge), pp. 363–375
- [54] M.A. Webster, J.D. Mollon *Nature* **349**:235–238

- [55] H.B. Barlow, P. Foldiak (1989) In: *The Computing Neuron* (Addison-Wesley, Reading, MA)
- [56] J.J. Atick, Z. Li, A.N. Redlich (1993) Vis. Res. **33**:123–129
- [57] T. Engen (1982) *The Perception of Odors* (Academic Press, New York)
- [58] E.H. Land (1977) Sci. Am. **237**:108–129
- [59] Z. Li, J.J. Atick Neural Comput. In press

# Detecting Coherence in Neuronal Data

Klaus Pawelzik<sup>1</sup>

with 17 figures

**Synopsis.** Spatially and temporally coherent activities emerge in many neuronal systems. The analysis of such responses in most cases is based on simple correlation techniques which cannot detect nonlinear relationships. In this contribution I review new approaches for the nonlinear analysis of coherent activities in neuronal signals. On the one hand I present model free approaches which are based on the general notion of statistical dependency and which apply to the neurophysiological observables spike activities and local field potentials, respectively. These approaches quantify coherence in information theoretical terms and can help to characterize the underlying dynamics. I show that the contributions to statistical dependency can be analyzed in a time resolved way and present a new method which allows identification of coherent episodes in a background of stochastic activity. On the other hand I present a model-dependent approach which is particularly well suited for the analysis of neuronal assemblies exhibiting emergent burst activities. It assumes a simple form of the network dynamics for which the parameters can be directly determined from experimental spike trains. This Ansatz deals with the fact that observable spike trains only stochastically reflect an underlying network dynamics. Despite the mathematical simplicity of this approach it determines important characteristics like memory and switching behavior of the underlying network synchronization dynamics. The methods are illustrated by multiunit activity and local field potential data from the visual cortex of the cat. Both approaches independently reveal that these data reflect a network dynamics which switches between essentially two states, an oscillatory and a stochastic one. The episodes within which oscillations and synchronizations occur are identified with high time resolution by either method from local field potentials as well as from multiunit activities. It turns out that synchronization across a cortical distance is quite a rare event and occurs within temporally coherent episodes only.

---

<sup>1</sup>Institute for Theoretical Physics, University of Frankfurt, Robert-Mayer-Str.  
8-10, D-60325 Frankfurt/Main, Germany.

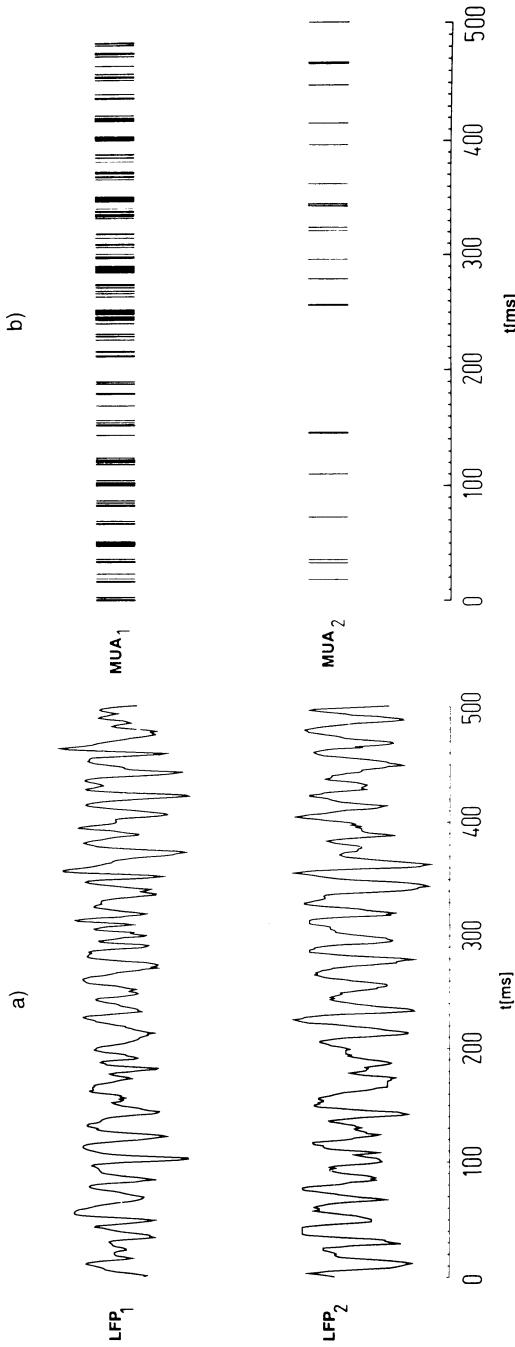
## 7.1 Introduction

Coherence is a concept which is of fundamental importance in scientific fields dealing with fluctuating quantities. While the term coherence quite commonly denotes the existence of relationships between or within variables of a system, its precise meaning varies strongly from field to field or even within some fields.

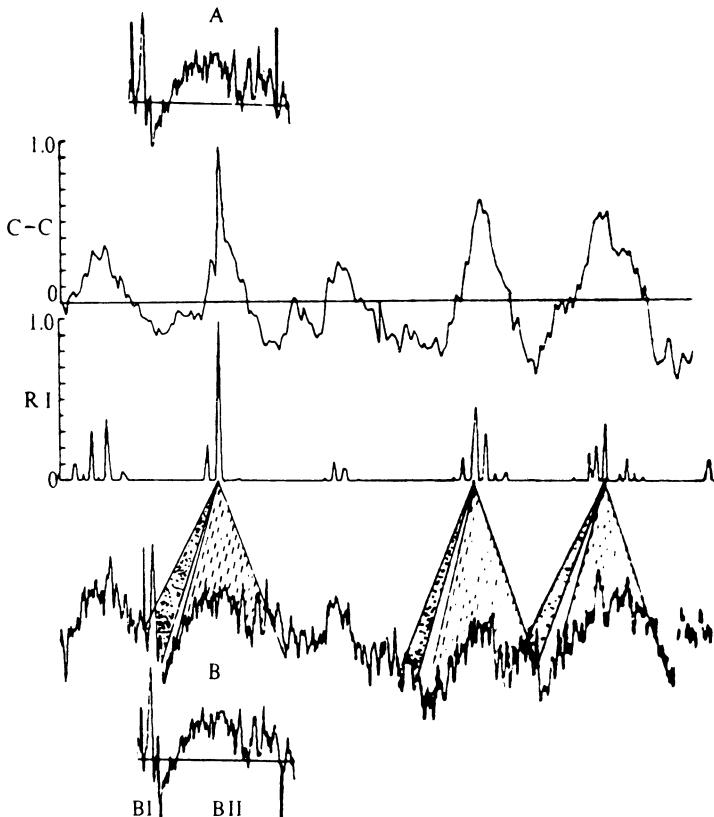
In brain science this lack of a generally accepted definition of coherence stands in contrast to the widespread belief that the detection and analysis of dependencies in neuronal activities forms the basis of an understanding of structural and functional aspects of the brain [1]. In particular one suspects that dependencies in neuronal responses reflect the layout of neuronal connectivities [2] and one expects that the search for coherent activities can help to analyze the data format of neuronal representations, that is, identify neuronal codes [3]. This latter potential role of coherent neuronal activity for brain function recently regained increased attention when Gray et al. [4] and others [5] found oscillatory responses in cat visual cortex which synchronized across a cortical distance when certain Gestalt properties were present in the stimuli. These findings backed earlier considerations of v.d. Malsburg [6] who proposed that a time structure of collective discharges of neuronal groups could be used to encode relations between different features, in this way providing an unambiguous representation of an object.

Figure 7.1 gives an impression of the observables that were measured in these experiments. It appears that the local field potentials (LFPs) as well as the multiunit activities (MUAs) are strongly fluctuating with only little obvious structure. The coherence within and the mutual coherence of these data have been termed “oscillatory” responses and “synchronicity,” respectively, states which have been established by means of correlation functions. These methods are well established and their significance can be controlled [1]; however, by definition they detect only specific dependencies which usually are linear. Another shortcoming of correlation methods is that they provide only little insight into characteristic properties of the underlying system, for example, the amount of memory involved. Furthermore correlation functions strongly rely on homogeneity of the data, that is, the constancy of the relations over time. The latter difficulty can in part be overcome for spike data if the coherency of the activities varies with a fixed phase relation to the stimulus [2]. If, however, coherent activities spontaneously emerge without fixed phase relation to a stimulus they are difficult to analyze with methods based on correlation functions. A time resolved determination of coherence has nevertheless been performed on local field potential data from cat visual cortex [7], with the main result that these signals are neither a permanent oscillation nor is there an ongoing synchronization. Rather the analysis indicate that the oscillatory phases as well as the synchronicity are very transient phenomena.

The notion of coherence which is related to correlation functions implies



**Fig. 7.1.** (a) Local field potentials (LFP) and (b) multiunit activities (MUA) from the visual cortex of the cat. An anesthetized cat was shown a long light bar which simultaneously crossed the receptive fields of the neurons at both electrodes denoted by 1 and 2.



**Fig. 7.2.** A method for recognition of expected temporal patterns (taken from [8]). In addition to simply correlating the templates (A and B) with the signal (C - C) the approach presented in this paper from 1972 already exploits the predictability of the temporal structure which leads to a temporally precise location of the template (RI).

order, for example, in the sense of periodicity or synchronicity and the like. This is in considerable contrast to the approach to coherence which I propose in this paper. The basic idea here is to consider events as coherent if their occurrence can be predicted to some extent on the basis of an *a priori* knowledge.

A simple example for this is given by the knowledge of the pattern itself. This has already been exploited in a paper published in 1972 where a nonlinear correlation technique was presented which somehow exploited the predictability of a signal containing the otherwise arbitrary pattern (Fig. 7.2).

The caveat with this method, however, is that it quantifies coherence only relative to an external *a priori* expectation, that is, it cannot detect unknown patterns which nevertheless might be predictable.

It is therefore necessary to have a notion of predictability which either relies exclusively on the statistical properties of the gathered data or additionally on general assumptions about the structure of the underlying system. These ways of exploiting knowledge about the system point to the two complementary strategies which I propose for the detection of coherence in this paper: model-free and model-dependent approaches.

In the model-free approach to coherent neuronal responses I exploit the notion of statistical dependency and quantify the corresponding degrees of predictability in terms of information theoretical quantities. I present a method based on this approach which not only quantifies the average degree of coherence in LFP data but also characterizes the dynamics of the underlying system. In particular my method not only determines the average degree of coherence, but furthermore estimates characteristic properties of the underlying dynamics as, for example, the amount of memory involved and identifies coherent episodes in the data.

The model-dependent approach serves essentially the same purposes as the model-free method, however, on the basis of MUA's. Its application, however, is restricted to the phenomenon of network burst synchronizations of the type observed in the cat as mentioned above. It assumes a renewal dynamics for the network and a stochastic coupling of individual neurons to the network states. Despite its simplicity it turns out that this model is phenomenologically adequate for the description of the dynamics observable in multiunit activities, and it leads to results concerning memory, switching, and synchronization of the network dynamics similar to those obtained from the model-free approach.

The paper is organized as follows: In Sec. 7.1.1 I review some notions of information theory which are illustrated with an application to unit correlations (correlograms). Here I show that the quantification of the coherence of spike events using mutual information naturally decomposes the overall dependencies into stimulus-locked and spontaneously emerging contributions. In Sec. 7.2 a model-free approach for the determination of coherence in continuous signals is introduced. It is based on the time resolved mutual information for which I present an efficient algorithm. This method is then applied to local field potential data from the cat visual cortex in Sec. 7.3 where lower bounds for the memory of the system are estimated and in which temporally and spatially coherent episodes in the data are identified. I complement this model-free approach with a hidden state model in Sec. 7.4. In Sec. 7.5 the parameters of this stochastic model for the network dynamics are extracted from the multiunit activities of the same experiment. Despite its simple structure this model then suffices for an analysis of the underlying system. The results are summarized and finally discussed in Sec. 7.6.

### 7.1.1 CORRELATIONS AND STATISTICAL DEPENDENCY

Coherence within and between experimental signals usually is quantified by means of correlation functions. Let  $x(t)$  and  $y(t)$  be two signals with zero mean from different aspects of a system, for example, from two electrodes. Correlations then are given by

$$C_{ij}(\tau) = \langle x^i(t)y^j(t+\tau) \rangle. \quad (7.1)$$

Here  $\langle \dots \rangle$  denotes the average over time  $t$  and  $i, j$  are integer exponents. In case the joint event  $\mathbf{z}(t, \tau) = (x(t), y(t + \tau))$  has an ergodic measure  $\rho(\mathbf{z})$  the time average can be replaced by an average over state space:

$$C_{ij}(\tau) = \int \rho(\mathbf{z}) x^i(t) y^j(t + \tau) d\mathbf{z}. \quad (7.2)$$

Correlations can be understood as estimators of predictability. On the other hand they are the contributions of order  $(i, j)$  of a model for the signal. Furthermore it is well known that the knowledge of all correlations  $C_{ij}$  allows the calculation of the density  $\rho$ . In practical cases, however, one can estimate only the lower moments (i.e.,  $i, j$  small  $\leq 3$  say), mainly because of problems of reliable estimation of the higher moments from finite data sets [9]. Taking into account only the contributions  $i = j = 1$  corresponds to a linear model for which the correlation function is equivalent to the power spectrum, while the higher moments can in principle be used for the modeling of nonlinearities [9].

The point here is to note that the use of only a small number of moments corresponds to simplifying assumptions about the system. In this section, however, I am interested in the detection of arbitrary dependencies without a priori assumptions about the underlying systems, and I therefore propose a model-free approach.

The basic idea is to directly analyze the distribution  $\rho(\mathbf{z})$  for statistical dependency. Clearly the distribution  $\rho(\mathbf{z})$  reflects statistical dependency of  $x$  and  $y$  if  $\rho(\mathbf{z})$  deviates from the null hypothesis that it factorizes, that is,  $\rho(\mathbf{z}) \neq \rho_x(x)\rho_y(y)$  where  $\rho_x(x) = \int \rho(\mathbf{z}) dy$  and  $\rho_y(y) = \int \rho(\mathbf{z}) dx$ .

A well known quantity measuring this deviation in an averaged way is the mutual information

$$M = \int \rho(\mathbf{z}) \log\left(\frac{\rho(\mathbf{z})}{\rho_x(x)\rho_y(y)}\right) d\mathbf{z} \quad (7.3)$$

which was introduced by Shannon in 1946 [11]. The mutual information is a special case of the Kullback–Leibler information [12]

$$K = \int \rho(\mathbf{z}) \log\left(\frac{\rho(\mathbf{z})}{\rho_0(\mathbf{z})}\right) d\mathbf{z} \geq 0 \quad (7.4)$$

a quantity which evaluates the deviation of a distribution  $\rho(\mathbf{z})$  from an arbitrary expectation represented by a normalized density  $\rho_0(\mathbf{z})$  which represents a priori knowledge.  $K = 0$  holds only when  $\rho(\mathbf{z}) = \rho_0(\mathbf{z})$  everywhere. Besides the positivity of these measures, the additivity with respect to different a priori measures motivates the use of the logarithm and justifies the name information:

$$\begin{aligned} K &= \int \rho(\mathbf{z}) \log\left(\frac{\rho(\mathbf{z})}{\rho_0(\mathbf{z})}\right) d\mathbf{z} = \int \rho(\mathbf{z}) \log\left(\frac{\rho(\mathbf{z})\rho_1(\mathbf{z})}{\rho_1(\mathbf{z})\rho_0(\mathbf{z})}\right) d\mathbf{z} \\ &= \int \rho(\mathbf{z}) \log\left(\frac{\rho(\mathbf{z})}{\rho_1(\mathbf{z})}\right) d\mathbf{z} + \int \rho(\mathbf{z}) \log\left(\frac{\rho_1(\mathbf{z})}{\rho_0(\mathbf{z})}\right) d\mathbf{z}. \end{aligned} \quad (7.5)$$

Although these measures and their convenient properties have long been well known, their potential for the analysis of dependencies in neuronal data has not always been recognized. One reason for this may be the suspicion that these measures are of little use in practice because of low statistical significance when estimated from experimental data. In fact, the estimation of densities from restricted amounts of data in general is a nontrivial problem. Recently, however, new methods have been developed in the field of nonlinear dynamics that allow a reliable estimation of mutual information even from small amounts of data and can also be applied when the states have dimensions larger than 2 [13]. While one such method will be discussed in the next section, in the following we demonstrate the potential of information theoretical notions for the quantification of coherence with the example of spike-spike correlations.

The dependency of events  $s_i^1, s_i^2 \in \{0, 1\}, i = 1, \dots, T$  in spike trains from two measurements usually is estimated by a counting statistics, the so-called correlogram

$$C_{11}(\tau) = \sum_{i=1}^T \delta_{s_i^1 s_{i+\tau}^2} \quad (7.6)$$

(for a review of correlation methods applied in the neurosciences see [1]).

If the spike trains  $s^x$  and  $s^y$  are from measurements taken simultaneously at different locations in the brain this measure is called “cross-correlogram” and I denote it by  $C^c$ . To estimate correlations within a spike train one uses  $s^1 = s^2$  in which case  $C$  is called auto-correlogram denoted here by  $C^a$ . A third application of correlograms is the estimation of correlations across trials, that is, the correlations which correspond to responses which appear with a fixed temporal relation relative to the stimulation. For this purpose one uses time series  $s^x$  and  $s^y$  which are from different repetitions of the same experiment such that the time index  $i$  corresponds to a time axis which is defined relative to a stimulation identical for each trial. In this case  $C$  is called “shift predictor” which I denote by  $C^s$ .

Such correlograms are used in many analyzes of neuronal activities as, for example, for the detection of oscillations in responses from cat visual

cortex. These data are interesting because here the shift predictors are rather flat while cross-correlograms from multiunit activities at different cortical locations are modulated (Fig. 7.1). This fact, among others, was claimed to indicate a spontaneous emergence of oscillations in the underlying neuronal substrate which in turn were considered to be functionally relevant as already mentioned in the Introduction.

The evaluation of correlograms in the simple form (Eq. (7.6)), however, does not give a quantitative account of the amount of coherence. Furthermore a large part of the correlations reflected by Eq. (7.6) can be expected on the basis of the firing rate and there may also be contributions to the cross-correlogram which correspond to activities occurring with a fixed temporal relationship to the stimulus onset which is obvious from a comparison of  $C^c$  and  $C^s$ . It is therefore desirable to have a quantitative interpretation of the spike–spike correlations and of its respective components. Several measures for these contributions have been proposed in the literature [1], some of them are discussed in [11]. Instead of reviewing these approaches in detail I here introduce an information theoretical alternative which has the desired properties: quantification of the dependency of spike events and decomposition of the contributions from spontaneous activities and from stimulus locked responses.

In a first step consider estimations of the joint probability for events  $s_t^1 = 1$  occurring at time  $t$  together with events  $s_{t+\tau}^2 = 1$  at a time  $t + \tau$ , that is, normalize the correlogram:

$$P_{11}(\tau) = \frac{1}{T} C_{11}(\tau), \quad (7.7)$$

and thereby obtain a quantity which is no longer explicitly dependent on the length of the time series. For low firing rates clearly the spikes are the most important events and information theoretical notions could be directly applied to  $P_{11}$  with a probability of  $1 - P_{11}$  for the nonoccurrence of the event (1, 1). In principle, however,  $s = 0$ , that is, no spike is an event and therefore in the general case one has to consider the distribution over the complete state space, that is, one has to estimate the normalized correlograms  $P_{s_1 s_2}(\tau)$  for all events  $(s^1, s^2)$ . In many cases it turns out that this extension of the state space captures a significant amount of the total statistical dependency which otherwise could not be detected.

Having estimated the joint probabilities  $P^c$  and  $P^s$  corresponding to the correlograms  $C^c$  and  $C^s$  from given time series it is straightforward to apply the information theoretical notions reviewed above. The total amount of spike–spike (better: event–event) coherence is given by the mutual information

$$M^T(\tau) = \sum_{s^1, s^2} P_{s_1 s_2}^c(\tau) \log\left(\frac{P_{s_1 s_2}^c(\tau)}{P_{s_1}^1 P_{s_2}^2}\right), \quad (7.8)$$

where  $P_{s^2}^1 = \sum_{s^2} P_{s_1 s_2}^c$  and  $P_{s^2}^2(\tau) = \sum_{s^1} P_{s_1 s_2}^c$  are the corresponding rates. Note that for data in which firing rates vary strongly, which is, for

example, the case at the stimulus onset, taking  $P_{sv}^y(\tau)$  instead of a constant rate helps to suppress systematic shifts.

Suitable for the evaluation of the contributions from spontaneous events is the Kullback information [12] which compares the distribution  $P^c$  with the null hypothesis that the contributions appear in a stimulus locked manner:

$$M^S(\tau) = \sum_{s^1, s^2} P_{s^1 s^2}^c(\tau) \log\left(\frac{P_{s^1 s^2}^c(\tau)}{P_{s^1 s^2}^s(\tau)}\right). \quad (7.9)$$

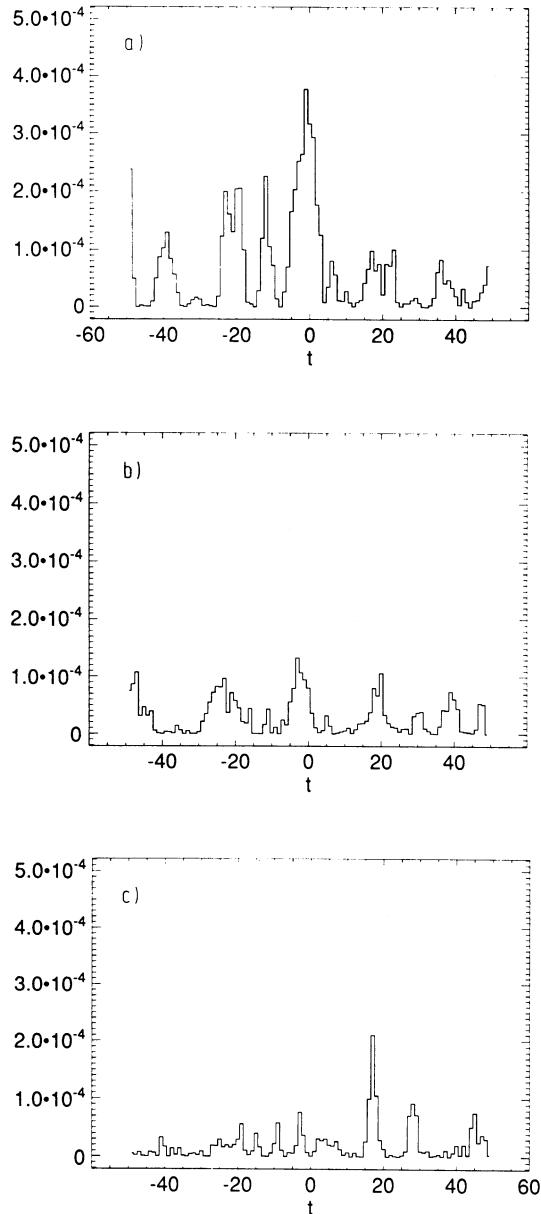
$M^T$  and  $M^S$  are always positive while the rest  $M^R = M^S - M^T$  can be negative, depending on which a priori measure contains more information relative to  $P^c$ :  $P^1 P^2$  corresponding to the Poisson assumption or the shift predictor  $P^s$ . Generally the shift predictor is expected to contain more information than the rates, that is,  $M^R > 0$ ; for finite data sets, however,  $M^R$  can sometimes become negative mainly because of a lack of statistics.

The results of an analysis of spike trains from the visual cortex of the cat with this method is shown in Fig. 7.3 (compare [4]). For a continuous long light bar the amount of information shared by spontaneous and synchronous events is significantly larger than for the case of stimulation with two shorter bars. For stimulations with bars moving in opposite directions these numbers become negligible indicating that there are no spontaneous synchronizations in this case despite a comparable firing rate.

It should be noted here that these results are not new in a qualitative sense. However, the information theoretical evaluation of correlations in spike trains helps to make quantitative claims about the relative amounts of coherent activity and therefore enables a comparison of the coherence emerging in different experiments or even different animals. Especially when firing rates vary strongly for different stimulus conditions, the method presented above gives a quantitative account of the absolute amount of spontaneously emerging coherent activities  $M^S$  in contrast to the relative contributions which usually are considered when regarding correlograms. Nevertheless, relative contributions can be estimated from the information theoretical quantities, for example, using  $M^S/M^T$ .

## 7.2 Time Resolved Detection of Coherence

The interpretation of event–event correlations as exemplified with spike trains above is only one example considering the applications of information theoretical notions in brain science. During the past decades information theoretical measures have been extensively applied, in particular for the analysis of neuronal codes [1, 14, 15]. This field is still developing, as is nicely demonstrated by a recent publication [16] in which the mutual information



**Fig. 7.3.** Not stimulus locked contributions (in natural units) to the mutual information of the spike trains from [4] under three stimulus conditions. (a) One long light bar moving across both receptive fields; (b) two short light bars moving colinearly over the respective receptive fields; and (c) two short light bars moving in opposite directions over the receptive fields.

between input signal and spike response is used to show that single sensory neurons essentially perform an inverse linear filtering with an astonishing precision ( $\geq 250$  bits/s).

Here, however, I am interested in another aspect of information theoretical methods, namely their potential for characterizing the dynamics. Some of these methods were developed in nonlinear dynamics, for example, for the determination of fractal dimensions of chaotic attractors [17] or for the identification of the number of independent degrees of freedom in a chaotic signal from a deterministic system [18]. For a stochastic system the notion of degrees of freedom is not adequate, but it can be replaced by the more general notion of memory. Because there has been only little evidence that chaos is relevant for the understanding of neuronal dynamics in the following only stochastic dynamical systems are considered. For these a method is presented which estimates the memory from measured signals. This method is equally useful for deterministic systems as well [13].

### 7.2.1 STATISTICAL DEPENDENCY OF CONTINUOUS SIGNALS

For spike trains the most straightforward strategy for an estimation of memory would be to extend the state space and to consider not only the values  $s_t$ , but instead, a sequence of events:

$$\vec{s}_t = (s_t, s_{t+1}, \dots, s_{t+m-1}) \quad (7.10)$$

for the construction of the state space. Memory then would be present in the data if the mutual information  $M(m)$  between  $\vec{s}_t$  and a subsequent event  $s_{t+\Delta t}$  ( $\Delta t \geq m$ ) increased with increasing  $m$ . The memory would be determined to be  $m^*$  if  $M(m^*) = M(m)$  for  $m \geq m^*$ .

The caveat with such an approach is that the amount of data necessary for a reliable estimation of probabilities grows exponentially with memory when statistical independency is to be detected. This is obvious because in this case the number of states with finite probability increases as  $2^{m+1}$  and one therefore would require time series of length  $T2^m$  when  $T$  is the length of a time series which is sufficient for a reliable detection of independence in the memoryless case where  $m = 1$  is sufficient. While  $T$  is rather large in case of spike trains for continuous signals the situation can be less dramatic because of the much higher information rates there.

Therefore an analysis of the memory with information theoretical methods is more promising with continuous signals like EEG or local field potentials than with discrete spike trains. In the following I present a method for the estimation of mutual information of continuous signals [13,19] which was originally developed in the field of nonlinear dynamics.

Let  $\xi_t, \chi_t \in \mathbb{R}$  denote different time series (e.g., of LFPs from two electrodes) of (quasi-) continuous observables, where  $t = 1, \dots, N$  denotes the

time index. Then I define “states” (or “patterns”) as vectors of successive values (in nonlinear dynamics known as delay coordinates)

$$\mathbf{x}_t = (\xi_t, \xi_{t+\tau}, \dots, \xi_{t+(m-1)\tau}), \quad (7.11)$$

$$\mathbf{y}_t = (\chi_t, \chi_{t+\tau}, \dots, \chi_{t+(n-1)\tau}). \quad (7.12)$$

While for deterministic systems the value of  $\tau$  can be chosen optimally [20], for an unknown source at least the sampling theorem should not be violated, that is,  $\tau < 1/2\nu$ , where  $\nu$  is the highest relevant frequency in the signal.

In a next step defined by  $\mathbf{z}_t = (\mathbf{x}_t, \mathbf{y}_t)$  the compound states for the co-occurrence of the states  $\mathbf{x}_t$  and  $\mathbf{y}_t$ . In order to calculate the mutual information between  $\mathbf{x}$  and  $\mathbf{y}$  one has to coarse grain the state spaces in a suitable way. An efficient coarse graining is given by (hyper-)boxes of a fixed size  $\varepsilon$  around the states. For the computation of statistical dependency within one time series choose  $\mathbf{y}_t = \mathbf{x}_{t+\Delta t}$  with  $\Delta t \leq m\tau$ , and for the determination of spatial coherence construct  $\mathbf{x}_t$  and  $\mathbf{y}_t$  from the two time series, respectively. With these ingredients one can now estimate the probabilities  $p$  of falling into a box with the temporal index  $i$  in the state spaces:

$$\begin{aligned} p_i^{\mathbf{x}} &= \frac{1}{N} \sum_{t'=1}^N \Theta(\varepsilon - |\mathbf{x}_t - \mathbf{x}_{t'}|), \\ p_i^{\mathbf{y}} &= \frac{1}{N} \sum_{t'=1}^N \Theta(\varepsilon - |\mathbf{y}_t - \mathbf{y}_{t'}|), \\ p_i^{\mathbf{z}} &= \frac{1}{N} \sum_{t'=1}^N \Theta(\varepsilon - |\mathbf{z}_t - \mathbf{z}_{t'}|), \end{aligned} \quad (7.13)$$

where  $\Theta$  is the Heaviside step function and  $|\dots|$  denotes the maximum norm. This procedure is equivalent to the calculation of correlation integrals which are used to determine the fractal dimensions of strange attractors.

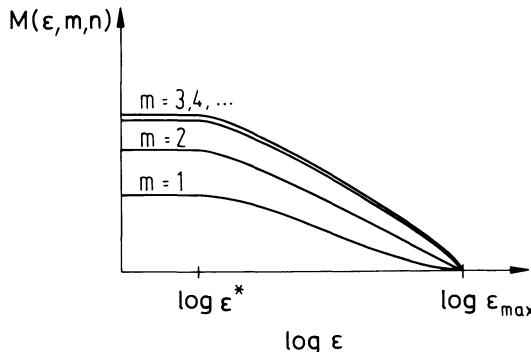
The mutual information is then given by

$$\mathbf{M} = \frac{1}{N} \sum_{t=1}^N M_t \quad (7.14)$$

with

$$M_t = \log \frac{p_i^{\mathbf{z}}}{p_i^{\mathbf{x}} p_i^{\mathbf{y}}}. \quad (7.15)$$

This definition might seem to differ from Shannon’s original formula because the state space average is replaced by a time average. When probabilities are estimated by counting points within boxes  $t$  (e.g., the box which is



**Fig. 7.4.** Dependency of mutual information on resolution  $\epsilon$  and memory  $m$ .  $M$  decreases monotonously with increasing  $\epsilon$  and for a noise level  $\epsilon^*$   $M$  becomes constant when  $\epsilon < \epsilon^*$ . For  $n = 1$  the dependency on  $m$  reflects the memory and if for  $n = m$  there is a scaling with a (negative) slope  $D_1$  this method identifies the presence of a global strange attractor [21].

visited at time  $t$ ) of a fixed partition of state space, these definitions, however, are equivalent because a given box contributes to the average as often as it is visited by the trajectory. Instead of a fixed partition, Eq. (7.14) applies a partition of state space into boxes which are centered around reference points  $z_t$  on the trajectory. These boxes can overlap and thereby provide better statistics than a fixed partition. Furthermore this overlap provides an average over the possibly different values of mutual information which one may obtain for fixed partitions (box counting) in which case  $M$  can in an unsystematic manner depend on the precise location of partition borders.

In principle  $\epsilon$  should be taken small in order to obtain the total amount of mutual information. An arbitrarily fine coarse graining, however, is not possible for finite amounts of data, and in addition,  $M$  may not even be well defined for  $\epsilon \rightarrow 0$ . The latter is, for example, the case if the support of  $P^Z$  does not have the same dimension as the direct product of the supports of  $P^X$  and  $P^Y$ . This is especially clear for identical random numbers which lie on the diagonal  $x_t = y_t$ . In this case  $M$  diverges like  $M \propto D \times \log(1/\epsilon)$  where  $D = 1$  is the dimension of the diagonal line. Besides, this effect makes the mutual information an important tool for estimations of fractal dimensions  $D$  in systems where several variables can be measured simultaneously [13,21].

Another reason for a careful consideration of the dependency of  $M$  on the discretization of the state space is the presence of additive noise of amplitude  $\epsilon^*$ . In this case no structures can be further resolved by using a finer graining and  $M(\epsilon)$  therefore does not increase for  $\epsilon < \epsilon^*$ . This discussion of the dependency of  $M$  on memory and discretization is summarized schematically in Fig. (7.4).

### 7.2.2 DETECTION OF COHERENT PATTERNS OF ACTIVITIES

A significant amount of average mutual information indicates *that* there is coherence within or between neuronal activities. It does not, however, tell *which* episodes in the signals are actually contributing. This question makes sense when the system is nonstationary. A particularly simple case of nonstationarity is a switching, either of the system parameters or of the couplings between different subsystems [22]. Another interesting kind of nonstationarity is the presence of temporal patterns in an otherwise noisy signal, a situation which, for example, is the case in speech data.

As an example for such a switching consider binary events  $s_t^i \in \{0, 1\}$  occurring at time  $t$  at two locations  $i = 1, 2$  (one could think of two coins flipped at equal times). Let the events be equally probable and independent of each other until a time  $t_0$ . Then let some coupling set in, such that the events become dependent, and let this coupling last until a time  $t_1$  say. While the following arguments hold for arbitrary coupling, I here only discuss  $s_t^1 = s_t^2, t \in [t_0, t_1]$ , that is, maximal coupling without loss of generality. The evaluation of the average mutual information of these time series will show that there is some statistical dependency and that no memory effects are present in this case. That there is some dependency of these time series will be reflected also by the linear correlation  $P_{11}$  in this case. If, however, one time series depends on a sequence of events in the other time series, and if the dependency is nonlinear it may be difficult to detect because the contributions to correlation functions of low order can then be negligible. While this is a general reason to prefer information theoretical methods over correlation functions I will present here a method based on mutual information which can help to identify the switching points  $t_0$  and  $t_1$  with a much higher time resolution than can be achieved by correlation-based methods.

Leaving out the average in Eq. (7.14),  $M_t$  gives the contributions to the mutual information  $M$  from every time step  $t$ . These contributions correspond to the states the system occupies at times  $t$  and which may reflect internal parameters and/or couplings. At the first sight a method based on  $M_t$  seems to provide no particular advantage for the detection of couplings in the example of coupled coins as compared to a correlation technique (e.g., by considering  $C_t = (s_t^1 - 1/2)(s_t^2 - 1/2)$ ) because such a method also only provides labels for the possible combinations of events (= joint states  $\vec{s}_t = (s_t^1, s_t^2)$ ). For the above case, for example, one has only two different values  $M_t = M_a$  if  $s_t^1 = s_t^2$  and  $M_t = M_b$  if  $s_t^1 \neq s_t^2$ , values which depend on the relative duration of the coupling. Similar to  $C_i$ ,  $M_t$  will fluctuate strongly in the uncoupled case  $t_1 < t < t_0$  because then all states  $\vec{s} = (s^1, s^2)$  are equally probable while during the coupling  $M_t = M_a$  holds.

The most straightforward way to suppress such fluctuations is to average

contributions from times  $t$  over a time window of length  $m = 2n + 1$  which for the correlation reads

$$C_\tau^m = \sum_{t=-n}^n C_{t+\tau}. \quad (7.16)$$

This procedure has actually been applied for correlations of local field potentials in order to detect nonstationarities of temporal and spatial coherence in these data [7]. The main disadvantage of this method is that the systematic fluctuations are suppressed rather weakly ( $\propto 1/\sqrt{m}$ ) at the cost of a time resolution of order  $m$ .

In contrast to suppressing fluctuations by averaging the consideration of the contributions  $M_t$  opens a way to achieve a much higher time resolution. This is done by artificially increasing the dimension of the state space using vectors of temporal sequences which for the above example means taking

$$\vec{s}_t^1 = (s_t^1, s_{t+1}^1, \dots, s_{t+m-1}^1) \quad (7.17)$$

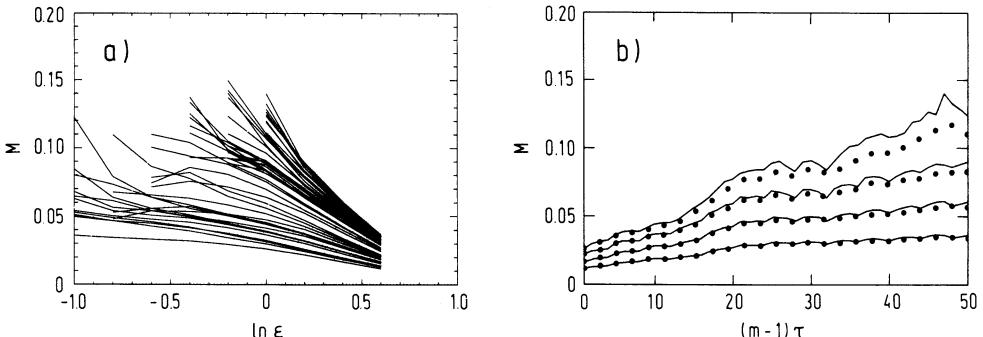
and

$$\vec{s}_t^2 = (s_t^2, s_{t+1}^2, \dots, s_{t+m-1}^2) \quad (7.18)$$

instead of the events  $s_t^{1,2}$  for the construction of the joint state  $\vec{s}_t = (s_t^1, s_t^2)$ . In the above example the number of states which can be met in the “uncoupled” situation grows like  $2^{2m}$  while the number of states which are visited during the coupling is only  $2^m$ . Therefore the probability to be in a state which corresponds to the coupling during the uncoupled phase decreases exponentially with  $m$ . A straightforward calculation shows that therefore also the fluctuations  $\sigma = \langle (M_t - M)^2 \rangle$  which are expected in an ensemble average decrease exponentially:  $\sigma \propto 2^{-m}$ . This means that one needs a much smaller value of  $m$  to obtain the same suppression of random fluctuations as compared to sliding average methods.

The deeper reason for this seemingly magic increase of resolution is the fact that for the estimation of the contribution  $M_t$  to the mutual information  $M$  an average over the whole time series has already taken place, namely the estimation of the joint probability  $P^{\vec{s}_t}$  with an accuracy which is only restricted by the total number of data points in the time series. This stands in contrast to sliding cross-correlation techniques which per definition only profit from a small fraction of the available data points which is of the order of  $m$ .

In more general cases (continuous-valued signals and lower degree of coherence) the resolution which can be achieved by including memory will be less dramatic; however, together with the ability of mutual information to detect nonlinear dependencies this method should be advantageous in these cases too [13]. Furthermore this method can also be combined with a sliding average without additional loss of time resolution if one uses a smoothing that has at most the temporal range of the memory included in



**Fig. 7.5.** Average mutual information  $M(\epsilon, m)$  vs  $\ln(\epsilon)$  (a) and vs  $m$  (b). Because of the relatively small number  $N = 10,000$  of data points the resolution of state space has to be chosen too coarse to detect the noise level. The dependence on memory  $(m-1)\tau$ , however, indicates memory effects of at least 25 ms irrespective of the choice of partition size ( $\ln(\epsilon) = 0.0, 0.2, 0.4, 0.6$ , from below) for  $\tau = 1$  (line) and  $\tau = 2$  (dots). A significant decrease of  $M$  was found for  $\tau \geq 5$ , in agreement with the sampling theorem (not shown).

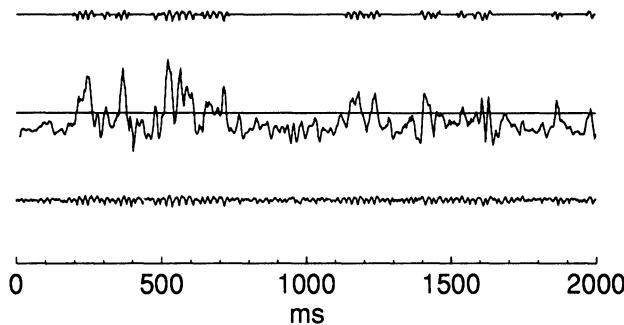
the construction of the states, a way to further suppress random fluctuations.

### 7.3 Memory and Switching in Local Field Potentials from Cat Visual Cortex

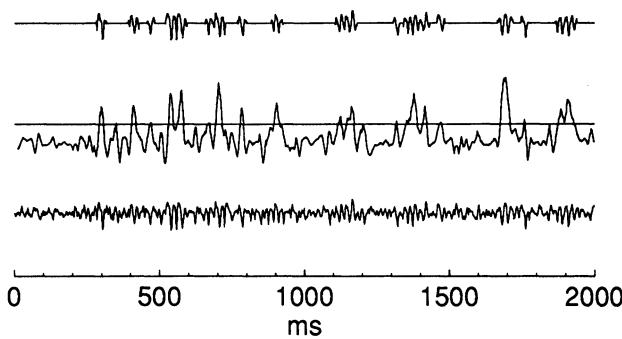
I now turn to the application of this method to LFPs of the long bar experiment [4] presented in the Introduction. The data of 10 repetitions of the stimulus presentation were used which provided  $N = 10,000$  values in total. Parts of the data are shown in Fig. 7.1.

#### 7.3.1 TEMPORAL COHERENCE OF LOCAL FIELD POTENTIALS

The method for estimating the mutual information is equivalent to evaluating correlation integrals which are used for the estimation of fractal dimensions of strange attractors too [24]. Applications of these algorithms to the LFP data reveal that no low-dimensional chaotic attractor is present [13]. Instead one finds that  $M$  increases most for  $(m-1)\tau \leq 25$  ms, indicating that the system is characterized by a stochastic process involving a memory of at least 25 ms (Fig. 7.5). Figures 7.6 and 7.7 show the results of a detection of activities coherent in time for two different electrodes. Plotted are 2000 ms of the data corresponding to two repetitions of the experiment. Periods of activity which lead to a value of  $M_t > 0.5$  were judged as tem-



**Fig. 7.6.** Analysis of temporal coherence in a time series of local field potentials. Plotted are (from below): the local field potential  $\xi_t$ , the time resolved mutual information  $M_t$  (computed with  $x_t = \xi_t$ ,  $m = 4$ ,  $\tau = 3$ ,  $y_t = \xi_{t+12}$ ,  $n = 4$ ), and the local field potential  $x_t'$  during the “windows of coherence” (that is, the values  $\xi_{t'}$  for which  $t' \in \{t' = t, t + 1, \dots, t + \Delta t/M_t > 0.5\}$ ), with the stochastic interplays being indicated by straight lines for  $\xi_t$ .

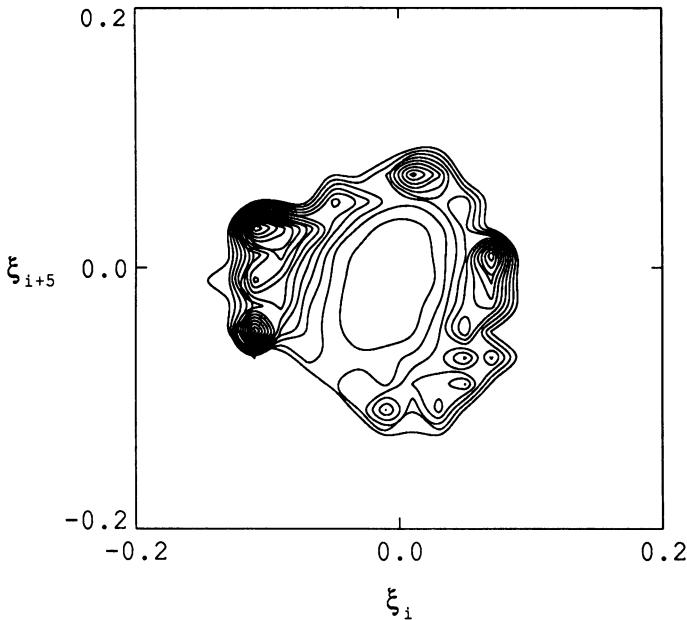


**Fig. 7.7.** As Fig. 7.6, but for the second electrode (i.e.,  $x_t = \chi_t$ ,  $y_t = \chi_{t+12}$ ).

porally coherent (this includes the width of the embedding time window). The methods indicate periods of increased temporal coherence, which are interrupted by incoherent periods.

The windows of coherence coincide with the parts of the data which a visual inspection would regard as oscillatory. Even though this aspect of the result seems rather trivial, I would like to remark that the results also rule out any hidden temporal coherence within the irregularly looking epochs. This comes about because the mutual information method does not rely on any specific waveform of the time series.

There are, however, large fluctuations in the contributions  $M_t$  and  $I$  therefore investigate the relation of the coherent episodes to the values of the LFP signal which is also a test for statistical significance. This is done by averaging the contributions  $M_t$  for the activities at times  $t$  and at times  $t' = t + T/4$ ; where  $T = 20$  ms is the period of a 50 Hz oscillation. If

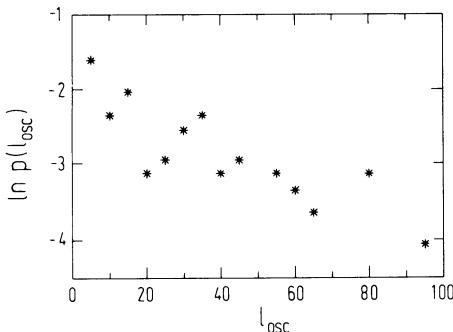


**Fig. 7.8.** Contour plot of the contributions to  $M$  averaged depending on pairs of successive values of the LFP  $\xi$  at time  $t$  and  $t+5$ . Contributions from a sinusoidal signal of period  $T$  would lie on a circle.

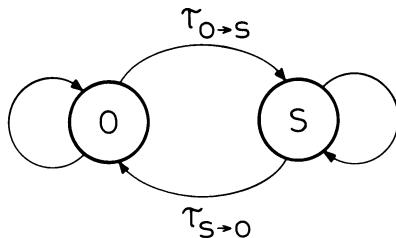
the fluctuations were random and not directly related to the activities at  $t$  and  $t'$  the average would disappear. If on the other hand the coherent episodes consisted of purely sinusoidal waves of 50 Hz one would see a circle in this representation (Fig. 7.8). In contrast to that one finds that the contributions to the mutual information averaged in this way are small following maximal amplitude (lower left corner) while they are positive during the rest of the cycle. This indicates a low degree of predictability for a short period of an otherwise periodic activity.

A preliminary interpretation of this picture is that there exists a cycle in the underlying system which is followed quite deterministically except for a rather short window. In this short epoch predictability is low, that is, the cycle can be left, and the system can switch into a state which is not coherent enough to contribute to our measure of predictability. The analysis of the multiunit activities which will be presented in Sec. 7.6 strongly supports this picture of the dynamics of the underlying neuronal system.

The nature of this switching is further investigated by estimating the length distribution of oscillatory (Fig. 7.9) and of the stochastic episodes (not shown) and one finds exponential tails of the distributions. This result indicates that the switching between coherent and stochastic phases of activity occurs at random with a decay time of 35 ms. A similar result is found for the stochastic episodes where the decay time was 75 ms (not



**Fig. 7.9.** Histogram of durations of oscillatory episodes. The decay appears to be exponential indicating a random switching from an oscillatory state with a decay time  $\tau_{o \rightarrow s} \sim 35$  ms.

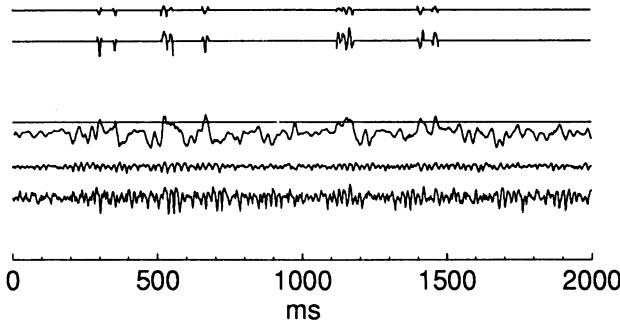


**Fig. 7.10.** A two state stochastic automaton illustrates the switching between phases of oscillatory (O) and phases of stochastic activity (S).

shown). On a more global level the system therefore seems to behave like a stochastic automaton with two states, a state of coherent/oscillatory behavior and a state of stochastic activity (Fig. 7.10).

### 7.3.2 SPATIAL COHERENCE OF LOCAL FIELD POTENTIALS

Figure 7.11 finally shows the amount of coherence of the activities of the two electrodes. Plotted are again 2000 ms of both LFP signals,  $M_t$  and the contributing windows (from below). Because the average mutual information for the LFPs from different cortical locations M is smaller indicating a weaker spatial coherence I here chose  $M_t > 0.2$  for the definition of the windows. All other parameters are as in Figs. 7.7 and 7.8. One finds that episodes of spatial coherence are relatively rare, and mostly appear where both LFPs are temporally very coherent (compare Figs. 7.7 and 7.8).



**Fig. 7.11.** *Results of an analysis of spatial coherence.* Plotted are (from below): 2000 ms of the LFP  $\chi$  at electrode 1, the corresponding LFP  $\xi$  at electrode 2, the time resolved mutual information  $M_t$  (computed with  $x_t = \chi_t, m = 4, \tau = 3, y_t = \xi_t, n = 4$ ), and contributing episodes (i.e.,  $x_{t'} = \chi_{t'}, y_{t'} = \xi_{t'}, t' \in \{t' = t, t+1, \dots, t + (m-1)\tau/M_t > 0.2\}$ ).

## 7.4 A Model-Dependent Approach

As demonstrated above, model-free approaches are useful for the identification of characteristic properties like memory or varying degrees of coherence in the dynamics of a neuronal system. Thereby one obtains global constraints to the class of models which are admissible for the description of the processes underlying the data. For a detailed comparison of specific models with the observed dynamics, however, such results might be too general.

At the other extreme one could attempt a direct comparison of a specific neuronal model with experimental data. Such a strategy, however, at best would be a fit of parameters and is in general unlikely to provide insights into basic mechanisms underlying possibly complex dynamics. Therefore I here pursue an intermediate approach. Motivated by the results from the model-free analysis I consider the system underlying the correlated responses in the cat visual cortex to be a stochastic process with memory. The mathematical framework which I propose here is very simple but nevertheless has enough structure to be used for a characterization of the memory, the switching between different phases of activity, and of the relation of single neurons to the overall network dynamics. The parameters of our Ansatz have an interpretation in neurophysiological terms. They characterize the observable spike activities from a small number of units as the outcome of a network dynamics which in itself is a Markov process to which the individual neurons are stochastically coupled. The analysis of a neuronal signal with this hidden Markov model goes beyond the single neuron doctrine in the sense that unit activities are considered to reflect a network dynamics. This aspect of our method can help to operationalize

the notion of a neuronal assembly which plays an important role in the discussion of distributed representations involving dynamics.

### 7.4.1 RENEWAL DYNAMICS

The stochastic dynamics of an isolated spiking neuron may be modeled by a renewal process [25]. For discretized neuron states such a description is equivalent to a stochastic automaton with a simple structure (Fig. 7.12(a)). In this model the neuron's state is given by the time since the last spike  $\phi$  and the parameters  $P_f$  are the probabilities for excitation, that is, for generating the next spike which can be derived from threshold dynamics [23]. This description neglects any memory in the neuron going beyond the last spike and is under this condition equivalent to the interspike interval distribution  $P_h$  with the relation

$$P_h(t) = P_f(t) \cdot \left[ 1 - \int_0^t P_h(t') dt' \right]. \quad (7.19)$$

From  $P_h(t)$ , the normalized autocorrelogram then can be calculated via

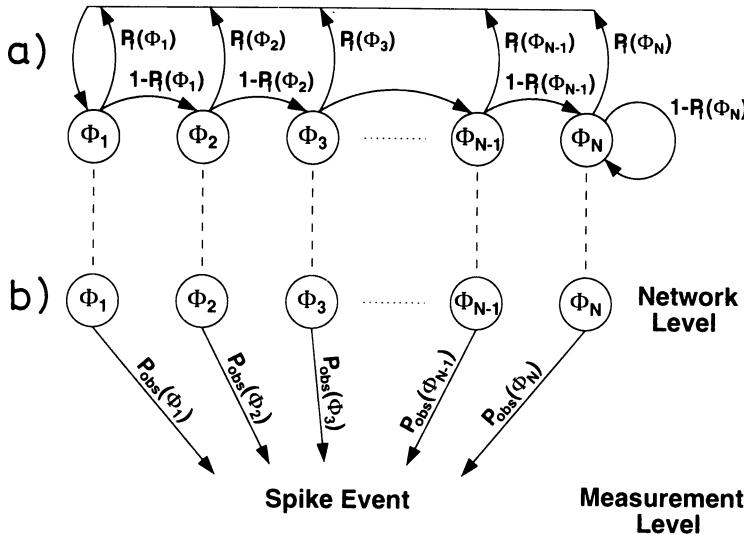
$$\hat{P}_{ii}(\tau) = P_h(\tau) + \int_0^\tau P_h(\tau) \hat{C}(\tau - t) dt. \quad (7.20)$$

For a simple threshold dynamics of a neuron  $P_f(\phi)$  has a monotonous and sigmoidal shape which reflects the refractoriness [23].

### 7.4.2 SWITCHING IN THE ASSEMBLY DYNAMICS

The Ansatz of a renewal process can also be applied for the description of the burst dynamics in networks of spiking neurons [23]. In that case  $\phi^b$  denotes the time since the last collective excitation and  $P_f^b(\phi^b)$  gives the probability for generating the next excitation (burst). Despite simple (e.g., refractory) elements the network dynamics can be more complex, as is reflected by a qualitative difference in the shape of the excitability  $P_f^b(\phi^b)$  for the network as compared to  $P_f(\phi)$  of the single neuron. For instance a pronounced maximum in  $P_f^b(\phi^b)$  at  $\phi^b = \tau_s$  indicates a switching between an oscillatory and a stochastic state in the system [28] which can emerge as a network property [26] and is hardly identifiable from the modulations in correlation functions alone [27]. Such a switching occurs with probability  $p_{o \rightarrow s}$  if the network fails to burst again (i.e., return to  $\phi_0$ ) before  $\tau_s$  and therefore goes into a state  $\phi^b > \tau_s$  with a reduced burst rate  $P_f$ . In order to calculate the switching probabilities I first introduce the stationary state distribution

$$\rho_s(\phi) = \rho_s(0) \exp \left[ - \int_0^\phi P_f(\phi') d\phi' \right] d\phi'. \quad (7.21)$$



**Fig. 7.12.** The hidden state model (discretized). (a) Renewal dynamics. States  $\phi$  are time steps since the last point event ( $\phi_1 = 0$ ). Parameters  $P_f(\phi)$  denote transition probabilities per time step for returns to  $\phi_1$ .  $P_f(\phi > T)$  is constant. (b) Network dynamics and measurement. The occurrence of a spike in a measurement is represented by the probabilities (per time step)  $P_{\text{obs}}(\phi^b)$ .

Clearly the probabilities to be in either macrostate are

$$P_{\text{osc}} = \int_0^{\tau_s} \rho_s(\Phi) d\Phi \quad (7.22)$$

and

$$P_{\text{sto}} = \int_{\tau_s}^{\infty} \rho_s(\Phi) d\Phi. \quad (7.23)$$

The transition rates between both are given by

$$P_{\text{osc} \rightarrow \text{sto}} = \frac{\rho_s(\tau_s)}{P_{\text{osc}}} \quad (7.24)$$

and

$$P_{\text{sto} \rightarrow \text{osc}} = \frac{\int_{\tau_s}^{\infty} P_f(\Phi) \rho_s(\Phi) d\Phi}{P_{\text{sto}}} = \frac{\rho_s(\tau_s)}{P_{\text{sto}}}. \quad (7.25)$$

The arbitrariness in choosing a proper  $\tau_s$  reflects the ambiguity in the allotment of the switching region to either phase, which a priori is not defined (for a more detailed discussion of this model see [28]).

### 7.4.3 NETWORK STATES ARE HIDDEN

A spike activity of an elementary neuron, however, only indirectly reflects the network excitation. In general, the neuron may remain silent or may

contribute with one or more spikes to a burst. Furthermore there may be some uncorrelated background activity independent of the network state. All this can easily be included into a hidden state model (HSM) by introducing the spike observability function  $P_{\text{obs}}(\phi^b)$  which denotes the probability for observing a spike when the network is in state  $\phi^b$  (Fig. 7.11(b)). Different neuron properties may then be reflected by different  $P_{\text{obs}}$ , for example, the average firing rate is proportional to  $P_{\text{obs}}$  leaving the time structure unchanged. While  $P_f^b$  and  $P_{\text{obs}}$  are useful by themselves for interpreting the network dynamics [23] I here only note that a given HSM predicts the statistical properties of spike trains. Let  $\{s_t^i\}, t = 1, \dots, T$  be spike trains from measurements at one or two locations  $i = 1, 2$ . Furthermore let the contributing neurons belong to the same network which is characterized by  $P_f^b$ . Then the correlograms  $P_\tau^{ij}$  are determined by

$$\hat{P}_{ij}(\tau) = \sum_{\phi} \sum_{\phi'} P_{\text{obs}}^i(\phi') \mathbf{M}_{\phi', \phi}^\tau P_{\text{obs}}^j(\phi) \rho_s(\phi), \quad (7.26)$$

where  $\mathbf{M}$  denotes the Markov matrix of the model. This formula predicts auto-correlograms ( $i = j$ ) on the basis of an underlying network renewal dynamics. A successful prediction of the cross-correlogram ( $i \neq j$ ) by Eq. (7.19) implies that not only the local dynamics is correctly characterized by the HSM, but furthermore that this dynamics is identical for the two locations, that is, that there is a unique neuronal assembly. The properties of the neurons which are localized at the electrodes ( $P_{\text{obs}}^1 \neq P_{\text{obs}}^2$ ), however, can nevertheless differ, thereby reflecting different degrees of excitability and coupling to the network state.

#### 7.4.4 MODEL IDENTIFICATION FROM MULTIUNIT ACTIVITIES

The hidden state model presented above belongs to the class of hidden Markov models. Hidden Markov models have the appealing feature that the set of parameters  $\lambda$  by which they are determined (in the case of the HSM):  $\lambda = (P_f^b, P_{\text{obs}})$  can be estimated from a signal by a well established method, the Baum-Welch algorithm [29]. Given an observation sequence  $O = O_1 O_2 O_3 \dots O_T$  and initial parameters  $\lambda$  this algorithm in each iteration step calculates a new parameter set  $\bar{\lambda}$  which increases the probability  $P(O|\lambda)$  that the HMM generates the sequence  $O$ :  $P(O|\bar{\lambda}) \geq P(O|\lambda)$ . In an intermediate step the probabilities  $\gamma_i(t) = P(O, \phi_i = q_t | \lambda)$  for all hidden state sequences  $q_1 q_2 q_3 \dots q_T$  with  $q_t \in \{\Phi_i | i = 1, \dots, N\}$  are computed on the basis of the previous and the subsequent observation sequences  $O = O_1 O_2 O_3 \dots O_t$  and  $O = O_t O_2 O_3 \dots O_T$ , respectively. The algorithm has converged if  $\bar{\lambda} = \lambda = \lambda^*$ . Due to the nonlinear optimization  $P(O|\lambda^*)$  may be just a local maximum and one has to decide whether  $\bar{\lambda}$  provides a sufficient approximation to the structure of the data by independent crite-

ria [28].  $P(O|\lambda)$  becomes exponentially small for long sequences, an effect which can be handled with an appropriate rescaling to avoid numerical failures. The algorithm can be extended to handle multiple independent sequences too, which is useful for taking advantage of repeated measurements of the same experiments for obtaining better statistics. As these methods are well established in speech recognition, I do not go into more detail here, but instead refer the reader to the literature (e.g., an excellent review of hidden Markov models can be found in [30]).

## 7.5 Memory and Switching in Multiunit Activities from Cat Visual Cortex

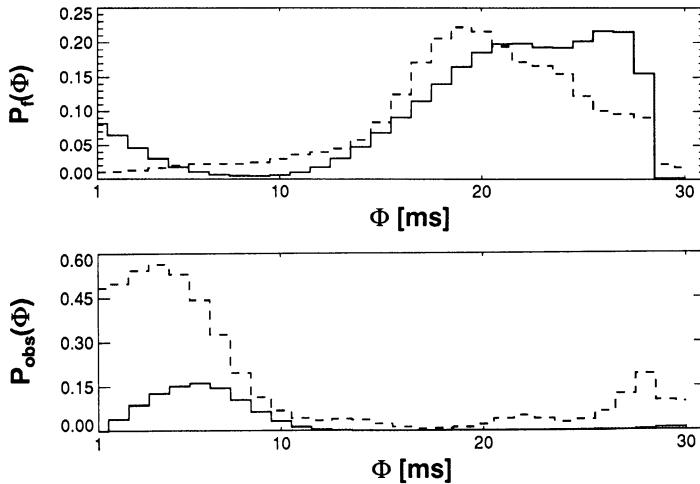
### 7.5.1 IDENTIFYING THE LOCAL POOL DYNAMICS

I illustrate the HSM presented above with multiunit activities from the visual cortex of the cat measured with two electrodes located 6 mm apart, the long bar experiment [4]. The parameters  $\lambda^*$  (i.e.,  $P_f^b$  and  $P_{\text{obs}}^i$ ) are estimated with the Baum–Welsh algorithm from 10 repetitions of the experiment. Figure 7.13 shows the result. One finds a nonmonotonous shape of  $P_f^b$  which indicates that the network dynamics includes switching between oscillatory and stochastic phases of activity rather than a purely refractory oscillation. The switching probabilities estimated from  $P_f^b$  agree well with the values estimated from the local field potentials. The shapes of  $P_{\text{obs}}^i$  give additional information about the width of a burst ( $\sim 5$  ms) and about the firing rate during the stochastic episodes ( $P_{\text{obs}}(\phi_{\text{sto}} = 30$  ms)).

The comparison of the auto-correlograms with the prediction (Eq. (7.26)) indicates that for coherent stimuli which are given by a long light bar in this experiment [4] the observed dynamics is captured quite accurately by the HSM (Fig. 7.14(a)). Furthermore one finds that for these stimulus conditions the assembly excitability functions  $P_f^b$  estimated from the two electrodes agree within error while the difference of the  $P_{\text{obs}}^i$  essentially reflects the difference of the firing rates of the contributing neurons.

### 7.5.2 TESTING THE ASSEMBLY HYPOTHESIS

A similarity of the  $P_f^b$ 's indicates comparability of the underlying dynamics but is not sufficient for concluding identity, that is, the presence of only one assembly. This, however, now can be tested quantitatively by comparing the cross-correlogram with the prediction of the HSM (Fig. 7.14(b)). For the mentioned stimulus conditions one finds agreement, however, with a slight overestimation of the modulation by the HSM. This shows that the assumption of a unique underlying dynamics was too strong (compare next



**Fig. 7.13.** Parameters of the hidden state model estimated from multiunit activities from the long bar experiment [4]. (a) The network excitability  $P_f$  has a nonmonotonous shape which indicates switching. (b) The observability  $P_{\text{obs}}$  shows that multiunit activities occur within  $\sim 5$  ms around the sync state in the network and that there is a finite firing rate in the stochastic state ( $P_{\text{obs}}(30 \text{ ms})$ ).

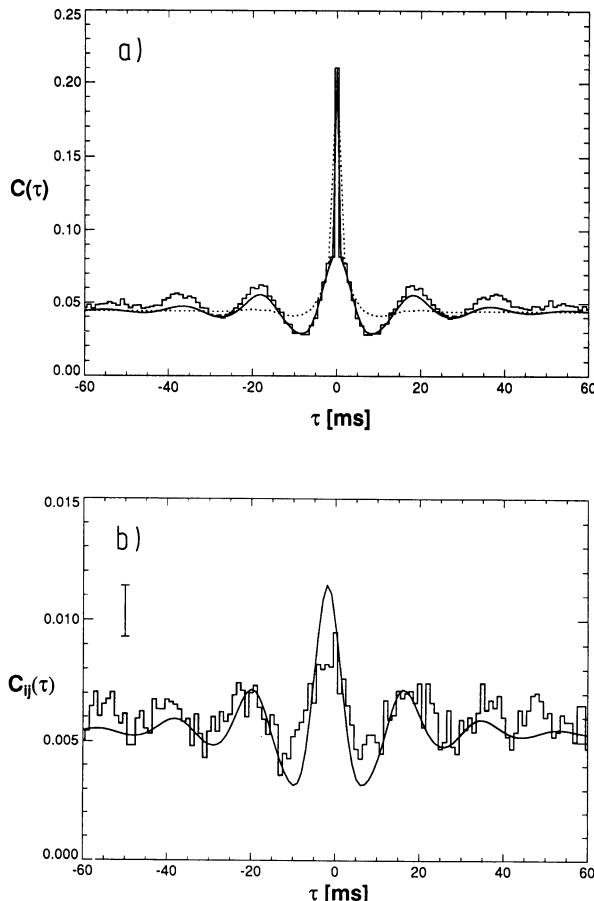
section). For other stimuli cross-correlograms become flatter, indicating that different assemblies are involved. This is quantitatively reflected in the HSM either by different  $P_f^b$  for the two MUA's, or, in case of similar burst excitabilities by a significant deviation of the cross-correlogram from the prediction by the model (not shown).

## 7.6 Reconstruction of Synchronous Network States

Once the consistency of the hidden state model with the local multiunit activity is established it can be used to retrospectively identify the time course of the internal states of the network. To this purpose consider the probabilities

$$\gamma_\phi(t) = P(\mathbf{O}, q_t = \phi | \lambda^*) \quad (7.27)$$

of being in the states  $\phi$  at times  $t$ , which are computed as part of the Baum–Welch procedure. The probability  $\gamma_0(t)$  to be in the burst state  $\phi = 0$  is of special interest in the context of network synchronization. Figure 7.15 shows these time-dependent rates for the two electrodes together with the MUAs from which they originally were obtained. These functions are smooth and exhibit prominent peaks which identify times of network synchronizations on the basis of quite stochastic spike events. Note that this method also



**Fig. 7.14.** (a) Auto-correlogram  $P_{11}$  (histogram) and prediction  $\hat{P}_{11}$  (smooth line) from the hidden state model. The functions agree except for a systematic shift in the true auto-correlogram which originates in the stimulus driven non-stationarity (shape of the PSTH). (b) Cross-correlogram  $P_{12}$  (histogram) and prediction  $\hat{P}_{12}$  (smooth line) from the hidden state model under the additional assumption of *unique* network dynamics, that is, one underlying dynamics expressed by the same  $P_f$  and identical network states for both measurement sites. The overestimation of actual cross-correlation indicates that the latter assumption is too strong.

applies to the case of low firing rates (Fig. 7.15(b)) where a finite probability for network synchronizations is estimated even at times where no event was detected. A careful comparison of Fig. 7.15(a) with Fig. 7.15(b) reveals that mutual synchronizations *across* the cortical sheet are quite precise but rare events (in the data shown they occur only during ms 40–120 and 310–350).

Synchronization as a network event might seem to be a rather artificial notion. Therefore it could be more interesting to have an estimate of the number of neurons active in the assembly at a given time. Cum grano salis this number should be proportional to the time-dependent rate  $\gamma(t)$  which originally is the probability to observe a spike in a particular time interval  $[t, t + \delta t]$ :

$$\gamma(t) = \sum_{\phi} \gamma_{\phi}(t) P_{\text{obs}}(\phi). \quad (7.28)$$

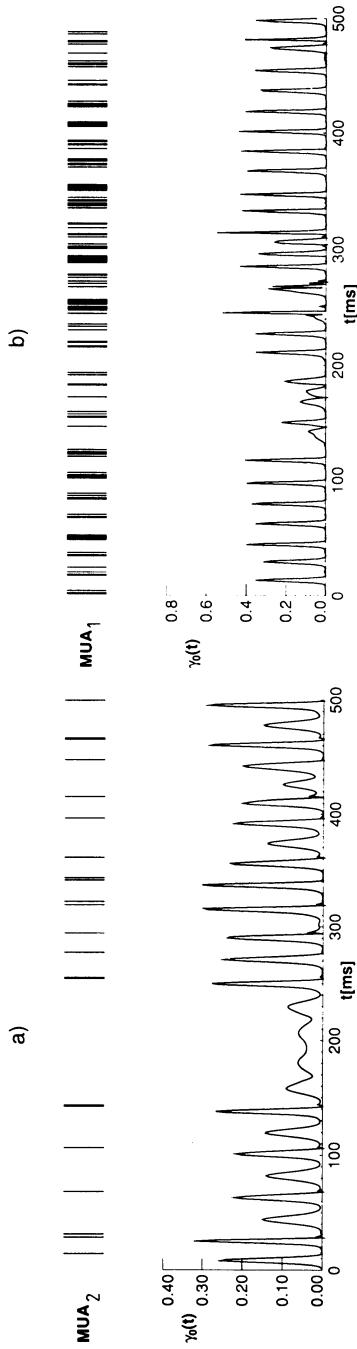
This function strongly resembles  $\gamma_0$ , but it is smoother and it has an offset which reflects the random background excitability of the neurons. Furthermore, it matches with the negative peaks in the local field potential to an astonishing degree, thereby underlining that LFPs reflect a local mass activity (Fig. 7.16) [31].

The significance of this method for the localization of the states of synchronization in an assembly is an open problem. For models which per construction obey the network renewal dynamics one can directly test the range of applicability [32]. For the data presented above one has the indirect evidence from the similarities of  $\gamma$  and the local field potentials that our method correctly identifies a mass action. More direct evidence would be provided by the simultaneous observation of a large portion of the neurons belonging to an assembly. Except for simple cases (e.g., the high frequency hippocampus oscillations [33]) such a control of the assembly activity is not possible. Therefore one here can only test whether the results have the correct statistical properties. For this purpose I calculate the cross-correlation of the probabilities  $P_{12}^{\gamma}(\tau) = < \gamma^1(t)\gamma^2(t + \tau) >$  which should give an estimate of the spike–spike joint probability if these rates caused the observable MUAs. When comparing  $P_{12}^{\gamma}$  with the true cross-correlogram one finds excellent agreement (Fig. 7.17). This result finally proves that the localization of local network synchronizations with our method is highly consistent with the data.

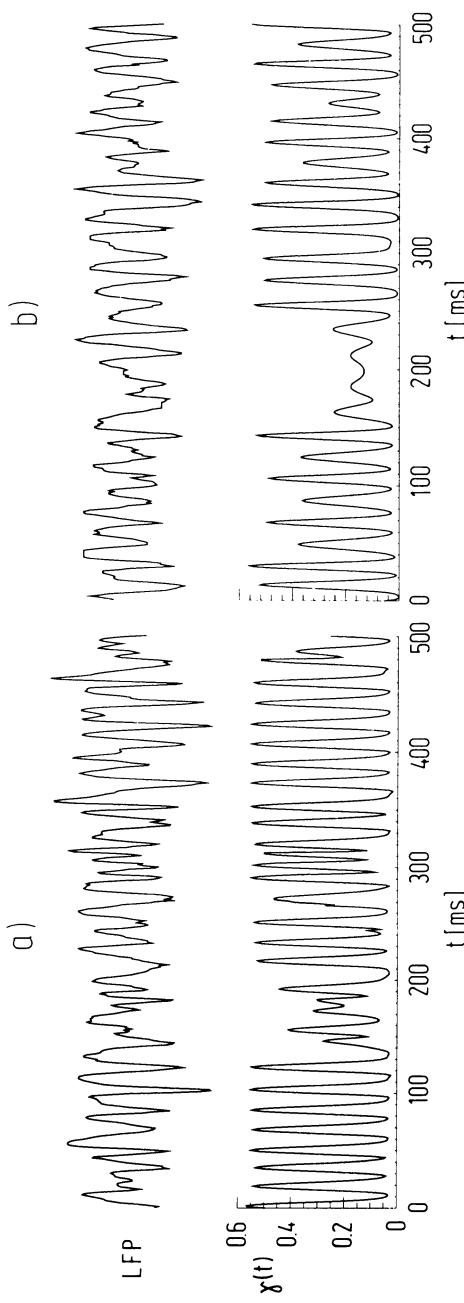
## 7.7 Summary

In this contribution I proposed to define coherence relative to the knowledge of an observer as the degree of predictability. Depending on the a priori expectations one then can distinguish model-free and model-dependent approaches for the detection of coherence in signals.

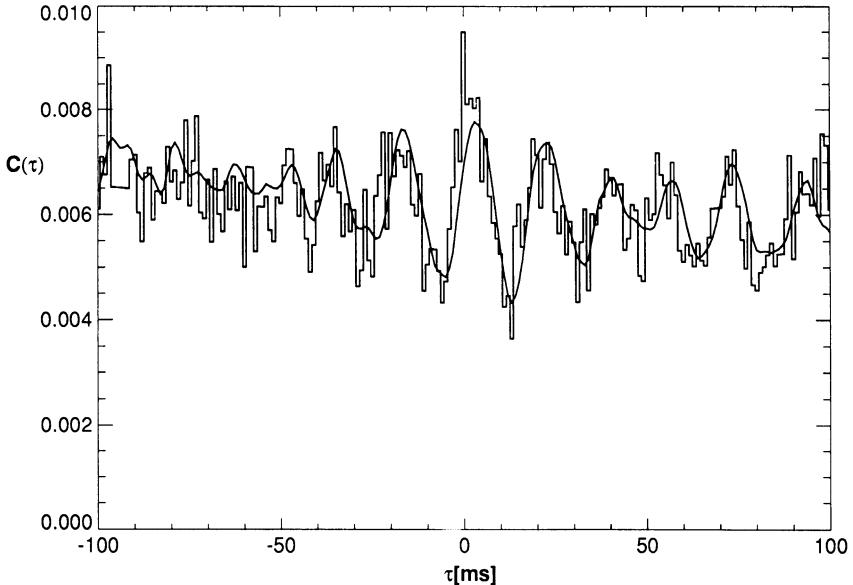
Model-free approaches rely only on the statistical features of the data.



**Fig. 7.15.** Multineuron activities and the corresponding probabilities for a synchronous network state  $\gamma_0^i(t)$  calculated from the hidden state models from the MUAs of electrode  $i = 1$  (b) and electrode  $i = 2$  (a).



**Fig. 7.16.** Local field potentials (LFP) and time-dependent firing rate  $\gamma^i(t)$  reconstructed with the hidden state model for electrode  $i = 1$  (a) and electrode  $i = 2$  (b). Negative values in the LFP closely correspond to peaks in  $\gamma$  even for electrode 1 where only a small number of spikes were available (compare Fig. 7.15).



**Fig. 7.17.** Cross-correlogram of multiunit activities  $P_{12}$  and cross-correlation of estimated network activities  $P_{12}^\gamma$ . The excellent agreement proves that the estimated rates consistently characterize the underlying processes with a high degree of precision.

Here information theoretical notions were considered which quantify coherence in the sense of statistical dependency. In particular I introduced the time resolved mutual information together with an efficient algorithm for its estimation. This provides a method for the time resolved identification of episodes in time series depending on their coherence in time and in space. This measure was applied to the continuous signals of local field potential data from the cat visual cortex and revealed that the coherent episodes (oscillations) in these signals alternate with less coherent episodes in a way which indicates switching between two dynamical states in the underlying network.

Complementary to model-free approaches, model-dependent methods exploit a priori assumptions about the structure of the system generating the signals. For network burst synchronizations a hidden state model was introduced which represents the network dynamics in terms of a renewal process to which the observable spike activities couple stochastically. The parameters of this simple hidden Markov Ansatz are directly obtainable from the neurophysiologically measured multiunit activities (MUA). For MUA from the visual cortex of the cat this method quantifies the memory and the switching in the underlying neuronal system. Similar to the model-free approach, this method identifies the dynamical network states which underly a given time series of MUAs, and, in particular, localizes network

synchronizations with a high temporal precision. Despite the simplicity of the Ansatz it is found to be quantitatively sufficient for a pheonomenological representation of the data obtained locally at one electrode. This fact, together with the assumption that unique network dynamics underlie the MUAs measured at electrodes spatially separated in the cortex, gives an upper bound for the possible outcome of correlation analysis of these data. These predictions overestimate the actual correlograms to some degree, which thereby quantifies the assumption that a unique assembly dynamics generates these data. Synchronization across a cortical distance turns out to be quite a rare event, a fact which is also evident from a comparison of the reconstructed network states.

I expect that methods for a detailed analysis of experimental data like those presented in this paper will enable better quantitative comparisons of models with their neuronal counterparts, which I believe to be a necessary prerequisite for the development of more quantitative theories of the brain.

*Acknowledgments.* Some of the presented methods were developed in collaboration with H.-U. Bauer, J. Deppisch, and F. Wolf. I also acknowledge most fruitful discussions with W. Singer, C. Gray, and P. König and thank them for their provision of the data. Part of this work was supported by the VW foundation (Grant No. I 65 961).

## REFERENCES

- [1] J.J. Eggermont (1990) *The Correlative Brain* (Springer, Berlin)
- [2] A.M.H.J. Aertsen, G.L. Gerstein, M.K. Habib, G. Palm (1989) Dynamics of neuronal firing correlation: Modulation of "effective connectivity." *J. Neurophys.* **61**:900
- [3] D.H. Perkel, T.H. Bullock (1968) Neural Coding. *Neurosci. Res. Prog. Bull.* **6**: No. 3
- [4] C.M. Gray, P. König, A.K. Engel, W. Singer (1989) Oscillatory responses in cat visual cortex exhibit intercolumnar synchronization which reflects global stimulus properties. *Nature* **338**:334
- [5] R. Eckhorn, R. Bauer, W. Jordan, M. Brosch, W. Kruse, M. Munk, H.J. Reitböck (1988) Coherent oscillations: A mechanism for feature linking in the visual cortex? *Biol. Cybern.* **60**:121
- [6] C.v.d. Malsburg (1994) The Correlation Theory of Brain Function, this book, Ch. 2. First appeared in Internal Report 81-2, Max-Planck-Inst. für Biophys. Chemie, Göttingen
- [7] C.M. Gray, A.K. Engel, P. König, W. Singer (1990) Stimulus-dependent neuronal oscillations in cat visual cortex: Receptive field properties and feature dependence. *Eur. J. Neurosci.* **2**:607-619

- [8] H. Weinberg, R. Cooper (1972) The use of correlational analysis for pattern recognition. *Nature* **238**:292
- [9] G. Palm, T. Poggio (1977) Wiener-like system identification in physiology. *J. Math. Biol.* **4**:375
- [10] G. Palm, A.M.H.J. Aertsen, G.L. Gertsein (1988) On the significance of correlations among neuronal spike trains. *Biol. Cybern.* **59**:1
- [11] C.E. Shannon (1948) The mathematical theory of communication. *Bell Syst. Tech. J.* **27**:379
- [12] S. Kullback, R.A. Leibler (1951) *Ann. Math. Stat.* **22**:79
- [13] K. Pawelzik (1991) *Nichtlineare Dynamik und Hirnaktivität* (Harri Deutsch, Frankfurt)
- [14] G. Hauske, E. Neuburger (1968) Gerichtete Informationsgrößen zur Analyse gekoppelter Verhaltensweisen, *Kybernetik* **4**:171
- [15] R. Eckhorn, O.-J. Glüscher, J. Kröller, K. Pellnitz, B. Pöpel (1976) Efficiency of different neuronal codes: Information transfer calculations for three different neuronal systems. *Biol. Cybern.* **22**:49
- [16] F. Rieke, D. Warland, W. Bialek (1993) Coding efficiency and information rates in sensory neurons. *Europhys. Lett.* **22**:151
- [17] P. Grassberger, T. Schreiber, C. Schaffrath (1991) Nonlinear time sequence analysis. WU B 91 - 14, Wuppertal University preprint
- [18] A.M. Fraser, H.L. Swinney (1986) *Phys. Rev. A* **33**:1134
- [19] K. Pawelzik, H.-U. Bauer, T. Geisel (1993) Switching between predictable and unpredictable states in data from cat visual cortex. *Proc. CNS 92 San Francisco* (Kluwer Academic, Dordrecht), in press
- [20] W. Liebert, K. Pawelzik, H.G. Schuster (1991) Optimal embeddings of chaotic attractors from topological considerations. *Europhys. Lett.* **14**:521
- [21] T. Buzug, K. Pawelzik, J. von Stamm, G. Pfister (1994) Mutual information and global strange attractors in Taylor-Couette flow. *Physica D*, **72**:343
- [22] J. Kohlmorgen, K.R. Müller, K. Pawelzik (1994) Competing predictors segment and identify switching dynamics. In: *ICANN 1994 Proceedings*, M. Marinaro, P.G. Morasso (Eds.) (Springer), pp. 1063–1066
- [23] K. Pawelzik, H.-U. Bauer, J. Deppisch, T. Geisel (1993) How oscillatory neuronal responses reflect bistability and switching of the hidden assembly dynamics. In: *Advances in Neural Information Processing Systems*, S.J. Hanson, J.D. Cowan, C.L. Giles (Eds.) (Morgan Kaufmann), Vol. 5, pp. 977–984
- [24] K. Pawelzik, H.G. Schuster (1987) Generalized dimensions and entropies from a measured time series. *Phys. Rev. A* **35**:481
- [25] D.H. Perkel, G.L. Gerstein, G.P. Moore (1967) Neuronal spike trains and stochastic point processes I. The single spike train, *Biophys. J.* **7**:391–418

- [26] H.-U. Bauer, K. Pawelzik (1993) Alternating oscillatory and stochastic dynamics in a model for a neuronal assembly. *Physica D*, **69**:380–393
- [27] C. Koch, H.-G. Schuster (1992) A simple network showing burst synchronization without frequency locking. *Neural Comput.* **4**:211–223
- [28] J. Deppisch, K. Pawelzik, T. Geisel (1994) Identification of network synchronization dynamics underlying correlated neuronal activity. *Biol. Cybern.* **71**:387–399
- [29] L.E. Baum, T. Petrie (1966) Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Stat.* **37**:1554–1563
- [30] L.R. Rabiner (1989) A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* **77**:257–286
- [31] W.J. Freeman (1975) *Mass Action in the Nervous System* (Academic Press, New York)
- [32] K. Pawelzik, J. Deppisch, in preparation
- [33] G. Buzsáki, Z. Horváth, R. Urioste, J. Hetke, K. Wise (1992) High frequency network oscillation in the hippocampus. *Science* **256**:1025–1027

# Hebbian Synaptic Plasticity: Evolution of the Contemporary Concept

Thomas H. Brown and Sumantra Chattarji

with 6 figures

**Synopsis.** Interest in the idea that certain forms or aspects of learning and memory result from use-dependent synaptic modifications has a long history [18]. Here we review some of the conceptual and experimental aspects of the type of synaptic learning mechanism originally suggested by the Canadian psychologist Donald Hebb (1949). We first summarize the contemporary concept of a Hebbian synaptic learning mechanism. Theoretical studies suggest that useful and potentially powerful forms of learning and self-organization can emerge in networks of elements that are interconnected by various formal representations of a Hebbian modification [2,49,60,69,70,93]. Interest in the computational aspects of Hebbian modification algorithms has been enhanced by the neurophysiological discovery [11] of a synaptic phenomenon in the hippocampus known as long-term potentiation (LTP), that is induced by a Hebbian mechanism. We review recent facts and hypotheses about LTP that are pertinent to contemporary interpretations of a Hebbian synaptic modification and describe the evolution of biophysical models of spines that account for the key features of LTP induction. Finally we review more recent evidence regarding variations and extensions of Hebb's original postulate for learning.

## 8.1 Concept of a Hebbian Synapse

In his 1949 book, *The Organization of Behavior*, Donald Hebb proposed that an important condition for triggering enhanced synaptic efficacy would be the repeated conjunction of presynaptic activity and the firing of the cell on to which this activity was afferent [43, p. 62]:

"When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic

change takes place in one or both cells such that *A*'s efficiency, as one of the cells firing *B*, is increased."

This idea has come to be known as "Hebb's postulate" for learning [18,97]. Hebb proposed this change as the basis of a memory formation and storage process that would cause enduring modifications in the elicited activity patterns of spatially distributed "nerve cell assemblies." Thus, it specifies the *location* of the modification and provides a qualitative statement of the *conditions* for change. In brief, a *Hebbian synapse* strengthens when the presynaptic and postsynaptic elements tend to be coactive.

Hebb's postulate can be seen as the synthesis of two earlier ideas about memory [18]. First, the American psychologist William James (1890) proposed that the most basic law of association was the *law of neural habit*, which maintained that if two "elementary brain processes" are active at the same time, then each will tend to "propagate its excitement into the other." James did not suggest a cellular correlate or substrate for these elementary brain processes. The second idea concerns the physical nature of the changes underlying memory. The Italian neuroanatomist Eugenio Tanzi (1893) proposed that the synapse might be the locus of the modification, an idea that was later advanced by Cajal (1911). Thus, Hebb's postulate can be seen as the synthesis of the law of neural habit and the synaptic hypothesis for memory.

### 8.1.1 CONTEMPORARY CONCEPT OF A HEBBIAN SYNAPTIC MODIFICATION

Since 1949 the concept of a Hebbian synapse has undergone considerable evolution to include several defining characteristics. These key features [18], which form the basis of the contemporary understanding of a Hebbian synaptic modification, are as follows:

(1) *Local mechanism.* The process is *local*, that is, all the information that is necessary for the modification is available to the synapse as a consequence of ordinary synaptic transmission. Thus, Hebbian synapses are said to enable an "unsupervised" form of learning—in the sense that a specific external "teacher" signal is not required to instruct change on an individual, synapse-by-synapse basis.

(2) *Interactive mechanism.* Whether a change occurs at a Hebbian synapse depends on activity on *both* sides of the synaptic cleft. It involves a true interaction, in the statistical sense, between presynaptic and postsynaptic activity. Thus, mechanisms that depend on either presynaptic (e.g., post-tetanic potentiation or PTP) or postsynaptic activity alone, or on a linear superposition of the consequences of presynaptic and postsynaptic activity, are termed "noninteractive" and would not qualify as Hebbian. This interactive requirement makes the mechanism fundamentally associative. The nature of the interaction can involve a *conjunctive-type* rule, where the simple co-occurrence of (some level) of pre- and postsynaptic activity (within

some short time interval) is sufficient to cause synaptic enhancement, or a *correlational-type* rule, where the synaptic enhancement depends on the covariance, or some other statistical measure of the association, between pre- and postsynaptic activity.

(3) *Time-dependent mechanism:* Modifications in a Hebbian synapse depend on the exact time of occurrence of pre- and postsynaptic activity. The timing of pre- and postsynaptic events play an essential role in determining the change. This is in contrast with many of the supervised learning schemes such as the backpropagation algorithm.

Thus, a Hebbian synapse may be defined as one that uses a time-dependent, highly local, and strongly interactive mechanism to increase synaptic efficacy as a function of the conjunction or correlation between pre- and postsynaptic activity.

Neither Hebb nor his colleagues of the day had the technology to conduct experiments to study Hebbian synaptic modifications. Thus, apart from providing a conceptual framework, Hebb's ideas had little impact on neurobiology. This changed with two important developments. First, even though experimental tools and preparations amenable to testing Hebb's ideas were lacking, this did not prevent theoretical explorations into its computational implications during the 1960s and 1970s. Second, emergence of more sophisticated physiological techniques in the 1980s eventually led to the discovery of Hebbian synaptic plasticity mechanisms in the hippocampus.

## 8.2 Experimental Evidence for Hebbian Synaptic Mechanisms

Hebbian synapses have now been shown to exist in the sense that one can experimentally demonstrate a use-dependent form of synaptic enhancement that is governed by a time-dependent, highly local, and strongly interactive mechanism. This mechanism appears to be responsible for at least one form of hippocampal synaptic plasticity called long-term potentiation (LTP). LTP is a use-dependent and persistent increase in synaptic strength that can be rapidly induced by brief periods of synaptic stimulation [11,12,16,18]. LTP was first discovered in a brain structure—the hippocampus—that has long been implicated in learning and memory.

There are at least two and possibly many more different LTP mechanisms or forms of LTP [17,56]. The form that has been most extensively studied is the variety that occurs in the Schaffer collateral/commissural (Sch/com) synaptic input to the pyramidal neurons of hippocampal region CA1. Several lines of evidence suggest that the induction of LTP in the CA1 Sch/com synapses satisfies the definition of a Hebbian modification

[16,18,41,57,78]. Here we summarize some of the key aspects of the physiology of LTP induction in the Sch/com synapses of the hippocampus.

### 8.2.1 INDUCTION OF HIPPOCAMPAL LTP

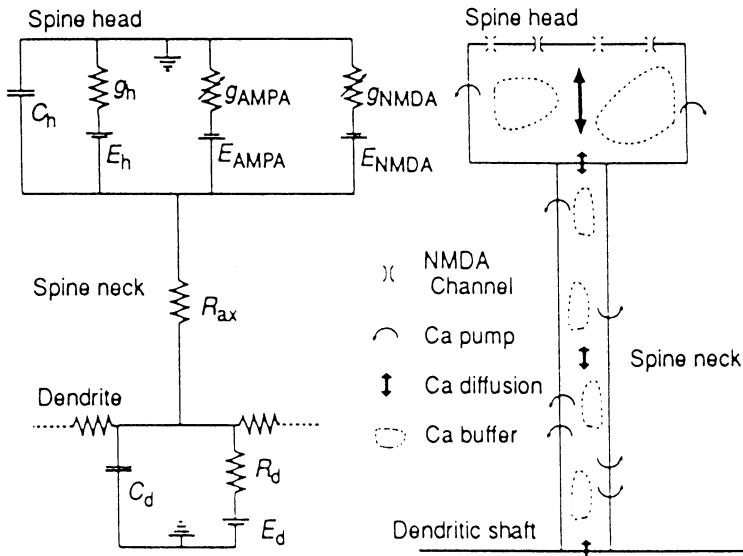
There is general agreement that the induction of the most commonly studied form of LTP at Sch/com synapses is controlled by the *N*-methyl-*D*-aspartate (NMDA) subclass of glutamate receptor. The working hypothesis is that  $\text{Ca}^{2+}$  influx through the NMDA receptor-gated channel and the resultant increase in postsynaptic  $\text{Ca}^{2+}$  are partly responsible for triggering the induction of LTP [15,16,17,31,33]. How does  $\text{Ca}^{2+}$  influx through the NMDA receptor-gated channel relate to the Hebbian nature of LTP induction? It turns out that this channel has just the kind of gating properties needed for a Hebb-type conjunctive mechanism.

The NMDA conductance is a function of both glutamate binding to the receptor site and voltage across the channel [12]. The voltage dependence results from the fact that the NMDA channels are normally partially blocked by  $\text{Mg}^{2+}$  and the fact that this channel block is gradually relieved as the membrane is depolarized [54,79,84]. Thus, the conjunction of presynaptic activity (to cause glutamate release) and a sufficient amount of postsynaptic depolarization (to relieve the channel block) allows  $\text{Ca}^{2+}$  influx into the postsynaptic cell.

Just as there are many ways of implementing a Hebbian algorithm theoretically, nature may have more than one way of designing a Hebbian synapse. The induction of LTP at the mossy fiber synapses in field CA3 is different from that observed in the CA1 region and dentate gyrus. In particular, induction of mossy fiber LTP does not require the activation of NMDA receptors [24,56]. Nevertheless, recent studies suggest that mossy fiber LTP depends on activation of postsynaptic voltage-dependent  $\text{Ca}^{2+}$  channels *and* high frequency synaptic activity [51,56]. Thus, although mossy fiber LTP is NMDA receptor independent, it may still require high frequency presynaptic stimulation in conjunction with postsynaptic depolarization, that is, it could still involve a Hebb-type conjunctive mechanism. Further experiments are required to resolve this issue, which is controversial [29,106].

## 8.3 Biophysical Models of LTP Induction

A quantitative model of the NMDA-dependent postsynaptic mechanism for LTP induction has been developed for the Sch/com synapses [18,105]. This biophysical model investigates electrical and  $\text{Ca}^{2+}$  dynamics following activation of NMDA receptors located on a dendritic spine (Fig. 8.1). The model is based on the widely accepted assumption that influx of  $\text{Ca}^{2+}$  through NMDA receptor-gated channels into the dendritic spine is the ex-



**Fig. 8.1.** First-generation biophysical model of spine (from [105]). The model has an electrical component (left) and a chemical component (right). The electrical component incorporates both an NMDA conductance and a non-NMDA conductance. In some simulations, the spine was attached to a 3000-compartment anatomically reconstructed neuron (Claiborne et al., 1991). The chemical component of the model simulated  $\text{Ca}^{2+}$  influx, transport, and buffering.  $\text{Ca}^{2+}$  entered the spine through channels located only on the distal spine head and then diffused along the length of the spine to the dendritic shaft. Diffusion was limited by binding to saturable membrane pumps and cytosolic buffers.

clusive trigger for LTP induction. Two other fundamental components of this model are calcium buffering and extrusion mechanisms throughout the spine head and neck and longitudinal calcium diffusion between the spine and the dendrite. Voltage gated  $\text{Ca}^{2+}$  channels were assumed in this model to be only located on the dendritic shaft.

The model (Fig. 8.1) consisted of separate electrical and  $\text{Ca}^{2+}$  components. The total synaptic current (Fig. 8.1) was the sum of separate NMDA and non-NMDA components. An alpha function [14],

$$I(t) = (E_{\text{syn}} - V_m)\kappa g_p t \exp(-t/t_p), \quad (8.1)$$

was used for the non-NMDA current, with  $\kappa = e/t_p$ ,  $e$  the base of the natural logarithm,  $t_p = 1.5$  ms,  $E_{\text{syn}} = 0$  mV, and the peak conductance  $g_p = 0.5$  nS [8,15]. The NMDA conductance was a function of both time and membrane potential. The voltage dependence was derived from a model in

which the binding rate constant of Mg<sup>2+</sup> to the site of channel block varied as a function of voltage. Specifically,

$$I(t) = (E_{\text{syn}} - V_m)g_n \frac{\exp(-t/\tau_1) - \exp(-t/\tau_2)}{1 + \eta[\text{Mg}] \exp(-\gamma V)}, \quad (8.2)$$

with  $\tau_1 = 80$  ms,  $\tau_2 = 0.67$  ms,  $\eta = 0.33/\text{mM}$ ,  $\gamma = 0.06/\text{mV}$ ,  $E_{\text{syn}} = 0$  mV, and  $g_n = 0.2$  nS. These parameters were based on voltage clamp studies of synaptic currents in the hippocampal slice and on single-channel recordings from cultured hippocampal neurons. Short trains of synaptic stimulations were modeled as the sum of these conductance waveforms.

To explore the spatio-temporal interactions between inputs, a simplified model of a whole hippocampal neuron was constructed. The neuronal structure consisted of a bifurcating apical branch and a single basal branch emerging from the soma. The following electrical parameters were used: membrane resistance,  $R_m = 20 \text{ k}\Omega \text{ cm}^2$ ; intracellular resistivity,  $R_i = 100 \Omega \text{ cm}$ ; and membrane capacitance,  $C_m = 1 \mu\text{F}/\text{cm}^2$ . A single synapse was placed on the head of a spine located halfway between the soma and the termination of the longer apical branch.

Ca<sup>2+</sup> entered through NMDA receptor-gated channels located on the distal part of a spine head. It then diffused along the length of the spine to the dendritic shaft. Ca<sup>2+</sup> was also buffered and pumped in each compartment. The [Ca<sup>2+</sup>] at the shaft was generally fixed at the assumed resting value of 0.05 μM. For purposes of Ca<sup>2+</sup> transport and buffering (Fig. 8.1), the spine geometry was modeled as a short cylindrical head at the end of a long cylindrical neck. Due to the lack of good quantitative evidence regarding the density and properties of Ca<sup>2+</sup> pumps in mammalian brain neurons, the authors modeled two separate pumps for Ca<sup>2+</sup> extrusion from the spine lumen. Both obeyed saturable first-order Michaelis-Menton kinetics:

$$\frac{\partial [\text{Ca}]_{\text{pump}}}{\partial t} = -K_{\max} P_s \frac{A}{V} \frac{[\text{Ca}]}{[\text{Ca}] + K_d} + J_{\text{leak}} \frac{A}{V}, \quad (8.3)$$

where  $K_{\max} = 0.2/\text{ms}$  is the Ca<sup>2+</sup> turnover rate,  $P_s$  is the surface density,  $A/V$  is the ratio of surface area to volume,  $J_{\text{leak}}$  is the leak flux of Ca<sup>2+</sup>, and  $K_d$  is the dissociation constant. A higher affinity ( $K_d = 0.5 \mu\text{M}$ ) lower-capacity pump was included at a uniform density ( $P_s = 5 \times 10^{-16} \mu\text{mol}/\mu\text{m}^2$ ) throughout the spine membrane. The density of a lower affinity ( $K_d = 20 \mu\text{M}$ ) higher-capacity pump was nonuniform. Finally, the model also included a fixed concentration (100 μM) of an immobile saturable buffer with a kinetic scheme based on one proposed for calmodulin (CaM). Calmodulin (CaM) is present at high concentrations at the postsynaptic density, and it has been suggested that a Ca<sup>2+</sup>-CaM-dependent protein kinase, CaMKII, may be important in LTP induction.

The equations describing electrical and chemical dynamics were advanced independently, and at the end of each time step electrical current was con-

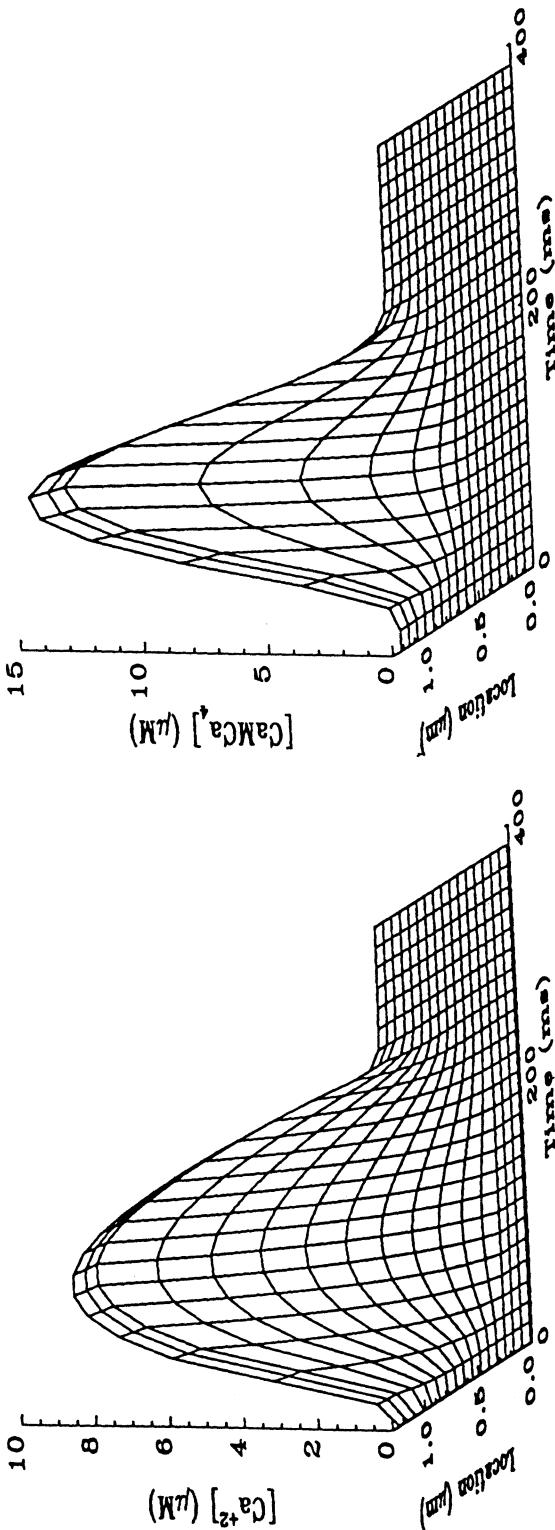
verted to ionic flux through Faraday's constant. These equations were discretized and then solved using alternating implicit-explicit steps [46]. A custom simulator, written in C, was used to simulate  $\text{Ca}^{2+}$  dynamics. When the full dendritic complexity of hippocampal neurons was included in the simulations [19,20], a customized version of NEURON, a compartmental simulator developed by Michael Hines [47], was used.

This model accounted for the key features of this form of LTP. First, it accounted for the Hebbian nature of LTP induction in these synapses, that is, the induction of LTP is dependent upon both presynaptic activity and some critical amount of postsynaptic depolarization. Second, the model accounted for the input specificity of LTP. The microphysiology of the spine restricted LTP induction to just those synapses that satisfied the conjunctive requirement. Third, the model was consistent with experimentally observed temporal contiguity requirements for LTP induction.

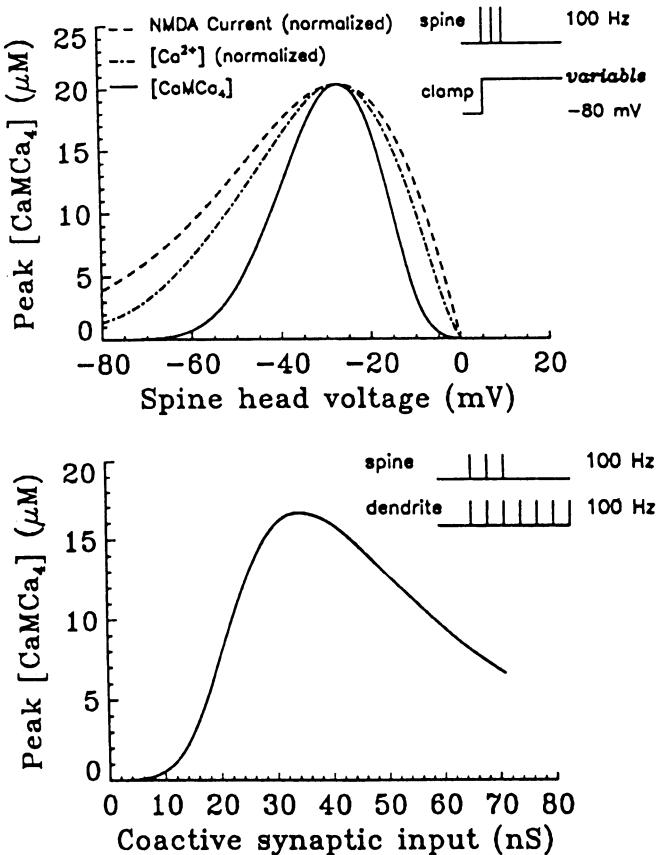
Computer simulations of the model suggested that spines served several critical functions during the induction of LTP. First, they amplify the increases in  $[\text{Ca}^{2+}]$  relative to what would occur if the same synaptic input were placed directly onto the dendritic shaft. This follows from the large surface-to-volume ratio coupled with restricted diffusion out of the spine.

Second, the large  $\text{Ca}^{2+}$  transient produced by synaptic activity is compartmentalized. The peak occurs adjacent to the subsynaptic membrane in the head of the spine and the  $\text{Ca}^{2+}$  transient is negligible in the dendritic shaft. This compartmentalization depended primarily upon the shape of the spine and the distribution of pumps. The slender spine neck slowed diffusion, and its high ratio of surface area to volume enhanced the effectiveness of membrane pumps. For the standard parameters used in the simulations, the  $[\text{Ca}^{2+}]$  in the spine head increased from its resting value of  $0.05 \mu\text{M}$  to almost  $10 \mu\text{M}$ , while at the dendrite it increased only to  $0.06 \mu\text{M}$  (Fig. 8.2, left). Compartmentalization of activated CaM (Fig. 8.2, right) was even more pronounced. Because each spine maintains its own microenvironment distinct from that of neighboring spines, synaptic enhancement can be highly selective.

Third, the spines isolate the biochemical machinery responsible for inducing LTP from increases in  $[\text{Ca}^{2+}]$  at the dendritic shaft caused by voltage-gated channels. Compartmentalization and isolation contribute to the input specificity and Hebbian nature of LTP. In the model, the coactivity requirement for LTP induction (Fig. 8.3) arose from the voltage dependence of the NMDA receptor-gated channel (Eq. (8.2)) and from the nonlinear relationship between synaptic current and peak  $[\text{Ca}^{2+}]$  at the spine head. The peak concentration of activated CaM during synaptic input increased from  $1.8 \times 10^{-3} \mu\text{M}$  at  $-80 \text{ mV}$  to  $20 \mu\text{M}$  at  $-30 \text{ mV}$  (Fig. 8.3, top). While the details of the voltage dependence shown in Fig. 8.3 depended on the assumptions about spine microphysiology, LTP induction retained its requirement for conjunctive activity—the defining feature of Hebbian LTP—over a very wide range of parameter values. Fourth, spines increase the non-linear rela-



**Fig. 8.2.** Spatio-temporal dynamics of  $\text{Ca}^{2+}$  and  $\text{CaM-Ca}_4$ . The time course of  $[\text{Ca}^{2+}]$  (left) and  $[\text{CaM-Ca}_4]$  (right) in a spine is shown. The axis labeled location indicates distance from dendritic shaft. A train of three presynaptic stimuli was applied at 100 Hz while postsynaptic voltage at the spine head was clamped to -40 mV. Changes in  $[\text{Ca}^{2+}]$  and  $[\text{CaM-Ca}_4]$  are restricted mainly to the spine head, where they are amplified. Note that the response of  $[\text{CaM-Ca}_4]$  is sharper than that of  $[\text{Ca}^{2+}]$ .



**Fig. 8.3. Effect of postsynaptic activity.** Peak [CaM–Ca<sub>4</sub>] in the subsynaptic compartment is shown after a train of three presynaptic stimuli applied at 100 Hz. In the first type of simulation, the postsynaptic membrane was clamped to potentials ranging from -80 to 0 mV (top). Peak [Ca<sup>2+</sup>] and NMDA current have been scaled to this graph. Note that [Ca<sup>2+</sup>] and [CaM–Ca<sub>4</sub>] are steeper functions of voltage than NMDA current. In the second kind of simulation, postsynaptic depolarization was provided by a coactive strong input (bottom). The strong input consisted of a 100 Hz train of stimuli applied at synapses uniformly distributed on the same apical branch as the spine. The strength of the strong input stimulus, defined as peak total conductance of the coactive synapses, was varied from 0 to 75 nS. Note that for very low values of strong input activity (only weak input; see the text), peak [CaM–Ca<sub>4</sub>] remains essentially unchanged from the resting value of [CaM–Ca<sub>4</sub>].

tionship between the strength of synaptic stimulation and the probability or magnitude of LTP induction—beyond what can be accounted for solely in terms of the voltage dependence of the NMDA receptor-gated channel, which is not very steep.

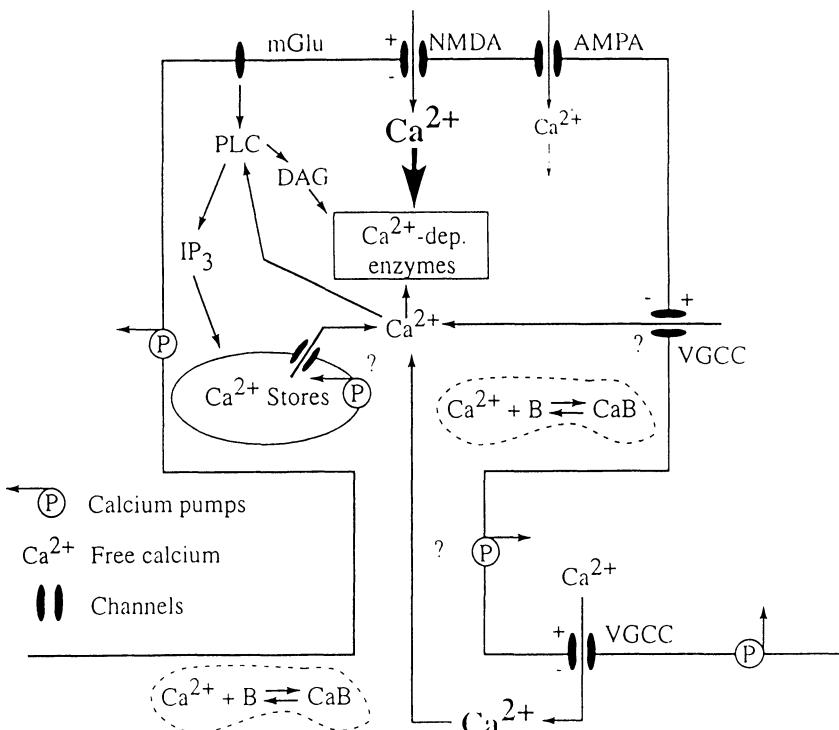
Recent developments in fluorescence imaging techniques and  $\text{Ca}^{2+}$ -sensitive indicators have provided experimental evidence in support of two aspects of the biophysical model discussed above—compartmentalization and isolation. Work by Muller and Connor (1991) suggested that synaptic stimulation of CA1 pyramidal neurons produces localized increases in  $\text{Ca}^{2+}$  that are sensitive to NMDA receptor antagonists and are compartmentalized within spines. Experiments by Guthrie et al. (1991) suggested that increases in dendritic  $\text{Ca}^{2+}$  fail to diffuse into the spines. In the model, isolation of the spine from dendritic  $\text{Ca}^{2+}$  was considered critical. Otherwise, strong postsynaptic depolarization alone would produce LTP in most of the synapses on the neuron. The lack of any such nonspecific effect has been demonstrated [57,58] and it is also known that suprathreshold postsynaptic depolarization does generate large increases in dendritic  $\text{Ca}^{2+}$  via voltage-gated calcium channels [81].

### 8.3.1 SECOND-GENERATION SPINE MODEL

Several lines of recent evidence, however, call into question some of the basic features of the model described above. First, there is now evidence that in many synaptic systems LTP can be induced via mechanisms that do not require the activation of NMDA receptors [56]. Second, the relative amount of  $\text{Ca}^{2+}$  influx into CA1 pyramidal neurons that is mediated by NMDA receptor-gated channels appears to be very small compared to the influx through voltage-gated calcium channels [52,81]. Finally, recent evidence suggests that NMDA receptor-mediated currents may not be sufficient for inducing LTP—activation of the metabotropic (i.e., second-messenger coupled) subtype of glutamate receptor may also be necessary [5].

More recently, experiments using fluorescence imaging techniques with confocal scanning laser microscopy (CLSM) have raised further questions about some of the assumptions underlying the model described above [53]. In particular, recent CLSM studies using confocal microscopy combined with simultaneous neurophysiology in the imaged neuron have examined the effects of direct depolarization of CA1 pyramidal neurons on  $\text{Ca}^{2+}$  signals within spines [52,53]. Contrary to predictions of the spine model discussed above, these recent experiments suggest that direct depolarization of the cell produces clear increases in the  $\text{Ca}^{2+}$  signals in spines—sometimes larger than the signal in the adjacent dendritic shaft.

These and other recent findings have led to the development of a second-generation biophysical model (Fig. 8.4) that recognizes more complex calcium dynamics and provides a platform for generating hypotheses that are now testable with more sophisticated technology [53]. The most obvious



**Fig. 8.4. Second-generation biophysical model of spine (from [53]).** The primary source of Ca<sup>2+</sup> for activating the calcium-dependent biochemical machinery responsible for triggering LTP is Ca<sup>2+</sup> entry through NMDA receptors. Entry of Ca<sup>2+</sup> through non-NMDA (AMPA) receptors or voltage-gated calcium channels is assumed to be normally subthreshold for LTP induction. Activation of metabotropic glutamate receptors (mGlu) is presumed to play an essential but still unknown role in the process. Activation of mGlu receptors activates phospholipase C (PLC), which has two important effects — it causes the release of Ca<sup>2+</sup> from intracellular stores via inositol triphosphate (IP<sub>3</sub>) and it generates diacylglycerol (DAG). Elevated intracellular [Ca<sup>2+</sup>] in turn can activate PLC, furnishing a possible positive feedback loop for propagating [Ca<sup>2+</sup>] waves, a process that might be primed by Ca<sup>2+</sup> influx through voltage-gated calcium channels (VGCCs) in the spine or shaft. Increased [Ca<sup>2+</sup>], in combination with DAG, can also activate protein kinase C, just one of the calcium-dependent protein kinases that has been proposed to participate in the regulation of synaptic strength. Another important calcium-dependent enzyme is calmodulin-dependent protein kinase.

addition to the previous model is the role of the metabotropic glutamate receptor in LTP induction [5]. Another obvious addition to the model concerns a possible role for voltage-gated calcium channels. Reluctance to consider the possible role of voltage-gated calcium channels in LTP induction stemmed from the requirement for input specificity.

One of several ways around this apparent dilemma is to assume that the amount of voltage-gated  $\text{Ca}^{2+}$  entry into spines is normally subthreshold for LTP induction and that glutamate receptors provide the most potent source of  $\text{Ca}^{2+}$  for triggering LTP, possibly by virtue of being strategically located relative to the substrate enzyme(s). In this case, it is still possible that  $\text{Ca}^{2+}$  entry through voltage-gated calcium channels acts synergistically, for example, by priming  $\text{Ca}^{2+}$ -dependent  $\text{Ca}^{2+}$  release mechanisms [40,85] that may contribute to or be necessary for LTP induction [5].

Preliminary results [53] from computer simulations based on the second-generation spine model (Fig. 8.4) suggest that it will prove fruitful to look beyond the NMDA receptor toward a scheme in which multiple sources of calcium and calcium-dependent processes interact within different spatio-temporal domains of the soma, dendritic shaft, and spine. Different locations, magnitudes, and durations of intracellular  $[\text{Ca}^{2+}]$  changes may play a critical role in determining the amplitude, time course, or even the polarity of the synaptic change, giving rise to more complex computations than are possible with a strictly Hebbian learning rule that is constrained by the properties of the NMDA receptor-gated channel. The second-generation computational model of the spine (Fig. 8.4) already incorporates features that will accommodate these more complex issues and is awaiting additional experimental evidence for calibration [53]. One of the key features that will be of importance in the second-generation spine model is the issue of going beyond the strictly Hebbian learning rule towards a more generalized Hebbian rule that allows bidirectional regulation of synaptic strength.

## 8.4 Bidirectional Regulation of Synaptic Strength

According to the above discussion, a strictly Hebbian synaptic modification makes use of a local, time-dependent, and interactive mechanism to *increase* synaptic strength as a function of correlation or conjunction between presynaptic and postsynaptic activities. Hebb, however, did not discuss the consequence of uncorrelated or negatively correlated pre- and postsynaptic activity. Hebb's original postulate did not explore the conditions under which synaptic efficacy could *decrease*. It is clear that, in the absence of a mechanism for synaptic weakening, the synaptic strength could tend to increase without bound, or else saturate at some asymptotic value.

One obvious problem with synapses that can only increase in efficacy is that the range of associative relationships which the memory system can encode is limited, an example being whether memory can be updated

when the relationship between two events changes. A second problem is that memory systems of this type can reliably store only a certain amount of information, and one question is whether the facility to decrease as well as increase synaptic strengths can improve performance. A third point is that LTP does not last indefinitely, that is, experiments in which the time course of LTP has been studied have typically found a steady return to baseline over time. Thus, processes that cause synaptic weakening could be relevant to all these points [91]. The existence of such a capacity in the hippocampus has long been suspected, and indeed is required by certain neural network models of associative memory and synaptic connectivity. This final section provides a brief overview of the theoretical justifications as well as experimental evidence for synaptic depression.

#### 8.4.1 THEORETICAL REPRESENTATIONS OF GENERALIZED HEBBIAN MODIFICATIONS

The emergence of theoretical studies involving “neural network” models brings us—decades after Cajal’s work and Hebb’s book—to a third and equally important conceptual issue. After *where* and *when* comes the issue of *how* synaptic plasticity enables networks of neurons or neuronlike processing elements to perform various cognitive operations. Most models of information storage in neural networks rely on changing the synaptic strengths, or weights, between model neurons [2,49,60,93]. The weights in these simplifying models are altered by learning rules or algorithms so that the network can later retrieve the stored information, or perform the desired task.

Probably the most thoroughly explored use of the Hebb rule in neural network models is the formation of associations between one stimulus or pattern of activity in one neural population and another [2,60]. The Hebb rule and variations on it have been appealing for use in such neural network models, because it provides a way of forming global associations between large-scale patterns of activity in assemblies of neurons using only the local information available at individual synapses [2,91,92].

The earliest models of associative memory were based on network models in which the output of a model neuron was assumed to be proportional to a linear sum of its inputs, each weighted by a synaptic strength. Thus,

$$V_B = \sum W_{BA} V_A, \quad (8.4)$$

where  $V_B$  are the firing rates of a group of  $M$  output cells,  $V_A$  are the firing rates of a group of  $N$  input cells, and  $W_{BA}$  is the synaptic strength between input cell  $A$  and output cell  $B$ .

The transformation between patterns of activity on the input vectors to patterns of activity on the output vectors is determined by the synaptic weight matrix  $W_{BA}$ . This matrix can be constructed from pairs of asso-

ciated input and output vectors using the simplest version of the Hebb rule:

$$\Delta W_{BA} = \epsilon V_B V_A, \quad (8.5)$$

where the strength of the learning  $\epsilon$  can be adjusted to scale the outputs to the desired values. This states that the variables relevant to synaptic change are the co-occurring activity levels, and that increases in synaptic strength are proportional to the product of the presynaptic and postsynaptic values. It should also be emphasized that this simple rule admits of many variations that still qualify as Hebbian.

This model of associative storage is simple and has several attractive features: first, the learning occurs in only one trial; second, the information is distributed over many synapses, so that recall is relatively immune to noise or damage; and third, input patterns similar to stored inputs will give output similar to stored outputs, a form of generalization. Furthermore, this linear model has been generalized to nonlinear models of associative memory, which has led to a new class of learning algorithms based on the principle of error correction [93].

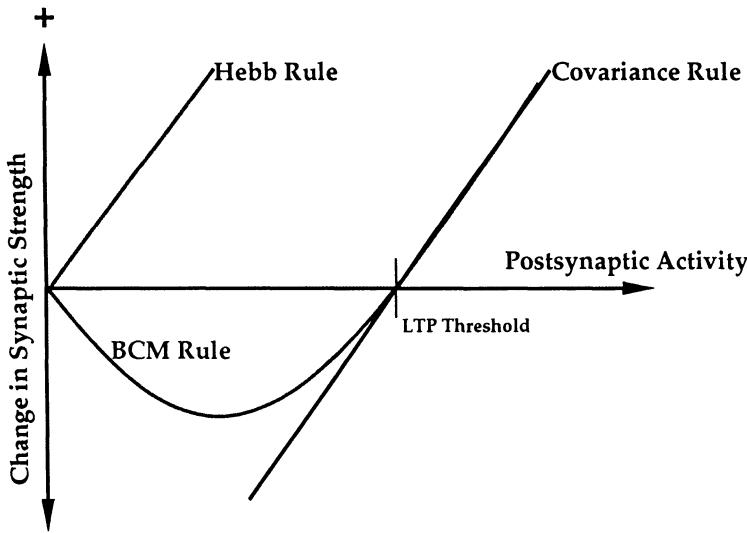
The matrix model of associative storage discussed above has some obvious limitations: first, items with input vectors that are similar, that is, with subpopulations of activated neurons having a significant overlap, will produce outputs that are mixtures of the stored outputs; but discriminations must often be made among similar inputs. Second, any learning system that uses a mechanism that can *only* increase the strengths of synapses will eventually degrade as all the synapses begin to saturate at their maximum values [89,90].

One way to prevent saturation of the synaptic strengths is to reduce the weights by nonspecific decay but this results in information loss at the same decay rate. Another approach is to renormalize the total synaptic weight of the entire terminal field from a single neuron to a constant value. Sejnowski [89,90] proposed a learning algorithm that accomplishes this by using the selective *decrease* of synaptic strength to accomplish optimal error correction learning based on storing the covariances between the pre- and postsynaptic neurons.

According to this *covariance* rule, the change in strength of a plastic synapse should be proportional to the covariance between presynaptic and postsynaptic firing:

$$\Delta W_{BA} = \epsilon (V_B - \langle V_B \rangle) (V_A - \langle V_A \rangle), \quad (8.6)$$

where  $\langle V_B \rangle$  are the average firing rates of the output neurons and  $\langle V_A \rangle$  are the average firing rates of the input neurons. Thus, the strength of the synapse should *increase* if the firing of the presynaptic and postsynaptic elements are *positively correlated*, *decrease* if they are *negatively correlated*, and remain *unchanged* if they are *uncorrelated*. The covariance rule is an extension of the Hebb rule and it is easy to show that traditional Hebbian



**Fig. 8.5.** Schematic drawing of the change in synaptic strength as a function of the postsynaptic activity for the Hebb rule, the covariance rule, and the BCM rule. Both covariance and BCM rules are variants of the Hebb rule and postulate a threshold above which there is LTP (positive change in synaptic strength) and below which there is LTD (negative change in synaptic strength).

synapses can be used to implement it. Taking a time average of the change in synaptic weight in Eq. (8.6):

$$\langle \Delta W_{BA} \rangle = \epsilon (\langle V_B V_A \rangle - \langle V_B \rangle \langle V_A \rangle). \quad (8.7)$$

The first term on the right-hand side has the same form as the simple Hebbian synapse in Eq. (8.5). The second term is a learning *threshold* that varies with the product of the time-averaged pre- and postsynaptic activity levels. This learning threshold ensures that no change in synaptic strength should occur if the average correlation between the pre- and postsynaptic activities is at chance level, that is, when there is no net covariance. The covariance rule achieves the optimal storage capacity for matrix associative memories. Both *heterosynaptic depression* and *homosynaptic depression* of synaptic strength are required by the covariance rule but performance almost as good can be obtained by using only one or the other [103].

Hebbian synaptic plasticity has been used to model the development of ocular dominance and orientation columns in visual cortex [67,68,80]. A network model of cortical development incorporating Hebbian potentiation as well as depression was proposed by Bienenstock, Cooper, and Munro [9]. Their model of synaptic plasticity, now known as the *BCM rule* (Fig. 8.5), strengthens the synapse when the average postsynaptic activity exceeds

a threshold and weakens the synapse when the activity falls below the threshold. In this respect, it resembles the covariance rule. Many others have explored variation of a generalized Hebbian synaptic modification that includes the combination of an interactive synaptic enhancement and some type of activity-dependent synaptic depression [7,18,32,38,63,67,68,87,94].

#### 8.4.2 EXPERIMENTAL EVIDENCE FOR LONG-TERM SYNAPTIC DEPRESSION

In spite of the obvious usefulness of a generalized Hebbian modification that includes both Hebb's original enhancement process plus one or more processes causing synaptic weakening, there has been a relative paucity of studies on synaptic depression compared to those on synaptic enhancement. Recently, several laboratories have provided evidence for several different forms of long-term synaptic depression (LTD) in the hippocampus and other cortical areas [3,4,10,25,34,48,50,59,66,82,91].

Before proceeding to describe the experiential evidence for synaptic depression, let us classify and define some of the terms related to use-dependent synaptic modifications. As already indicated, *long-term synaptic potentiation* (LTP) refers to a rapid and persistent *increase* in synaptic strength induced by synaptic stimulation. Similarly, *long-term synaptic depression* (LTD), as the flip side of LTP, refers to a rapid and persistent *decrease* in synaptic strength caused by synaptic stimulation. The terms LTP and LTD do not imply any particular mechanism. Rather, they are generic categories of synaptic activity modification relationships. There may be several varieties of LTP and LTD (Fig. 8.6), and both phenomena may be present at the same synapse. If a synaptic modification occurs in a stimulated pathway (Fig. 8.6), the phenomenon is referred to as a *homosynaptic* change (e.g., homosynaptic LTP, homosynaptic LTD, Fig. 8.6). If stimulation of one pathway causes a change in a pathway that was not stimulated,

---

**Fig. 8.6.** *Use-dependent synaptic changes.* Each neuron is shown receiving two nonoverlapping synaptic inputs (long horizontal lines). The small vertical lines represent afferent stimuli applied to the two synaptic inputs at either high (closely spaced) or low (sparsely spaced) frequency. The resultant change in synaptic strength is indicated by a downward arrow (LTD), upward arrow (LTP), or 0 (no change) placed next to the inputs. (A) Homosynaptic LTP; (B) homosynaptic LTD; (C) associative LTP, the input receiving the high frequency burst is also potentiated (not shown); (D) associative LTD, the input receiving the high frequency burst is potentiated (not shown); (E) heterosynaptic LTD, the input receiving high frequency stimuli is potentiated (not shown); (F) cerebellar LTD. Details of the stimulus protocols are described in the text.

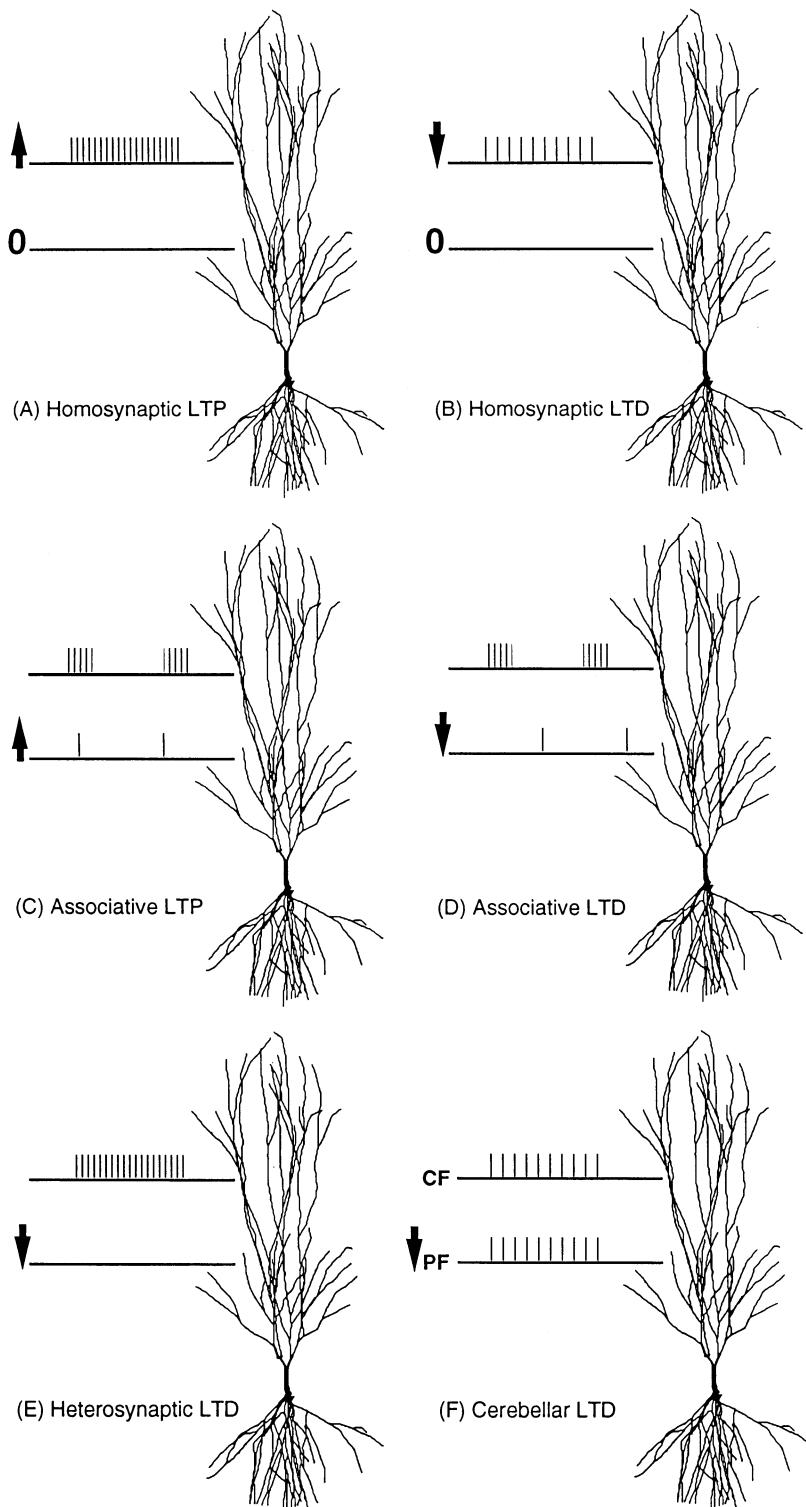


Fig. 8.6.

the result is referred to as a *heterosynaptic* change (e.g., heterosynaptic LTD, Fig. 8.6). In other words, unlike heterosynaptic changes, homosynaptic modifications are input specific or restricted to the stimulate input only. A third kind of activity modification relationship is referred to as an *associative* change (e.g., associative LTP, associative LTD, Fig. 8.6), which is a synaptic change that occurs in one stimulated pathway as a function of the interaction between that pathway and another pathway. In other words, the synaptic modification qualifies as an associative change only if stimulation of either pathway alone does not elicit the desired result either homosynaptically or heterosynaptically.

Heterosynaptic LTD was originally observed as a correlate of homosynaptic LTP induced in the Schaffer collateral inputs to CA1 pyramidal cells in the hippocampal slice [36,73]. Over the years, a more detailed description of heterosynaptic LTD has emerged from studies in hippocampal dentate gyrus [1,26,27,62] and area CA3 [13]. These studies suggest that the induction of heterosynaptic LTD requires strong postsynaptic depolarization caused by synaptic activation and consequent increase in postsynaptic  $\text{Ca}^{2+}$  at neighboring inactive sites [28,86,104]. At present, there is no experimental evidence to address either the molecular mechanisms of heterosynaptic LTD expression or the intermediate steps that may link a moderate rise in postsynaptic  $\text{Ca}^{2+}$  at inactive synapses to the expression of this form of LTD.

Associative LTD has been reported in hippocampal fields CA1, CA3, and dentate gyrus [24,25,26,95]. In area CA1 of the hippocampus, associative LTD is produced when a low frequency test input is negatively correlated in time with a high frequency conditioning input [95]. In these experiments, a conditioning pathway received high frequency bursts (each burst had 5 pulses at a frequency of 100 Hz, the interburst interval was 200 ms) and a separate test pathway received low frequency stimuli at 5 Hz. Each shock of the test input was either superimposed on the middle of each burst of the conditioning input (in-phase), or occurred symmetrically between the bursts (out-of-phase) (Fig. 8.6). The in-phase stimulation induced associative LTP of the test pathway, in agreement with previous results. The new finding was that the out-of-phase stimulation induced associative LTD of the test pathway. This form of associative LTD was suggested to result from pairing of the test input with hyperpolarization caused by the out-of-phase conditioning input, since low frequency stimuli to the test pathway had no effect when applied alone, whereas pairing the same low frequency stimuli with postsynaptic hyperpolarization induced LTD of that pathway [95]. Similar to the CA1 area, CA3 commissural/associational synapses receiving low frequency stimuli out-of-phase with high frequency mossy fiber bursts showed associative LTD [24]. However, the reverse protocol fails to elicit LTD of the mossy fiber pathway [24]. Associative LTD has also been demonstrated at the perforant path-dentate gyrus synapse in the hippocampus.

Whereas associative LTD is not induced in naive perforant path inputs by an out-of-phase stimulation paradigm [26], this same stimulus protocol is capable of inducing associative LTD when the test pathway was previously "primed" with a low frequency (5 Hz) stimulation [27].

There is now evidence for homosynaptic LTD in the hippocampus [34,82] and other brain regions [4,10,66]. Recently Dudek and Bear (1992) and Mulkey and Malenka (1992) have shown that in the CA1 region of the hippocampus, prolonged, low frequency afferent stimulation (1–5 Hz for 5–15 min) produces a saturable and stable homosynaptic LTD. This form of homosynaptic LTD may be induced in previously potentiated pathways, and may be reversed by induction of LTP [35,82]. Surprisingly, the induction of this form of LTD is quite similar to LTP in that it requires activation of NMDA receptors and is blocked by strong hyperpolarization or buffering postsynaptic  $\text{Ca}^{2+}$ . However, it is likely that the low frequency stimulation which results in homosynaptic LTD causes a smaller postsynaptic  $\text{Ca}^{2+}$  transient than the high frequency stimulation which induces homosynaptic LTP. The same type low frequency stimulation can also induce homosynaptic LTD layer IV of the visual cortex [59]. Furthermore, this form of homosynaptic LTD in the hippocampus has much in common with previous reports of "depotentiation" following LTP [96].

Homosynaptic LTD can also be induced by high frequency stimulation (typically 50–100 Hz for 1–5 s) at nonhippocampal synapses (see Linden, 1994 for a detailed review). High frequency stimulation has been reported to be effective at striatal synapses [23,72]. Both low [59] and high [3,48] frequency stimulation protocols have been used at neocortical synapses. While there appears to be a wide range of stimulation protocols that are capable of inducing homosynaptic LTD in a number of brain structures, it is likely that a moderate increase in postsynaptic  $\text{Ca}^{2+}$  or a related co-factor is a critical component in the intracellular biochemistry underlying homosynaptic LTD.

In the cerebellum, Ito and co-workers [50] first described an anti-Hebbian form of LTD that is elicited at the parallel fiber (PF)–Purkinje cell (PC) synapse when parallel fiber and climbing fiber (CF) inputs are activated in phase and at low frequency (1–4 Hz). This form of cerebellar LTD is input specific and is not induced by PF or CF stimulation alone. Several findings suggest that an absolute requirement for cerebellar LTD is a rise in postsynaptic  $\text{Ca}^{2+}$  [61,64–66,88].

Thus, there is increasing evidence suggesting that postsynaptic  $\text{Ca}^{2+}$  signals may play an important role in the induction of LTP as well as LTD in the hippocampus, cerebral cortex, and the cerebellum. A common theme that emerges from the experimental results to date is that the magnitude and spatio-temporal characteristics of postsynaptic signals may determine the direction of synaptic change, although the specific characteristics of this  $\text{Ca}^{2+}$  signal may vary in different regions of the brain.

## 8.5 Interaction Between Dendritic Signaling and Hebbian Learning

Of particular relevance to theoretical investigations of NMDA receptor-dependent synaptic modifications is the fact that one key postsynaptic component of the interactive mechanism appears to be the *local voltage* at the subsynaptic membrane—as opposed to a more global postsynaptic parameter. The difference between a local and a global postsynaptic activity term is important because hippocampal neurons are not isopotential in response to synaptic inputs [15,21,30,76,100,101]. Synaptic signals can in fact suffer extreme attenuation as they spread passively from one point on the dendritic tree to another [21,30,76,100,101].

Computer simulations of realistic representations of hippocampal neurons containing Hebbian synapses have supported the prediction [15] that electrotonic structure is an important determinant of the manner in which Hebbian synapses self-organize in hippocampal neurons [19,20,21,74,75,101]. The results demonstrated that the electrotonic structure causes two specific kinds of self-organization, neither of which occur in an isopotential processing element. The first of these learning phenomena is termed *pattern selection* and the second is called *cluster formation* [19,20,74,75]. Pattern selection results from an increase or decrease of the average synaptic conductance for the affected patterns. Cluster formation is the emergence of mosaics of synapses that have similar conductances. The manner in which electrotonic structure affect both phenomena has been treated formally [101]. The results show quantitatively how the synaptic strengths are regulated by spatio-temporal pattern of presynaptic activity filtered by the electrotonic structure of the neuron. Spatio-temporal patterns can also be learned in a Hebbian way in a network with a distribution of delays [39,44,45,102].

More recently, the simulations have been extended to nonlinear models of hippocampal neurons that incorporate realistic active membrane properties [37]. Both pattern selection and cluster formation persist in the nonlinear case, although the dynamics are more complex. Making the transition to the nonlinear case will require a strategy similar to the linear case—extensive simulations will be required to gain some intuition about the parameter space and then a more formal analysis can follow.

Understanding the complexities of the nonlinear case will also require additional experimental specifications regarding the spatial distribution of active membrane. This was clear in our earlier simulations [105] which showed that the depolarization at the subsynaptic membrane due to a postsynaptic train of fast somatic action potentials was not sufficient, when paired with a weak presynaptic stimulus, to induce LTP. However, the results were quite different when a uniform density of low-threshold dendritic  $\text{Ca}^{2+}$  channels was included in the model, fast somatic action potentials activated the

slower dendritic  $\text{Ca}^{2+}$  currents. The depolarization mediated by these dendritic  $\text{Ca}^{2+}$  channels in response to somatic sodium spikes was sufficient, when paired with a weak stimulus, for the induction of LTP.

These simulations raised several questions with important theoretical implications—to what extent can backpropagating electrical activity supply the depolarization required for the induction of Hebbian synaptic enhancement? Recent experiments using simultaneous whole-cell recordings from the soma and the apical dendrite or axon of the same cell, indicate that synaptically evoked action potentials are initiated first in the axon and then actively propagate back into the dendritic tree [98]. Imaging studies have also shown that action potentials elicited in the soma in fact can result in considerable influx of  $\text{Ca}^{2+}$  in the dendrite [53,81]. Thus, backpropagating action potentials may trigger voltage-gated  $\text{Ca}^{2+}$  influx, resulting in prolonged postsynaptic depolarization and increase in postsynaptic  $\text{Ca}^{2+}$  levels, both of which would seem to be relevant to synaptic plasticity.

In conclusion, to appreciate Hebbian learning more fully we need to understand the interactions between dendritic signaling and the synaptic modification mechanisms. This requires a more complete picture of the microdomains of postsynaptic  $\text{Ca}^{2+}$  signals and other second messengers [52,53]. We then need to combine this information with detailed knowledge of the distribution of active membrane conductances on the dendritic tree. Fortunately, the relevant technology for addressing these issues is rapidly emerging [52,81,83,98].

*Acknowledgments.* This work was supported by the Office of Naval Research.

## REFERENCES

- [1] Abraham, W. C., Goddard, G. V. (1983) Asymmetric relationships between homosynaptic long-term potentiation and heterosynaptic long-term depression. *Nature* **305**: 717–719
- [2] Anderson, J. A. (1985) What Hebb synapses build. In: *Synaptic Modification, Neuron Selectivity, and Nervous System Organization*, W. B. Levy, J. A. Anderson, S. Lehmkuhle (Eds.) (Erlbaum, Hillsdale, NJ), pp. 153–173
- [3] Artola, A., Brocher, S., Singer, W. (1990) Different voltage-dependent thresholds for the induction of long-term depression and long-term potentiation in slices of the rat visual cortex. *Nature* **347**: 69–72
- [4] Artola, A., Singer, W. (1993) Long-term depression of excitatory synaptic transmission and its relationship to long-term potentiation. *Trends Neurosci.* **16**: 480–487
- [5] Bashir, Z. I., Bortolotto, Z. A., Davies, C. H., Berretta, N., Irving, A. J., Seal, A. J., Henley, J. M., Jane, D. E., Watkins, J. C., Collingridge, G. L.

- (1993) Induction of LTP in the hippocampus needs synaptic activation of glutamate metabotropic receptors. *Nature* **363**: 347–350
- [6] Bear, M. F., Cooper, L. N., Ebner, F. F. (1987) A physiological basis for a theory of synapse modification. *Science* **237**: 42–48
  - [7] Bear, M. F., Cooper, L. N. (1990) Molecular mechanisms for synaptic modification in the visual cortex: Interaction between theory and experiment. In: *Neuroscience and Connectionist Theory*, M. A. Gluck, D. E. Rumelhart (Eds.) (Erlbaum, Hillsdale, NJ), pp. 65–93
  - [8] Bekkers, J. M., Stevens, C. F. (1989) NMDA and non-NMDA receptors are co-localized at individual excitatory synapses in cultured rat hippocampus. *Nature* **341**: 230–233
  - [9] Bienenstock, E. L., Cooper, L. N., Munro, P. W. (1982) Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *J. Neurosci.* **2**: 32–48
  - [10] Bindman, L., Christofi, G., Murphy, K., Nowicky, A. (1991) Long-term potentiation (LTP) and depression (LTD) in the neocortex and hippocampus: an overview. In: *Aspects of Synaptic Transmission*, T. W. Stone (Ed.) (Taylor Francis, London), pp. 3–25
  - [11] Bliss, T. V. P., Lomo, T. (1973) Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *J. Physiol.* **232**: 331–356
  - [12] Bliss, T. V. P., Collingridge, G. L. (1993) A synaptic model of memory: Long-term potentiation in the hippocampus. *Nature* **361**: 31–39
  - [13] Bradler, J. E., Barrioanevo, G. (1989) Long-term potentiation in hippocampal CA3 neurons: Tetanized input regulates heterosynaptic efficacy. *Synapse* **4**: 132–142
  - [14] Brown, T. H., Johnston, D. (1983) Voltage-clamp analysis of mossy fiber synaptic input to hippocampal neurons. *J. Neurophysiol.* **50**: 487–507
  - [15] Brown, T. H., Chang, V. C., Ganong, A. H., Keenan, C. L., Kelso, S. R. (1988a) Biophysical properties of dendrites and spines that may control the induction and expression of long-term synaptic potentiation. In: *Long-term Potentiation: From Biophysics to Behavior*, P. W. Landfield, S. Deadwyler (Eds.) (Alan R. Liss, New York), pp. 201–264
  - [16] Brown, T. H., Chapman, P. F., Kairiss, E. W., Keenan, C. L. (1988b) Long-term synaptic potentiation. *Science* **242**: 724–728
  - [17] Brown, T. H., Ganong, A. H., Kairiss, E. W., Keenan, C. L., Kelso, S. R. (1989) Long-term potentiation in two synaptic systems of the hippocampal brain slice. In: *Neural Models of Plasticity*, J. H. Byrne, W. O. Berry (Eds.) (Academic Press, New York), pp. 266–306
  - [18] Brown, T. H., Ganong, A. H., Kairiss, E. W., Keenan, C. L. (1990) Hebbian synapses: Biophysical mechanisms and algorithms. *Annu. Rev. Neurosci.* **13**: 475–511

- [19] Brown, T. H., Mainen, Z. F., Zador, A. M., Claiborne, B. J. (1991a) Self-organization of Hebbian synapses in hippocampal neurons. In: *Advances in Neural Information Processing Systems*, R. P. Lippmann, J. E. Moody, D. J. Touretzky (Eds.) (Morgan Kaufmann, San Mateo, CA), pp. 39–45
- [20] Brown, T. H., Zador, A. M., Mainen, Z. F., Claiborne, B. J. (1991b) Hebbian modifications in hippocampal neurons. In: *Long-Term Potentiation: A Debate of Current Issues*, J. Davis, M. Baudry (Eds.) (MIT Press, Cambridge), pp. 357–389
- [21] Brown, T. H., Zador, A. M., Mainen, Z. F., Claiborne, B. J. (1992) Hebbian computations in hippocampal dendrites and spines. In: *Single Neuron Computation*, T. McKenna, J. Davis, S. F. Zornetzer (Eds.) (Academic Press, New York), pp. 81–116
- [22] Cajal, S. Ramon y. (1911) *Histologie du Systeme Nerveux* (Maloine, Paris), Vol. 2
- [23] Calabresi, P., Maj, R., Pisani, A., Mercuri, N. B., Bernardi, G. (1992) Long-term synaptic depression in the striatum: Physiological and pharmacological characterization. *J. Neurosci.* **12**: 4224–4233
- [24] Chattarji, S., Stanton, P. K., Sejnowski, T. J. (1989) Commissural synapses, but not mossy fiber synapses, in hippocampal field CA3 exhibit associative long-term potentiation and depression. *Brain Res.* **495**: 145–150
- [25] Chattarji, S. (1992) Properties of long-term synaptic depression in area CA1 of the hippocampus. Ph. D. thesis, Johns Hopkins University
- [26] Christie, B. R., Abraham, W. C. (1992a) NMDA-dependent heterosynaptic long-term depression in the dentate gyrus of anesthetized rats. *Synapse* **10**: 1–6
- [27] Christie, B. R., Abraham, W. C. (1992b) Priming of associative long-term depression in the dentate gyrus by theta frequency synaptic activity. *Neuron* **9**: 79–84
- [28] Christofi, G., Nowicky, A. V., Bolsover, S. R., Bindman, L. J. (1993) The postsynaptic induction of nonassociative long-term depression of excitatory synaptic transmission in rat hippocampal slices. *J. Neurophysiol.* **69**: 219–229
- [29] Claiborne, B. J., Xiang, Z., and Brown, T. H. (1993) Hippocampal circuitry complicates analysis of long-term potentiation in mossy fiber synapses. *Hippocampus* **3**: 115–122
- [30] Claiborne, B. J., Zador, A. M., Mainen, Z. F., Brown, T. H. (1992) Computational models of hippocampal neurons. In: *Single Neuron Computation*, T. McKenna, J. Davis, S. F. Zornetzer (Eds.) (Academic Press, New York), pp. 61–79
- [31] Collingridge, G. L., Bliss, T. V. P. (1987) NMDA receptors: Their role in long-term potentiation. *Trends Neurosci.* **10**: 288–293
- [32] Constantine-Paton, M., Cline, H. T., Debski, E. (1990) Patterned activity, synaptic convergence, and the NMDA receptor in developing visual pathways. *Annu. Rev. Neurosci.* **13**: 129–154

- [33] Cotman, C. W., Monaghan, D. T., Ganong, A. H. (1988) Excitatory amino acid neurotransmission: NMDA receptors and Hebb-type synaptic plasticity. *Annu. Rev. Neurosci.* **11**: 61–80
- [34] Dudek, S. M., Bear, M. F. (1992) Homosynaptic long-term depression in area CA1 of hippocampus and effects of *N*-methyl-*D*-aspartate receptor blockade. *Proc. Natl. Acad. Sci. U.S.A.* **89**: 4363–4367
- [35] Dudek, S. M., Bear, M. F. (1993) Bidirectional long-term modification of synaptic effectiveness in the adult and immature hippocampus. *J. Neurosci.* **13**: 2910–2918
- [36] Dunwiddie, T., Lynch, G. (1978) Long-term potentiation and depression of synaptic responses in the rat hippocampus: Localization and frequency dependency. *J. Physiol.* **276**: 353–367
- [37] Fisher, S. A., Jaffe, D. B., Claiborne, B. J., Brown, T. H. (1993) Self-organization of Hebbian synapses on a biophysical model of a hippocampal neuron. *Soc. Neurosci. Abstr.* **19**: 808
- [38] Fregnac, Y., Shultz, D. (1991) Models of synaptic plasticity and cellular learning in mammalian visual cortex. In: *Advances in Neural and Behavioral Development*, V. Casagrande, P. Shinkman (Eds.) (Neural Ablex Publishers), pp. 97–109
- [39] Gerstner, W., Ritz, R., van Hemmen, J. L. (1993) Why spikes? Hebbian learning and retrieval of time resolved excitation patterns. *Biol. Cybern.* **69**: 503–515
- [40] Gill, D. L., Udeda, T., Chueh, S. H., Noel, M. W. (1986) Ca release from endoplasmic reticulum is mediated by a guanine nucleotide regulatory mechanism. *Nature* **320**: 461–464
- [41] Gustafsson, B., Wigstrom, H., Abraham, W., Huang, C. (1987) Long-term synaptic potentiation in the hippocampus using depolarizing current pulses as the conditioning stimulus to single volley synaptic potentials. *J. Neurosci.* **7**: 774–780
- [42] Guthrie, P. B., Segal, M., Kater, S. B. (1991) Independent regulation of calcium revealed by imaging dendritic spines. *Nature* **354**: 76–80
- [43] Hebb, D. O. (1949) *The Organization of Behavior* (Wiley, New York)
- [44] Herz, A. V. M., Sulzer, B., Kühn, R., van Hemmen, J. L. (1988a) The Hebb rule: Storing static and dynamic objects in an associative neural network. *Europhys. Lett.* **7**: 663–669
- [45] Herz, A. V. M., Sulzer, B., Kühn, R., van Hemmen, J. L. (1988b) Hebbian learning reconsidered: Representation of static and dynamic objects. *Biol. Cybern.* **60**: 457–467
- [46] Hines, M. (1984) Efficient computation of branched nerve equations. *Int. J. Biomed. Comput.* **15**: 69–76
- [47] Hines, M. (1980) A program for simulation of nerve equations with branching geometries. *Int. J. Biomed. Comput.* **24**: 55–68

- [48] Hirsch, J. C., Crepel, F. (1990) Use-dependent changes in synaptic efficacy in rat prefrontal neurons in vitro. *J. Physiol.* **427**: 31–49
- [49] Hopfield, J. J., Tank, D. W. (1986) Computing with neural circuits: A model. *Science* **233**: 625–633
- [50] Ito, M. (1989) Long-term depression. *Annu. Rev. Neurosci.* **12**: 85–102
- [51] Jaffe, D. B., Johnston, D., (1990) Induction of long-term potentiation at hippocampal mossy fiber synapses follows a Hebbian rule. *J. Neurophysiol.* **64**: 948–960
- [52] Jaffe, D. B., Brown, T. H. (1994) Confocal imaging of dendritic calcium transients in hippocampal brain slices with simultaneous current and voltage clamp recordings. *Microscopy Res. Technol.*, in press
- [53] Jaffe, D. B., Fisher, S. A., Brown, T. H. (1994) Confocal laser scanning microscopy reveals voltage-gated calcium signals within hippocampal dendritic spines. *J. Neurobiol.*, in press
- [54] Jahr, C. E., Stevens, C. F. (1987) Glutamate activates multiple single channel conductances in hippocampal neurons. *Nature* **325**: 522–525
- [55] James, W. (1890) *Psychology: Briefer Course* (Harvard University Press, Cambridge)
- [56] Johnston, D., Williams, S., Jaffe, D. B., Gray, R. (1992) NMDA receptor-independent long-term potentiation. *Annu. Rev. Physiol.* **54**: 489–505
- [57] Kelso, S. R., Brown, T. H. (1986) Differential conditioning of associative synaptic enhancement in hippocampal brain slices. *Science* **232**: 85–87
- [58] Kelso, S. R., Ganong, A. H., Brown, T. H. (1986) Hebbian synapses in hippocampus, 1986. *Proc. Natl. Acad. Sci. U.S.A.* **83**: 5326–5330
- [59] Kirkwood, A., Dudek, S. M., Gold, J. T., Aizenman, C. D., Bear, M. F. (1993) Common forms of synaptic plasticity in the hippocampus and neocortex in vitro. *Science* **260**: 1518–1521
- [60] Kohonen, T. (1984) *Self-Organization and Associative Memory* (Springer-Verlag, Berlin)
- [61] Konnerth, A., Dreessen, J., Augustine, G. J. (1992) Brief dendritic calcium signals initiate long-lasting synaptic depression in cerebellar Purkinje cells. *Proc. Natl. Acad. Sci. U.S.A.* **89**: 7051–7055
- [62] Levy, W. B., Steward, O. (1979) Synapses as associative memory elements in the hippocampal formation. *Brain Res.* **175**: 233–245
- [63] Levy, W. B., Burger, B. (1987) *IEEE 1st International Conference on Neural Networks, San Diego* (IEEE, San Diego), Vol. 4, pp. 11–15
- [64] Linden, D. J., Connor, J. A. (1991) Participation of postsynaptic PKC in cerebellar long-term depression in culture. *Science* **254**: 1656–1659
- [65] Linden, D. J., Dickinson, M. H., Smeyne, M., Connor, J. A. (1991) A long-term depression of AMPA currents in cultured cerebellar Purkinje neurons. *Neuron* **7**: 81–89

- [66] Linden, D. J. (1994) Long-term synaptic depression in the mammalian brain. *Neuron* **12**: 457–472
- [67] Linsker, R. (1986a) Emergence of orientation columns. *Proc. Natl. Acad. Sci. U.S.A.* **83**: 8779–8783
- [68] Linsker, R. (1986b) From basic network principles to neural architecture. *Proc. Natl. Acad. Sci. U.S.A.* **83**: 7508–7512
- [69] Linsker, R. (1988a) Towards and organizing principle for a layered perceptual network. In: *Neural Information Processing Systems*, D. Andersen (Ed.) (American Institute of Physics, New York), pp. 485–494
- [70] Linsker, R. (1988b) Self-organization in a perceptual network. *Computer* **21**: 105–117
- [71] Lisman, J. A. (1989) A mechanism for the Hebb and anti-Hebb processes underlying learning and memory, 1989. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 9574–9578
- [72] Lovinger, D. M., Tyler, E. C., Merritt, A. (1993) Short- and long-term synaptic depression in rat neostriatum. *J. Neurophysiol.* **70**: 1937–1949
- [73] Lynch, G. S., Dunwiddie, T., Gribkoff, V. (1977) Heterosynaptic depression: A postsynaptic correlate of long-term potentiation. *Nature* **266**: 737–739
- [74] Mainen, Z. F., Zador, A. M., Claiborne, B. J., Brown, T. H. (1990) Hebbian synapses induce feature mosaics in hippocampal dendrites. *Soc. Neurosci. Abstr.* **16**: 492
- [75] Mainen, Z. F., Claiborne, B. J., Brown, T. H. (1991) A novel role for synaptic competition in the development of cortical lamination. *Soc. Neurosci. Abstr.* **17**: 759
- [76] Mainen, Z. F., Zador, A. M., Claiborne, B. J., Brown, T. H. (1994) Electrotonic structure of hippocampal CA1 pyramidal neurons based on three-dimensional reconstructions. *J. Neurophysiol.*, in preparation
- [77] Malenka, R. C. (1993) Long-term depression: Not so depressing after all. *Proc. Natl. Acad. Sci. U.S.A.* **90**: 3121–3123
- [78] Malinow, R., Miller, J. P. (1986) Postsynaptic hyperpolarization during conditioning reversibly blocks induction of long-term potentiation. *Nature* **320**: 529–530
- [79] Mayer, M. L., Westbrook, G. L., Guthrie, P. B. (1984) Voltage-dependent block by  $Mg^{2+}$  of NMDA responses in spinal cord neurones. *Nature* **309**: 261–263
- [80] Miller, K. D., Keller, J. B., Stryker, M. (1989) Ocular dominance column development: Analysis and simulation. *Science* **245**: 605–615
- [81] Miyakawa, H., Ross, W. N., Jaffe, D. B., Callaway, J. C., Lasser-Ross, N., Lisman, J. E., Johnston, D. (1992) Synaptically activated increases in  $Ca^{2+}$  concentration in hippocampal CA1 pyramidal cells are primarily due to voltage-gated  $Ca^{2+}$  channels. *Neuron* **9**: 1163–1173

- [82] Mulkey, R. M., Malenka, R. C. (1992) Mechanisms underlying induction of homosynaptic long-term depression in area CA1 of the hippocampus. *Neuron* **9**: 967–975
- [83] Muller, W., Connor, J. A. (1991) Dendritic spines as individual neuronal compartments for synaptic  $\text{Ca}^{2+}$  responses. *Nature* **354**: 73–76
- [84] Nowak, L., Bregestovski, P., Ascher, P., Herbet, A., Prochiantz, A. (1984) Magnesium gates glutamate-activated channels in mouse central neurones. *Nature* **307**: 462–465
- [85] Penner, R., Fasolato, C., Hoth, M. (1993) Calcium influx and its control by calcium release. *Curr. Opinion Neurobiol.* **3**: 368–374
- [86] Pockett, S., Brookes, N. H., Bindman, L. J. (1990) Long-term depression at synapses in slices of rat hippocampus can be induced by bursts of postsynaptic activity. *Exp. Brain Res.* **80**: 196–200
- [87] Reiter, H. O., Stryker, M. P. (1988) Neural plasticity without postsynaptic action potentials: Less active inputs become dominant when kitten visual cortical cells are pharmacologically inhibited. *Proc. Natl. Acad. Sci. U.S.A.* **85**: 3623–3627
- [88] Sakurai, M. (1990) Calcium is an intracellular mediator of the climbing fiber in induction of cerebellar long-term depression. *Proc. Natl. Acad. Sci. U.S.A.* **87**: 3383–3385
- [89] Sejnowski, T. J. (1977a) Storing covariance with nonlinearly interacting neurons. *J. Math. Biol.* **4**: 303–321
- [90] Sejnowski, T. J. (1977b) Statistical constraints on synaptic plasticity. *J. Theor. Biol.* **69**: 385–389
- [91] Sejnowski, T. J., Chattarji, S., Stanton, P. K. (1990) Homosynaptic long-term depression in hippocampus and neocortex. *Seminars Neurosci.* **2**: 355–363
- [92] Sejnowski, T. J., Chattarji, S., Stanton, P. K. (1989) Induction of synaptic plasticity by Hebbian covariance in the hippocampus. In: *The Computing Neuron*, R. Durbin, C. Miall, G. Mitchison (Eds.) (Addison-Wesley, Reading, MA), pp. 105–124
- [93] Sejnowski, T. J., Tesauro, G. (1989) The Hebb rule for synaptic plasticity: Algorithms and implementations. In: *Neural Models of Plasticity*, J. H. Byrne, W. O. Berry (Eds.) (Academic Press, New York), pp. 94–103
- [94] Shatz, C. J., Stryker, M. P. (1978) Ocular dominance in layer IV of the cat's visual cortex and the effects of monocular deprivation. *J. Physiol.* **281**: 267–283
- [95] Stanton, P. K., Sejnowski, T. J. (1989) Associative long-term depression in the hippocampus induced by Hebbian covariance. *Nature* **229**: 215–218
- [96] Staubli, U., Lynch, G. (1990) Stable depression of potentiated synaptic responses in the hippocampus with 1–5 Hz stimulation. *Brain Res.* **513**: 113–118

- [97] Stent, G. S. (1973) A physiological mechanism for Hebb's postulate of learning. Proc. Natl. Acad. Sci. U.S.A. **70**: 997–1001
- [98] Stuart, G. J., Sakmann, B. (1994) Active propagation of somatic action potentials into neocortical pyramidal cell dendrites. Nature **367**: 69–72
- [99] Tanzi, E. (1893) I fatti e la induzioni nell 'odierna istologia del sistema nervoso. Riv. Sper. Freniatr. Med. Leg. Alienazioni Ment. Soc. Ital. Psichiatria **19**: 419–472
- [100] Tsai, K. Y., Carnevale, N. T., Claiborne, B. J., Brown, T. H. (1993) Morphoelectronic transforms in three classes of rat hippocampal neurons. Soc. Neurosci. Abst. **19**: 1522
- [101] Tsai, K. Y., Carnevale, N. T., Brown, T. H. (1994) Hebbian learning is jointly controlled by electrotonic and input structure. Network **5**: 1–19
- [102] van Hemmen, J. L., Gerstner, W., Herz, A. V. M., Kuehn, R., Sulzer, B., Vaas, M. (1990) Encoding and decoding of patterns which are correlated in space and time. In: *Konnektionismus in Artificial Intelligence und Kognitionsforschung*, G. Dorffner (Ed.) (Springer-Verlag, Berlin), pp. 153–162
- [103] Willshaw, D., Dayan, P. (1990) Optimal plasticity from matrix memories: What goes up must come down. Neural Comput. **2**: 85–93
- [104] Wickens, J. R., Abraham, W. C. (1991) The involvement of *L*-type calcium channels in heterosynaptic long-term depression in the hippocampus. Neurosci. Lett. **130**: 128–132
- [105] Zador, A., Koch, C., Brown, T. H. (1990), Biophysical model of a Hebbian synapse. Proc. Natl. Acad. Sci. U.S.A. **87**: 6718–6722
- [106] Zalutsky, R. A., Nicoll, R. A. (1990) Comparison of two forms of long-term potentiation in single hippocampal neurons. Science **248**: 1619–1624

# Reentry and Dynamical Interactions of Cortical Networks

Olaf Sporns, Giulio Tononi, and Gerald M. Edelman

with 12 figures

**Synopsis.** Recent advances in our understanding of the functional organization of the cerebral cortex pose the problem of how cortical neural activity is integrated. Even within one sensory domain, such as vision, there is a multitude of functionally segregated areas. However, visual perception always appears unitary and coherent. In order to achieve perceptual coherence as well as an integrated behavioral response to stimuli composed of many elementary features and attributes, these distributed neural signals must be put together. In this chapter, we present a set of computer simulations based on the anatomy and physiology of the cortex that address the problem of cortical integration. We propose that integration takes place at many different levels of organization: locally within populations of neurons, within a given cortical area (linking) and across several areas (binding). We discuss some stringent temporal constraints for neural integration and the possible effects of patterns of correlations on behavior.

## 9.1 Introduction

The mammalian cerebral cortex is partitioned into several areas, each specialized to carry out a relatively specific function [58,60]. A recent survey [21] of the visual cortex of the macaque monkey listed over 30 specialized areas in that part of the cortex alone. In view of this high degree of functional segregation the question arises how neural activity in the various areas is integrated to form a unified percept of the world and to allow coherent behavioral responses. This problem of *cortical integration* has been of central concern in theoretical accounts of brain function [11,16,17,20,22,46,53,54,60]. The purpose of this chapter is to present a se-

ries of realistic computer models of cortical networks that demonstrate how cortical integration might be achieved.

Most earlier attempts to explain integration have started from the notion that the visual cortex is organized hierarchically with a progressive increase in the specificity of its neurons from the sensory periphery to more central areas. According to one version of the model, perceptual integration is achieved by the confluence of diverse processing streams at a very high hierarchical level or "master area." However, while there are areas of the brain that receive convergent inputs from multiple modalities or submodalities [11], no single area has been identified that receives a sufficiently rich set of afferents to allow it to carry out a universal integrative function.

An alternative view of integration takes into account the cooperative interactions between different areas of the brain. This view is consistent with a characteristic anatomic feature of cortical organization: the abundance of reciprocal pathways linking functionally segregated areas. In the visual system, such reciprocal pathways exist between areas at the same or different hierarchical levels [21]. In addition neurons within each area are often densely and reciprocally interconnected [34]. What is the function of reciprocal or return projections in the cortex? Some return projections are anatomically diffuse [59] and may serve a modulatory function, while others can directly drive neurons in lower areas [40]. Edelman (1978) proposed that reciprocal pathways form the substrate of cortical integration by allowing the dynamic, bidirectional exchange of neural signals between areas in a process called *reentry*. Reentry has several important characteristics: It occurs between neuronal populations either within the same cortical map or between separate maps. It involves ongoing and parallel exchange of signals between the reentrantly linked cell populations. The anatomic projections subserving reentry are usually ordered, both locally (e.g., in terms of arborization patterns of axons) as well as globally (e.g., in terms of topography). Computer modeling of reentrant projections has shown that there are at least two ways in which reentry can act. It can directly influence the response properties of neurons (*constructive* function; e.g., [22]) or it can give rise to patterns of temporal correlations (*correlative* function; e.g., [49]).

Recently, microelectrode recordings in the cat visual cortex have provided new evidence for the correlative function of reentry by revealing stimulus-dependent patterns of correlated neural activity. Gray and Singer (1989) reported that orientation-selective neurons show oscillatory discharges at around 40 Hz when presented with an optimally oriented stimulus. These oscillations can be observed in single cells [31], as well as local populations of cells. Such populations, characterized by shared receptive field properties and temporally correlated discharge patterns, have been called *neuronal groups* [15]. Oscillatory activity of cortical neurons in the 40 Hz frequency range has also been reported in motor cortex [41], inferotemporal cortex [42], and visual cortex of the monkey [14,38]. 40 Hz electrical activity has

also been recorded from human subjects [39,45] in various states of sleep and wakefulness.

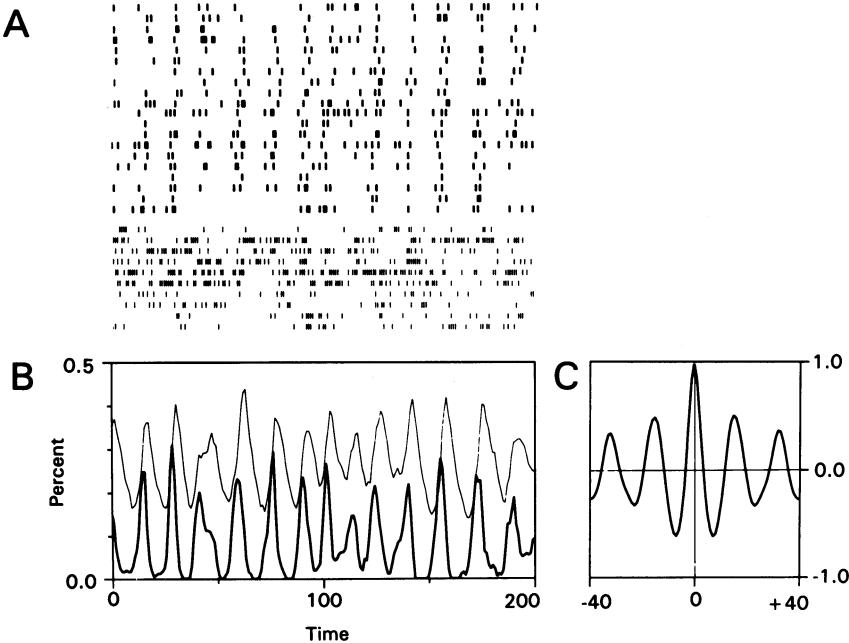
Further experiments revealed widespread patterns of correlations in the cat visual system and provided support for the notion that these patterns depend on reentrant connectivity. When a single long light bar is moved across the receptive fields of spatially separated neurons with similar orientation specificity, cross-correlations reveal that their oscillatory responses are synchronized [12,27]. The synchronization becomes weaker if a gap is inserted into the stimulus contour and it disappears completely if two parts of the contour are moved separately and in opposite directions. Synchrony is established rapidly, often within 100 ms, and single episodes of coherency last for 50 to 500 ms. Frequency and phase of the oscillations vary within the range of 40–60 Hz and  $\pm 3$  ms, respectively [28]. Synchronized neural activity was observed between cortical areas V1 and V2 [12,43] and between striate and extrastriate cortical areas [18]. Engel et al. (1991b) have shown stimulus-dependent correlations between neurons located in the two hemispheres of cat visual cortex in response to two simultaneously presented stimuli, one on each side of the visual midline. These correlations disappear when the corpus callosum is transected, an indication that reentrant cortico-cortical pathways are responsible for their generation. After transection, the hemispheres continue to show neuronal activity at normal levels. This supports the view that mean activity rates alone are an insufficient indicator of cortical integration, and that the temporal characteristics of neuronal firing play an important role in this process.

In this chapter we will review a series of computer simulations which address the problem of cortical integration. First, we show how populations of neurons may interact locally to produce correlated neuronal discharges. Then we discuss some characteristic dynamic phenomena that result from coupling two or more of these groups. We address a specific problem in visual perception, that of perceptual grouping and figure-ground segregation; we discuss how temporal correlations might serve as the neural basis for these perceptual phenomena. We also consider the effect of temporal correlations on behavioral responses. Finally, we briefly describe some key results obtained with a detailed model of the primate visual system.

## 9.2 Models of Cortical Integration

### 9.2.1 DYNAMIC BEHAVIOR OF SINGLE NEURONAL GROUPS

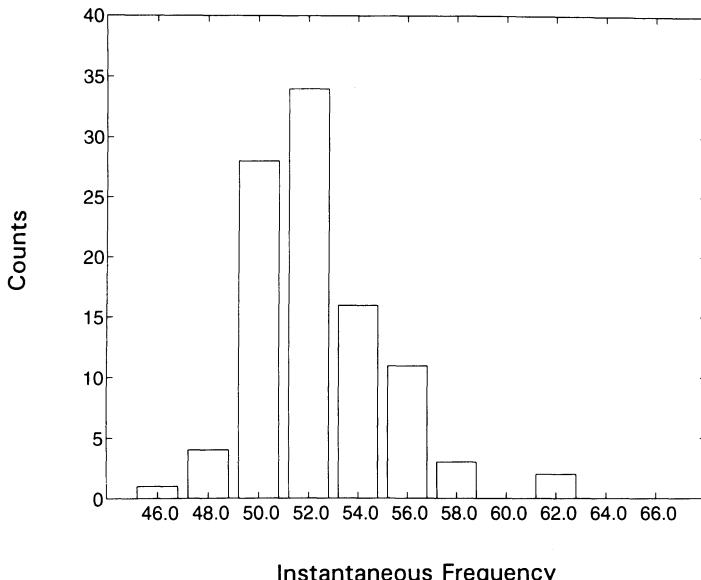
We have modeled neuronal groups in several computer simulations [49–51]. In the models, neuronal groups contain both excitatory and inhibitory neurons; there are connections within the excitatory cell population as well as reciprocal connections linking the excitatory and inhibitory populations.



**Fig. 9.1.** *Rhythmic discharges in a simulated neuronal group.* The group responds to an optimally oriented moving light bar and contains 160 excitatory and 80 inhibitory cells (as in [49]). (A) Firing patterns of 20 randomly selected excitatory (top) and 10 randomly selected inhibitory cells (bottom) for a time period of 200 ms. Over this short time period, individual excitatory cells show a fairly irregular discharge pattern. It is not immediately obvious that these cells have oscillatory characteristics. In the model, inhibitory neurons have a higher spontaneous firing rate which further obscures their oscillatory discharge pattern. (B) Population activity (analogous to a local field potential) of the same neuronal group recorded simultaneously with the single neurons shown in (A). The activity of both excitatory and inhibitory populations (thick and thin traces, respectively) shows a pronounced oscillatory time course. The inhibitory population trails the excitatory population by about 5 ms (about one quarter of the oscillation period). Note that the frequency shows considerable variations. (C) Auto-correlation of the excitatory cell population for the same time period shown in (B).

The connectivity pattern is sparse and random; for example, a given excitatory cell connects to 10% (chosen at random) of all other excitatory cells within the group. This local anatomical pattern is consistent with statistical data obtained from real cortical circuits [52].

The discharge characteristics of a neuronal group are largely determined by cooperative interactions. In the simulations, locally correlated oscillatory activity is generated by the interactions between excitatory and inhibitory cells within a group (Fig. 9.1). The mean oscillation frequency depends critically on the temporal delay introduced by the recurrent inhibitory



**Fig. 9.2.** Histogram distribution of the instantaneous oscillation frequency of a neuronal group (see Fig. 9.1). The instantaneous frequency was obtained from a continuous time series of rhythmic activity under optimal stimulation. The population activity of the group was recorded for 20,000 time steps (ms) and the time series was divided into nonoverlapping episodes of 200 ms each. For each episode the auto-correlogram was computed and a Gabor function was fitted to it (using nonlinear regression according to Levenberg–Marquardt). The instantaneous frequency was derived from the period of the fitted Gabor function. Note that the instantaneous frequency varies significantly and shows a unimodal normal distribution. Different neuronal groups with random connectivity show different distributions (see Fig. 9.3(B) at coupling strength 0.0).

connections. In accordance with experimental results [28], the frequency of the oscillations when estimated for a short time interval of 200 ms (“instantaneous frequency”) varies significantly (Fig. 9.2). This is due to the fact that each neuronal group acts as a population oscillator, composed of many sparsely connected and partly independent neurons. Over short periods of time the coherent activity of such neuronal groups is statistically more reliable than the activity of each of its constituent neurons. Because cortico-cortical connections tend to be ordered (often preserving topography) and axonal arbors of cortical neurons often show local branching with clustered end terminals, a neuronal group in another cortical region is likely to receive numerous correlated inputs from another group that shows locally synchronized activity. Thus, local synchronization in populations of neurons facilitates the “transmission” of correlations between different cortical regions and is a first, elementary step for establishing functionally significant correlations over long distances.

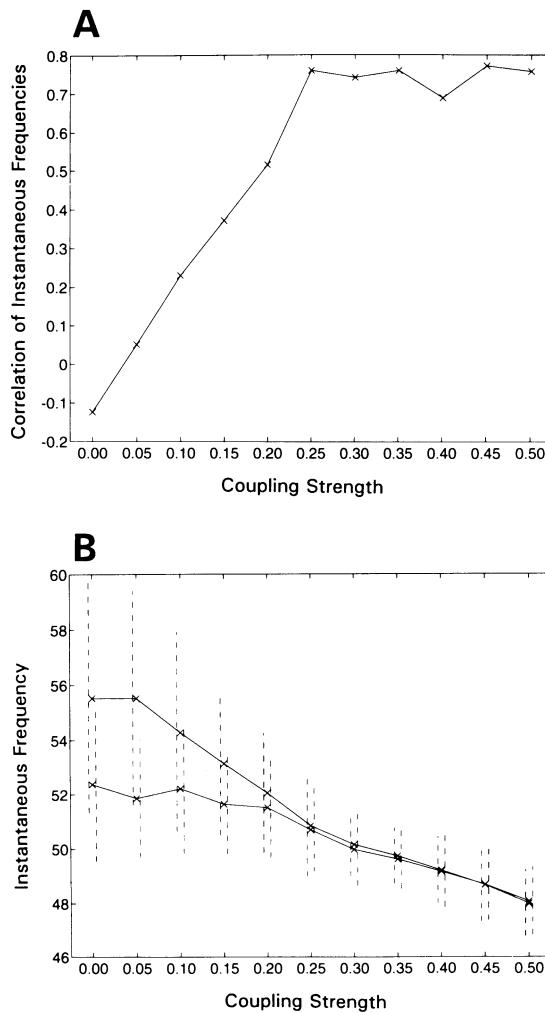
Another important observation is that neuronal groups with sparse and random intrinsic connectivity display dynamic behavior that is highly variable across a population of individual examples. As a result of stochastic processes in neural development, cortical neuronal groups are generated according to a gross anatomic plan but individual groups differ in the precise pattern of termination of connections. In simulations, such variable groups show different mean frequencies and frequency dispersions in the absence of mutual coupling (see Fig. 9.3(B)). In short, dynamic variability is the result of underlying structural variability. Such dynamic variability is of functional importance because it prevents accidental phase locking of groups that are not interacting or that are responding to common input.

### 9.2.2 COUPLED NEURONAL GROUPS

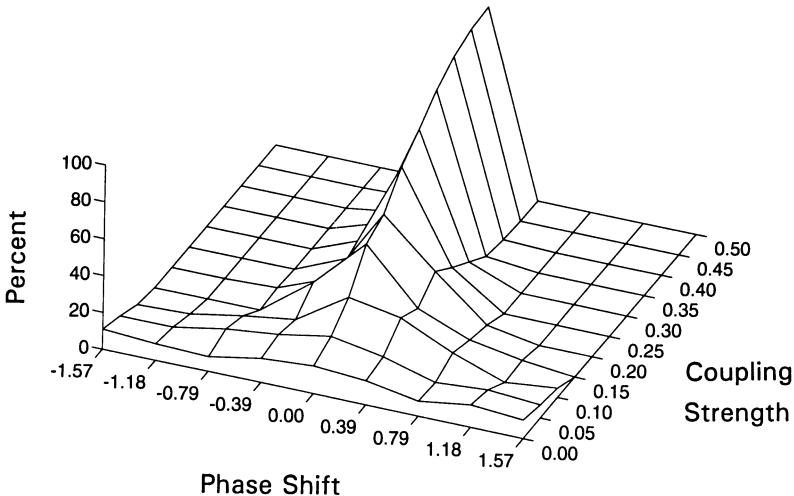
#### Dynamic Effects of Coupling Two Groups

When two neuronal groups are reentrantly coupled (with fixed coupling strength) a variety of new dynamic phenomena emerge. Given a constant input and reciprocal coupling (in the absence of any conduction delays) the groups will synchronize their rhythmic discharge patterns and become temporally correlated. Synchronization requires only weak coupling (relative to the strength of local interactions). In the simulations, the instantaneous frequencies of two coupled groups become more similar as coupling is increased. Furthermore, the group's individual frequencies as well as their frequency dispersions decrease with increasing coupling strength [51, Fig. 9.3]. Thus, even when coupled to other groups, a given neuronal group will display dynamic variability in its oscillatory discharge pattern, although this variability is somewhat reduced. As observed by Gray et al. (1992), the phase shifts of the cross-correlations of two coupled groups are normally distributed with a mean phase shift of zero (Fig. 9.4). At very low coupling strength (or in the absence of coupling) all phase shifts are equally probable and the distribution is flat. Stronger coupling is accompanied by a reduced dispersion of phase shifts and a pronounced phase locking with zero phase lag.

Conduction delays in the reentrant connections linking the two groups influence the distribution of phase shifts significantly (Fig. 9.5). For small delays (1 to 4 ms) the distribution remains centered around a phase shift of zero. At around 5 ms (corresponding to about one-quarter of the 20 ms oscillation period) a transition occurs and for delays above 5 ms most phase shifts are around  $\pm(\pi/2)$ . According to these simulations phase shifts are not equiprobable even with arbitrary conduction delays; phase shifts of 0 or  $\pi/2$  are more likely to occur than  $\pi/4$ . In the brain, correlations (even those found over long distances, e.g., cross-callosally; see [19]) most often have a mean phase shift of 0 ms. On the other hand, the conductance velocity of many cortical fibers and pathways seems to be in the critical range of 5 ms



**Fig. 9.3.** (A) Correlation coefficients of the instantaneous frequencies of two reentrantly coupled neuronal groups as a function of coupling strength (normalized connection strength of reentrant connections). The instantaneous frequencies were obtained as described in Fig. 9.2. (B) Mean instantaneous frequency and frequency dispersion (vertical dashed bars) for two reentrantly coupled neuronal groups as a function of coupling strength. Same data as in (A). Both mean frequency and dispersion become more similar for the two groups and decrease for higher coupling strengths.



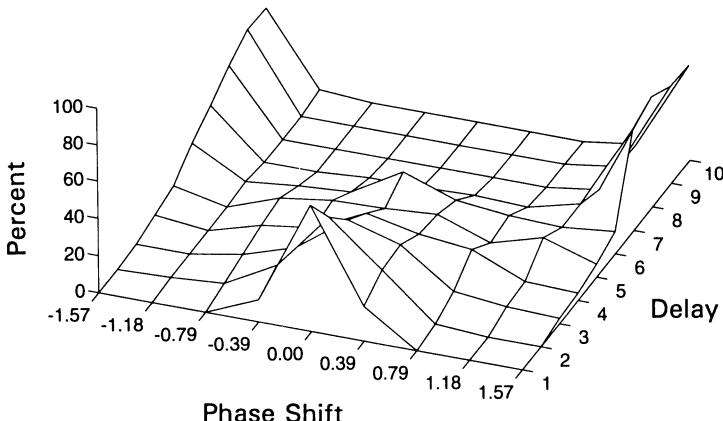
**Fig. 9.4.** Distribution of phase shifts for two reentrantly coupled neuronal groups as a function of coupling strength. Phase shifts are obtained using fitted Gabor functions (as described in Fig. 9.2) and plotted from  $-\pi/2$  to  $+\pi/2$  (corresponding to approximately  $-10$  and  $+10$  ms). Distributions are unimodal, normal, and centered around 0. At high coupling strengths, the two groups are strongly phase-locked with a phase shift of 0. As in Figs. 9.2 and 9.3, there are no conduction delays in the reentrant connections.

or more (e.g., [10]). Zero phase shifts could be achieved by the presence of at least some proportion of fast axons, or by other mechanisms including common input or short-term synaptic modification.

### Linking and Binding

In the cortex, neuronal groups may interact within a single area or between different areas. In the first case, the responses of neuronal groups (often belonging to the same feature domain such as orientation) are linked by means of reentrant connections within a single cortical area. A simple example is the linking of neuronal responses to oriented line segments forming an extended contour. We distinguish this level of *linking* (by intra-areal reentry) from that of *binding*. Binding occurs between neuronal groups in different cortical areas (often belonging to different feature domains) by interareal reentry. An example is the integration of neuronal responses to a particular stimulus contour with those to its direction of movement or to its color. This distinction of multiple levels of integration (linking and binding) closely corresponds to levels of anatomic organization (intra-areal and interareal connectivity).

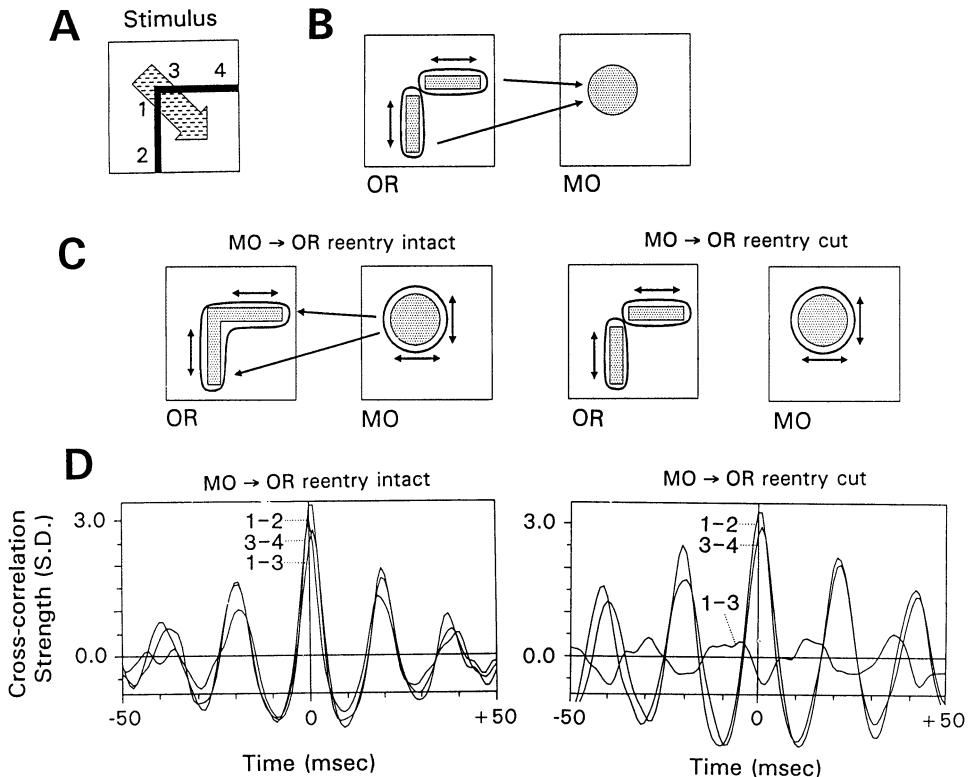
We have used neuronal groups arranged in topographic maps of visual space and linked by anatomically ordered reentrant connections to model



**Fig. 9.5.** Effect of allowing conduction delays in reentrant connections. Shown is the distribution of phase shifts for two reentrantly coupled neuronal groups as a function of delay (in ms) for a constant coupling strength of 0.25.

some functional aspects of the visual system [49,50]. The first model we want to discuss in this chapter ([49], see Fig. 9.6) exemplifies the emergence of temporal correlations both within and between visual maps (linking and binding). The model consists of two areas (called OR and MO) containing neuronal groups responsive to different stimulus attributes (orientation and movement, respectively) but sharing common primary dimensions (gross topography of visual space). As described above, each neuronal group is explicitly modeled as a local population of excitatory and inhibitory neurons. Upon external stimulation, local recurrent network interactions give rise to oscillatory activity. Groups at different positions within the separate visual maps are connected by two kinds of reentrant connections. Intra-areal reentrant connections link adjacent groups (which have overlapping receptive fields) of the same stimulus specificity. In the model, there are no connections between groups of orthogonal specificity (e.g., horizontal and vertical orientations). In addition, there are interareal reentrant connections between groups within OR and MO. These connections are divergent-convergent in both directions; therefore groups in MO have larger receptive fields and in turn project back onto a larger region of OR. As a result of the spatial arrangement of their afferent connectivity, groups in MO are selective for the direction of motion of a spatially extended stimulus contour (pattern motion; [2]); in contrast, groups in OR are only selective for the motion of short line segments (component motion).

Reentrant interactions within OR give rise to temporal correlations between neighboring as well as distant groups with a near zero phase lag. In the model as well as in experiments [26], correlations are found between neuronal groups that have nonoverlapping receptive fields if a long, continuous moving bar is presented. These correlations are weakened if two collinear



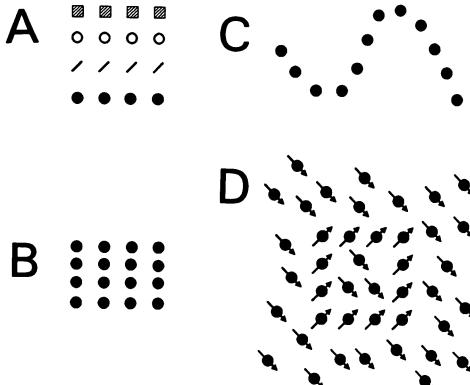
**Fig. 9.6.** Schematic diagram (A-C) and examples of neuronal responses (D) in the model of [49]. Both of the model's two areas (OR and MO) contain intra-areal reentrant connections, and the two areas are linked by interareal reentrant connections. (A) The stimulus is an extended contour shaped like a corner moving right down. (B) and (C) schematically show different stages of linking and binding; active regions of the network are stippled and coherent domains are enclosed in thick outlines. (B) Due to the pattern of selective connectivity in OR, direction selective neurons correlate their activities for each of the legs of the corner stimulus (linking). Pattern motion selective neurons in MO respond to the overall motion of the corner; this is brought about by convergent connections from many direction-selective neurons in OR. (C, left panel) Intra-areal reentrant connections within MO lead to locally coherent activity (linking). Divergent reentrant connections from MO to OR establish correlations between these two areas (binding). Furthermore, they enhance the local linking process in OR and establish correlations between the neuronal responses to the two legs of the moving corner. This example shows that linking and binding can interact across multiple areas. (C, right panel) If the reentrant backconnections from MO to OR are cut the two legs of the corner stimulus cannot be linked in OR. (D) Cross-correlations between neuronal groups responding to parts of the stimulus indicated in (A), with MO → OR connections intact (left panel) and cut (right panel). Cross-correlation strength is computed from scaled cross-correlation functions and is expressed in standard deviations (S.D.). Data plotted are averages over 10 presentations of the stimulus.

short bars (separated by a narrow gap) are moved separately, but with the same velocity. The correlations disappear altogether if the gap between the two bars is wider than the lateral extent of reentrant connections in OR. This is a relatively simple example of linking (the establishment of correlations among features belonging to the same feature domain, in this case orientation), demonstrating the dependence of a pattern of correlations on a global property (i.e., continuity) of a visual stimulus. Several other simulation studies have shown similar results (e.g., [13,23,32,37,47,48]). The interareal reentrant connections between OR and MO can also give rise to temporal correlations. If, for example, the model is presented with a moving contour shaped like a corner, OR responds to the orientation and precise position of the parts of the contour, but does not respond to the pattern motion of the entire stimulus. By contrast, MO responds to the direction of motion of the entire corner (pattern motion), but not to the orientation and precise position of the parts of which it is composed. The integration of the stimulus as a “moving corner” is achieved through the establishment of coherent oscillations between the two maps, made possible by interareal reentry. Cutting the reentrant  $\text{MO} \rightarrow \text{OR}$  projections (Figs. 9.6(C) and 9.6(D)) abolishes temporal correlations between MO and OR, as well as some correlations within OR. Thus, the model is able to establish correlations between different stimulus attributes, orientation, and motion (an example of binding).

An important aspect of the model is that local linking of responses in one area may depend on global interactions with another area (binding). The synthesis of coherent neuronal responses to an extended contour in OR occurs only if appropriate reentrant inputs from MO can be utilized (Figs. 9.6(C) and 9.6(D)). This is an example of how responses in “lower” visual areas can be modified by inputs from “higher” visual areas. Another important aspect of the model is the dependence of patterns of temporal correlations on underlying patterns of anatomical connectivity. This aspect was explored more extensively in a subsequent model [50].

### 9.2.3 FIGURE-GROUND SEGREGATION

An interconnected network of many neuronal groups can produce a very large number of patterns of temporal correlations that depend on its anatomical connectivity and its inputs. Based on our models we have suggested that temporal correlations as established by reentrant signaling may provide a key to the solution of a classical problem in visual perception, that of *perceptual grouping* and *figure-ground segregation*. These two processes, both of fundamental importance in perceptual organization, refer to the ability to group together elementary features into discrete objects and to segregate these objects from each other and from the background. At the beginning of this century, Gestalt psychologists thoroughly investigated the factors influencing grouping and the distinction between figure and ground



**Fig. 9.7.** Examples of Gestalt laws. The panels show grouping and segregation of elementary stimulus features according to the laws of (A) similarity, (B) proximity, (C) continuity, and (D) common motion. In (D), arrows indicate the movement of two separate populations of dots.

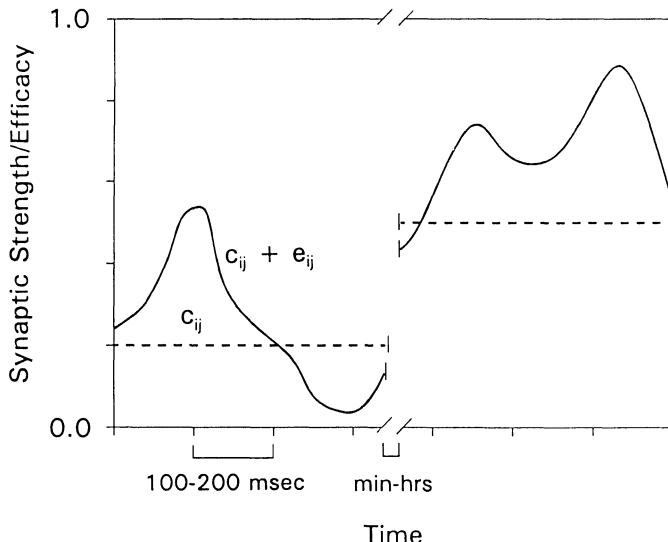
(e.g., [33,35,36]). They described a number of perceptual laws (“Gestalt laws”), such as those of similarity, proximity, continuity, and common motion (see Figs. 9.7(A)–9.7(D)). For example, looking at the stimulus in Fig. 9.7(A), we tend to perceptually group together similar stimulus elements and perceive a horizontal organization.

The role of correlations in sensory segmentation has been explored in a number of models (e.g., [30,44,56,57]). Our model [50] of perceptual grouping and figure-ground segregation in vision consists of an array of orientation- and direction-selective neuronal groups forming a map of visual space. These groups are linked by a pattern of reentrant connections similar to that existing in visual cortex. There are preferential connections between neuronal groups with similar orientation and direction selectivity, and connection density falls off with distance. An important feature of the model is the distinction between synaptic strength  $c_{ij}$  and synaptic efficacy  $e_{ij}$  (Fig. 9.8). Changes in synaptic strength are generally considered to represent long-term synaptic plasticity; in the present model all strengths  $c_{ij}$  remain fixed. However, the synaptic efficacies  $e_{ij}$  of reentrant (and local excitatory–excitatory) connections can change on a short time scale (within tens of milliseconds).

The activity of single cells is computed as

$$s_i(t) = \left[ \sum_j (c_{ij} + e_{ij}) \cdot s_j \right]_\theta + N + \omega s_i(t-1) \quad (9.1)$$

with  $s_i, s_j$  = states of cells  $i, j$  ( $0 \leq s_i \leq 1$ );  $t$  = time (one iteration in the computer simulations is taken to correspond to 1 ms of real time);  $c_{ij}$  = baseline synaptic strength; and  $e_{ij}$  = fast-changing synaptic efficacy be-



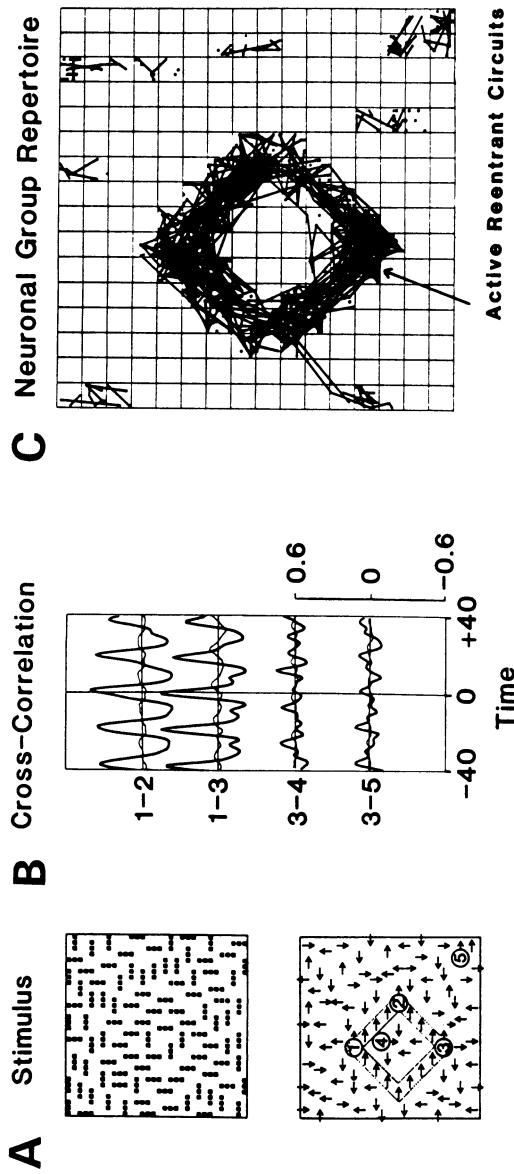
**Fig. 9.8.** Schematic diagram of the time courses of synaptic strengths ( $c_{ij}$ ) and efficacy ( $e_{ij}$ ). Efficacy and strength change on time scales of hundreds of milliseconds and of minutes to hours, respectively.

tween unit  $j$  (source) and unit  $i$  (target;  $0 \leq c_{ij} + e_{ij} \leq 1$  for excitatory connections);  $j$  = index over individual afferent connections;  $\Theta$  = input threshold (inputs set to 0, if below  $\Theta$ ;  $\Theta$  subtracted from inputs otherwise);  $N$  = Gaussian noise;  $\omega$  = decay coefficient. At the beginning of a simulation, each synapse is assigned a baseline  $c_{ij}$  value, which is nonzero and remains constant throughout the run, and an  $e_{ij}$  value, which is zero. For reentrant connections and excitatory-excitatory connections within a neuronal group, the synaptic efficacy  $e_{ij}$  is allowed to change as a result of pre- and postsynaptic activity, according to

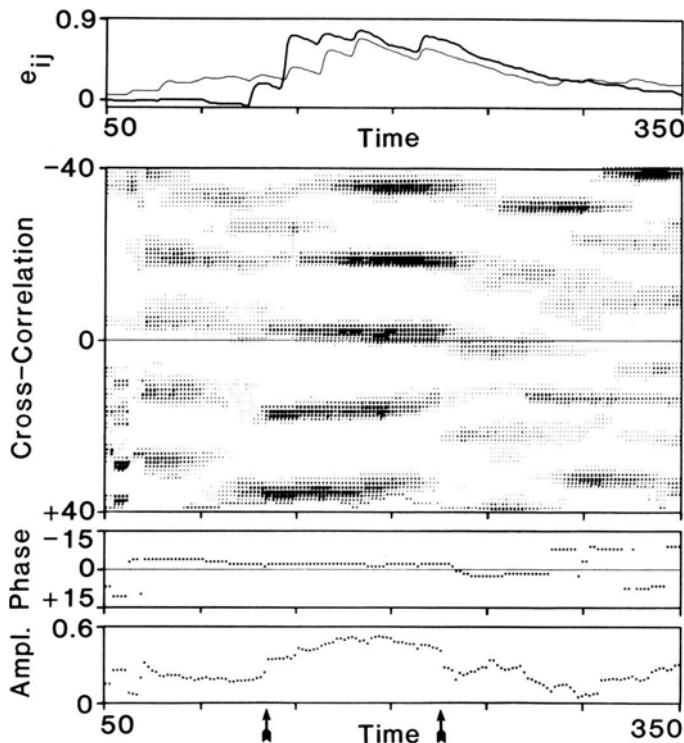
$$e_{ij}(t+1) = (1 - \gamma) \cdot e_{ij}(t) + \delta \cdot \phi(e_{ij}) \cdot (\bar{s}_i - \theta_I) \cdot (\bar{s}_j - \theta_J) \cdot R \quad (9.2)$$

with  $\delta$  = amplification factor, a parameter which adjusts the overall rate of synaptic change;  $\gamma$  = decay constant for synaptic efficacy;  $\bar{s}_i, \bar{s}_j$  = time-averaged activity of cells  $i, j$ ;  $\theta_I, \theta_J$  = amplification thresholds relating to post- and presynaptic activity, respectively;  $R = 0$  if  $\bar{s}_i < \theta_I$  and  $\bar{s}_j < \theta_J$  and  $R = 1$  otherwise;  $\phi(x)$  = decreasing sigmoidal function.

In general, the efficacy  $e_{ij}$  among cells or groups that show correlated activity (i.e., both  $\bar{s}_i$  and  $\bar{s}_j$  are above their respective amplification thresholds) increases. In turn, enhanced synaptic efficacy has the effect of rapidly amplifying and stabilizing correlations. On the other hand, synaptic efficacies between uncorrelated cells or groups decrease, resulting in further decorrelation of their activities. Changes in synaptic efficacy are transient; in the absence of activity, the efficacy rapidly returns to its resting value



**Fig. 9.9. An example of grouping and segmentation in the model of [51].** (A) Stimulus presented to the model, consisting of a diamond-shaped figure, composed of vertically oriented bars, and a set of randomly oriented bars forming a background. In the top panel the bars are shown at their starting positions, the bottom panel shows their corresponding directions of movement indicated by arrows. Encircled numbers with arrows in the bottom panel refer to the locations of recorded neuronal activity. The corresponding cross-correlations are displayed in (B). “Electrodes” 1 and 2 recorded from neurons responding to the figure, electrodes 3, 4, and 5 from neurons responding to the background. (B) Cross-correlograms of neuronal responses to the stimulus shown in (A). Cross-correlograms are computed over a 100 ms sample period and subsequently averaged over 10 trials. Numbers refer to the locations of neuronal groups within the direction-selective repertoires (see A). Four correlograms are shown, computed between 201 and 300 ms after stimulus onset. The correlograms are scaled and shift predictors (thin lines, averaged over nine shifts) are displayed for comparison. (C) Establishment of active reentrant circuits among neuronal groups selective for vertically oriented bars moving to the right, 250 ms after the onset of the stimulus shown in (A). Black lines indicate strongly enhanced functional connections.



**Fig. 9.10.** Comparison between time courses of synaptic efficacy  $e_{ij}$ , cross-correlation strengths, amplitude, and phase, all measured between two neuronal groups activated by a stimulus similar to the one shown in Fig. 9.9(A). Stimulus onset is at 0 ms. Top panel: time course of synaptic efficacy for two individual connections (thick and thin lines) linking cells in two neuronal groups responding to different parts of the stimulus. (In this case,  $e_{ij}$  varies between  $-0.1$  and  $0.9$ , since  $c_{ij}$  was set to  $0.1$ , see Eq. (9.2)). Middle panel: time-stepped cross-correlation function for the same nearby groups during the same trial. The cross-correlation function is evaluated every 2 ms, starting with 1–100 ms (i.e., centered on 50 ms). The cross-correlation amplitude is coded as stippling density. Bottom panels: phase and amplitude of the central peak of the time-stepped cross-correlation function. Note the simultaneous occurrence (within the time period marked by arrows) of enhanced synaptic efficacy, increased cross-correlation amplitude, and near-zero phase lag.

(within 100–200 ms for the value of  $\gamma$  chosen in our simulations). Rapid and transient changes in synaptic efficacy result in activity-dependent modulations of functional patterns of connectivity on a time scale of hundreds of milliseconds. Aertsen has pointed out [9] that short-term changes in the efficacy of connections can be achieved either by an actual biophysical change at the synapse or by dynamic influences exerted by other groups of neurons. From our perspective this underscores that, irrespective of actual biophysical mechanisms at the synaptic level, short-term modifications should be a general and widespread phenomenon in the cerebral cortex. Evidence for voltage-dependent short-term increases in synaptic efficacy has been found for horizontal (intra-areal) connections in cat visual cortex [24,29]. Fast modulation of synaptic efficacy has also been observed in other parts of the nervous system (e.g., [3–5,25]).

Figure 9.9 shows an example of the responses of the model. The groups responding to the bars which composed the object are rapidly linked by coherent oscillations and are segregated from those groups responding to elements of the background. We found that the ability to establish specific linking (grouping) is directly related to the ability of achieving segmentation. Accordingly, there is no coherency among groups responding to elements of the figure and others responding to elements of the background; the latter include elements that are moving in the same direction as the figure, but are placed some distance away. In the model, synchronization after stimulus onset is rapid and—in accordance with perceptual data—occurs usually within 100–200 ms. Multiple coherent episodes of varying length may occur at different times in different trials. Furthermore, synchrony is transient (with coherent episodes lasting between 100–500 ms), and its offset is fast, as would be required by the fact that the visual scene continuously changes due to eye movements. In the model, episodes of correlated activity coincide with the transient enhancement of reentrant connectivity due to short-term changes in synaptic efficacy ([50], Fig. 9.10).

This computer model shows that, at least in principle, the neural basis for the integration and segregation of elementary features into objects and background might be represented by the pattern of temporal correlations among neuronal groups mediated by reentry. In addition, since the resulting grouping and segregation are consistent with the Gestalt laws of continuity, proximity, similarity, common orientation, and common motion (Table 9.1), it suggests that the neural basis for these laws is to be found implicitly in the specific pattern of reentrant connectivity incorporated into the architecture. Perceptual grouping in the auditory domain follows certain “perceptual laws” not unlike those governing visual grouping. It might be suggested that the neural basis for auditory grouping is represented by patterns of correlated activity. Auditory grouping laws might be the result of a specific pattern of reentrant connections in auditory cortex.

**Table 9.1.** Gestalt laws and their possible neural bases in visual cortex.

Gestalt law	Neural bases	
	Structural	Dynamic
Similarity (common orientation)	Preferential intra-areal connections between neuronal groups of similar specificity (e.g., for orientation) <sup>a,b</sup>	Linking
Proximity	Decreasing density with distance and limited range of intra-areal connections <sup>a,b</sup>	Linking
Continuity	Same as proximity	Linking
Common motion	Preferential intra-areal connections <sup>a</sup> Reentrant (interareal) connections from pattern-motion selective cells <sup>b</sup>	Linking Linking and binding

<sup>a</sup>Sporns et al., 1991a.

<sup>b</sup>Sporns et al., 1989.

#### 9.2.4 COOPERATIVE INTERACTIONS AMONG MULTIPLE CORTICAL AREAS

Recently, we generalized our simulation results obtained from the models previously discussed [49,51] by simultaneously implementing linking and binding as well as constructive and correlative functions of reentry in a model of multiple areas of the visual system [54]. With this model, we attempted to explain the unity of visual perception across submodalities, and to establish a link between cooperative effects in cortical networks and behavior.

We introduced a computational scheme that deals explicitly and efficiently with short-term correlations among large numbers of units. In the model, cortical maps are made up of basic units that stand for populations of neurons organized in groups. The activity of a unit thus reflects the temporal discharge pattern of an entire group rather than that of an individual cell. At each computer iteration (corresponding to several tens of milliseconds) the state of each unit  $i$  is characterized by an activation variable  $s_i(t)$  ( $0 \leq s_i(t) < 1$ ), and a phase variable  $p_i(t)$  (circular,  $0 \leq p_i(t) < 2\pi$ ). While  $s_i$  indicates the *activity level*,  $p_i$  indicates *when* the unit fires within the given iteration. Thus, we approximate the activity profile of a group of neurons for a given iteration by using an amplitude and a phase variable. In our scheme, correlations between two or more units can be segregated into two components, long-term (hundreds of milliseconds or seconds) and short-term (tens of milliseconds). While the long-term component is reflected in the covariance between activity variables over *many* iterations, the short-term component is reflected by phase differences at *each* iteration. As in the model of figure-ground segregation we incorporated a mechanism

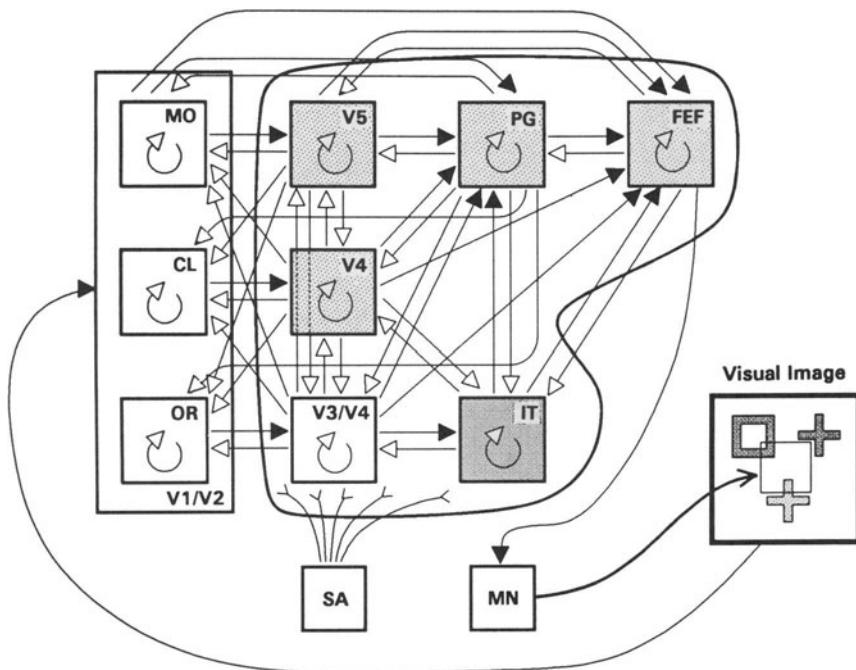
of short-term plasticity (voltage-dependent effects) for many of the visual pathways.

The model receives visual input from a color camera and contains nine functionally segregated areas divided into three parallel anatomical streams for form, color, and motion. Each of the streams is organized hierarchically; areas at “lower” levels contain units with small receptive fields that respond to simple stimulus features (such as orientation), while areas at higher levels contain units with large receptive fields that respond to complex features (such as shape). All areas contain intra-areal connections linking units of similar specificity. The areas are connected by mostly reciprocal pathways. Altogether, 10,000 units are linked by about 1,000,000 connections between areas at different levels (forward and backward), areas at the same level (lateral), and within an area (intrinsic). Forward connections tend to increase functional specialization, while backward connections tend to terminate more diffusely and are mostly voltage-dependent. A detailed description of the model is given in Tononi et al. ([54]; for a schematic diagram illustrating the basic anatomy see Fig. 9.11).

We used the model to investigate a number of problems all of which are related to the overall theme of how integrated cortical function is achieved. We studied two psychophysical phenomena (the perception of form from motion and motion capture) which, respectively, illustrate the constructive and correlative functions of reentry. We showed that the reentrant interactions between the motion and form streams can be used to construct responses to oriented lines from moving random dot fields (form from motion). In the model, the activity of a “higher” area of the motion stream (V5) is reentered into a “lower” area of the form stream (V1) and modifies its responses to an incoming stimulus. Motion capture involves the illusory attribution of motion signals from a moving object to another object (defined by chromatic or color boundaries) that is actually stationary. Based on the model, we propose that the basis for the perceptual effect of motion capture is the emergence of short-term correlations between units in the motion and color streams as a result of reentrant connections linking these streams.

The correlative properties of the model are further explored in simulations involving all three streams. When presented with a single object, the model solves the so-called “binding problem” and displays coherent unit activity both within and between different areas, including a nontopographic one (see figures in [54]). Two or more objects can be simultaneously differentiated. Coherent unit activity depends on the presence of reentrant pathways giving rise to widespread cooperative interactions among areas. In order to efficiently evaluate patterns of correlations we simultaneously display short-term correlations in a correlation matrix. Consecutive displays of the correlation matrix reflect the functional connectivity of areas and units within areas over time (Fig. 9.12).

A key characteristic of this model is that successful integration is linked



**Fig. 9.11. Architecture of a model of the visual system [54].** Segregated visual maps are indicated as boxes, pathways (composed of many thousands of individual connections) are indicated as arrows. The model comprises three parallel streams involved in the analysis of visual motion (top row), color (middle row), and form (bottom row). Areas are finely (no shading) or coarsely topographic (light shading) or nontopographic (heavy shading). The visual image (sampled by a color CCD camera) is indicated at the extreme right. The output of the system (simulated saccade movements under the control of eye motoneurons MN) is indicated at the bottom. Filled arrows indicate voltage-independent pathways, unfilled arrows indicate voltage-dependent pathways (incorporating short-term synaptic plasticity). Curved arrows within boxes indicate intra-areal connections. The box labeled SA refers to the diffusely projecting saliency system used in the behavioral paradigm; the general area of projection is outlined. V1/V2, V3/V4, V4, and V5 = corresponding visual cortical areas in the macaque monkey; MO, CL, OR = motion-, color-, and orientation-selective parts of V1/V2; PG = parietal cortex; FEF = frontal eye fields; IT = inferotemporal cortex; SA = saliency system; MN = eye motor neurons.

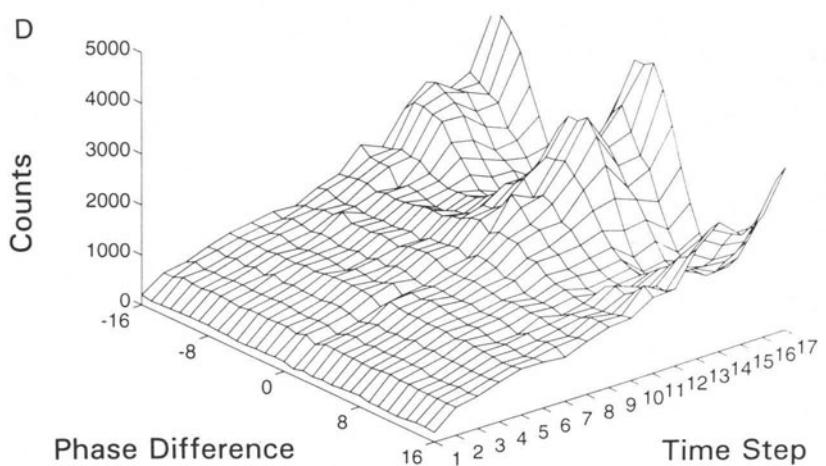
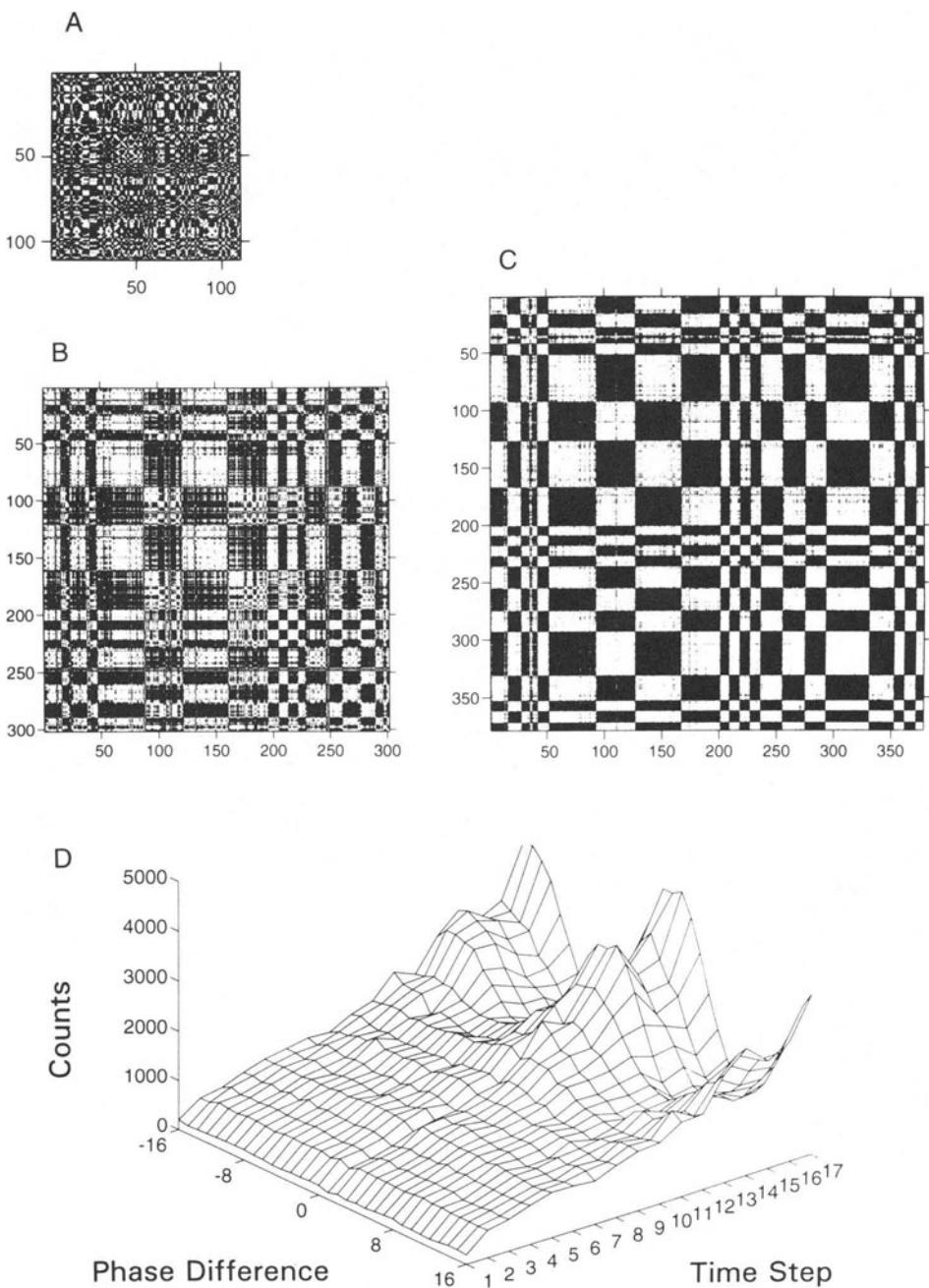


Fig. 9.12.

to an observable output, a simulated saccade response (for another example of how an integrated state can affect behavior see [53]). This eliminates the problem of deducing potential outputs by interpreting specific patterns of neural activity and correlations from the point of view of a privileged observer; it also shows that temporal correlations are not merely epiphenomena of neural activity. If one assumes that neurons are sensitive to temporal correlations in their inputs [1], it can be demonstrated [53,54] that behavioral responses may depend specifically on patterns of correlations.

The saccade response is also used as a basis for conditioning. Reward for a correct discrimination response is mediated by activation of a saliency system that resembles diffuse projection systems in the brain, such as the monoaminergic and cholinergic systems. The simulated diffuse release of a modulatory substance from the saliency system regulates long-term synaptic changes in multiple cortical pathways. After conditioning, the model performs a behavioral discrimination of objects that requires the integration through reentry of distributed information regarding shape, color, and location. This behavior depends critically upon the presence of short-term temporal correlations brought about by reentry and it does not require integration by a hierarchically superordinate area.

### 9.3 Summary and Conclusions

In this chapter we have reviewed a variety of anatomically and physiologically detailed models of the visual cortex that provide new insight into a fundamental problem of brain function, the problem of cortical integration. Central to the solution of this problem are the patterns of temporal corre-

---

**Fig. 9.12.** (A–C) Matrices of phase differences obtained from the model of the visual cortex [54]. All active units ( $s_i > 0.2$ ) of the system are tabulated on the vertical and horizontal axes and their phase difference ( $p_i - p_j$ ) is displayed (coded in shades of gray, white = 0, black =  $\pm 16$ ; note that  $p_i$  is circular and runs from  $0 \leq p_i < 1$ , divided in 32 bins). The model responded to a set of two objects, a cross and a square (same as in Fig. 9.9 in [54]), displayed are time steps 1 (A), 10 (B), and 17 (C). In (A), not many units are responding and phase differences are randomly and fairly uniformly distributed. In (B), more units have become active and coherent domains of units with phase difference 0 begin to form. In (C), a checkerboard pattern indicates that the active units have split into two coherent, but mutually uncorrelated, populations. (D) Distribution of phase differences over time from the onset of stimulation (time steps 1 to 17, same run as in A–C). The distribution is flat at the beginning of the run; at the end most phase differences are either 0 or  $\pm 16$ .

lations that result from the cooperative reentrant interactions within and between multiple brain areas.

### 9.3.1 CORTICAL INTEGRATION AT MULTIPLE LEVELS

The models illustrate that cortical integration takes place at multiple levels of organization: locally within neuronal groups, as well as within and between cortical areas. These various levels are mutually interdependent. The cooperative interactions of neurons within groups are important because in order to express the linking and binding of, for example, object attributes, temporal correlations need to be statistically significant and be able to affect other sets of neurons. Significant correlations between small sets of single neurons spanning multiple cortical areas could not be formed or maintained over time. All our models have shown that dynamic integration is strongly dependent on underlying anatomy. The dynamic distinction of linking and binding closely follows the anatomic distinction of intramap and intermap reentry. Other levels that we have not taken into account involve the binding of neural signals relating different sensory modalities, or the binding of neuronal activity related to memory and actual sensory input.

### 9.3.2 TEMPORAL DYNAMICS

Cortical integration has to meet important temporal constraints. Speaking somewhat loosely, the “perceptual time scale” is in the range of hundreds of milliseconds. For instance, in vision, the minimum duration of a perceptual experience appears to be around 100 ms [8]. Simple visual stimuli as well as complex visual scenes can be perceived within this time span [7]. According to these and other psychophysical results, any mechanism accounting for perceptual integration must be fast, on the order of a few hundreds of milliseconds. As both experimental and modeling studies have shown, short-term correlations provide a means to achieve integration within the limits of the perceptual time scale. We have suggested that the transient and dynamic character of perceptual integration is the result of changing patterns of short-term correlations and short-term modification of synaptic efficacies.

### 9.3.3 CORRELATIONS AND BEHAVIOR

While it is generally accepted that changes in the firing frequency of neurons can influence and control behavior, it is less obvious that changing patterns of temporal correlations (even in the absence of any changes in firing frequency) can do the same. Experiments as well as modeling studies have suggested that individual cortical neurons are sensitive to temporal correlations in their inputs [1,6]; thus, the timing of activity in one set of

neurons can have a well-defined effect on the activity of other neurons and can ultimately influence behavior. In some of our models [53,54] patterns of correlations were crucial in determining behavioral responses to complex stimuli. Some evidence [55] has pointed to a role for correlated neuronal activity in the control of behavior.

### 9.3.4 FUNCTIONS OF REENTRY

Our model of the visual cortex [54] shows that integration involves cooperative interactions of neurons both within and between cortical areas. We proposed that in integrating multiple cortical areas reentry can operate in two fairly distinct modes, constructive and correlative reentry. Reentrant inputs can modify the response properties of neurons directly and thus help to construct new response properties. An example of this was the generation of responses to form-from-motion boundaries in V1/V2 by reentrant inputs from V5. Reentrant interactions can also give rise to temporal correlations both within and between cortical areas; several examples have been given in this chapter. While having distinct functional characteristics these two modes of reentry are not mutually exclusive; correlations may give rise to subtle changes in neuronal response properties, which may in turn affect patterns of correlations.

In summary, computer simulations of cortical networks suggest that the reentrant interactions within and between segregated areas form an important neural basis for cortical integration.

*Acknowledgments.* This work was carried out as part of the Institute Fellows in Theoretical Neurobiology program at The Neurosciences Institute, which is supported through the Neurosciences Research Foundation. The Foundation received major support from the J.D. and C.T. MacArthur Foundation, the Lucille P. Markey Charitable Trust, and the van Ameringen Foundation. O.S. is a W.M. Keck Foundation Fellow.

## REFERENCES

- [1] Abeles, M. (1991) *Corticonics* (Cambridge University Press, Cambridge)
- [2] Adelson, E.H., Movshon, J.A. (1982) Phenomenal coherence of moving visual patterns. *Nature* **300**:523–525
- [3] Aertsen, A.M.H.J., Gerstein, G.L., Habib, M.K., Palm, G. (1989) Dynamics of neuronal firing correlations: Modulation of “effective connectivity.” *J. Neurophysiol.* **61**:900–917
- [4] Aertsen, A., Vaadia, E., Abeles, M., Ahissar, E., Bergman, H., Karmon, B., Lavner, Y., Margalit, E., Nelken, I., Rotter, S. (1991) Neural interactions in the frontal cortex of a behaving monkey: Signs of dependence on stimulus context and behavioral state. *J. Hirnforschung.* **32**:735–743

- [5] Ahissar, M., Ahissar, E., Bergman, H., Vaadia, E. (1992) Encoding of sound-source location and movement: Activity of single neurons and interactions between adjacent neurons in the monkey auditory cortex. *J. Neurophysiol.* **67**:203–215
- [6] Bernander, O., Douglas, R.J., Martin, K.A.C., Koch, C. (1991) Synaptic background activity influences spatiotemporal integration in single pyramidal cells. *Proc. Natl. Acad. Sci. U.S.A.* **88**:11,569–11,573
- [7] Biederman, I., Mezzanotte, R.J., Rabinowitz, J.C. (1982) Scene perception: Detecting and judging objects undergoing relational violations. *Cog. Psychol.* **14**:143–177
- [8] Blumenthal, A.L. (1977) *The Process of Cognition* (Prentice Hall, Englewood Cliffs)
- [9] Boven, K.-H., Aertsen, A. (1990) Dynamics of activity in neuronal networks give rise to fast modulations of functional connectivity. In: *Parallel Processing in Neural Systems and Computers*, R. Eckmiller, G. Hartmann, G. Hauske (Eds.) (Elsevier, Amsterdam), pp. 53–56
- [10] Bullier, J., McCourt, M.E., Henry, G.H. (1988) Physiological studies on the feedback connection to the striate cortex from cortical areas 18 and 19 of the cat. *Exp. Brain Res.* **70**:90–98
- [11] Damasio, A.R. (1989) The brain binds entities and events by multiregional activation from convergence zones. *Neural Comput.* **1**:123–132
- [12] Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., Reitboeck, H.J. (1988) Coherent oscillations: A mechanism of feature linking in the visual cortex? Multiple electrode and correlation analyses in the cat. *Biol. Cybern.* **60**:121–130
- [13] Eckhorn, R., Reitboeck, H.J., Arndt, M., Dicke, P. (1990) Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex. *Neural Comput.* **2**:293–307
- [14] Eckhorn, R., Frien, A., Bauer, R., Woelbern, T., Kehr, H. (1993) High frequency (60–90 Hz) oscillations in primary visual cortex of awake monkey. *NeuroReport* **4**:243–246
- [15] Edelman, G.M. (1978) Group selection and phasic reentrant signaling: A theory of higher brain function. In: *The Mindful Brain*, G.M. Edelman, V.B. Mountcastle (Eds.) (MIT Press, Cambridge, MA), pp. 51–100
- [16] Edelman, G.M. (1987) *Neural Darwinism* (Basic Books, New York)
- [17] Edelman, G.M. (1989) *The Remembered Present* (Basic Books, New York)
- [18] Engel, A.K., Kreiter, A.K., König, P., Singer, W. (1991a) Synchronization of oscillatory neuronal responses between striate and extrastriate visual cortical areas of the cat. *Proc. Natl. Acad. Sci. U.S.A.* **88**:6048–6052
- [19] Engel, A.K., König, P., Kreiter, A.K., Singer, W. (1991b) Interhemispheric synchronization of oscillatory neuronal responses in cat visual cortex. *Science* **252**:1177–1179

- [20] Engel, A.K., König, P., Kreiter, A.K., Schillen, T.B., Singer, W. (1992) Temporal coding in the visual cortex: New vistas on integration in the nervous system. *Trends Neurosci.* **15**:218–226
- [21] Felleman, D.J., Van Essen, D.C. (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex* **1**:1–47
- [22] Finkel, L.H., Edelman, G.M. (1989) The integration of distributed cortical systems by reentry: A computer simulation of interactive functionally segregated visual areas. *J. Neurosci.* **9**:3188–3208
- [23] Gerstner, W., Ritz, R., van Hemmen, L. (1993) A biologically motivated and analytically soluble model of collective oscillations in the cortex. *Biol. Cyber.* **68**:363–374
- [24] Gilbert, C.D. (1992) Horizontal integration and cortical dynamics. *Neuron* **9**:1–13
- [25] Gochin, P.M., Miller, E.K., Gross, C.G., Gerstein, G.L. (1991) Functional interactions among neurons in inferior temporal cortex of the awake macaque Exp. Brain Res. **84**:505–516
- [26] Gray, C.M., Singer, W. (1989) Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl. Acad. Sci. U.S.A.* **86**:1698–1702
- [27] Gray, C.M., König, P., Engel, A.K., Singer, W. (1989) Oscillatory responses in cat visual cortex exhibit intercolumnar synchronization which reflects global stimulus properties. *Nature* **338**:334–337
- [28] Gray, C.M., Engel, A.K., König, P., Singer, W. (1992) Synchronization of oscillatory neuronal responses in cat striate cortex: Temporal properties. *Visual Neurosci.* **8**:337–347
- [29] Hirsch, J.A., Gilbert, C.D. (1991) Synaptic physiology of horizontal connections in the cat's visual cortex. *J. Neurosci.* **11**:1800–1809
- [30] Hummel, J.E., Biederman, I. (1992) Dynamic binding in a neural network for shape recognition. *Psychol. Rev.* **99**:480–517
- [31] Jagadeesh, B., Gray, C.M., Ferster, D. (1992) Visually evoked oscillations of membrane potential in cells of cat visual cortex. *Science* **257**:552–554
- [32] Kammen, D.M., Holmes, P.J., Koch, C. (1989) Cortical architecture and oscillations in neuronal networks: Feedback versus local coupling. In: *Models of Brain Function*, R.M.J. Cotterill (Ed.) (Cambridge University Press, Cambridge), pp. 273–284
- [33] Kanizsa, G. (1979) *Organization in Vision: Essays on Gestalt Perception* (Praeger, New York)
- [34] Kisvarday, Z.F., Eysel, U.T. (1992) Cellular organization of reciprocal patchy networks in layer III of cat visual cortex (Area 17). *Neuroscience* **46**:275–286
- [35] Koffka, K. (1935) *Principles of Gestalt Psychology* (Harcourt, New York)
- [36] Köhler, W. (1947) *Gestalt psychology* (Liveright, New York)

- [37] König, P., Schillen, T.B. (1991) Stimulus-dependent assembly formation of oscillatory responses. I. Synchronization. *Neural Comput.* **3**:155–166
- [38] Kreiter, A.K., Singer, W. (1992) Oscillatory neuronal responses in the visual cortex of the awake macaque monkey. *Eur. J. Neurosci.* **4**:369–375
- [39] Llinas, R., Ribary, U. (1993) Coherent 40 Hz oscillation characterizes dream state in humans. *Proc. Natl. Acad. Sci. U.S.A.* **90**:2078–2081
- [40] Mignard, M., Malpeli, J.G. (1991) Paths of information flow through visual cortex. *Science* **251**:49–1251
- [41] Murthy, V.N., Fetz, E.E. (1992) Coherent 25 to 35 Hz oscillations in the sensorimotor cortex of awake behaving monkeys. *Proc. Natl. Acad. Sci. U.S.A.* **89**:5670–5674
- [42] Nakamura, K., Mikami, A., Kubota, K. (1992) Oscillatory neuronal activity related to visual short-term memory in monkey temporal pole. *NeuroReport* **3**:117–120
- [43] Nelson, J.I., Salin, P.A., Munk, M.H.-J., Arzi, M., Bullier, J. (1992) Spatial and temporal coherence in cortico-cortical connections: A cross-correlation study in areas 17 and 18 in the cat. *Visual Neurosci.* **9**:21–37
- [44] Neven, H., Aertsen, A. (1992) Rate coherence and event coherence in the visual cortex: A neuronal model of object recognition. *Biol. Cyber.* **67**:309–322
- [45] Ribary, U., Ioannides, A.A., Singh, K.D., Hasson, R., Bolton, J.P.R., Lado, F., Mogilner, A., Llinas, R. (1991) Magnetic field tomography of coherent thalamocortical 40 Hz oscillations in humans. *Proc. Natl. Acad. Sci. U.S.A.* **88**:11037–11041
- [46] Sherrington, C. (1947) *The Integrative Action of the Nervous System* (Yale University Press, New Haven, CT)
- [47] Sompolinsky, H., Golomb, D., Kleinfeld, D. (1990) Global processing of visual stimuli in a network of coupled oscillators. *Proc. Natl. Acad. Sci. U.S.A.* **87**:7200–7204
- [48] Sompolinsky, H., Golomb, D., Kleinfeld, D. (1991) Cooperative dynamics in visual processing. *Phys. Rev. A* **43**:6990–7011
- [49] Sporns, O., Gally, J.A., Reeke, G.N., Jr., Edelman, G.M. (1989) Reentrant signaling among simulated neuronal groups leads to coherency in their oscillatory activity. *Proc. Natl. Acad. Sci. U.S.A.* **86**:7265–7269
- [50] Sporns, O., Tononi, G., Edelman, G.M. (1991a) Modeling perceptual grouping and figure-ground segregation by means of active reentrant connections. *Proc. Natl. Acad. Sci. U.S.A.* **88**:129–133
- [51] Sporns, O., Tononi, G., Edelman, G.M. (1991b) Dynamic interactions of neuronal groups and the problem of cortical integration. In: *Nonlinear Dynamics and Neural Networks*, H.G. Schuster (Ed.) (Verlag Chemie, Weinheim, Germany), pp. 205–240

- [52] Stevens, C. (1989) How cortical interconnectedness varies with network size. *Neural Comput.* **1**:473–479
- [53] Tononi, G., Sporns, O., Edelman, G.M. (1992a) The problem of neural integration: Induced rhythms and short-term correlations. In: *Induced Rhythms in the Brain*, E. Basar, T.H. Bullock (Eds.) (Birkhäuser, Boston), pp. 367–395
- [54] Tononi, G., Sporns, O., Edelman, G.M. (1992b) Reentry and the problem of integrating multiple cortical areas: Simulation of dynamic integration in the visual system. *Cerebral Cortex* **2**:310–335
- [55] Vaadia, E., Ahissar, E., Bergman, H., Lavner, Y. (1991) Correlated activity of neurons: A neural code for higher brain functions? In: *Neuronal Cooperativity*, J. Krüger (Ed.) (Springer-Verlag, Berlin), pp. 249–279
- [56] von der Malsburg, C., Buhmann, J. (1992) Sensory segmentation with coupled neural oscillators. *Biol. Cybern.* **67**:233–242
- [57] von der Malsburg, C., Schneider, W. (1986) A neural cocktail party processor. *Biol. Cybern.* **54**:29–40
- [58] Zeki, S. (1978) Functional specialization in the visual cortex of the rhesus monkey. *Nature* **274**:423–428
- [59] Zeki, S., Shipp, S. (1988) The functional logic of cortical connections. *Nature* **335**:311–317
- [60] Zeki, S. (1990) Parallelism and functional specialization in human visual cortex. *Cold Spring Harbor Symp. Quant. Biol.* **55**:651–661

# Index

## A

Absolute refractory period, 178  
Absolute refractory time, 40  
Action potential, 36  
Adaptation, 247  
Adaptive neuron, 44  
Additive noise, 265  
After-hyperpolarizing potential, 40  
 $\alpha$  function, 181  
Amblyopia, 156  
Area 17, 147  
Assembly, 142  
Assembly code, 145  
Associative memory, 102, 299  
Associative recall, 202  
Associative retrieval, 175  
Asynchrony, 104, 110  
Attention, 162  
Auto-correlation, 124, 126  
Auto-correlogram, 147, 259  
Axonal transmission, 37  
Axonal transmission delay, 180

## B

Background activity, 194  
Basket cell, 176  
Baum-Welch algorithm, 275  
Binding, 156, 322  
Binding by synchrony, 146  
Binding problem, 143, 201  
Bistability, 187  
Box counting, 265  
Bulbar activities, 230  
Bulbar oscillations, 225  
Bulbar oscillatory time scale, 236  
Burst, 126, 165, 225

Burst of spikes, 185  
Burst structure, 196  
Burst synchronization, 176  
Bursting neuron, 45  
Bursting period, 185, 187

## C

$\text{Ca}^{2+}$  channels, 306  
 $\text{Ca}^{2+}$  conductance, 164  
 $\text{Ca}^{2+}$  dynamics, 290  
Calcium channels, 297, 298  
Calcium dynamics, 296  
Calmodulin, 292, 297  
Cardinal cells, 101  
Cell assemblies, 101  
Central feedback, 225, 236  
Coherence, 176  
Coherent firing, 68  
Coherent oscillation, 4, 26, 187, 227  
Collective firing, 188  
Collective oscillation, 177, 187  
Color constancy, 248  
Color hue shifts, 247  
Columnar organization, 176  
Columns, 206  
Compartments, 37  
Composite elements, 110  
Constructive reentry, 337  
Context sensitive binding, 206  
Control of action, 99, 111  
Corpus callosum, 198  
Correlation, 104, 106, 110, 113, 258, 323  
Correlation functions, 254  
Correlation theory, 95, 104, 116

- C**  
 Correlative reentry, 337  
 Correlogram, 259  
 Cortical integration, 315  
 Covariance rule, 300  
 Cross-correlation, 126, 129  
 Cross-correlogram, 259  
 Cross-enhancement, 238
- D**  
 Damped oscillator, 227  
 Delays, 158  
 Dendritic integration, 37  
 Dendritic spine, 290  
 Dendritic summation, 212  
 Dendrodendritic synapses, 223  
 Distribution of delay times, 183
- E**  
 Ensemble Ansatz, 176  
 Ensemble code, 2  
 Excitatory postsynaptic potential (EPSP), 47, 178, 181  
 Exhalation, 225  
 Extracellular recording, 122
- F**  
 Fast-changing synaptic efficacy, 326  
 Fatigue, 235  
 Feature detector, 101,  
 Feature linking, 176, 201  
 Features, 143  
 Feedback, 206, 209  
 Feedback connection, 144  
 Figure-ground, 102, 103, 104, 113  
 Figure-ground problem, 95  
 Figure-ground segregation, 176, 325  
 Firing probability, 49  
 Firing rate, 2, 36, 121, 130, 132, 226  
 Firing threshold, 178, 226  
 Frequency locking, 6
- G**  
 Gain function, 36  
 Gestalt laws, 331  
 Gestalt properties, 254  
 Global coherence, 196  
 Global oscillation, 232  
 Globally locked oscillation, 190  
 Glomeruli, 223  
 Glutamate, 290, 297, 298  
 Grandmother cell, 2, 101  
 Granule cells, 223
- H**  
 Hebb rule, 299, 300  
 Hebbian learning, 72, 74  
 Hebbian learning rule, 179  
 Hebbian synapse, 287, 288, 289  
 Hebb's postulate, 288  
 Hemispheres, 198  
 Hidden Markov model, 272  
 Hidden state model, 275  
 Hippocampus, 289  
 Hodgkin-Huxley model, 35  
 Homosynaptic potentiation, 165  
 Hopf bifurcation, 187  
 Hopfield model, 201  
 Hypercolumn, 26, 97
- I**  
 Incoherent firing, 55  
 Inhalation, 225  
 Inhibition, 178, 226  
 Inhibitory delay loop, 185  
 Inhibitory feedback, 159, 185  
 Inhibitory postsynaptic potential (IPSP), 47, 178  
 Integrate-and-fire model, 38  
 Interhemispheric delay, 198  
 Internal noise, 178  
 Interval distribution, 43  
 Intracellular recording, 122  
 Invariance, 99

Invariant, 103, 113

Ion conductance, 35

Ionic currents, 35

## J

Joint peri stimulus time histogram,  
131

## K

Kullback-Leibler information, 258

Kuramoto model, 6

## L

Large-deviations technique, 30

Layered network, 206

Linking, 322

Localization theory, 96, 97

Locked oscillations, 187

Locking, 187, 212

Locking condition, 189

Long-term depression, *see* LTD

Long-term memory, 96, 105

Long-term potentiation, *see* LTP  
*entries*

Long-term synaptic depression, *see*  
LTD

LTD, 164, 302, 304, 305

LTP, 164, 289, 302

LTP induction, 290, 291, 292, 293,  
298

Lyapunov function, 7

## M

Mammalian bulb, 222

Mass activity, 157

Master area, 316

Mean activity, 183

Mean-field theory, 184

Mean firing rate, 43, 177, 186

Membrane potential, 178, 226

Memory, 99, 100, 114, 263

*N*-methyl-*D*-aspartate, *see* NMDA

Micropipette, 122

Mitral cells, 223

Mode, 111

Moving targets, 162

Multiunit activity, 152

Mutual information, 258

## N

Neural code, 2

Neuronal groups, 316

NMDA, 164, 165, 290, 291, 292,  
293, 296, 297

Noise, 191

Non-Gaussian, 243

Nonlinear correlation technique,  
256

Nonlinear oscillator, 183, 227

## O

Object detection, 238

Object localization, 245

Object-specific cell, 144

Odor, 221

Odor fluctuation analysis, 238

Odor identity information, 239

Odor pattern, 226

Odor recognition, 232

Odor segmentation, 235, 239

Odor-specific canceling, 236

Odor strength, 232, 234

Olfactory adaptation, 222, 235

Olfactory bulb, 158, 221, 223

Olfactory cross-adaptation, 237

Olfactory system, 221

Onset of synchrony, 160

Orientation preference, 152

Oscillation, 126, 157, 164, 225

Oscillation pattern, 232

Oscillation period, 188

Oscillator models, 4

Oscillatory retrieval, 212

Oscillatory states, 187

Overlap, 57, 183

**P**

Pacemaker neuron, 159  
 Partial phase locking, 20  
 Patch electrode, 122  
 Pattern, 179  
 Pattern classifier, 227  
 Pattern segmentation, 176, 201  
 Perceptron, 103  
 Perceptual constancy, 248  
 Perceptual deficits, 156  
 Perceptual grouping, 325  
 Perceptual shifts, 247  
 Peri event time histogram, 130  
 Phase lag, 149  
 Phase locking, 4  
 Phase-shifted oscillation, 200  
 Point process, 124, 129  
 Poisson process, 124  
 Population, 129  
 Population code, 142  
 Poststimulus time histogram (PSTH),  
     97, 130  
 Postsynaptic potential, 100, 105  
 Predictability, 256  
 Principle components, 243  
 Probability of firing, 178  
 Psychophysics, 237  
 Pulse code, 2  
 Pyramidal neuron, 176

**R**

Raster dot display, 130  
 Rate function, 124, 129  
 Rates, 317  
 Receptive field, 112, 150  
 Receptor, 290  
 Receptor cells, 223  
 Reciprocal pathways, 316  
 Recognition, 161  
 Recognition times, 160  
 Reentry, 316  
 Refractoriness, 179  
 Refractory function, 40  
 Refractory period, 124

Relational code, 142  
 Renewal dynamics, 273  
 Renewal process, 124  
 Representation, 98, 113  
 Respiration, 225  
 Response function, 180  
 Response synchronization, 156,  
     157  
 Retrieval states, 186

**S**

Segmentation, 156, 201, 204, 222,  
     240  
 Self-organization, 99  
 Self-organizing systems, 111  
 Sensitivity enhancement, 238  
 Sensory information processing,  
     221  
 Shift predictor, 259  
 Short-term plasticity, 332  
 Single cell code, 142  
 Single neuron Ansatz, 176  
 Single unit activity, 123  
 Spatial coherence, 271  
 Spatial redundancies, 247  
 Spatio-temporal oscillatory activities, 222  
 Spatio-temporal patterns, 72  
 Spectral reflectance, 248  
 Spike generation, 37, 39  
 Spike raster, 196, 204  
 Spike response model (SRM), 39,  
     176, 177  
 Spike train, 2, 36, 104, 196  
 Spike transmission, 46  
 Spikes, 97, 123  
 Spiking probability, 179  
 Spiking statistics, 42  
 Spindles, 196  
 Stability, 17, 61, 70  
 Stationary overlap, 186  
 Stationary pattern, 175  
 Stationary state, 184, 210  
 Statistical dependency, 257

Stereo-olfactory mechanism, 245  
 Stimulus dependence, 152  
 Strabismic amblyopia, 156  
 Strabismus, 155  
 Sublattice phase locking, 20  
 Sublattices, 10, 184  
 Subnetwork, 111, 112, 114  
 Superposition problem, 164  
 Survivor function, 42  
 Synaptic adaptation, 241  
 Synaptic connection, 105  
 Synaptic depression, 299, 302  
 Synaptic enhancement, 289  
 Synaptic gain, 163  
 Synaptic modulation, 95, 105  
 Synaptic plasticity, 95, 100, 102,  
     105, 114, 163, 287, 299,  
     301  
 Synaptic strength, 239, 289, 298,  
     299  
 Synaptic weight, 100, 299  
 Synchronization, 146, 151, 317  
 Synchronization probability, 155  
 Synchronizing connection, 155  
 Synchrony, 104, 110, 157  
 Synfire chains, 82, 134, 135

**T**  
 Telencephalon, 221  
 Temporal code, 145  
 Temporal coherence, 145, 269

Temporal fluctuation analysis, 245  
 Temporal precision, 165  
 Temporal window, 165  
 Thalamic oscillations, 153  
 Threshold for odor detection, 236  
 Time resolution, 72  
 Time structure, 104, 106  
 Topological network, 108

**U**

Universality classes, 21

**V**

Visual scene, 164

**W**

Waterfall effect, 247  
 Weak locking, 190, 201, 209  
 Weak locking scenario, 26, 212  
 Weakly locked, 196  
 Whole cell recording, 122

**X**

*XY* ferromagnet, 8

**Z**

Zero phase lag, 158